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

Final Program

Program: November 17–22, 2019
Exhibits: November 18–21, 2019

Colorado Convention Center
Denver, CO
November 17, 2019

International Conference of
High Performance Computing (HPC)
Networking, Storage, and Analytics (SC19)
Colorado Convention Center
700 14th Street
Denver, Colorado 80202

Greetings:

It is my pleasure to welcome you to the “Mile High City.” We are excited that you have chosen our beautiful city to host your 2019 conference.

Over the next several days, you will have an opportunity to explore the latest innovations, research, and passion for high performance computing. Networking, finding solutions and applications, and reviewing new programs and services are available. The work that you do is so important to the way we all navigate the world because technology is critical to every facet of our lives.

We invite you to explore our city and experience all it has to offer. From seven professional sports teams to the nation’s second-largest performing arts complex, the mile-long 16th Street Mall and our many museums and cultural amenities, Denver is filled with attractions to entertain you. As an added feature, we are located just minutes from the adventure and beauty of the Rocky Mountains.

Neighborhood destinations such as Cherry Creek, Highlands, LoDo, Uptown, Five Points, Golden Triangle and RiNo are filled with outstanding restaurants and one-of-a-kind shops, galleries and more. And once the sun sets, Denver is an epicenter for live music and entertainment. Whether you visit one of our awe-inspiring mountain parks like Red Rocks Park & Amphitheatre, explore the newly revitalized Union Station, or take a spin on a B-cycle through our 85 miles of off-street bike paths, you’ll find Denver is bursting with activity year-round.

Best wishes for an informative and enjoyable conference and please plan to return to Denver soon.

Respectfully,

Michael B. Hancock
Mayor
General Information

**On-Site Registration and Conference Store**

Lobby F
Saturday, November 16, 1–6 pm
Sunday, November 17, 7 am–6 pm
Monday, November 18, 7 am–9 pm
Tuesday, November 19, 7:30 am–6 pm
Wednesday, November 20, 7:30 am–6 pm
Thursday, November 21, 7:30 am–5 pm
Friday, November 22, 8–11 am

**First Time at the SC Conference**

Four Seasons Ballroom 1-2-3
Monday, November 18, 4:30–5:15 pm

The SC19 Conference extends a warm welcome to all first-time attendees. A session on Monday afternoon November 18 will provide an overview of the conference, describe what each type of badge permits, and make suggestions about interesting activities in which to participate so you can organize your conference experience.

**Information Booths**

**Main Information Booth**

Lobby F
Saturday, November 16, 1–5 pm
Sunday, November 17, 8 am–6 pm
Monday, November 18, 8 am–7 pm
Tuesday, November 19, 8 am–6 pm
Wednesday, November 20, 8 am–6 pm
Thursday, November 21, 8 am–6 pm
Friday, November 22, 8:30 am–noon

**Secondary Information Booth**

D Concourse
Sunday, November 17, 8 am–5 pm
Monday, November 18, 8 am–5 pm
Tuesday, November 19, 8 am–6 pm
Wednesday, November 20, 8 am–4 pm
**ATM/Banks/Currency Exchange**

**U.S. Bank ATMs**
Lobby F
Directly behind Main Information Booth
Colorado Convention Center
700 14th St, Denver, CO 80202

**Colorado Credit Union**
717 17th St, Denver, CO 80202
(303) 978-2274

**Colorado National Bank**
700 17th St, Denver, CO 80202
(720) 214-0770

**Business Centers**

**UPS**
Lobby F
Colorado Convention Center
700 14th St, Denver, CO 80202
store6611@theupsstore.com
(720) 904-2300

Monday–Friday: 8 am–6 pm
Saturday: 9 am–3 pm
Sunday: Closed

**FedEx Office**
Hyatt Regency Hotel
650 15th St, Denver, CO 80202
(303) 534-2198

Monday–Friday: 7 am–7 pm
Saturday: 9 am–4 pm
Sunday: Closed

**Coat and Bag Check**
Lobby A
Saturday, November 16, 11 am–8 pm
Sunday, November 17, 7 am–6 pm
Monday, November 18, 7 am–10 pm
Tuesday, November 19, 7 am–9 pm
Wednesday, November 20, 7 am–8 pm
Thursday, November 21, 7 am–7 pm
Friday, November 22, 7 am–1 pm

**Exhibits**
Exhibit Halls A, B, E and F
Tuesday, November 19, 10 am–6 pm
Wednesday, November 20, 10 am–6 pm
Thursday, November 21, 10 am–3 pm

**Family Resources**

**Parents Room**
Room 608
Saturday, November 16, 7 am–9 pm
Sunday, November 17, 8 am–6 pm
Monday, November 18, 8 am–9 pm
Tuesday, November 19, 8 am–7 pm
Wednesday, November 20, 8 am–7 pm
Thursday, November 21, 8 am–7 pm
Friday, November 22, 8 am–noon

Sometimes you just need a quiet place to be with your child. At the convention center, we have a Parents Room with private feeding areas, refrigerators, changing tables, lockers to store your personal items, and plush chairs. We hope this room provides you a comfortable place to spend time with your child while you are away from home.

The Parents Room is intended for parents with infants. This will ensure peace and quiet for parents caring for their babies.
On-Site Child Care
Rooms 610, 612
Sunday, November 17, 8 am–6 pm
Monday, November 18, 8 am–9 pm
Tuesday, November 19, 8 am–7 pm
Wednesday, November 20, 8 am–7 pm
Thursday, November 21, 8 am–6 pm

Child care services, provided by KiddieCorp, are available in the convention center to registered attendees and exhibitors. For a small fee, children between the ages of 6 months and 12 years can participate.

We recommend registering children for childcare in advance as space may fill up.

On-Site Child Care Registration
There may be limited availability for drop-ins depending on the day, time, and age. Attendees are welcome to drop by the Child Care room and check the schedule.

Family Day
Exhibit Halls A, B, E and F
Wednesday, November 20, 4–6 pm

Welcome families! Any registered attendee can register for a Family Day pass. Attendees must check in their family members at the Assisted Registration Desk in Lobby A. On Wednesday from 4–6 pm, come to the Exhibit Hall and show your family around. Pass on your love of SC and HPC! Take pictures and share with @SCInclusivity!

First Aid
The first aid station is in Lobby E near Room 507.

Lost and Found
Located at the Exhibitor Approved Contractor (EAC) desk next to Coat and Bag Check in Lobby A.

Quiet Room
Mezzanine A1, Upper Level
Access via elevators from Lobby A
Saturday, November 16, 8:30 am–5 pm
Sunday, November 17, 8 am–6 pm
Monday, November 18, 8 am–9 pm
Tuesday, November 19, 8 am–7 pm
Wednesday, November 20, 8 am–7 pm
Thursday, November 21, 8 am–6 pm
Friday, November 22, 8 am–noon

The quiet room is a physical space where conversation and interaction are not allowed, where attendees can go for quiet, prayer, meditation, or similar activities.

Receptions
At SC19, you’ll have the opportunity to attend several receptions associated with key conference elements. Enjoy these times to meet and socialize with presenters and fellow attendees.

Exhibitors Reception
Wynkoop Brewery
1634 18th Street, Denver, CO 80202
Sunday, November 17, 6–9 pm

Food, refreshments, and entertainment for all registered exhibitors. Take a break and get something to eat before completing your booth setup.

Admission requires an Exhibitor badge, an event ticket, and government-issued ID.

Grand Opening Gala Reception
Exhibit Halls A, B, E and F
Monday, November 18, 7–9 pm

SC19 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, Exhibitor and Students@SC registrants.
Posters Reception

E Concourse
Tuesday, November 19, 5:15–7 pm

Celebrate the opening of the SC19 Posters and meet this year’s poster authors. Attendees with a Technical Program badge enjoy complimentary refreshments and appetizers.

Technical Program Reception

Wings Over the Rockies Air & Space Museum
7711 East Academy Blvd, Denver, CO 80230
Thursday, November 21, 6–9pm

The annual reunion of the people who create, operate, and use the world’s highest-performance computing systems. Drink a toast to another year of SC!

Admission requires a Technical Program badge, an event ticket, and government-issued ID. Event tickets for guests without a Technical Program badge can be purchased at the SC Conference store.

SC20 Information

Members of next year’s SC committee will be available in the SC20 preview booth to offer information and discuss next year’s SC conference in Atlanta, Georgia. Stop by for a copy of next year’s Call for Participation and pick up some free gifts!

SC20 Preview Booth

Lobby F
Tuesday, November 19, 10 am–5 pm
Wednesday, November 20, 10 am–5 pm
Thursday, November 21, 10 am–3 pm

SC20 Preview

Mile High Ballroom
Thursday, November 21, 8:30–8:35 am

Security

The SC19 security office is located in Room 204.

Visit Denver Booth

Information specialists can assist with providing information on attractions, dining, transportation options, maps, brochures, and a myriad of other helpful hints for navigating Denver.

Main Lobby

Across from the Blue Bear
Sunday, November 17, 8 am–6 pm
Monday, November 18, 8 am–7:30 pm
Tuesday, November 19, 8 am–6 pm
Wednesday, November 20, 8 am–6 pm
Thursday, November 21, 8:30 am–5 pm
Friday, November 22, 8–11 am

Security

The SC19 security office is located in Room 204.
## Technical Program (TP)

- Awards (Tue–Wed)
- Awards Ceremony (Thu)
- Birds of a Feather (Tue–Thu)
- Doctoral Showcase Presentations (Thu)
- Early Career (Mon)
- Exhibitor Forum (Tue–Thu)
- Exhibits (Tue–Thu)
- Grand Opening Gala Reception (Mon)
- HPC Impact Showcase (Tue–Thu)
- HPC Is Now Plenary (Mon)
- Invited Talks (Tue–Thu)
- Keynote (Tue)
- Panels (Tue–Fri)
- Papers (Tue–Thu)
- Posters Display (Tue–Thu)
- Posters Reception (Tue)
- Research Posters Presentations (Wed)
- Scientific Visualization & Data Analytics Showcase Presentations (Wed)
- Student Cluster Competition (Mon–Thu)
- Technical Program Reception (Thu)
- One copy of the SC19 Proceedings

## Tutorials (TUT)

- HPC Is Now Plenary (Mon)
- Keynote (Tue)
- Panels (Friday only)
- Tutorial Lunch (Sun and/or Mon)
- Tutorials (Sun and/or Mon)
- Electronic access to all tutorial notes

## Workshops (W)

- HPC Is Now Plenary (Mon)
- Keynote (Tue)
- Panels (Friday only)
- Workshops (Sun, Mon, Fri)

## Exhibitor 24-hour Access (EX)

- Awards Ceremony (Thu)
- Birds of a Feather (Tue–Thu)
- Exhibitor Forum (Tue–Thu)
- Exhibitor Reception (Sun)
- Exhibits (Tue–Thu)
- Grand Opening Gala Reception (Mon)
- HPC Impact Showcase (Tue–Thu)
- HPC Is Now Plenary (Mon)
- Keynote (Tue)
- Panels (Friday only)
- Posters Display (Tue–Thu)
- Student Cluster Competition (Mon–Thu)

## Exhibits Hall Only (EXH)

- Awards Ceremony (Thu)
- Birds of a Feather (Tue–Thu)
- Exhibitor Forum (Tue–Thu)
- Exhibits (Tue–Thu)
- HPC Impact Showcase (Tue–Thu)
- HPC Is Now Plenary (Mon)
- Keynote (Tue)
- Panels (Friday only)
- Posters Display (Tue–Thu)
- Student Cluster Competition (Mon–Thu)
# Table of Contents

- ACM Gordon Bell Finalist
- ACM Student Research Competition: Graduate Posters
- ACM Student Research Competition: Undergraduate Posters
- Awards Presentation
- Birds of a Feather
- Computing for Change
- Doctoral Showcase
- Early Career Program
- Exhibitor Forum
- Exhibits
- HPC Impact Showcase
- HPC Is Now Plenary
- Invited Talk
- Keynote
- Navigating SC19
- Panel
- Paper
- Reception
- Research Posters
- SC Theater
- SC20
- Scientific Visualization & Data Analytics Showcase
- SCinet
- Students@SC
- Student Cluster Competition
- Student Job/Opportunity Fair
- Technology Challenge
- Test of Time
- Tutorial
- Workshop
A Data-Centric Approach to Extreme-Scale Ab Initio Dissipative Quantum Transport Simulations
Alexandros Nikolaos Ziogas (ETH Zurich)

The computational efficiency of a state of the art ab initio quantum transport (QT) solver, capable of revealing the coupled electrothermal properties of atomically-resolved nano-transistors, has been improved by up to two orders of magnitude through a data centric reorganization of the application. The approach yields coarse- and fine-grained data-movement characteristics that can be used for performance and communication modeling, communication avoidance, and dataflow transformations. The resulting code has been tuned for two top-6 hybrid supercomputers, reaching a sustained performance of 85.45 Pflop/s on 4,560 nodes of Summit (42.55% of the peak) in double precision, and 90.89 Pflop/s in mixed precision. These computational achievements enable the restructured QT simulator to treat realistic nanoelectronic devices made of more than 10,000 atoms within a 14× shorter duration than the original code needs to handle a system with 1,000 atoms, on the same number of CPUs/GPUs and with the same physical accuracy.

Fast, Scalable and Accurate Finite-Element Based Ab Initio Calculations Using Mixed Precision Computing: 46 PFLOPS Simulation of a Metallic Dislocation System
Vikram Gavini (University of Michigan)

Accurate large-scale first principles calculations based on density functional theory (DFT) in metallic systems are prohibitively expensive due to the asymptotic cubic scaling computational complexity with number of electrons. Using algorithmic advances in employing finite-element discretization for DFT (DFT-FE) in conjunction with efficient computational methodologies and mixed precision strategies, we delay the onset of this cubic scaling by significantly reducing the computational prefactor while increasing the arithmetic intensity and lowering the data movement costs. This has enabled fast, accurate, and massively parallel DFT calculations on large-scale metallic systems on
both many-core and heterogeneous architectures, with time-to-solution being an order of magnitude faster than state-of-the-art plane-wave DFT codes. We demonstrate an unprecedented sustained performance of 46 PFLOPS (27.8% peak FP64 performance) on a dislocation system in Magnesium containing 105,080 electrons using 3,800 GPU nodes of Summit supercomputer, which is the highest performance to date among DFT codes.
ACM Student Research Competition: Graduate Posters

Tuesday, November 19
8:30 am - 5:00 pm

ACM Student Research Competition Posters Display

Poster 12: Investigating Input Sensitivity in GPU Graph Analytics with Machine Learning
Jacob M. Hope (Texas State University), Trisha Nag (Texas Tech University)

Graph algorithms are at the core of data-intensive applications. As such, efficient graph processing is of great importance. Irregularity in real-world graphs can make performance unpredictable and non-portable across different inputs and architectures. Given a type of graph, the same optimized implementation of an algorithm can produce performance numbers that differ by orders-of-magnitude. We conduct extensive analysis on a set of 1238 graphs to identify input-sensitive performance inefficiencies, including two that have not been studied: (i) register pressure and (ii) CPU-GPU data movement via demand paging. We then build a multiclass decision tree classifier that characterizes the irregular properties of graphs from our data and maps them to an optimal control parameterization at the compiler, system and algorithmic layers, that yield the highest overall algorithmic performance. We then integrate the classifier into a system where it will process a new graph and generate a kernel on the predicted optimal configuration.

Poster 13: Scalable Load-Balancing Concurrent Queues on Many-Core Architectures
Caleb Lehman (Ohio State University)

As the core counts of computing platforms continue to rise, parallel runtime systems with support for very fine-grained tasks become increasingly necessary to fully utilize the available resources. A critical feature of such task-based parallel runtime systems is the ability to balance work evenly and quickly across available cores. We highlight this by studying XTask, a custom parallel runtime system based on XQueue, which is a novel lock-less concurrent queuing system with relaxed
ordering semantics that is geared to realizing scalability to hundreds of concurrent threads. We
demonstrate the lack of adequate load balancing in the original XQueue design and present several
solutions for improving load balancing. We also evaluate the corresponding improvements in
performance on two sample workloads, computation of Fibonacci numbers and computation of
Cholesky factorization. Finally, we compare the performance of several versions of XTask along with
several implementations of the popular OpenMP runtime system.

Poster 6: Analyzing the Performance of ZFP Compressed Arrays on HPC Kernels
Pavlo D. Triantafyllides (Clemson University)

Per-node memory capacity limits the maximal problem size of HPC applications. Naïve data
compression alleviates the memory impact, but requires full decompression before the data is
accessed. ZFP compressed arrays reduce the memory footprint, by independently compressing
data in fixed sized blocks. Thus, decompressing individual blocks and caching them enables random
access and a reduction in decompressions on the critical path. The performance of ZFP compressed
arrays is dependent on several key variables: software cache size, cache policy, and compression
rate. In this poster, we explore the sensitivity of these ZFP parameters on runtime performance for
the matrix-matrix multiplication algorithm. Results show that selection of cache size, policy, and rate
yields 8% performance improvement over the default ZFP configuration.

Poster 1: Early Experiences on OpenPOWER Architecture: Analysis of Billion-Scale Atomistic
Datasets
Yuya Kawakami (Grinnell College, Oak Ridge National Laboratory)

We present performance analysis on OpenPOWER architecture of an algorithm to generate
transversal views of atomistic models. The algorithm was implemented with data parallel primitives
in NVIDIA Thrust for architecture portability. We report performance results on IBM Power9 CPUs
(OpenMP, Intel Threading Blocks) and NVIDIA Volta GPUs (single and multi GPU). We also evaluate
CUDA unified memory performance, exposed by NVIDIA RAPIDS Memory Manager library (RMM).

Poster 2: An Efficient Parallel Algorithm for Dominator Detection
Daniel Giger (University of Massachusetts, Lowell), Hang Liu (Stevens Institute of Technology)

In graph theory, a vertex v dominates a vertex u if every path from the entry vertex to u must go
through vertex v. This algorithm is called dominator detection and holds a wide range of
applications, such as compiler design, circuit testing, and social network analysis. While the
performance of many other graph algorithms soars with respect to the increase of the hardware parallelism, dominator detection algorithm experiences very little advancement due to the hardship of parallelism. This work thus introduces an efficient parallel dominator detection algorithm that is inspired by Breadth-First Search (BFS), which bests SEMI-NCA on large graphs.

Poster 9: Machine Specific Symbolic Code Generation
Robert King (University of Utah)

New discoveries in science and engineering are primarily driven by numerical simulations of underlying governing equations. Manually written codes to evaluate the main computational kernels lack portability, prone to human errors, ability to perform code optimizations due to the complexity of the underlying equations. In this work we present a symbolic code generation framework, which generates architecture optimized code for different platforms. As the driving application, we primarily use computational relativity where computations of Einstein equations become complicated due to the presence of curvature in spacetime.

The presented framework is based on SymPy with additional modules written to handle complicated partial differential equations (PDEs). The symbolically written equations are converted to a computational graph, which enables to perform architecture (cache, register optimizations) and language specific (SIMD vectorization, CUDA) optimizations. By computing common subexpressions, we reduce the number of compute operations needed, by storing them in temporary variables.

Poster 5: Evaluating Lossy Compressors for Inline Compression
Donald T. Elmore (Clemson University)

HPC applications require massive amounts of data. The data required is growing faster than memory capabilities. An example of this is pySDC, a framework for solving collocation problems iteratively using parallel-in-time methods. pySDC requires storing and exchange of 3D volume data for each parallel point in time. We evaluate several state-of-the-art lossy compressors such as SZ and ZFP for their applicability to inline compression for pySDC. We evaluate the compressors based on compression ratio, compression bandwidth, decompression bandwidth, and overall error introduced.

This poster evaluates state-of-the-art lossy compressors for use in pySDC; shows lossy data compression is an effective tool for reducing memory requirements for pySDC; and highlights current compression/decompression bandwidth is not fast enough for inline lossy compression yet.
Results show using SZ with an error bound of 1e-5, we reduce the memory footprint by a factor of 311.99 while maintaining an acceptable level of loss.

**Poster 3: A Deep Learning Approach to Noise Prediction and Circuit Optimization for Near-Term Quantum Devices**  
*Alexander Zlokapa (California Institute of Technology), Alexandru Gheorghiu (California Institute of Technology)*

Noisy intermediate-scale quantum (NISQ) devices face challenges in achieving high-fidelity computations due to hardware-specific noise. As a basis for noise mitigation, we develop a convolutional neural network noise model to estimate the difference in noise between a given pair of equivalent quantum circuits. On a classically simulated dataset of 1.6 million pairs of quantum circuits with a simplified noise model calibrated to IBM Q hardware, the deep learning approach shows a significant improvement in noise prediction over linear gate count models. A greedy peephole optimization procedure is proposed to minimize noise using the deep learning model as an objective function, showing further improvement in noise mitigation compared to commonly used gate count minimization heuristics.

**Poster 10: Learning Page Access Patterns for Algorithms Programmed with GPU UVM**  
*Bennett Cooper (Clemson University), Derek Rodriguez (Clemson University)*

We aim to mitigate the performance bottleneck of migrating data between host and device memory in GPU applications by accurately predicting application access patterns using deep neural networks. We model the memory access pattern of any given application by collecting page faults that trigger data migration to the GPU and feed this time series as input to a neural network that outputs the next several page faults. We evaluate these predictions on the basis of what makes a useful prefetch in our context for the GPU. Current work has looked at trivial GPU applications, such as matrix operations, and moving toward real, complex applications. Our work will be presented by bringing attention to the predictive capability of our neural network on the current applications tested.

**Poster 8: Mitigating Communication Bottlenecks in MPI-Based Distributed Learning**  
*Abdullah B. Nauman (Ward Melville High School, Brookhaven National Laboratory)*

Current commercial and scientific facilities generate and maintain vast amounts of complex data. While machine learning (ML) techniques can provide crucial insight, developing these models is
often impractical on a single process. Distributed learning techniques mitigate this problem; however, current models contain significant performance bottlenecks. Here, we conduct a detailed performance analysis of MPI_Learn, a widespread distributed ML framework for high-energy physics (HEP) applications, on the Summit supercomputer, by training a network to classify simulated collision events from high-energy particle detectors at the CERN Large Hadron Collider (LHC).

We conclude that these bottlenecks occur as a result of increasing communication time between the different processes, and to mitigate the bottlenecks we propose the implementation of a new distributed algorithm for stochastic gradient descent (SGD). We provide a proof of concept by demonstrating better scalability with results on 250 GPUs, and with hyperparameter optimization, show a ten-fold decrease in training time.

**Hearing Single- and Multi-Threaded Program Behavior**  
Mark Wissink (Calvin University), Joel Adams (Calvin University)

There are many examples of visual program representation, but there is an unexplored area of sonic representation—audialization—of programs. The Thread Safe Audio Library (TSAL) provides the necessary tools to create such program audializations. TSAL is a platform-independent, object-oriented C++ library that provides thread safe classes for synthesizing sound from a program. By adding TSAL calls to a sequential or parallel program, its behavior can be represented with sound. Audializations require a unique interpretation that may greatly benefit auditory learners and other programmers who are looking for a different way to profile a sequential or parallel program.

**Poster 11: Walking the Cost-Accuracy Tightrope: Balancing Trade-Offs in Data-Intensive Genomics**  
Kathryn Leung (Princeton University), Meghan Kimball (DePaul University)

Scientific applications often exhibit a trade-off between cost and accuracy. However, measuring and predicting cost and accuracy in a way that users can understand these trade-offs is challenging. To address these needs, we present predictive cost and accuracy models for data-intensive genomics applications. We use these models to create a trade-off graph, which researchers can use to selectively trade-off cost and accuracy.

**Poster 31: Designing High-Performance Erasure Coding Schemes for Next-Generation Storage Systems**
Erasure Coding (EC) NIC offload is a promising technology for designing next-generation distributed storage systems. To alleviate the EC overhead, various hardware-based Erasure Coding schemes have been proposed to leverage the advanced compute capabilities on modern HPC clusters. However, this poster has identified some critical limitations of current-generation EC NIC offload schemes on modern SmartNICs and the missing of exploiting heterogeneous hardware available on modern HPC clusters in parallel. To address these limitations, this poster first proposes a unified benchmark suite to benchmark, measure and characterize hardware-optimized erasure coders. Then, it proposes a Multi-Rail EC concept which enables upper-layer applications to leverage heterogeneous hardware to perform EC operations simultaneously. Finally, it proposes a new EC NIC offload paradigm based on the tripartite graph model, namely TriEC. TriEC supports both encode-and-send and receive-and-decode offload primitives efficiently.

Poster 27: A Framework for Measuring Hardware Gather-Scatter Support
Patrick Lavin (Georgia Institute of Technology), Jeffrey Young (Georgia Institute of Technology), Richard Vuduc (Georgia Institute of Technology)

In recent years, we have seen the re-addition of vector units to CPUs. While these units easily give speedups for easily vectorized applications with dense memory access, it can be hard to characterize how different access patterns will affect the performance of vectorized code.

We have developed Spatter, a benchmark which allows us to test and investigate the gather-scatter units available on current and upcoming hardware. The information that Spatter reveals to users is of use to everyone from hardware vendors who wish to compare gather-scatter units across platforms, to application developers who wish to test memory access pattern performance in their vectorized code.

Poster 30: A Fast and Efficient Incremental Approach toward Dynamic Community Detection
Neda Zarayeneh (Washington State University)

Community detection is a discovery tool to identify groups of vertices which are tightly connected, called communities. Most of the literature on this subject caters to the static use-case where the underlying network does not change. However, many emerging real-world use-cases give rise to a need to incorporate dynamic graphs as inputs.

We present a fast and efficient incremental approach toward dynamic community detection. The key
contribution is a generic technique called \( \Delta \)-screening, which examines the most recent batch of changes made to an input graph and selects a subset of vertices to reevaluate for potential community (re)assignment. This technique can be incorporated into any of the community detection methods that use modularity as its objective function for clustering. For demonstration purposes, we incorporated the technique into two well-known community detection tools. Our experiments demonstrate that our approach is able to generate performance speedups without compromising on the output quality.

**Poster 29: Optimal Routing for a Family of Scalable Interconnection Networks**  
Zhipeng Xu (Stony Brook University; Sun Yat-sen University, Guangzhou, China)

We propose a scheme to construct a family of large and high-performance interconnection networks that are scalable, low-radix, minimum diameters. These networks, whose diameters grow linearly as their sizes grow exponentially, are generated by using the Cartesian products of smaller optimal networks of minimum diameters. For the smaller base networks, we design the vertex-balanced routing algorithm by considering the forwarding pressure at each vertex. Comparative benchmarks on a Beowulf cluster show significant improvement in performance after using the new routing algorithm. Each node of the new network generated from base graphs with low-diameter can also sustain balanced forwarding loadings if we apply optimal routing algorithms to the base network. Simulation results for larger networks show that the optimal routing algorithms achieve the gain of communication performance.

**Poster 16: perf-taint: Taint Analysis for Automatic Many-Parameter Performance Modeling**  
Marcin Copik (ETH Zurich)

Performance modeling is a well-known technique for understanding the scaling behavior of an application. Although the modeling process is today often automatic, it still relies on a domain expert selecting program parameters and deciding relevant sampling intervals. Since existing empirical methods attempt blackbox modeling, the decision on which parameters influence a selected part of the program is based on measured data, making empirical modeling sensitive to human errors and instrumentation noise. We introduce a hybrid analysis to mitigate the current limitations of empirical modeling, combining the confidence of static analysis with the ability of dynamic taint analysis to capture the effects of control-flow and memory operations. We construct models of computation and communication volumes that help the modeler to remove effects of noise and improve the correctness of estimated models. Our automatic analysis prunes irrelevant program parameters and brings an understanding of parameter dependencies which helps in designing the experiment.
Poster 24: Fingerprinting Anomalous Computation with RNN for GPU-Accelerated HPC Machines
Pengfei Zou (Clemson University)

This paper presents a workload classification framework that discriminates illicit computation from authorized workloads on GPU-accelerated HPC systems. As such systems become more and more powerful, they are exploited by attackers to run malicious and for-profit programs that typically require extremely high computing ability to be successful. Our classification framework leverages the distinctive signatures between illicit and authorized workloads, and explore machine learning methods to learn the workloads and classify them. The framework uses lightweight, non-intrusive workload profiling to collect model input data, and explores multiple machine learning methods, particularly recurrent neural network (RNN) that is suitable for online anomalous workload detection. Evaluation results on four generations of GPU machines demonstrate that the workload classification framework can tell apart the illicit authorized workloads with high accuracy of over 95%. The collected dataset, detection framework, and neural network models will be made available on GitHub.

Poster 33: Enhancing Neural Architecture Search with Speciation and Inter-Epoch Crossover
Matthew E. Baughman (University of Chicago)

As deep learning continues to expand into new areas of application, the demand for efficient use of our HPC resources increase. For new problem domains, new model architectures are developed through a neural architecture search (NAS), which consist of iteratively training many neural networks. To combat the computational waste and maximize compute efficiency for NAS, we demonstrate that the use of genetic algorithms with speciation can be used to both shorten training time and increase accuracy at each iteration.

Poster 14: Data Reuse Analysis for GPU Offloading Using OpenMP
Alok Mishra (Stony Brook University, Lawrence Livermore National Laboratory)

More researchers and developers desire to port their applications to GPU-based clusters, due to their abundant parallelism and energy efficiency. Unfortunately porting or writing an application for accelerators, such as GPUs, requires extensive knowledge of the underlying architectures, the application/algorithm and the interfacing programming model (e.g. OpenMP). Often applications spend a significant portion of their execution time on data transfer. Exploiting data reuse opportunities in an application can reduce its overall execution time. In this research, we present an approach to automatically recognize data reuse opportunities in an application which uses OpenMP for exploiting
GPU parallelism, and consequently insert pertinent code to take advantage of data reuse on GPU. Using our approach we were able to retain reused data on the GPU and reduce the overall execution time of multiple benchmark application.

Poster 17: Exploiting Multi-Resource Scheduling for HPC
Yuping Fan (Illinois Institute of Technology)

High-performance computing (HPC) is undergoing significant changes. Next generation HPC systems are equipped with diverse global/local resources. HPC job scheduler plays a crucial role in efficient use of resources. However, traditional job schedulers are single-objective and fail to efficient use of other resources. In our previous work, we present a job scheduling framework named BBSched to schedule CPUs and burst buffers. As we are heading toward exascale computing, a variety of heterogeneous resources are deployed in HPC systems. In this poster, we extend BBSched for managing multiple resources beyond CPUs and burst buffers. We formulate multi-resource scheduling as a general multi-objective optimization (MOO) problem, present a heuristic method to solve the NP-hard MOO problem, and provide a preliminary evaluation for scheduling up to ten resources. The proposed multi-resource scheduling design is intended to enhance Cobalt, a production job scheduler deployed on HPC systems at Argonne Leadership Computing Facility (ALCF).

Poster 15: Cost-Aware Cloudlet Placement in Edge Computing Systems
Dixit Bhatta (University of Delaware)

A well-known challenge in Edge Computing is strategic placement of cloudlets. The fundamental goals of this challenge are to minimize the cloudlet deployment cost and to guarantee minimum latency to the users of edge services. We address this challenge by designing a cost-aware cloudlet placement approach that that ensures user latency requirements while covering all devices in the service region. We first mathematically formulate the problem as a multi-objective integer programming model in a general deployment scenario, which is computationally NP-hard. We then propose a genetic algorithm-based approach, GACP, to find heuristic solutions in significantly reduced time. We investigate the effectiveness of GACP by performing extensive experiments on multiple deployment scenarios based on New York City OpenData. The results presented in the poster show that our approach obtains close to optimal cost solutions with significant time reduction.

Poster 19: Accelerating Real-World Stencil Computations Using Temporal Blocking: Handling Sparse Sources and Receivers
George Bisbas (Imperial College, London)

This paper concerns performance optimization in finite-difference solvers found in seismic imaging. We tackle a complicating factor that is arising in seismic inversion problems: the addition of waves injected from sources distributed sparsely over the 2D/3D domain, and the need for receivers that interpolate data measurements at a set of points, again distributed sparsely across the domain. In this work we show how to overcome this limitation. We introduce an algorithm for a loop nest transformation policy applied to wave propagation in order to improve data locality and optimize our cache memory use. Our algorithm uses an inspector/executor scheme capable of inspecting measurement operations at sparse locations and then compute the field updates. The ultimate goal of this ongoing work is to automate this scheme for stencil codes generated by Devito.

Poster 23: PERQ: Fair and Efficient Power Management of Power-Constrained Large-Scale Computing Systems
Tirthak Patel (Northeastern University), Devesh Tiwari (Northeastern University)

Large-scale computing systems are becoming increasingly more power-constrained, but these systems employ hardware over-provisioning to achieve higher system throughput because HPC applications often do not consume the peak power capacity of nodes. Unfortunately, focusing on system throughput alone can lead to severe unfairness among multiple concurrently-running applications. This paper introduces PERQ, a new feedback-based principled approach to improve system throughput while achieving fairness among concurrent applications.

Poster 18: Using Non Volatile Memories to Build Energy- and Cost-Efficient Clusters
Onkar Patil (North Carolina State University)

Non-Volatile Memory (NVM) is a byte-addressable, high capacity, high latency, and persistent form of memory that can extend the primary memory hierarchy by another level. It allows clusters to have significantly greater memory capacity per node. Intel’s Optane DC Persistent Memory Module (PMM) is a NVM device that can be used to increase the memory density of high performance computing (HPC) system. With higher memory density, we can run scientific computing applications with larger problem sizes on fewer compute nodes than on current HPC systems. Smaller HPC clusters will reduce the cost of running scientific simulations. We test our hypothesis by running a HPC application with large problem sizes on a single node with NVM and on multiple nodes connected by a high speed interconnect with equivalent amount of volatile memory. We compare the performance characteristics and power consumption of both systems.
Poster 22: Fast Profiling-Based Performance Modeling of Distributed GPU Applications
Jaemin Choi (University of Illinois, Lawrence Livermore National Laboratory)

An increasing number of applications utilize GPUs to accelerate computation, with MPI responsible for communication in distributed environments. Existing performance models only focus on either modeling GPU kernels or MPI communication; few that do model the entire application are often too specialized for a single application and require extensive input from the programmer.

To be able to quickly model different types of distributed GPU applications, we propose a profiling-based methodology for creating performance models. We build upon the roofline performance model for GPU kernels and analytical models for MPI communication, with a significant reduction in profiling time. We also develop a benchmark to model 3D halo exchange that occurs in many scientific applications. Our proposed model for the main iteration loops of MiniFE achieves 6-7% prediction error on LLNL Lassen and 1-2% error on PSC Bridges, with minimal code inspection required to model MPI communication.

Poster 26: Neural Networks for the Benchmarking of Detection Algorithms
Silvia Miramontes (Lawrence Berkeley National Laboratory; University of California, Berkeley)

There are several automated methods to detect objects from grayscale images. However, materials scientists still lack basic tools to compare different detection results, particularly when working with microtomography. This poster introduces FibCAM, a convolutional neural network (CNN)-based method using TensorFlow that allows benchmarking fiber detection algorithms. Our contribution is three-fold: (a) the design of a computational framework to compare automated fiber detection models with curated datasets through classification; (b) lossless data reduction by embedding prior knowledge into data-driven models; (c) a scheme to decompose computation into embarrassingly parallel processes for future analysis at scale. Our results show how FibCAM classifies different structures, and how it illustrates the material's composition and frequency distribution of microstructures for improved interpretability of machine learning models. The proposed algorithms support probing the specimen content from gigabyte-sized volumes and enable pinpointing inconsistencies between real structures known a priori and results derived from automated detections.

Poster 21: Optimization of GPU Kernels for Sparse Matrix Computations in Hypre
Chaoyu Zhang (Arkansas State University)
The acceleration of sparse matrix computations on GPUs can significantly enhance the performance of iterative methods for solving linear systems. In this work, we consider the kernels of Sparse Matrix Vector Multiplications (SpMV), Sparse Triangular Matrix Solves (SpTrSv) and Sparse Matrix Matrix Multiplications (SpMM), which are often demanded by Algebraic Multigrid (AMG) solvers. With the CUDA and the hardware support of the Volta GPUs on Sierra, the existing kernels should be further optimized to fully take the advantage of the new hardware, and the optimizations have shown significant performance improvement. The presented kernels have been put in HYPRE for solving large scale linear systems on HPC equipped with GPUs. These shared-memory kernels for single GPU are the building blocks of distributed matrix operations required by the solver across multiple GPUs and compute nodes. The implementations of these kernels in Hypre and the code optimizations will be discussed.

**Poster 20: From IoT to Cloud: Research Platform for IoT/Cloud Experiments**

*Jinfeng Lin (University of Notre Dame)*

IoT studies leverage a wide range of lightweight hardware for collecting and processing data in the field. Limited by the resources on the devices, IoT systems have to interoperate with cloud platforms for addressing computation intensive tasks such as image processing, application backend supporting and centralized data storage. Therefore, a testbed for IoT/Cloud experiments should provide infrastructure for IoT to cloud communication, computation deployment, and hardware resource management. With these functionalities, users can focus on research problems without distraction from manually constructing experiment environments. Though cloud providers such as Google, Amazon, and Microsoft all provide IoT to Cloud solutions in general, this commercial model is not entirely compatible with research purpose. We propose a framework named Chameleon IoT testbed (CHIoT) that extends the Chameleon bare-metal cloud resources to support general IoT to Cloud experiments.

**Poster 32: OSU INAM: A Profiling and Visualization Tool for Scalable and In-Depth Analysis of High-Performance GPU-Enabled HPC Clusters**

*Pouya Kousha (Ohio State University)*

The lack of low-overhead and scalable monitoring tools have prevented a comprehensive study of efficiency and utilization of emerging NVLink-enabled GPU clusters. We address this by proposing and designing an in-depth, real-time analysis, profiling, and visualization tool for high-performance GPU-enabled clusters with NVLinks on the top of the OSU INAM. The proposed tool is capable of presenting a unified and holistic view of MPI-level and fabric level information for emerging NVLink-enabled high-performance GPU clusters. It also provides insights into the efficiency and utilization
of underlying interconnects for different communication patterns. We also designed a low overhead and scalable modules to discover the fabric topology and gather fabric metrics by using different levels of threading, bulk insertions and deletions for storage, and using parallel components for fabric discovery and port metric inquiry.

Zheng Miao (Clemson University)

High-Performance Computing systems must simultaneously address both resilience and power. In heterogeneous systems, the trade-offs between resilience and energy-efficiency are more complex for applications using both CPUs and GPUs. A deep understanding of the interplay among energy efficiency, resilience, and performance is required for heterogeneous systems to address them simultaneously.

In this work, we present a new framework for resilient and energy-efficient computing in GPU-accelerated systems. This framework supports partial or full redundancy and checkpointing for resilience, and provides users with flexible hardware resource selection, adjustable precision and power management to improve performance and energy-efficiency. We further perform CUDA-aware MPI to reduce resilience overhead, mainly in message communication between GPUs. Using CG as an example, we show that our framework provides about 40% time and 45% energy savings, comparing to simple extension of RedMPI, a redundancy based resilience framework for homogeneous CPU systems.

**Poster 25: Leveraging Smart Data Transfer and Metadirective in Adaptive Computing**
Anjia Wang (University of North Carolina, Charlotte; Lawrence Livermore National Laboratory)

In this work, we propose smart data transfer (SDT) along with extensions to metadirective and map constructs in OpenMP 5.0 to improve adaptive computing. The Smith-Waterman algorithm is used as an example, whose naïve implementation does not conduct data transfer efficiently. SDT is used to solve this issue with the following advantages: (1) SDT only transfers necessary data to GPU instead of the whole data set, resulting in 4.5x of speedup in our initial experiments. (2) Depending on computing vs. data transfer requirements of a program, SDT will transfer the output of each iteration from GPU to host either immediately or all together after the last GPU kernel call. (3) It supports large data exceeding GPU device memory’s size via data tiling. We propose to extend metadirective’s context selector to obtain similar improvement by enabling target enter/exit data and on-demand data access.
Poster 12: Investigating Input Sensitivity in GPU Graph Analytics with Machine Learning
Jacob M. Hope (Texas State University), Trisha Nag (Texas Tech University)

Graph algorithms are at the core of data-intensive applications. As such, efficient graph processing is of great importance. Irregularity in real-world graphs can make performance unpredictable and non-portable across different inputs and architectures. Given a type of graph, the same optimized implementation of an algorithm can produce performance numbers that differ by orders-of-magnitude. We conduct extensive analysis on a set of 1238 graphs to identify input-sensitive performance inefficiencies, including two that have not been studied: (i) register pressure and (ii) CPU-GPU data movement via demand paging. We then build a multiclass decision tree classifier that characterizes the irregular properties of graphs from our data and maps them to an optimal control parameterization at the compiler, system and algorithmic layers, that yield the highest overall algorithmic performance. We then integrate the classifier into a system where it will process a new graph and generate a kernel on the predicted optimal configuration.

Poster 13: Scalable Load-Balancing Concurrent Queues on Many-Core Architectures
Caleb Lehman (Ohio State University)

As the core counts of computing platforms continue to rise, parallel runtime systems with support for very fine-grained tasks become increasingly necessary to fully utilize the available resources. A critical feature of such task-based parallel runtime systems is the ability to balance work evenly and quickly across available cores. We highlight this by studying XTask, a custom parallel runtime system based on XQueue, which is a novel lock-less concurrent queuing system with relaxed ordering semantics that is geared to realizing scalability to hundreds of concurrent threads. We demonstrate the lack of adequate load balancing in the original XQueue design and present several
solutions for improving load balancing. We also evaluate the corresponding improvements in performance on two sample workloads, computation of Fibonacci numbers and computation of Cholesky factorization. Finally, we compare the performance of several versions of XTask along with several implementations of the popular OpenMP runtime system.

Poster 6: Analyzing the Performance of ZFP Compressed Arrays on HPC Kernels
Pavlo D. Triantafyllides (Clemson University)

Per-node memory capacity limits the maximal problem size of HPC applications. Naïve data compression alleviates the memory impact, but requires full decompression before the data is accessed. ZFP compressed arrays reduce the memory footprint, by independently compressing data in fixed sized blocks. Thus, decompressing individual blocks and caching them enables random access and a reduction in decompressions on the critical path. The performance of ZFP compressed arrays is dependent on several key variables: software cache size, cache policy, and compression rate. In this poster, we explore the sensitivity of these ZFP parameters on runtime performance for the matrix-matrix multiplication algorithm. Results show that selection of cache size, policy, and rate yields 8% performance improvement over the default ZFP configuration.

Poster 1: Early Experiences on OpenPOWER Architecture: Analysis of Billion-Scale Atomistic Datasets
Yuya Kawakami (Grinnell College, Oak Ridge National Laboratory)

We present performance analysis on OpenPOWER architecture of an algorithm to generate transversal views of atomistic models. The algorithm was implemented with data parallel primitives in NVIDIA Thrust for architecture portability. We report performance results on IBM Power9 CPUs (OpenMP, Intel Threading Blocks) and NVIDIA Volta GPUs (single and multi GPU). We also evaluate CUDA unified memory performance, exposed by NVIDIA RAPIDS Memory Manager library (RMM).

Poster 2: An Efficient Parallel Algorithm for Dominator Detection
Daniel Giger (University of Massachusetts, Lowell), Hang Liu (Stevens Institute of Technology)

In graph theory, a vertex v dominates a vertex u if every path from the entry vertex to u must go through vertex v. This algorithm is called dominator detection and holds a wide range of applications, such as compiler design, circuit testing, and social network analysis. While the performance of many other graph algorithms soars with respect to the increase of the hardware parallelism, dominator detection algorithm experiences very little advancement due to the hardship
of parallelism. This work thus introduces an efficient parallel dominator detection algorithm that is inspired by Breadth-First Search (BFS), which bests SEMI-NCA on large graphs.

**Poster 9: Machine Specific Symbolic Code Generation**  
Robert King (University of Utah)

New discoveries in science and engineering are primarily driven by numerical simulations of underlying governing equations. Manually written codes to evaluate the main computational kernels lack portability, prone to human errors, ability to perform code optimizations due to the complexity of the underlying equations. In this work we present a symbolic code generation framework, which generates architecture optimized code for different platforms. As the driving application, we primarily use computational relativity where computations of Einstein equations become complicated due to the presence of curvature in spacetime.

The presented framework is based on SymPy with additional modules written to handle complicated partial differential equations (PDEs). The symbolically written equations are converted to a computational graph, which enables to perform architecture (cache, register optimizations) and language specific (SIMD vectorization, CUDA) optimizations. By computing common subexpressions, we reduce the number of compute operations needed, by storing them in temporary variables.

**Poster 5: Evaluating Lossy Compressors for Inline Compression**  
Donald T. Elmore (Clemson University)

HPC applications require massive amounts of data. The data required is growing faster than memory capabilities. An example of this is pySDC, a framework for solving collocation problems iteratively using parallel-in-time methods. pySDC requires storing and exchange of 3D volume data for each parallel point in time. We evaluate several state-of-the-art lossy compressors such as SZ and ZFP for their applicability to inline compression for pySDC. We evaluate the compressors based on compression ratio, compression bandwidth, decompression bandwidth, and overall error introduced.

This poster evaluates state-of-the-art lossy compressors for use in pySDC; shows lossy data compression is an effective tool for reducing memory requirements for pySDC; and highlights current compression/decompression bandwidth is not fast enough for inline lossy compression yet. Results show using SZ with an error bound of 1e-5, we reduce the memory footprint by a factor of 311.99 while maintaining an acceptable level of loss.
**Poster 3: A Deep Learning Approach to Noise Prediction and Circuit Optimization for Near-Term Quantum Devices**  
Alexander Zlokapa (California Institute of Technology), Alexandru Gheorghiu (California Institute of Technology)

Noisy intermediate-scale quantum (NISQ) devices face challenges in achieving high-fidelity computations due to hardware-specific noise. As a basis for noise mitigation, we develop a convolutional neural network noise model to estimate the difference in noise between a given pair of equivalent quantum circuits. On a classically simulated dataset of 1.6 million pairs of quantum circuits with a simplified noise model calibrated to IBM Q hardware, the deep learning approach shows a significant improvement in noise prediction over linear gate count models. A greedy peephole optimization procedure is proposed to minimize noise using the deep learning model as an objective function, showing further improvement in noise mitigation compared to commonly used gate count minimization heuristics.

**Poster 10: Learning Page Access Patterns for Algorithms Programmed with GPU UVM**  
Bennett Cooper (Clemson University), Derek Rodriguez (Clemson University)

We aim to mitigate the performance bottleneck of migrating data between host and device memory in GPU applications by accurately predicting application access patterns using deep neural networks. We model the memory access pattern of any given application by collecting page faults that trigger data migration to the GPU and feed this time series as input to a neural network that outputs the next several page faults. We evaluate these predictions on the basis of what makes a useful prefetch in our context for the GPU. Current work has looked at trivial GPU applications, such as matrix operations, and moving toward real, complex applications. Our work will be presented by bringing attention to the predictive capability of our neural network on the current applications tested.

**Poster 8: Mitigating Communication Bottlenecks in MPI-Based Distributed Learning**  
Abdullah B. Nauman (Ward Melville High School, Brookhaven National Laboratory)

Current commercial and scientific facilities generate and maintain vast amounts of complex data. While machine learning (ML) techniques can provide crucial insight, developing these models is often impractical on a single process. Distributed learning techniques mitigate this problem; however, current models contain significant performance bottlenecks. Here, we conduct a detailed
performance analysis of MPI_Learn, a widespread distributed ML framework for high-energy physics (HEP) applications, on the Summit supercomputer, by training a network to classify simulated collision events from high-energy particle detectors at the CERN Large Hadron Collider (LHC).

We conclude that these bottlenecks occur as a result of increasing communication time between the different processes, and to mitigate the bottlenecks we propose the implementation of a new distributed algorithm for stochastic gradient descent (SGD). We provide a proof of concept by demonstrating better scalability with results on 250 GPUs, and with hyperparameter optimization, show a ten-fold decrease in training time.

Hearing Single- and Multi-Threaded Program Behavior
Mark Wissink (Calvin University), Joel Adams (Calvin University)

There are many examples of visual program representation, but there is an unexplored area of sonic representation—audialization—of programs. The Thread Safe Audio Library (TSAL) provides the necessary tools to create such program audializations. TSAL is a platform-independent, object-oriented C++ library that provides thread safe classes for synthesizing sound from a program. By adding TSAL calls to a sequential or parallel program, its behavior can be represented with sound. Audializations require a unique interpretation that may greatly benefit auditory learners and other programmers who are looking for a different way to profile a sequential or parallel program.

Poster 11: Walking the Cost-Accuracy Tightrope: Balancing Trade-Offs in Data-Intensive Genomics
Kathryn Leung (Princeton University), Meghan Kimball (DePaul University)

Scientific applications often exhibit a trade-off between cost and accuracy. However, measuring and predicting cost and accuracy in a way that users can understand these trade-offs is challenging. To address these needs, we present predictive cost and accuracy models for data-intensive genomics applications. We use these models to create a trade-off graph, which researchers can use to selectively trade-off cost and accuracy.

Poster 31: Designing High-Performance Erasure Coding Schemes for Next-Generation Storage Systems
Haiyang Shi (Ohio State University), Xiaoyi Lu (Ohio State University)
Erasure Coding (EC) NIC offload is a promising technology for designing next-generation distributed storage systems. To alleviate the EC overhead, various hardware-based Erasure Coding schemes have been proposed to leverage the advanced compute capabilities on modern HPC clusters. However, this poster has identified some critical limitations of current-generation EC NIC offload schemes on modern SmartNICs and the missing of exploiting heterogeneous hardware available on modern HPC clusters in parallel. To address these limitations, this poster first proposes a unified benchmark suite to benchmark, measure and characterize hardware-optimized erasure coders. Then, it proposes a Multi-Rail EC concept which enables upper-layer applications to leverage heterogeneous hardware to perform EC operations simultaneously. Finally, it proposes a new EC NIC offload paradigm based on the tripartite graph model, namely TriEC. TriEC supports both encode-and-send and receive-and-decode offload primitives efficiently.

Poster 27: A Framework for Measuring Hardware Gather-Scatter Support
Patrick Lavin (Georgia Institute of Technology), Jeffrey Young (Georgia Institute of Technology), Richard Vuduc (Georgia Institute of Technology)

In recent years, we have seen the re-addition of vector units to CPUs. While these units easily give speedups for easily vectorized applications with dense memory access, it can be hard to characterize how different access patterns will affect the performance of vectorized code.

We have developed Spatter, a benchmark which allows us to test and investigate the gather-scatter units available on current and upcoming hardware. The information that Spatter reveals to users is of use to everyone from hardware vendors who wish to compare gather-scatter units across platforms, to application developers who wish to test memory access pattern performance in their vectorized code.

Poster 30: A Fast and Efficient Incremental Approach toward Dynamic Community Detection
Neda Zarayeneh (Washington State University)

Community detection is a discovery tool to identify groups of vertices which are tightly connected, called communities. Most of the literature on this subject caters to the static use-case where the underlying network does not change. However, many emerging real-world use-cases give rise to a need to incorporate dynamic graphs as inputs.

We present a fast and efficient incremental approach toward dynamic community detection. The key contribution is a generic technique called Ω-screening, which examines the most recent batch of changes made to an input graph and selects a subset of vertices to reevaluate for potential
community (re)assignment. This technique can be incorporated into any of the community detection methods that use modularity as its objective function for clustering. For demonstration purposes, we incorporated the technique into two well-known community detection tools. Our experiments demonstrate that our approach is able to generate performance speedups without compromising on the output quality.

**Poster 29: Optimal Routing for a Family of Scalable Interconnection Networks**  
Zhipeng Xu (Stony Brook University; Sun Yat-sen University, Guangzhou, China)

We propose a scheme to construct a family of large and high-performance interconnection networks that are scalable, low-radix, minimum diameters. These networks, whose diameters grow linearly as their sizes grow exponentially, are generated by using the Cartesian products of smaller optimal networks of minimum diameters. For the smaller base networks, we design the vertex-balanced routing algorithm by considering the forwarding pressure at each vertex. Comparative benchmarks on a Beowulf cluster show significant improvement in performance after using the new routing algorithm. Each node of the new network generated from base graphs with low-diameter can also sustain balanced forwarding loadings if we apply optimal routing algorithms to the base network. Simulation results for larger networks show that the optimal routing algorithms achieve the gain of communication performance.

**Poster 16: perf-taint: Taint Analysis for Automatic Many-Parameter Performance Modeling**  
Marcin Copik (ETH Zurich)

Performance modeling is a well-known technique for understanding the scaling behavior of an application. Although the modeling process is today often automatic, it still relies on a domain expert selecting program parameters and deciding relevant sampling intervals. Since existing empirical methods attempt blackbox modeling, the decision on which parameters influence a selected part of the program is based on measured data, making empirical modeling sensitive to human errors and instrumentation noise. We introduce a hybrid analysis to mitigate the current limitations of empirical modeling, combining the confidence of static analysis with the ability of dynamic taint analysis to capture the effects of control-flow and memory operations. We construct models of computation and communication volumes that help the modeler to remove effects of noise and improve the correctness of estimated models. Our automatic analysis prunes irrelevant program parameters and brings an understanding of parameter dependencies which helps in designing the experiment.

**Poster 24: Fingerprinting Anomalous Computation with RNN for GPU-Accelerated HPC**
This paper presents a workload classification framework that discriminates illicit computation from authorized workloads on GPU-accelerated HPC systems. As such systems become more and more powerful, they are exploited by attackers to run malicious and for-profit programs that typically require extremely high computing ability to be successful. Our classification framework leverages the distinctive signatures between illicit and authorized workloads, and explore machine learning methods to learn the workloads and classify them. The framework uses lightweight, non-intrusive workload profiling to collect model input data, and explores multiple machine learning methods, particularly recurrent neural network (RNN) that is suitable for online anomalous workload detection. Evaluation results on four generations of GPU machines demonstrate that the workload classification framework can tell apart the illicit authorized workloads with high accuracy of over 95%. The collected dataset, detection framework, and neural network models will be made available on GitHub.

**Poster 33: Enhancing Neural Architecture Search with Speciation and Inter-Epoch Crossover**

Matthew E. Baughman (University of Chicago)

As deep learning continues to expand into new areas of application, the demand for efficient use of our HPC resources increase. For new problem domains, new model architectures are developed through a neural architecture search (NAS), which consist of iteratively training many neural networks. To combat the computational waste and maximize compute efficiency for NAS, we demonstrate that the use of genetic algorithms with speciation can be used to both shorten training time and increase accuracy at each iteration.

**Poster 14: Data Reuse Analysis for GPU Offloading Using OpenMP**

Alok Mishra (Stony Brook University, Lawrence Livermore National Laboratory)

More researchers and developers desire to port their applications to GPU-based clusters, due to their abundant parallelism and energy efficiency. Unfortunately porting or writing an application for accelerators, such as GPUs, requires extensive knowledge of the underlying architectures, the application/algorithm and the interfacing programming model (e.g. OpenMP). Often applications spend a significant portion of their execution time on data transfer. Exploiting data reuse opportunities in an application can reduce its overall execution time. In this research, we present an approach to automatically recognize data reuse opportunities in an application which uses OpenMP for exploiting GPU parallelism, and consequently insert pertinent code to take advantage of data reuse on GPU. Using our approach we were able to retain reused data on the GPU and reduce the overall execution
time of multiple benchmark application.

**Poster 17: Exploiting Multi-Resource Scheduling for HPC**
Yuping Fan (Illinois Institute of Technology)

High-performance computing (HPC) is undergoing significant changes. Next generation HPC systems are equipped with diverse global/local resources. HPC job scheduler plays a crucial role in efficient use of resources. However, traditional job schedulers are single-objective and fail to efficiently use other resources. In our previous work, we present a job scheduling framework named BBSched to schedule CPUs and burst buffers. As we are heading toward exascale computing, a variety of heterogeneous resources are deployed in HPC systems. In this poster, we extend BBSched for managing multiple resources beyond CPUs and burst buffers. We formulate multi-resource scheduling as a general multi-objective optimization (MOO) problem, present a heuristic method to solve the NP-hard MOO problem, and provide a preliminary evaluation for scheduling up to ten resources. The proposed multi-resource scheduling design is intended to enhance Cobalt, a production job scheduler deployed on HPC systems at Argonne Leadership Computing Facility (ALCF).

**Poster 15: Cost-Aware Cloudlet Placement in Edge Computing Systems**
Dixit Bhatta (University of Delaware)

A well-known challenge in Edge Computing is strategic placement of cloudlets. The fundamental goals of this challenge are to minimize the cloudlet deployment cost and to guarantee minimum latency to the users of edge services. We address this challenge by designing a cost-aware cloudlet placement approach that ensures user latency requirements while covering all devices in the service region. We first mathematically formulate the problem as a multi-objective integer programming model in a general deployment scenario, which is computationally NP-hard. We then propose a genetic algorithm-based approach, GACP, to find heuristic solutions in significantly reduced time. We investigate the effectiveness of GACP by performing extensive experiments on multiple deployment scenarios based on New York City OpenData. The results presented in the poster show that our approach obtains close to optimal cost solutions with significant time reduction.

**Poster 19: Accelerating Real-World Stencil Computations Using Temporal Blocking: Handling Sparse Sources and Receivers**
George Bisbas (Imperial College, London)
This paper concerns performance optimization in finite-difference solvers found in seismic imaging. We tackle a complicating factor that is arising in seismic inversion problems: the addition of waves injected from sources distributed sparsely over the 2D/3D domain, and the need for receivers that interpolate data measurements at a set of points, again distributed sparsely across the domain. In this work we show how to overcome this limitation. We introduce an algorithm for a loop nest transformation policy applied to wave propagation in order to improve data locality and optimize our cache memory use. Our algorithm uses an inspector/executor scheme capable of inspecting measurement operations at sparse locations and then compute the field updates. The ultimate goal of this ongoing work is to automate this scheme for stencil codes generated by Devito.

Poster 23: PERQ: Fair and Efficient Power Management of Power-Constrained Large-Scale Computing Systems
Tirthak Patel (Northeastern University), Devesh Tiwari (Northeastern University)

Large-scale computing systems are becoming increasingly more power-constrained, but these systems employ hardware over-provisioning to achieve higher system throughput because HPC applications often do not consume the peak power capacity of nodes. Unfortunately, focusing on system throughput alone can lead to severe unfairness among multiple concurrently-running applications. This paper introduces PERQ, a new feedback-based principled approach to improve system throughput while achieving fairness among concurrent applications.

Poster 18: Using Non Volatile Memories to Build Energy- and Cost-Efficient Clusters
Onkar Patil (North Carolina State University)

Non-Volatile Memory (NVM) is a byte-addressable, high capacity, high latency, and persistent form of memory that can extend the primary memory hierarchy by another level. It allows clusters to have significantly greater memory capacity per node. Intel’s Optane DC Persistent Memory Module (PMM) is a NVM device that can be used to increase the memory density of high performance computing (HPC) system. With higher memory density, we can run scientific computing applications with larger problem sizes on fewer compute nodes than on current HPC systems. Smaller HPC clusters will reduce the cost of running scientific simulations. We test our hypothesis by running a HPC application with large problem sizes on a single node with NVM and on multiple nodes connected by a high speed interconnect with equivalent amount of volatile memory. We compare the performance characteristics and power consumption of both systems.

Poster 22: Fast Profiling-Based Performance Modeling of Distributed GPU Applications
An increasing number of applications utilize GPUs to accelerate computation, with MPI responsible for communication in distributed environments. Existing performance models only focus on either modeling GPU kernels or MPI communication; few that do model the entire application are often too specialized for a single application and require extensive input from the programmer.

To be able to quickly model different types of distributed GPU applications, we propose a profiling-based methodology for creating performance models. We build upon the roofline performance model for GPU kernels and analytical models for MPI communication, with a significant reduction in profiling time. We also develop a benchmark to model 3D halo exchange that occurs in many scientific applications. Our proposed model for the main iteration loops of MiniFE achieves 6-7% prediction error on LLNL Lassen and 1-2% error on PSC Bridges, with minimal code inspection required to model MPI communication.

Poster 26: Neural Networks for the Benchmarking of Detection Algorithms
Silvia Miramontes (Lawrence Berkeley National Laboratory; University of California, Berkeley)

There are several automated methods to detect objects from grayscale images. However, materials scientists still lack basic tools to compare different detection results, particularly when working with microtomography. This poster introduces FibCAM, a convolutional neural network (CNN)-based method using TensorFlow that allows benchmarking fiber detection algorithms. Our contribution is three-fold: (a) the design of a computational framework to compare automated fiber detection models with curated datasets through classification; (b) lossless data reduction by embedding prior knowledge into data-driven models; (c) a scheme to decompose computation into embarrassingly parallel processes for future analysis at scale. Our results show how FibCAM classifies different structures, and how it illustrates the material’s composition and frequency distribution of microstructures for improved interpretability of machine learning models. The proposed algorithms support probing the specimen content from gigabyte-sized volumes and enable pinpointing inconsistencies between real structures known a priori and results derived from automated detections.

Poster 21: Optimization of GPU Kernels for Sparse Matrix Computations in Hypre
Chaoyu Zhang (Arkansas State University)

The acceleration of sparse matrix computations on GPUs can significantly enhance the performance of iterative methods for solving linear systems. In this work, we consider the kernels of Sparse Matrix
Vector Multiplications (SpMV), Sparse Triangular Matrix Solves (SpTrSv) and Sparse Matrix Matrix Multiplications (SpMM), which are often demanded by Algebraic Multigrid (AMG) solvers. With the CUDA and the hardware support of the Volta GPUs on Sierra, the existing kernels should be further optimized to fully take the advantage of the new hardware, and the optimizations have shown significant performance improvement. The presented kernels have been put in HYPRE for solving large scale linear systems on HPC equipped with GPUs. These shared-memory kernels for single GPU are the building blocks of distributed matrix operations required by the solver across multiple GPUs and compute nodes. The implementations of these kernels in Hypre and the code optimizations will be discussed.

Poster 20: From IoT to Cloud: Research Platform for IoT/Cloud Experiments
Jinfeng Lin (University of Notre Dame)

IoT studies leverage a wide range of lightweight hardware for collecting and processing data in the field. Limited by the resources on the devices, IoT systems have to interoperate with cloud platforms for addressing computation intensive tasks such as image processing, application backend supporting and centralized data storage. Therefore, a testbed for IoT/Cloud experiments should provide infrastructure for IoT to cloud communication, computation deployment, and hardware resource management. With these functionalities, users can focus on research problems without distraction from manually constructing experiment environments. Though cloud providers such as Google, Amazon, and Microsoft all provide IoT to Cloud solutions in general, this commercial model is not entirely compatible with research purpose. We propose a framework named Chameleon IoT testbed (CHIoT) that extends the Chameleon bare-metal cloud resources to support general IoT to Cloud experiments.

Poster 32: OSU INAM: A Profiling and Visualization Tool for Scalable and In-Depth Analysis of High-Performance GPU-Enabled HPC Clusters
Pouya Kousha (Ohio State University)

The lack of low-overhead and scalable monitoring tools have prevented a comprehensive study of efficiency and utilization of emerging NVLink-enabled GPU clusters. We address this by proposing and designing an in-depth, real-time analysis, profiling, and visualization tool for high-performance GPU-enabled clusters with NVLinks on the top of the OSU INAM. The proposed tool is capable of presenting a unified and holistic view of MPI-level and fabric level information for emerging NVLink-enabled high-performance GPU clusters. It also provides insights into the efficiency and utilization of underlying interconnects for different communication patterns. We also designed a low overhead and scalable modules to discover the fabric topology and gather fabric metrics by using different
levels of threading, bulk insertions and deletions for storage, and using parallel components for fabric discovery and port metric inquiry.

Zheng Miao (Clemson University)

High-Performance Computing systems must simultaneously address both resilience and power. In heterogeneous systems, the trade-offs between resilience and energy-efficiency are more complex for applications using both CPUs and GPUs. A deep understanding of the interplay among energy efficiency, resilience, and performance is required for heterogeneous systems to address them simultaneously.

In this work, we present a new framework for resilient and energy-efficient computing in GPU-accelerated systems. This framework supports partial or full redundancy and checkpointing for resilience, and provides users with flexible hardware resource selection, adjustable precision and power management to improve performance and energy-efficiency. We further perform CUDA-aware MPI to reduce resilience overhead, mainly in message communication between GPUs. Using CG as an example, we show that our framework provides about 40% time and 45% energy savings, comparing to simple extension of RedMPI, a redundancy based resilience framework for homogeneous CPU systems.

**Poster 25: Leveraging Smart Data Transfer and Metadirective in Adaptive Computing**  
Anjia Wang (University of North Carolina, Charlotte; Lawrence Livermore National Laboratory)

In this work, we propose smart data transfer (SDT) along with extensions to metadirective and map constructs in OpenMP 5.0 to improve adaptive computing. The Smith-Waterman algorithm is used as an example, whose naïve implementation does not conduct data transfer efficiently. SDT is used to solve this issue with the following advantages: (1) SDT only transfers necessary data to GPU instead of the whole data set, resulting in 4.5x of speedup in our initial experiments. (2) Depending on computing vs. data transfer requirements of a program, SDT will transfer the output of each iteration from GPU to host either immediately or all together after the last GPU kernel call. (3) It supports large data exceeding GPU device memory’s size via data tiling. We propose to extend metadirective’s context selector to obtain similar improvement by enabling target enter/exit data and on-demand data access.
Poster 12: Investigating Input Sensitivity in GPU Graph Analytics with Machine Learning
Jacob M. Hope (Texas State University), Trisha Nag (Texas Tech University)

Graph algorithms are at the core of data-intensive applications. As such, efficient graph processing is of great importance. Irregularity in real-world graphs can make performance unpredictable and non-portable across different inputs and architectures. Given a type of graph, the same optimized implementation of an algorithm can produce performance numbers that differ by orders-of-magnitude. We conduct extensive analysis on a set of 1238 graphs to identify input-sensitive performance inefficiencies, including two that have not been studied: (i) register pressure and (ii) CPU-GPU data movement via demand paging. We then build a multiclass decision tree classifier that characterizes the irregular properties of graphs from our data and maps them to an optimal control parameterization at the compiler, system and algorithmic layers, that yield the highest overall algorithmic performance. We then integrate the classifier into a system where it will process a new graph and generate a kernel on the predicted optimal configuration.

Poster 13: Scalable Load-Balancing Concurrent Queues on Many-Core Architectures
Caleb Lehman (Ohio State University)

As the core counts of computing platforms continue to rise, parallel runtime systems with support for very fine-grained tasks become increasingly necessary to fully utilize the available resources. A critical feature of such task-based parallel runtime systems is the ability to balance work evenly and quickly across available cores. We highlight this by studying XTask, a custom parallel runtime system based on XQueue, which is a novel lock-less concurrent queuing system with relaxed ordering semantics that is geared to realizing scalability to hundreds of concurrent threads. We demonstrate the lack of adequate load balancing in the original XQueue design and present several solutions for improving load balancing. We also evaluate the corresponding improvements in
performance on two sample workloads, computation of Fibonacci numbers and computation of Cholesky factorization. Finally, we compare the performance of several versions of XTask along with several implementations of the popular OpenMP runtime system.

Poster 6: Analyzing the Performance of ZFP Compressed Arrays on HPC Kernels
Pavlo D. Triantafyllides (Clemson University)

Per-node memory capacity limits the maximal problem size of HPC applications. Naïve data compression alleviates the memory impact, but requires full decompression before the data is accessed. ZFP compressed arrays reduce the memory footprint, by independently compressing data in fixed sized blocks. Thus, decompressing individual blocks and caching them enables random access and a reduction in decompressions on the critical path. The performance of ZFP compressed arrays is dependent on several key variables: software cache size, cache policy, and compression rate. In this poster, we explore the sensitivity of these ZFP parameters on runtime performance for the matrix-matrix multiplication algorithm. Results show that selection of cache size, policy, and rate yields 8% performance improvement over the default ZFP configuration.

Poster 1: Early Experiences on OpenPOWER Architecture: Analysis of Billion-Scale Atomistic Datasets
Yuya Kawakami (Grinnell College, Oak Ridge National Laboratory)

We present performance analysis on OpenPOWER architecture of an algorithm to generate transversal views of atomistic models. The algorithm was implemented with data parallel primitives in NVIDIA Thrust for architecture portability. We report performance results on IBM Power9 CPUs (OpenMP, Intel Threading Blocks) and NVIDIA Volta GPUs (single and multi GPU). We also evaluate CUDA unified memory performance, exposed by NVIDIA RAPIDS Memory Manager library (RMM).

Poster 2: An Efficient Parallel Algorithm for Dominator Detection
Daniel Giger (University of Massachusetts, Lowell), Hang Liu (Stevens Institute of Technology)

In graph theory, a vertex v dominates a vertex u if every path from the entry vertex to u must go through vertex v. This algorithm is called dominator detection and holds a wide range of applications, such as compiler design, circuit testing, and social network analysis. While the performance of many other graph algorithms soars with respect to the increase of the hardware parallelism, dominator detection algorithm experiences very little advancement due to the hardship of parallelism. This work thus introduces an efficient parallel dominator detection algorithm that is
inspired by Breadth-First Search (BFS), which bests SEMI-NCA on large graphs.

**Poster 9: Machine Specific Symbolic Code Generation**  
Robert King (University of Utah)

New discoveries in science and engineering are primarily driven by numerical simulations of underlying governing equations. Manually written codes to evaluate the main computational kernels lack portability, prone to human errors, ability to perform code optimizations due to the complexity of the underlying equations. In this work we present a symbolic code generation framework, which generates architecture optimized code for different platforms. As the driving application, we primarily use computational relativity where computations of Einstein equations become complicated due to the presence of curvature in spacetime.

The presented framework is based on SymPy with additional modules written to handle complicated partial differential equations (PDEs). The symbolically written equations are converted to a computational graph, which enables to perform architecture (cache, register optimizations) and language specific (SIMD vectorization, CUDA) optimizations. By computing common subexpressions, we reduce the number of compute operations needed, by storing them in temporary variables.

**Poster 5: Evaluating Lossy Compressors for Inline Compression**  
Donald T. Elmore (Clemson University)

HPC applications require massive amounts of data. The data required is growing faster than memory capabilities. An example of this is pySDC, a framework for solving collocation problems iteratively using parallel-in-time methods. pySDC requires storing and exchange of 3D volume data for each parallel point in time. We evaluate several state-of-the-art lossy compressors such as SZ and ZFP for their applicability to inline compression for pySDC. We evaluate the compressors based on compression ratio, compression bandwidth, decompression bandwidth, and overall error introduced.

This poster evaluates state-of-the-art lossy compressors for use in pySDC; shows lossy data compression is an effective tool for reducing memory requirements for pySDC; and highlights current compression/decompression bandwidth is not fast enough for inline lossy compression yet. Results show using SZ with an error bound of 1e-5, we reduce the memory footprint by a factor of 311.99 while maintaining an acceptable level of loss.
Poster 3: A Deep Learning Approach to Noise Prediction and Circuit Optimization for Near-Term Quantum Devices
Alexander Zlokapa (California Institute of Technology), Alexandru Gheorghiu (California Institute of Technology)

Noisy intermediate-scale quantum (NISQ) devices face challenges in achieving high-fidelity computations due to hardware-specific noise. As a basis for noise mitigation, we develop a convolutional neural network noise model to estimate the difference in noise between a given pair of equivalent quantum circuits. On a classically simulated dataset of 1.6 million pairs of quantum circuits with a simplified noise model calibrated to IBM Q hardware, the deep learning approach shows a significant improvement in noise prediction over linear gate count models. A greedy peephole optimization procedure is proposed to minimize noise using the deep learning model as an objective function, showing further improvement in noise mitigation compared to commonly used gate count minimization heuristics.

Poster 10: Learning Page Access Patterns for Algorithms Programmed with GPU UVM
Bennett Cooper (Clemson University), Derek Rodriguez (Clemson University)

We aim to mitigate the performance bottleneck of migrating data between host and device memory in GPU applications by accurately predicting application access patterns using deep neural networks. We model the memory access pattern of any given application by collecting page faults that trigger data migration to the GPU and feed this time series as input to a neural network that outputs the next several page faults. We evaluate these predictions on the basis of what makes a useful prefetch in our context for the GPU. Current work has looked at trivial GPU applications, such as matrix operations, and moving toward real, complex applications. Our work will be presented by bringing attention to the predictive capability of our neural network on the current applications tested.

Poster 8: Mitigating Communication Bottlenecks in MPI-Based Distributed Learning
Abdullah B. Nauman (Ward Melville High School, Brookhaven National Laboratory)

Current commercial and scientific facilities generate and maintain vast amounts of complex data. While machine learning (ML) techniques can provide crucial insight, developing these models is often impractical on a single process. Distributed learning techniques mitigate this problem; however, current models contain significant performance bottlenecks. Here, we conduct a detailed performance analysis of MPI_Learn, a widespread distributed ML framework for high-energy
physics (HEP) applications, on the Summit supercomputer, by training a network to classify simulated collision events from high-energy particle detectors at the CERN Large Hadron Collider (LHC).

We conclude that these bottlenecks occur as a result of increasing communication time between the different processes, and to mitigate the bottlenecks we propose the implementation of a new distributed algorithm for stochastic gradient descent (SGD). We provide a proof of concept by demonstrating better scalability with results on 250 GPUs, and with hyperparameter optimization, show a ten-fold decrease in training time.

**Hearing Single- and Multi-Threaded Program Behavior**  
*Mark Wissink (Calvin University), Joel Adams (Calvin University)*

There are many examples of visual program representation, but there is an unexplored area of sonic representation—audialization—of programs. The Thread Safe Audio Library (TSAL) provides the necessary tools to create such program audializations. TSAL is a platform-independent, object-oriented C++ library that provides thread safe classes for synthesizing sound from a program. By adding TSAL calls to a sequential or parallel program, its behavior can be represented with sound. Audializations require a unique interpretation that may greatly benefit auditory learners and other programmers who are looking for a different way to profile a sequential or parallel program.

**Poster 11: Walking the Cost-Accuracy Tightrope: Balancing Trade-Offs in Data-Intensive Genomics**  
*Kathryn Leung (Princeton University), Meghan Kimball (DePaul University)*

Scientific applications often exhibit a trade-off between cost and accuracy. However, measuring and predicting cost and accuracy in a way that users can understand these trade-offs is challenging. To address these needs, we present predictive cost and accuracy models for data-intensive genomics applications. We use these models to create a trade-off graph, which researchers can use to selectively trade-off cost and accuracy.

**Poster 31: Designing High-Performance Erasure Coding Schemes for Next-Generation Storage Systems**  
*Haiyang Shi (Ohio State University), Xiaoyi Lu (Ohio State University)*

Erasure Coding (EC) NIC offload is a promising technology for designing next-generation distributed
storage systems. To alleviate the EC overhead, various hardware-based Erasure Coding schemes have been proposed to leverage the advanced compute capabilities on modern HPC clusters. However, this poster has identified some critical limitations of current-generation EC NIC offload schemes on modern SmartNICs and the missing of exploiting heterogeneous hardware available on modern HPC clusters in parallel. To address these limitations, this poster first proposes a unified benchmark suite to benchmark, measure and characterize hardware-optimized erasure coders. Then, it proposes a Multi-Rail EC concept which enables upper-layer applications to leverage heterogeneous hardware to perform EC operations simultaneously. Finally, it proposes a new EC NIC offload paradigm based on the tripartite graph model, namely TriEC. TriEC supports both encode-and-send and receive-and-decode offload primitives efficiently.

Poster 27: A Framework for Measuring Hardware Gather-Scatter Support
Patrick Lavin (Georgia Institute of Technology), Jeffrey Young (Georgia Institute of Technology), Richard Vuduc (Georgia Institute of Technology)

In recent years, we have seen the re-addition of vector units to CPUs. While these units easily give speedups for easily vectorized applications with dense memory access, it can be hard to characterize how different access patterns will affect the performance of vectorized code.

We have developed Spatter, a benchmark which allows us to test and investigate the gather-scatter units available on current and upcoming hardware. The information that Spatter reveals to users is of use to everyone from hardware vendors who wish to compare gather-scatter units across platforms, to application developers who wish to test memory access pattern performance in their vectorized code.

Poster 30: A Fast and Efficient Incremental Approach toward Dynamic Community Detection
Neda Zarayeneh (Washington State University)

Community detection is a discovery tool to identify groups of vertices which are tightly connected, called communities. Most of the literature on this subject caters to the static use-case where the underlying network does not change. However, many emerging real-world use-cases give rise to a need to incorporate dynamic graphs as inputs.

We present a fast and efficient incremental approach toward dynamic community detection. The key contribution is a generic technique called $\Omega$-screening, which examines the most recent batch of changes made to an input graph and selects a subset of vertices to reevaluate for potential community (re)assignment. This technique can be incorporated into any of the community detection
methods that use modularity as its objective function for clustering. For demonstration purposes, we incorporated the technique into two well-known community detection tools. Our experiments demonstrate that our approach is able to generate performance speedups without compromising on the output quality.

**Poster 29: Optimal Routing for a Family of Scalable Interconnection Networks**  
Zhipeng Xu (Stony Brook University; Sun Yat-sen University, Guangzhou, China)

We propose a scheme to construct a family of large and high-performance interconnection networks that are scalable, low-radix, minimum diameters. These networks, whose diameters grow linearly as their sizes grow exponentially, are generated by using the Cartesian products of smaller optimal networks of minimum diameters. For the smaller base networks, we design the vertex-balanced routing algorithm by considering the forwarding pressure at each vertex. Comparative benchmarks on a Beowulf cluster show significant improvement in performance after using the new routing algorithm. Each node of the new network generated from base graphs with low-diameter can also sustain balanced forwarding loadings if we apply optimal routing algorithms to the base network. Simulation results for larger networks show that the optimal routing algorithms achieve the gain of communication performance.

**Poster 16: perf-taint: Taint Analysis for Automatic Many-Parameter Performance Modeling**  
Marcin Copik (ETH Zurich)

Performance modeling is a well-known technique for understanding the scaling behavior of an application. Although the modeling process is today often automatic, it still relies on a domain expert selecting program parameters and deciding relevant sampling intervals. Since existing empirical methods attempt blackbox modeling, the decision on which parameters influence a selected part of the program is based on measured data, making empirical modeling sensitive to human errors and instrumentation noise. We introduce a hybrid analysis to mitigate the current limitations of empirical modeling, combining the confidence of static analysis with the ability of dynamic taint analysis to capture the effects of control-flow and memory operations. We construct models of computation and communication volumes that help the modeler to remove effects of noise and improve the correctness of estimated models. Our automatic analysis prunes irrelevant program parameters and brings an understanding of parameter dependencies which helps in designing the experiment.

**Poster 24: Fingerprinting Anomalous Computation with RNN for GPU-Accelerated HPC Machines**
Pengfei Zou (Clemson University)

This paper presents a workload classification framework that discriminates illicit computation from authorized workloads on GPU-accelerated HPC systems. As such systems become more and more powerful, they are exploited by attackers to run malicious and for-profit programs that typically require extremely high computing ability to be successful. Our classification framework leverages the distinctive signatures between illicit and authorized workloads, and explores machine learning methods to learn the workloads and classify them. The framework uses lightweight, non-intrusive workload profiling to collect model input data, and explores multiple machine learning methods, particularly recurrent neural network (RNN) that is suitable for online anomalous workload detection. Evaluation results on four generations of GPU machines demonstrate that the workload classification framework can tell apart the illicit authorized workloads with high accuracy of over 95%. The collected dataset, detection framework, and neural network models will be made available on GitHub.

Poster 33: Enhancing Neural Architecture Search with Speciation and Inter-Epoch Crossover
Matthew E. Baughman (University of Chicago)

As deep learning continues to expand into new areas of application, the demand for efficient use of our HPC resources increase. For new problem domains, new model architectures are developed through a neural architecture search (NAS), which consist of iteratively training many neural networks. To combat the computational waste and maximize compute efficiency for NAS, we demonstrate that the use of genetic algorithms with speciation can be used to both shorten training time and increase accuracy at each iteration.

Poster 14: Data Reuse Analysis for GPU Offloading Using OpenMP
Alok Mishra (Stony Brook University, Lawrence Livermore National Laboratory)

More researchers and developers desire to port their applications to GPU-based clusters, due to their abundant parallelism and energy efficiency. Unfortunately porting or writing an application for accelerators, such as GPUs, requires extensive knowledge of the underlying architectures, the application/algorithm and the interfacing programming model (e.g. OpenMP). Often applications spend a significant portion of their execution time on data transfer. Exploiting data reuse opportunities in an application can reduce its overall execution time. In this research, we present an approach to automatically recognize data reuse opportunities in an application which uses OpenMP for exploiting GPU parallelism, and consequently insert pertinent code to take advantage of data reuse on GPU. Using our approach we were able to retain reused data on the GPU and reduce the overall execution time of multiple benchmark application.
Poster 17: Exploiting Multi-Resource Scheduling for HPC
Yuping Fan (Illinois Institute of Technology)

High-performance computing (HPC) is undergoing significant changes. Next generation HPC systems are equipped with diverse global/local resources. HPC job scheduler plays a crucial role in efficient use of resources. However, traditional job schedulers are single-objective and fail to efficient use of other resources. In our previous work, we present a job scheduling framework named BBSched to schedule CPUs and burst buffers. As we are heading toward exascale computing, a variety of heterogeneous resources are deployed in HPC systems. In this poster, we extend BBSched for managing multiple resources beyond CPUs and burst buffers. We formulate multi-resource scheduling as a general multi-objective optimization (MOO) problem, present a heuristic method to solve the NP-hard MOO problem, and provide a preliminary evaluation for scheduling up to ten resources. The proposed multi-resource scheduling design is intended to enhance Cobalt, a production job scheduler deployed on HPC systems at Argonne Leadership Computing Facility (ALCF).

Poster 15: Cost-Aware Cloudlet Placement in Edge Computing Systems
Dixit Bhatta (University of Delaware)

A well-known challenge in Edge Computing is strategic placement of cloudlets. The fundamental goals of this challenge are to minimize the cloudlet deployment cost and to guarantee minimum latency to the users of edge services. We address this challenge by designing a cost-aware cloudlet placement approach that that ensures user latency requirements while covering all devices in the service region. We first mathematically formulate the problem as a multi-objective integer programming model in a general deployment scenario, which is computationally NP-hard. We then propose a genetic algorithm-based approach, GACP, to find heuristic solutions in significantly reduced time. We investigate the effectiveness of GACP by performing extensive experiments on multiple deployment scenarios based on New York City OpenData. The results presented in the poster show that our approach obtains close to optimal cost solutions with significant time reduction.

Poster 19: Accelerating Real-World Stencil Computations Using Temporal Blocking: Handling Sparse Sources and Receivers
George Bisbas (Imperial College, London)

This paper concerns performance optimization in finite-difference solvers found in seismic imaging.
We tackle a complicating factor that is arising in seismic inversion problems: the addition of waves injected from sources distributed sparsely over the 2D/3D domain, and the need for receivers that interpolate data measurements at a set of points, again distributed sparsely across the domain. In this work we show how to overcome this limitation. We introduce an algorithm for a loop nest transformation policy applied to wave propagation in order to improve data locality and optimize our cache memory use. Our algorithm uses an inspector/executor scheme capable of inspecting measurement operations at sparse locations and then compute the field updates. The ultimate goal of this ongoing work is to automate this scheme for stencil codes generated by Devito.

*Tirthak Patel (Northeastern University), Devesh Tiwari (Northeastern University)*

Large-scale computing systems are becoming increasingly more power-constrained, but these systems employ hardware over-provisioning to achieve higher system throughput because HPC applications often do not consume the peak power capacity of nodes. Unfortunately, focusing on system throughput alone can lead to severe unfairness among multiple concurrently-running applications. This paper introduces PERQ, a new feedback-based principled approach to improve system throughput while achieving fairness among concurrent applications.

**Poster 18: Using Non Volatile Memories to Build Energy- and Cost-Efficient Clusters**  
*Onkar Patil (North Carolina State University)*

Non-Volatile Memory (NVM) is a byte-addressable, high capacity, high latency, and persistent form of memory that can extend the primary memory hierarchy by another level. It allows clusters to have significantly greater memory capacity per node. Intel’s Optane DC Persistent Memory Module (PMM) is a NVM device that can be used to increase the memory density of high performance computing (HPC) system. With higher memory density, we can run scientific computing applications with larger problem sizes on fewer compute nodes than on current HPC systems. Smaller HPC clusters will reduce the cost of running scientific simulations. We test our hypothesis by running a HPC application with large problem sizes on a single node with NVM and on multiple nodes connected by a high speed interconnect with equivalent amount of volatile memory. We compare the performance characteristics and power consumption of both systems.

**Poster 22: Fast Profiling-Based Performance Modeling of Distributed GPU Applications**  
*Jaemin Choi (University of Illinois, Lawrence Livermore National Laboratory)*
An increasing number of applications utilize GPUs to accelerate computation, with MPI responsible for communication in distributed environments. Existing performance models only focus on either modeling GPU kernels or MPI communication; few that do model the entire application are often too specialized for a single application and require extensive input from the programmer.

To be able to quickly model different types of distributed GPU applications, we propose a profiling-based methodology for creating performance models. We build upon the roofline performance model for GPU kernels and analytical models for MPI communication, with a significant reduction in profiling time. We also develop a benchmark to model 3D halo exchange that occurs in many scientific applications. Our proposed model for the main iteration loops of MiniFE achieves 6-7% prediction error on LLNL Lassen and 1-2% error on PSC Bridges, with minimal code inspection required to model MPI communication.

**Poster 26: Neural Networks for the Benchmarking of Detection Algorithms**
Silvia Miramontes (Lawrence Berkeley National Laboratory; University of California, Berkeley)

There are several automated methods to detect objects from grayscale images. However, materials scientists still lack basic tools to compare different detection results, particularly when working with microtomography. This poster introduces FibCAM, a convolutional neural network (CNN)-based method using TensorFlow that allows benchmarking fiber detection algorithms. Our contribution is three-fold: (a) the design of a computational framework to compare automated fiber detection models with curated datasets through classification; (b) lossless data reduction by embedding prior knowledge into data-driven models; (c) a scheme to decompose computation into embarrassingly parallel processes for future analysis at scale. Our results show how FibCAM classifies different structures, and how it illustrates the material’s composition and frequency distribution of microstructures for improved interpretability of machine learning models. The proposed algorithms support probing the specimen content from gigabyte-sized volumes and enable pinpointing inconsistencies between real structures known a priori and results derived from automated detections.

**Poster 21: Optimization of GPU Kernels for Sparse Matrix Computations in Hypre**
Chaoyu Zhang (Arkansas State University)

The acceleration of sparse matrix computations on GPUs can significantly enhance the performance of iterative methods for solving linear systems. In this work, we consider the kernels of Sparse Matrix Vector Multiplications (SpMV), Sparse Triangular Matrix Solves (SpTrSv) and Sparse Matrix Matrix
Multiplications (SpMM), which are often demanded by Algebraic Multigrid (AMG) solvers. With the CUDA and the hardware support of the Volta GPUs on Sierra, the existing kernels should be further optimized to fully take the advantage of the new hardware, and the optimizations have shown significant performance improvement. The presented kernels have been put in HYPRE for solving large scale linear systems on HPC equipped with GPUs. These shared-memory kernels for single GPU are the building blocks of distributed matrix operations required by the solver across multiple GPUs and compute nodes. The implementations of these kernels in Hypre and the code optimizations will be discussed.

Poster 20: From IoT to Cloud: Research Platform for IoT/Cloud Experiments
Jinfeng Lin (University of Notre Dame)

IoT studies leverage a wide range of lightweight hardware for collecting and processing data in the field. Limited by the resources on the devices, IoT systems have to interoperate with cloud platforms for addressing computation intensive tasks such as image processing, application backend supporting and centralized data storage. Therefore, a testbed for IoT/Cloud experiments should provide infrastructure for IoT to cloud communication, computation deployment, and hardware resource management. With these functionalities, users can focus on research problems without distraction from manually constructing experiment environments. Though cloud providers such as Google, Amazon, and Microsoft all provide IoT to Cloud solutions in general, this commercial model is not entirely compatible with research purpose. We propose a framework named Chameleon IoT testbed (CHIoT) that extends the Chameleon bare-metal cloud resources to support general IoT to Cloud experiments.

Poster 32: OSU INAM: A Profiling and Visualization Tool for Scalable and In-Depth Analysis of High-Performance GPU-Enabled HPC Clusters
Pouya Kousha (Ohio State University)

The lack of low-overhead and scalable monitoring tools have prevented a comprehensive study of efficiency and utilization of emerging NVLink-enabled GPU clusters. We address this by proposing and designing an in-depth, real-time analysis, profiling, and visualization tool for high-performance GPU-enabled clusters with NVLinks on the top of the OSU INAM. The proposed tool is capable of presenting a unified and holistic view of MPI-level and fabric level information for emerging NVLink-enabled high-performance GPU clusters. It also provides insights into the efficiency and utilization of underlying interconnects for different communication patterns. We also designed a low overhead and scalable modules to discover the fabric topology and gather fabric metrics by using different levels of threading, bulk insertions and deletions for storage, and using parallel components for
Zheng Miao (Clemson University)

High-Performance Computing systems must simultaneously address both resilience and power. In heterogeneous systems, the trade-offs between resilience and energy-efficiency are more complex for applications using both CPUs and GPUs. A deep understanding of the interplay among energy efficiency, resilience, and performance is required for heterogeneous systems to address them simultaneously.

In this work, we present a new framework for resilient and energy-efficient computing in GPU-accelerated systems. This framework supports partial or full redundancy and checkpointing for resilience, and provides users with flexible hardware resource selection, adjustable precision and power management to improve performance and energy-efficiency. We further perform CUDA-aware MPI to reduce resilience overhead, mainly in message communication between GPUs. Using CG as an example, we show that our framework provides about 40% time and 45% energy savings, comparing to simple extension of RedMPI, a redundancy based resilience framework for homogeneous CPU systems.

Poster 25: Leveraging Smart Data Transfer and Metadirective in Adaptive Computing
Anjia Wang (University of North Carolina, Charlotte; Lawrence Livermore National Laboratory)

In this work, we propose smart data transfer (SDT) along with extensions to metadirective and map constructs in OpenMP 5.0 to improve adaptive computing. The Smith-Waterman algorithm is used as an example, whose naïve implementation does not conduct data transfer efficiently. SDT is used to solve this issue with the following advantages: (1) SDT only transfers necessary data to GPU instead of the whole data set, resulting in 4.5x of speedup in our initial experiments. (2) Depending on computing vs. data transfer requirements of a program, SDT will transfer the output of each iteration from GPU to host either immediately or all together after the last GPU kernel call. (3) It supports large data exceeding GPU device memory’s size via data tiling. We propose to extend metadirective’s context selector to obtain similar improvement by enabling target enter/exit data and on-demand data access.
ACM Student Research Competition: Undergraduate Posters

Tuesday, November 19

8:30 am - 5:00 pm

ACM Student Research Competition Posters Display

Poster 12: Investigating Input Sensitivity in GPU Graph Analytics with Machine Learning
Jacob M. Hope (Texas State University), Trisha Nag (Texas Tech University)

Graph algorithms are at the core of data-intensive applications. As such, efficient graph processing is of great importance. Irregularity in real-world graphs can make performance unpredictable and non-portable across different inputs and architectures. Given a type of graph, the same optimized implementation of an algorithm can produce performance numbers that differ by orders-of-magnitude. We conduct extensive analysis on a set of 1238 graphs to identify input-sensitive performance inefficiencies, including two that have not been studied: (i) register pressure and (ii) CPU-GPU data movement via demand paging. We then build a multiclass decision tree classifier that characterizes the irregular properties of graphs from our data and maps them to an optimal control parameterization at the compiler, system and algorithmic layers, that yield the highest overall algorithmic performance. We then integrate the classifier into a system where it will process a new graph and generate a kernel on the predicted optimal configuration.

Poster 13: Scalable Load-Balancing Concurrent Queues on Many-Core Architectures
Caleb Lehman (Ohio State University)

As the core counts of computing platforms continue to rise, parallel runtime systems with support for very fine-grained tasks become increasingly necessary to fully utilize the available resources. A critical feature of such task-based parallel runtime systems is the ability to balance work evenly and quickly across available cores. We highlight this by studying XTask, a custom parallel runtime system based on XQueue, which is a novel lock-less concurrent queuing system with relaxed
ordering semantics that is geared to realizing scalability to hundreds of concurrent threads. We demonstrate the lack of adequate load balancing in the original XQueue design and present several solutions for improving load balancing. We also evaluate the corresponding improvements in performance on two sample workloads, computation of Fibonacci numbers and computation of Cholesky factorization. Finally, we compare the performance of several versions of XTask along with several implementations of the popular OpenMP runtime system.

Poster 6: Analyzing the Performance of ZFP Compressed Arrays on HPC Kernels
Pavlo D. Triantafyllides (Clemson University)

Per-node memory capacity limits the maximal problem size of HPC applications. Naïve data compression alleviates the memory impact, but requires full decompression before the data is accessed. ZFP compressed arrays reduce the memory footprint, by independently compressing data in fixed sized blocks. Thus, decompressing individual blocks and caching them enables random access and a reduction in decompressions on the critical path. The performance of ZFP compressed arrays is dependent on several key variables: software cache size, cache policy, and compression rate. In this poster, we explore the sensitivity of these ZFP parameters on runtime performance for the matrix-matrix multiplication algorithm. Results show that selection of cache size, policy, and rate yields 8% performance improvement over the default ZFP configuration.

Poster 1: Early Experiences on OpenPOWER Architecture: Analysis of Billion-Scale Atomistic Datasets
Yuya Kawakami (Grinnell College, Oak Ridge National Laboratory)

We present performance analysis on OpenPOWER architecture of an algorithm to generate transversal views of atomistic models. The algorithm was implemented with data parallel primitives in NVIDIA Thrust for architecture portability. We report performance results on IBM Power9 CPUs (OpenMP, Intel Threading Blocks) and NVIDIA Volta GPUs (single and multi GPU). We also evaluate CUDA unified memory performance, exposed by NVIDIA RAPIDS Memory Manager library (RMM).

Poster 2: An Efficient Parallel Algorithm for Dominator Detection
Daniel Giger (University of Massachusetts, Lowell), Hang Liu (Stevens Institute of Technology)

In graph theory, a vertex v dominates a vertex u if every path from the entry vertex to u must go through vertex v. This algorithm is called dominator detection and holds a wide range of applications, such as compiler design, circuit testing, and social network analysis. While the
performance of many other graph algorithms soars with respect to the increase of the hardware parallelism, dominator detection algorithm experiences very little advancement due to the hardship of parallelism. This work thus introduces an efficient parallel dominator detection algorithm that is inspired by Breadth-First Search (BFS), which bests SEMI-NCA on large graphs.

**Poster 9: Machine Specific Symbolic Code Generation**  
Robert King (University of Utah)

New discoveries in science and engineering are primarily driven by numerical simulations of underlying governing equations. Manually written codes to evaluate the main computational kernels lack portability, prone to human errors, ability to perform code optimizations due to the complexity of the underlying equations. In this work we present a symbolic code generation framework, which generates architecture optimized code for different platforms. As the driving application, we primarily use computational relativity where computations of Einstein equations become complicated due to the presence of curvature in spacetime.

The presented framework is based on SymPy with additional modules written to handle complicated partial differential equations (PDEs). The symbolically written equations are converted to a computational graph, which enables to perform architecture (cache, register optimizations) and language specific (SIMD vectorization, CUDA) optimizations. By computing common subexpressions, we reduce the number of compute operations needed, by storing them in temporary variables.

**Poster 5: Evaluating Lossy Compressors for Inline Compression**  
Donald T. Elmore (Clemson University)

HPC applications require massive amounts of data. The data required is growing faster than memory capabilities. An example of this is pySDC, a framework for solving collocation problems iteratively using parallel-in-time methods. pySDC requires storing and exchange of 3D volume data for each parallel point in time. We evaluate several state-of-the-art lossy compressors such as SZ and ZFP for their applicability to inline compression for pySDC. We evaluate the compressors based on compression ratio, compression bandwidth, decompression bandwidth, and overall error introduced.

This poster evaluates state-of-the-art lossy compressors for use in pySDC; shows lossy data compression is an effective tool for reducing memory requirements for pySDC; and highlights current compression/decompression bandwidth is not fast enough for inline lossy compression yet.
Results show using SZ with an error bound of 1e-5, we reduce the memory footprint by a factor of 311.99 while maintaining an acceptable level of loss.

Poster 3: A Deep Learning Approach to Noise Prediction and Circuit Optimization for Near-Term Quantum Devices
Alexander Zlokapa (California Institute of Technology), Alexandru Gheorghiu (California Institute of Technology)

Noisy intermediate-scale quantum (NISQ) devices face challenges in achieving high-fidelity computations due to hardware-specific noise. As a basis for noise mitigation, we develop a convolutional neural network noise model to estimate the difference in noise between a given pair of equivalent quantum circuits. On a classically simulated dataset of 1.6 million pairs of quantum circuits with a simplified noise model calibrated to IBM Q hardware, the deep learning approach shows a significant improvement in noise prediction over linear gate count models. A greedy peephole optimization procedure is proposed to minimize noise using the deep learning model as an objective function, showing further improvement in noise mitigation compared to commonly used gate count minimization heuristics.

Poster 10: Learning Page Access Patterns for Algorithms Programmed with GPU UVM
Bennett Cooper (Clemson University), Derek Rodriguez (Clemson University)

We aim to mitigate the performance bottleneck of migrating data between host and device memory in GPU applications by accurately predicting application access patterns using deep neural networks. We model the memory access pattern of any given application by collecting page faults that trigger data migration to the GPU and feed this time series as input to a neural network that outputs the next several page faults. We evaluate these predictions on the basis of what makes a useful prefetch in our context for the GPU. Current work has looked at trivial GPU applications, such as matrix operations, and moving toward real, complex applications. Our work will be presented by bringing attention to the predictive capability of our neural network on the current applications tested.

Poster 8: Mitigating Communication Bottlenecks in MPI-Based Distributed Learning
Abdullah B. Nauman (Ward Melville High School, Brookhaven National Laboratory)

Current commercial and scientific facilities generate and maintain vast amounts of complex data. While machine learning (ML) techniques can provide crucial insight, developing these models is
often impractical on a single process. Distributed learning techniques mitigate this problem; however, current models contain significant performance bottlenecks. Here, we conduct a detailed performance analysis of MPI_Learn, a widespread distributed ML framework for high-energy physics (HEP) applications, on the Summit supercomputer, by training a network to classify simulated collision events from high-energy particle detectors at the CERN Large Hadron Collider (LHC).

We conclude that these bottlenecks occur as a result of increasing communication time between the different processes, and to mitigate the bottlenecks we propose the implementation of a new distributed algorithm for stochastic gradient descent (SGD). We provide a proof of concept by demonstrating better scalability with results on 250 GPUs, and with hyperparameter optimization, show a ten-fold decrease in training time.

Hearing Single- and Multi-Threaded Program Behavior
Mark Wissink (Calvin University), Joel Adams (Calvin University)

There are many examples of visual program representation, but there is an unexplored area of sonic representation—audialization—of programs. The Thread Safe Audio Library (TSAL) provides the necessary tools to create such program audializations. TSAL is a platform-independent, object-oriented C++ library that provides thread safe classes for synthesizing sound from a program. By adding TSAL calls to a sequential or parallel program, its behavior can be represented with sound. Audializations require a unique interpretation that may greatly benefit auditory learners and other programmers who are looking for a different way to profile a sequential or parallel program.

Poster 11: Walking the Cost-Accuracy Tightrope: Balancing Trade-Offs in Data-Intensive Genomics
Kathryn Leung (Princeton University), Meghan Kimball (DePaul University)

Scientific applications often exhibit a trade-off between cost and accuracy. However, measuring and predicting cost and accuracy in a way that users can understand these trade-offs is challenging. To address these needs, we present predictive cost and accuracy models for data-intensive genomics applications. We use these models to create a trade-off graph, which researchers can use to selectively trade-off cost and accuracy.

Poster 31: Designing High-Performance Erasure Coding Schemes for Next-Generation Storage Systems
Haiyang Shi (Ohio State University), Xiaoyi Lu (Ohio State University)

Erasure Coding (EC) NIC offload is a promising technology for designing next-generation distributed storage systems. To alleviate the EC overhead, various hardware-based Erasure Coding schemes have been proposed to leverage the advanced compute capabilities on modern HPC clusters. However, this poster has identified some critical limitations of current-generation EC NIC offload schemes on modern SmartNICS and the missing of exploiting heterogeneous hardware available on modern HPC clusters in parallel. To address these limitations, this poster first proposes a unified benchmark suite to benchmark, measure and characterize hardware-optimized erasure coders. Then, it proposes a Multi-Rail EC concept which enables upper-layer applications to leverage heterogeneous hardware to perform EC operations simultaneously. Finally, it proposes a new EC NIC offload paradigm based on the tripartite graph model, namely TriEC. TriEC supports both encode-and-send and receive-and-decode offload primitives efficiently.

Poster 27: A Framework for Measuring Hardware Gather-Scatter Support
Patrick Lavin (Georgia Institute of Technology), Jeffrey Young (Georgia Institute of Technology), Richard Vuduc (Georgia Institute of Technology)

In recent years, we have seen the re-addition of vector units to CPUs. While these units easily give speedups for easily vectorized applications with dense memory access, it can be hard to characterize how different access patterns will affect the performance of vectorized code.

We have developed Spatter, a benchmark which allows us to test and investigate the gather-scatter units available on current and upcoming hardware. The information that Spatter reveals to users is of use to everyone from hardware vendors who wish to compare gather-scatter units across platforms, to application developers who wish to test memory access pattern performance in their vectorized code.

Poster 30: A Fast and Efficient Incremental Approach toward Dynamic Community Detection
Neda Zarayeneh (Washington State University)

Community detection is a discovery tool to identify groups of vertices which are tightly connected, called communities. Most of the literature on this subject caters to the static use-case where the underlying network does not change. However, many emerging real-world use-cases give rise to a need to incorporate dynamic graphs as inputs.

We present a fast and efficient incremental approach toward dynamic community detection. The key
A contribution is a generic technique called ∆-screening, which examines the most recent batch of changes made to an input graph and selects a subset of vertices to reevaluate for potential community (re)assignment. This technique can be incorporated into any of the community detection methods that use modularity as its objective function for clustering. For demonstration purposes, we incorporated the technique into two well-known community detection tools. Our experiments demonstrate that our approach is able to generate performance speedups without compromising on the output quality.

**Poster 29: Optimal Routing for a Family of Scalable Interconnection Networks**  
Zhipeng Xu (Stony Brook University; Sun Yat-sen University, Guangzhou, China)

We propose a scheme to construct a family of large and high-performance interconnection networks that are scalable, low-radix, minimum diameters. These networks, whose diameters grow linearly as their sizes grow exponentially, are generated by using the Cartesian products of smaller optimal networks of minimum diameters. For the smaller base networks, we design the vertex-balanced routing algorithm by considering the forwarding pressure at each vertex. Comparative benchmarks on a Beowulf cluster show significant improvement in performance after using the new routing algorithm. Each node of the new network generated from base graphs with low-diameter can also sustain balanced forwarding loadings if we apply optimal routing algorithms to the base network. Simulation results for larger networks show that the optimal routing algorithms achieve the gain of communication performance.

**Poster 16: perf-taint: Taint Analysis for Automatic Many-Parameter Performance Modeling**  
Marcin Copik (ETH Zurich)

Performance modeling is a well-known technique for understanding the scaling behavior of an application. Although the modeling process is today often automatic, it still relies on a domain expert selecting program parameters and deciding relevant sampling intervals. Since existing empirical methods attempt blackbox modeling, the decision on which parameters influence a selected part of the program is based on measured data, making empirical modeling sensitive to human errors and instrumentation noise. We introduce a hybrid analysis to mitigate the current limitations of empirical modeling, combining the confidence of static analysis with the ability of dynamic taint analysis to capture the effects of control-flow and memory operations. We construct models of computation and communication volumes that help the modeler to remove effects of noise and improve the correctness of estimated models. Our automatic analysis prunes irrelevant program parameters and brings an understanding of parameter dependencies which helps in designing the experiment.
**Poster 24: Fingerprinting Anomalous Computation with RNN for GPU-Accelerated HPC Machines**  
*Pengfei Zou (Clemson University)*

This paper presents a workload classification framework that discriminates illicit computation from authorized workloads on GPU-accelerated HPC systems. As such systems become more and more powerful, they are exploited by attackers to run malicious and for-profit programs that typically require extremely high computing ability to be successful. Our classification framework leverages the distinctive signatures between illicit and authorized workloads, and explore machine learning methods to learn the workloads and classify them. The framework uses lightweight, non-intrusive workload profiling to collect model input data, and explores multiple machine learning methods, particularly recurrent neural network (RNN) that is suitable for online anomalous workload detection. Evaluation results on four generations of GPU machines demonstrate that the workload classification framework can tell apart the illicit authorized workloads with high accuracy of over 95%. The collected dataset, detection framework, and neural network models will be made available on GitHub.

**Poster 33: Enhancing Neural Architecture Search with Speciation and Inter-Epoch Crossover**  
*Matthew E. Baughman (University of Chicago)*

As deep learning continues to expand into new areas of application, the demand for efficient use of our HPC resources increase. For new problem domains, new model architectures are developed through a neural architecture search (NAS), which consist of iteratively training many neural networks. To combat the computational waste and maximize compute efficiency for NAS, we demonstrate that the use of genetic algorithms with speciation can be used to both shorten training time and increase accuracy at each iteration.

**Poster 14: Data Reuse Analysis for GPU Offloading Using OpenMP**  
*Alok Mishra (Stony Brook University, Lawrence Livermore National Laboratory)*

More researchers and developers desire to port their applications to GPU-based clusters, due to their abundant parallelism and energy efficiency. Unfortunately porting or writing an application for accelerators, such as GPUs, requires extensive knowledge of the underlying architectures, the application/algorithm and the interfacing programming model (e.g. OpenMP). Often applications spend a significant portion of their execution time on data transfer. Exploiting data reuse opportunities in an application can reduce its overall execution time. In this research, we present an approach to automatically recognize data reuse opportunities in an application which uses OpenMP for exploiting
GPU parallelism, and consequently insert pertinent code to take advantage of data reuse on GPU. Using our approach we were able to retain reused data on the GPU and reduce the overall execution time of multiple benchmark application.

**Poster 17: Exploiting Multi-Resource Scheduling for HPC**  
Yuping Fan (Illinois Institute of Technology)

High-performance computing (HPC) is undergoing significant changes. Next generation HPC systems are equipped with diverse global/local resources. HPC job scheduler plays a crucial role in efficient use of resources. However, traditional job schedulers are single-objective and fail to efficiently use of other resources. In our previous work, we present a job scheduling framework named BBSched to schedule CPUs and burst buffers. As we are heading toward exascale computing, a variety of heterogeneous resources are deployed in HPC systems. In this poster, we extend BBSched for managing multiple resources beyond CPUs and burst buffers. We formulate multi-resource scheduling as a general multi-objective optimization (MOO) problem, present a heuristic method to solve the NP-hard MOO problem, and provide a preliminary evaluation for scheduling up to ten resources. The proposed multi-resource scheduling design is intended to enhance Cobalt, a production job scheduler deployed on HPC systems at Argonne Leadership Computing Facility (ALCF).

**Poster 15: Cost-Aware Cloudlet Placement in Edge Computing Systems**  
Dixit Bhatta (University of Delaware)

A well-known challenge in Edge Computing is strategic placement of cloudlets. The fundamental goals of this challenge are to minimize the cloudlet deployment cost and to guarantee minimum latency to the users of edge services. We address this challenge by designing a cost-aware cloudlet placement approach that that ensures user latency requirements while covering all devices in the service region. We first mathematically formulate the problem as a multi-objective integer programming model in a general deployment scenario, which is computationally NP-hard. We then propose a genetic algorithm-based approach, GACP, to find heuristic solutions in significantly reduced time. We investigate the effectiveness of GACP by performing extensive experiments on multiple deployment scenarios based on New York City OpenData. The results presented in the poster show that our approach obtains close to optimal cost solutions with significant time reduction.

**Poster 19: Accelerating Real-World Stencil Computations Using Temporal Blocking: Handling Sparse Sources and Receivers**
George Bisbas (Imperial College, London)

This paper concerns performance optimization in finite-difference solvers found in seismic imaging. We tackle a complicating factor that is arising in seismic inversion problems: the addition of waves injected from sources distributed sparsely over the 2D/3D domain, and the need for receivers that interpolate data measurements at a set of points, again distributed sparsely across the domain. In this work we show how to overcome this limitation. We introduce an algorithm for a loop nest transformation policy applied to wave propagation in order to improve data locality and optimize our cache memory use. Our algorithm uses an inspector/executor scheme capable of inspecting measurement operations at sparse locations and then compute the field updates. The ultimate goal of this ongoing work is to automate this scheme for stencil codes generated by Devito.

Poster 23: PERQ: Fair and Efficient Power Management of Power-Constrained Large-Scale Computing Systems
Tirthak Patel (Northeastern University), Devesh Tiwari (Northeastern University)

Large-scale computing systems are becoming increasingly more power-constrained, but these systems employ hardware over-provisioning to achieve higher system throughput because HPC applications often do not consume the peak power capacity of nodes. Unfortunately, focusing on system throughput alone can lead to severe unfairness among multiple concurrently-running applications. This paper introduces PERQ, a new feedback-based principled approach to improve system throughput while achieving fairness among concurrent applications.

Poster 18: Using Non Volatile Memories to Build Energy- and Cost-Efficient Clusters
Onkar Patil (North Carolina State University)

Non-Volatile Memory (NVM) is a byte-addressable, high capacity, high latency, and persistent form of memory that can extend the primary memory hierarchy by another level. It allows clusters to have significantly greater memory capacity per node. Intel’s Optane DC Persistent Memory Module (PMM) is a NVM device that can be used to increase the memory density of high performance computing (HPC) system. With higher memory density, we can run scientific computing applications with larger problem sizes on fewer compute nodes than on current HPC systems. Smaller HPC clusters will reduce the cost of running scientific simulations. We test our hypothesis by running a HPC application with large problem sizes on a single node with NVM and on multiple nodes connected by a high speed interconnect with equivalent amount of volatile memory. We compare the performance characteristics and power consumption of both systems.
**Poster 22: Fast Profiling-Based Performance Modeling of Distributed GPU Applications**  
*Jaemin Choi (University of Illinois, Lawrence Livermore National Laboratory)*

An increasing number of applications utilize GPUs to accelerate computation, with MPI responsible for communication in distributed environments. Existing performance models only focus on either modeling GPU kernels or MPI communication; few that do model the entire application are often too specialized for a single application and require extensive input from the programmer.

To be able to quickly model different types of distributed GPU applications, we propose a profiling-based methodology for creating performance models. We build upon the roofline performance model for GPU kernels and analytical models for MPI communication, with a significant reduction in profiling time. We also develop a benchmark to model 3D halo exchange that occurs in many scientific applications. Our proposed model for the main iteration loops of MiniFE achieves 6-7% prediction error on LLNL Lassen and 1-2% error on PSC Bridges, with minimal code inspection required to model MPI communication.

**Poster 26: Neural Networks for the Benchmarking of Detection Algorithms**  
*Silvia Miramontes (Lawrence Berkeley National Laboratory; University of California, Berkeley)*

There are several automated methods to detect objects from grayscale images. However, materials scientists still lack basic tools to compare different detection results, particularly when working with microtomography. This poster introduces FibCAM, a convolutional neural network (CNN)-based method using TensorFlow that allows benchmarking fiber detection algorithms. Our contribution is three-fold: (a) the design of a computational framework to compare automated fiber detection models with curated datasets through classification; (b) lossless data reduction by embedding prior knowledge into data-driven models; (c) a scheme to decompose computation into embarrassingly parallel processes for future analysis at scale. Our results show how FibCAM classifies different structures, and how it illustrates the material's composition and frequency distribution of microstructures for improved interpretability of machine learning models. The proposed algorithms support probing the specimen content from gigabyte-sized volumes and enable pinpointing inconsistencies between real structures known a priori and results derived from automated detections.

**Poster 21: Optimization of GPU Kernels for Sparse Matrix Computations in Hypre**  
*Chaoyu Zhang (Arkansas State University)*
The acceleration of sparse matrix computations on GPUs can significantly enhance the performance of iterative methods for solving linear systems. In this work, we consider the kernels of Sparse Matrix Vector Multiplications (SpMV), Sparse Triangular Matrix Solves (SpTrSv) and Sparse Matrix Matrix Multiplications (SpMM), which are often demanded by Algebraic Multigrid (AMG) solvers. With the CUDA and the hardware support of the Volta GPUs on Sierra, the existing kernels should be further optimized to fully take the advantage of the new hardware, and the optimizations have shown significant performance improvement. The presented kernels have been put in HYPRE for solving large scale linear systems on HPC equipped with GPUs. These shared-memory kernels for single GPU are the building blocks of distributed matrix operations required by the solver across multiple GPUs and compute nodes. The implementations of these kernels in Hypre and the code optimizations will be discussed.

**Poster 20: From IoT to Cloud: Research Platform for IoT/Cloud Experiments**
*Jinfeng Lin (University of Notre Dame)*

IoT studies leverage a wide range of lightweight hardware for collecting and processing data in the field. Limited by the resources on the devices, IoT systems have to interoperate with cloud platforms for addressing computation intensive tasks such as image processing, application backend supporting and centralized data storage. Therefore, a testbed for IoT/Cloud experiments should provide infrastructure for IoT to cloud communication, computation deployment, and hardware resource management. With these functionalities, users can focus on research problems without distraction from manually constructing experiment environments. Though cloud providers such as Google, Amazon, and Microsoft all provide IoT to Cloud solutions in general, this commercial model is not entirely compatible with research purpose. We propose a framework named Chameleon IoT testbed (CHIoT) that extends the Chameleon bare-metal cloud resources to support general IoT to Cloud experiments.

**Poster 32: OSU INAM: A Profiling and Visualization Tool for Scalable and In-Depth Analysis of High-Performance GPU-Enabled HPC Clusters**
*Pouya Kousha (Ohio State University)*

The lack of low-overhead and scalable monitoring tools have prevented a comprehensive study of efficiency and utilization of emerging NVLink-enabled GPU clusters. We address this by proposing and designing an in-depth, real-time analysis, profiling, and visualization tool for high-performance GPU-enabled clusters with NVLinks on the top of the OSU INAM. The proposed tool is capable of presenting a unified and holistic view of MPI-level and fabric level information for emerging NVLink-enabled high-performance GPU clusters. It also provides insights into the efficiency and utilization
of underlying interconnects for different communication patterns. We also designed a low overhead and scalable modules to discover the fabric topology and gather fabric metrics by using different levels of threading, bulk insertions and deletions for storage, and using parallel components for fabric discovery and port metric inquiry.

Zheng Miao (Clemson University)

High-Performance Computing systems must simultaneously address both resilience and power. In heterogeneous systems, the trade-offs between resilience and energy-efficiency are more complex for applications using both CPUs and GPUs. A deep understanding of the interplay among energy efficiency, resilience, and performance is required for heterogeneous systems to address them simultaneously.

In this work, we present a new framework for resilient and energy-efficient computing in GPU-accelerated systems. This framework supports partial or full redundancy and checkpointing for resilience, and provides users with flexible hardware resource selection, adjustable precision and power management to improve performance and energy-efficiency. We further perform CUDA-aware MPI to reduce resilience overhead, mainly in message communication between GPUs. Using CG as an example, we show that our framework provides about 40% time and 45% energy savings, comparing to simple extension of RedMPI, a redundancy based resilience framework for homogeneous CPU systems.

Poster 25: Leveraging Smart Data Transfer and Metadirective in Adaptive Computing
Anjia Wang (University of North Carolina, Charlotte; Lawrence Livermore National Laboratory)

In this work, we propose smart data transfer (SDT) along with extensions to metadirective and map constructs in OpenMP 5.0 to improve adaptive computing. The Smith-Waterman algorithm is used as an example, whose naïve implementation does not conduct data transfer efficiently. SDT is used to solve this issue with the following advantages: (1) SDT only transfers necessary data to GPU instead of the whole data set, resulting in 4.5x of speedup in our initial experiments. (2) Depending on computing vs. data transfer requirements of a program, SDT will transfer the output of each iteration from GPU to host either immediately or all together after the last GPU kernel call. (3) It supports large data exceeding GPU device memory’s size via data tiling. We propose to extend metadirective’s context selector to obtain similar improvement by enabling target enter/exit data and on-demand data access.
5:15 pm - 7:00 pm

Poster Reception

Wednesday, November 20

8:30 am - 5:00 pm

ACM Student Research Competition Posters Display

Poster 12: Investigating Input Sensitivity in GPU Graph Analytics with Machine Learning

Jacob M. Hope (Texas State University), Trisha Nag (Texas Tech University)

Graph algorithms are at the core of data-intensive applications. As such, efficient graph processing is of great importance. Irregularity in real-world graphs can make performance unpredictable and non-portable across different inputs and architectures. Given a type of graph, the same optimized implementation of an algorithm can produce performance numbers that differ by orders-of-magnitude. We conduct extensive analysis on a set of 1238 graphs to identify input-sensitive performance inefficiencies, including two that have not been studied: (i) register pressure and (ii) CPU-GPU data movement via demand paging. We then build a multiclass decision tree classifier that characterizes the irregular properties of graphs from our data and maps them to an optimal control parameterization at the compiler, system and algorithmic layers, that yield the highest overall algorithmic performance. We then integrate the classifier into a system where it will process a new graph and generate a kernel on the predicted optimal configuration.

Poster 13: Scalable Load-Balancing Concurrent Queues on Many-Core Architectures

Caleb Lehman (Ohio State University)

As the core counts of computing platforms continue to rise, parallel runtime systems with support for very fine-grained tasks become increasingly necessary to fully utilize the available resources. A critical feature of such task-based parallel runtime systems is the ability to balance work evenly and quickly across available cores. We highlight this by studying XTask, a custom parallel runtime system based on XQueue, which is a novel lock-less concurrent queuing system with relaxed ordering semantics that is geared to realizing scalability to hundreds of concurrent threads. We demonstrate the lack of adequate load balancing in the original XQueue design and present several
solutions for improving load balancing. We also evaluate the corresponding improvements in performance on two sample workloads, computation of Fibonacci numbers and computation of Cholesky factorization. Finally, we compare the performance of several versions of XTask along with several implementations of the popular OpenMP runtime system.

**Poster 6: Analyzing the Performance of ZFP Compressed Arrays on HPC Kernels**  
Pavlo D. Triantafyllides (Clemson University)

Per-node memory capacity limits the maximal problem size of HPC applications. Naïve data compression alleviates the memory impact, but requires full decompression before the data is accessed. ZFP compressed arrays reduce the memory footprint, by independently compressing data in fixed sized blocks. Thus, decompressing individual blocks and caching them enables random access and a reduction in decompressions on the critical path. The performance of ZFP compressed arrays is dependent on several key variables: software cache size, cache policy, and compression rate. In this poster, we explore the sensitivity of these ZFP parameters on runtime performance for the matrix-matrix multiplication algorithm. Results show that selection of cache size, policy, and rate yields 8% performance improvement over the default ZFP configuration.

**Poster 1: Early Experiences on OpenPOWER Architecture: Analysis of Billion-Scale Atomistic Datasets**  
Yuya Kawakami (Grinnell College, Oak Ridge National Laboratory)

We present performance analysis on OpenPOWER architecture of an algorithm to generate transversal views of atomistic models. The algorithm was implemented with data parallel primitives in NVIDIA Thrust for architecture portability. We report performance results on IBM Power9 CPUs (OpenMP, Intel Threading Blocks) and NVIDIA Volta GPUs (single and multi GPU). We also evaluate CUDA unified memory performance, exposed by NVIDIA RAPIDS Memory Manager library (RMM).

**Poster 2: An Efficient Parallel Algorithm for Dominator Detection**  
Daniel Giger (University of Massachusetts, Lowell), Hang Liu (Stevens Institute of Technology)

In graph theory, a vertex v dominates a vertex u if every path from the entry vertex to u must go through vertex v. This algorithm is called dominator detection and holds a wide range of applications, such as compiler design, circuit testing, and social network analysis. While the performance of many other graph algorithms soars with respect to the increase of the hardware parallelism, dominator detection algorithm experiences very little advancement due to the hardship
of parallelism. This work thus introduces an efficient parallel dominator detection algorithm that is inspired by Breadth-First Search (BFS), which bests SEMI-NCA on large graphs.

**Poster 9: Machine Specific Symbolic Code Generation**
Robert King (University of Utah)

New discoveries in science and engineering are primarily driven by numerical simulations of underlying governing equations. Manually written codes to evaluate the main computational kernels lack portability, prone to human errors, ability to perform code optimizations due to the complexity of the underlying equations. In this work we present a symbolic code generation framework, which generates architecture optimized code for different platforms. As the driving application, we primarily use computational relativity where computations of Einstein equations become complicated due to the presence of curvature in spacetime.

The presented framework is based on SymPy with additional modules written to handle complicated partial differential equations (PDEs). The symbolically written equations are converted to a computational graph, which enables to perform architecture (cache, register optimizations) and language specific (SIMD vectorization, CUDA) optimizations. By computing common subexpressions, we reduce the number of compute operations needed, by storing them in temporary variables.

**Poster 5: Evaluating Lossy Compressors for Inline Compression**
Donald T. Elmore (Clemson University)

HPC applications require massive amounts of data. The data required is growing faster than memory capabilities. An example of this is pySDC, a framework for solving collocation problems iteratively using parallel-in-time methods. pySDC requires storing and exchange of 3D volume data for each parallel point in time. We evaluate several state-of-the-art lossy compressors such as SZ and ZFP for their applicability to inline compression for pySDC. We evaluate the compressors based on compression ratio, compression bandwidth, decompression bandwidth, and overall error introduced.

This poster evaluates state-of-the-art lossy compressors for use in pySDC; shows lossy data compression is an effective tool for reducing memory requirements for pySDC; and highlights current compression/decompression bandwidth is not fast enough for inline lossy compression yet. Results show using SZ with an error bound of 1e-5, we reduce the memory footprint by a factor of 311.99 while maintaining an acceptable level of loss.
Poster 3: A Deep Learning Approach to Noise Prediction and Circuit Optimization for Near-Term Quantum Devices

Alexander Zlokapa (California Institute of Technology), Alexandru Gheorghiu (California Institute of Technology)

Noisy intermediate-scale quantum (NISQ) devices face challenges in achieving high-fidelity computations due to hardware-specific noise. As a basis for noise mitigation, we develop a convolutional neural network noise model to estimate the difference in noise between a given pair of equivalent quantum circuits. On a classically simulated dataset of 1.6 million pairs of quantum circuits with a simplified noise model calibrated to IBM Q hardware, the deep learning approach shows a significant improvement in noise prediction over linear gate count models. A greedy peephole optimization procedure is proposed to minimize noise using the deep learning model as an objective function, showing further improvement in noise mitigation compared to commonly used gate count minimization heuristics.

Poster 10: Learning Page Access Patterns for Algorithms Programmed with GPU UVM

Bennett Cooper (Clemson University), Derek Rodriguez (Clemson University)

We aim to mitigate the performance bottleneck of migrating data between host and device memory in GPU applications by accurately predicting application access patterns using deep neural networks. We model the memory access pattern of any given application by collecting page faults that trigger data migration to the GPU and feed this time series as input to a neural network that outputs the next several page faults. We evaluate these predictions on the basis of what makes a useful prefetch in our context for the GPU. Current work has looked at trivial GPU applications, such as matrix operations, and moving toward real, complex applications. Our work will be presented by bringing attention to the predictive capability of our neural network on the current applications tested.

Poster 8: Mitigating Communication Bottlenecks in MPI-Based Distributed Learning

Abdullah B. Nauman (Ward Melville High School, Brookhaven National Laboratory)

Current commercial and scientific facilities generate and maintain vast amounts of complex data. While machine learning (ML) techniques can provide crucial insight, developing these models is often impractical on a single process. Distributed learning techniques mitigate this problem; however, current models contain significant performance bottlenecks. Here, we conduct a detailed
performance analysis of MPI_Learn, a widespread distributed ML framework for high-energy physics (HEP) applications, on the Summit supercomputer, by training a network to classify simulated collision events from high-energy particle detectors at the CERN Large Hadron Collider (LHC).

We conclude that these bottlenecks occur as a result of increasing communication time between the different processes, and to mitigate the bottlenecks we propose the implementation of a new distributed algorithm for stochastic gradient descent (SGD). We provide a proof of concept by demonstrating better scalability with results on 250 GPUs, and with hyperparameter optimization, show a ten-fold decrease in training time.

**Hearing Single- and Multi-Threaded Program Behavior**

Mark Wissink (Calvin University), Joel Adams (Calvin University)

There are many examples of visual program representation, but there is an unexplored area of sonic representation—audialization—of programs. The Thread Safe Audio Library (TSAL) provides the necessary tools to create such program audializations. TSAL is a platform-independent, object-oriented C++ library that provides thread safe classes for synthesizing sound from a program. By adding TSAL calls to a sequential or parallel program, its behavior can be represented with sound. Audializations require a unique interpretation that may greatly benefit auditory learners and other programmers who are looking for a different way to profile a sequential or parallel program.

**Poster 11: Walking the Cost-Accuracy Tightrope: Balancing Trade-Offs in Data-Intensive Genomics**

Kathryn Leung (Princeton University), Meghan Kimball (DePaul University)

Scientific applications often exhibit a trade-off between cost and accuracy. However, measuring and predicting cost and accuracy in a way that users can understand these trade-offs is challenging. To address these needs, we present predictive cost and accuracy models for data-intensive genomics applications. We use these models to create a trade-off graph, which researchers can use to selectively trade-off cost and accuracy.

**Poster 31: Designing High-Performance Erasure Coding Schemes for Next-Generation Storage Systems**

Haiyang Shi (Ohio State University), Xiaoyi Lu (Ohio State University)
Erasure Coding (EC) NIC offload is a promising technology for designing next-generation distributed storage systems. To alleviate the EC overhead, various hardware-based Erasure Coding schemes have been proposed to leverage the advanced compute capabilities on modern HPC clusters. However, this poster has identified some critical limitations of current-generation EC NIC offload schemes on modern SmartNICs and the missing of exploiting heterogeneous hardware available on modern HPC clusters in parallel. To address these limitations, this poster first proposes a unified benchmark suite to benchmark, measure and characterize hardware-optimized erasure coders. Then, it proposes a Multi-Rail EC concept which enables upper-layer applications to leverage heterogeneous hardware to perform EC operations simultaneously. Finally, it proposes a new EC NIC offload paradigm based on the tripartite graph model, namely TriEC. TriEC supports both encode-and-send and receive-and-decode offload primitives efficiently.

Poster 27: A Framework for Measuring Hardware Gather-Scatter Support
Patrick Lavin (Georgia Institute of Technology), Jeffrey Young (Georgia Institute of Technology), Richard Vuduc (Georgia Institute of Technology)

In recent years, we have seen the re-addition of vector units to CPUs. While these units easily give speedups for easily vectorized applications with dense memory access, it can be hard to characterize how different access patterns will affect the performance of vectorized code.

We have developed Spatter, a benchmark which allows us to test and investigate the gather-scatter units available on current and upcoming hardware. The information that Spatter reveals to users is of use to everyone from hardware vendors who wish to compare gather-scatter units across platforms, to application developers who wish to test memory access pattern performance in their vectorized code.

Poster 30: A Fast and Efficient Incremental Approach toward Dynamic Community Detection
Neda Zarayeneh (Washington State University)

Community detection is a discovery tool to identify groups of vertices which are tightly connected, called communities. Most of the literature on this subject caters to the static use-case where the underlying network does not change. However, many emerging real-world use-cases give rise to a need to incorporate dynamic graphs as inputs.

We present a fast and efficient incremental approach toward dynamic community detection. The key contribution is a generic technique called Ω-screening, which examines the most recent batch of changes made to an input graph and selects a subset of vertices to reevaluate for potential
community (re)assignment. This technique can be incorporated into any of the community detection methods that use modularity as its objective function for clustering. For demonstration purposes, we incorporated the technique into two well-known community detection tools. Our experiments demonstrate that our approach is able to generate performance speedups without compromising on the output quality.

**Poster 29: Optimal Routing for a Family of Scalable Interconnection Networks**  
Zhipeng Xu (Stony Brook University; Sun Yat-sen University, Guangzhou, China)

We propose a scheme to construct a family of large and high-performance interconnection networks that are scalable, low-radix, minimum diameters. These networks, whose diameters grow linearly as their sizes grow exponentially, are generated by using the Cartesian products of smaller optimal networks of minimum diameters. For the smaller base networks, we design the vertex-balanced routing algorithm by considering the forwarding pressure at each vertex. Comparative benchmarks on a Beowulf cluster show significant improvement in performance after using the new routing algorithm. Each node of the new network generated from base graphs with low-diameter can also sustain balanced forwarding loadings if we apply optimal routing algorithms to the base network. Simulation results for larger networks show that the optimal routing algorithms achieve the gain of communication performance.

**Poster 16: perf-taint: Taint Analysis for Automatic Many-Parameter Performance Modeling**  
Marcin Copik (ETH Zurich)

Performance modeling is a well-known technique for understanding the scaling behavior of an application. Although the modeling process is today often automatic, it still relies on a domain expert selecting program parameters and deciding relevant sampling intervals. Since existing empirical methods attempt blackbox modeling, the decision on which parameters influence a selected part of the program is based on measured data, making empirical modeling sensitive to human errors and instrumentation noise. We introduce a hybrid analysis to mitigate the current limitations of empirical modeling, combining the confidence of static analysis with the ability of dynamic taint analysis to capture the effects of control-flow and memory operations. We construct models of computation and communication volumes that help the modeler to remove effects of noise and improve the correctness of estimated models. Our automatic analysis prunes irrelevant program parameters and brings an understanding of parameter dependencies which helps in designing the experiment.

**Poster 24: Fingerprinting Anomalous Computation with RNN for GPU-Accelerated HPC**
Machines
Pengfei Zou (Clemson University)

This paper presents a workload classification framework that discriminates illicit computation from authorized workloads on GPU-accelerated HPC systems. As such systems become more and more powerful, they are exploited by attackers to run malicious and for-profit programs that typically require extremely high computing ability to be successful. Our classification framework leverages the distinctive signatures between illicit and authorized workloads, and explore machine learning methods to learn the workloads and classify them. The framework uses lightweight, non-intrusive workload profiling to collect model input data, and explores multiple machine learning methods, particularly recurrent neural network (RNN) that is suitable for online anomalous workload detection. Evaluation results on four generations of GPU machines demonstrate that the workload classification framework can tell apart the illicit authorized workloads with high accuracy of over 95%. The collected dataset, detection framework, and neural network models will be made available on GitHub.

Poster 33: Enhancing Neural Architecture Search with Speciation and Inter-Epoch Crossover
Matthew E. Baughman (University of Chicago)

As deep learning continues to expand into new areas of application, the demand for efficient use of our HPC resources increase. For new problem domains, new model architectures are developed through a neural architecture search (NAS), which consist of iteratively training many neural networks. To combat the computational waste and maximize compute efficiency for NAS, we demonstrate that the use of genetic algorithms with speciation can be used to both shorten training time and increase accuracy at each iteration.

Poster 14: Data Reuse Analysis for GPU Offloading Using OpenMP
Alok Mishra (Stony Brook University, Lawrence Livermore National Laboratory)

More researchers and developers desire to port their applications to GPU-based clusters, due to their abundant parallelism and energy efficiency. Unfortunately porting or writing an application for accelerators, such as GPUs, requires extensive knowledge of the underlying architectures, the application/algorithm and the interfacing programming model (e.g. OpenMP). Often applications spend a significant portion of their execution time on data transfer. Exploiting data reuse opportunities in an application can reduce its overall execution time. In this research, we present an approach to automatically recognize data reuse opportunities in an application which uses OpenMP for exploiting GPU parallelism, and consequently insert pertinent code to take advantage of data reuse on GPU. Using our approach we were able to retain reused data on the GPU and reduce the overall execution
time of multiple benchmark application.

**Poster 17: Exploiting Multi-Resource Scheduling for HPC**
Yuping Fan (Illinois Institute of Technology)

High-performance computing (HPC) is undergoing significant changes. Next generation HPC systems are equipped with diverse global/local resources. HPC job scheduler plays a crucial role in efficient use of resources. However, traditional job schedulers are single-objective and fail to efficient use of other resources. In our previous work, we present a job scheduling framework named BBSched to schedule CPUs and burst buffers. As we are heading toward exascale computing, a variety of heterogeneous resources are deployed in HPC systems. In this poster, we extend BBSched for managing multiple resources beyond CPUs and burst buffers. We formulate multi-resource scheduling as a general multi-objective optimization (MOO) problem, present a heuristic method to solve the NP-hard MOO problem, and provide a preliminary evaluation for scheduling up to ten resources. The proposed multi-resource scheduling design is intended to enhance Cobalt, a production job scheduler deployed on HPC systems at Argonne Leadership Computing Facility (ALCF).

**Poster 15: Cost-Aware Cloudlet Placement in Edge Computing Systems**
Dixit Bhatta (University of Delaware)

A well-known challenge in Edge Computing is strategic placement of cloudlets. The fundamental goals of this challenge are to minimize the cloudlet deployment cost and to guarantee minimum latency to the users of edge services. We address this challenge by designing a cost-aware cloudlet placement approach that ensures user latency requirements while covering all devices in the service region. We first mathematically formulate the problem as a multi-objective integer programming model in a general deployment scenario, which is computationally NP-hard. We then propose a genetic algorithm-based approach, GACP, to find heuristic solutions in significantly reduced time. We investigate the effectiveness of GACP by performing extensive experiments on multiple deployment scenarios based on New York City OpenData. The results presented in the poster show that our approach obtains close to optimal cost solutions with significant time reduction.

**Poster 19: Accelerating Real-World Stencil Computations Using Temporal Blocking: Handling Sparse Sources and Receivers**
George Bisbas (Imperial College, London)
This paper concerns performance optimization in finite-difference solvers found in seismic imaging. We tackle a complicating factor that is arising in seismic inversion problems: the addition of waves injected from sources distributed sparsely over the 2D/3D domain, and the need for receivers that interpolate data measurements at a set of points, again distributed sparsely across the domain. In this work we show how to overcome this limitation. We introduce an algorithm for a loop nest transformation policy applied to wave propagation in order to improve data locality and optimize our cache memory use. Our algorithm uses an inspector/executor scheme capable of inspecting measurement operations at sparse locations and then compute the field updates. The ultimate goal of this ongoing work is to automate this scheme for stencil codes generated by Devito.


Tirthak Patel (Northeastern University), Devesh Tiwari (Northeastern University)

Large-scale computing systems are becoming increasingly more power-constrained, but these systems employ hardware over-provisioning to achieve higher system throughput because HPC applications often do not consume the peak power capacity of nodes. Unfortunately, focusing on system throughput alone can lead to severe unfairness among multiple concurrently-running applications. This paper introduces PERQ, a new feedback-based principled approach to improve system throughput while achieving fairness among concurrent applications.

**Poster 18: Using Non Volatile Memories to Build Energy- and Cost-Efficient Clusters**

Onkar Patil (North Carolina State University)

Non-Volatile Memory (NVM) is a byte-addressable, high capacity, high latency, and persistent form of memory that can extend the primary memory hierarchy by another level. It allows clusters to have significantly greater memory capacity per node. Intel’s Optane DC Persistent Memory Module (PMM) is a NVM device that can be used to increase the memory density of high performance computing (HPC) system. With higher memory density, we can run scientific computing applications with larger problem sizes on fewer compute nodes than on current HPC systems. Smaller HPC clusters will reduce the cost of running scientific simulations. We test our hypothesis by running a HPC application with large problem sizes on a single node with NVM and on multiple nodes connected by a high speed interconnect with equivalent amount of volatile memory. We compare the performance characteristics and power consumption of both systems.

**Poster 22: Fast Profiling-Based Performance Modeling of Distributed GPU Applications**
An increasing number of applications utilize GPUs to accelerate computation, with MPI responsible for communication in distributed environments. Existing performance models only focus on either modeling GPU kernels or MPI communication; few that do model the entire application are often too specialized for a single application and require extensive input from the programmer.

To be able to quickly model different types of distributed GPU applications, we propose a profiling-based methodology for creating performance models. We build upon the roofline performance model for GPU kernels and analytical models for MPI communication, with a significant reduction in profiling time. We also develop a benchmark to model 3D halo exchange that occurs in many scientific applications. Our proposed model for the main iteration loops of MiniFE achieves 6-7% prediction error on LLNL Lassen and 1-2% error on PSC Bridges, with minimal code inspection required to model MPI communication.

**Poster 26: Neural Networks for the Benchmarking of Detection Algorithms**  
Silvia Miramontes (Lawrence Berkeley National Laboratory; University of California, Berkeley)

There are several automated methods to detect objects from grayscale images. However, materials scientists still lack basic tools to compare different detection results, particularly when working with microtomography. This poster introduces FibCAM, a convolutional neural network (CNN)-based method using TensorFlow that allows benchmarking fiber detection algorithms. Our contribution is three-fold: (a) the design of a computational framework to compare automated fiber detection models with curated datasets through classification; (b) lossless data reduction by embedding prior knowledge into data-driven models; (c) a scheme to decompose computation into embarrassingly parallel processes for future analysis at scale. Our results show how FibCAM classifies different structures, and how it illustrates the material’s composition and frequency distribution of microstructures for improved interpretability of machine learning models. The proposed algorithms support probing the specimen content from gigabyte-sized volumes and enable pinpointing inconsistencies between real structures known a priori and results derived from automated detections.

**Poster 21: Optimization of GPU Kernels for Sparse Matrix Computations in Hypre**  
Chaoyu Zhang (Arkansas State University)

The acceleration of sparse matrix computations on GPUs can significantly enhance the performance of iterative methods for solving linear systems. In this work, we consider the kernels of Sparse Matrix
Vector Multiplications (SpMV), Sparse Triangular Matrix Solves (SpTrSv) and Sparse Matrix Matrix Multiplications (SpMM), which are often demanded by Algebraic Multigrid (AMG) solvers. With the CUDA and the hardware support of the Volta GPUs on Sierra, the existing kernels should be further optimized to fully take the advantage of the new hardware, and the optimizations have shown significant performance improvement. The presented kernels have been put in HYPRE for solving large scale linear systems on HPC equipped with GPUs. These shared-memory kernels for single GPU are the building blocks of distributed matrix operations required by the solver across multiple GPUs and compute nodes. The implementations of these kernels in Hypre and the code optimizations will be discussed.

**Poster 20: From IoT to Cloud: Research Platform for IoT/Cloud Experiments**

*Jinfeng Lin (University of Notre Dame)*

IoT studies leverage a wide range of lightweight hardware for collecting and processing data in the field. Limited by the resources on the devices, IoT systems have to interoperate with cloud platforms for addressing computation intensive tasks such as image processing, application backend supporting and centralized data storage. Therefore, a testbed for IoT/Cloud experiments should provide infrastructure for IoT to cloud communication, computation deployment, and hardware resource management. With these functionalities, users can focus on research problems without distraction from manually constructing experiment environments. Though cloud providers such as Google, Amazon, and Microsoft all provide IoT to Cloud solutions in general, this commercial model is not entirely compatible with research purpose. We propose a framework named Chameleon IoT testbed (CHIoT) that extends the Chameleon bare-metal cloud resources to support general IoT to Cloud experiments.

**Poster 32: OSU INAM: A Profiling and Visualization Tool for Scalable and In-Depth Analysis of High-Performance GPU-Enabled HPC Clusters**

*Pouya Kousha (Ohio State University)*

The lack of low-overhead and scalable monitoring tools have prevented a comprehensive study of efficiency and utilization of emerging NVLink-enabled GPU clusters. We address this by proposing and designing an in-depth, real-time analysis, profiling, and visualization tool for high-performance GPU-enabled clusters with NVLinks on the top of the OSU INAM. The proposed tool is capable of presenting a unified and holistic view of MPI-level and fabric level information for emerging NVLink-enabled high-performance GPU clusters. It also provides insights into the efficiency and utilization of underlying interconnects for different communication patterns. We also designed a low overhead and scalable modules to discover the fabric topology and gather fabric metrics by using different
levels of threading, bulk insertions and deletions for storage, and using parallel components for fabric discovery and port metric inquiry.

Zheng Miao (Clemson University)

High-Performance Computing systems must simultaneously address both resilience and power. In heterogeneous systems, the trade-offs between resilience and energy-efficiency are more complex for applications using both CPUs and GPUs. A deep understanding of the interplay among energy efficiency, resilience, and performance is required for heterogeneous systems to address them simultaneously.

In this work, we present a new framework for resilient and energy-efficient computing in GPU-accelerated systems. This framework supports partial or full redundancy and checkpointing for resilience, and provides users with flexible hardware resource selection, adjustable precision and power management to improve performance and energy-efficiency. We further perform CUDA-aware MPI to reduce resilience overhead, mainly in message communication between GPUs. Using CG as an example, we show that our framework provides about 40% time and 45% energy savings, comparing to simple extension of RedMPI, a redundancy based resilience framework for homogeneous CPU systems.

**Poster 25: Leveraging Smart Data Transfer and Metadirective in Adaptive Computing**
Anjia Wang (University of North Carolina, Charlotte; Lawrence Livermore National Laboratory)

In this work, we propose smart data transfer (SDT) along with extensions to metadirective and map constructs in OpenMP 5.0 to improve adaptive computing. The Smith-Waterman algorithm is used as an example, whose naïve implementation does not conduct data transfer efficiently. SDT is used to solve this issue with the following advantages: (1) SDT only transfers necessary data to GPU instead of the whole data set, resulting in 4.5x of speedup in our initial experiments. (2) Depending on computing vs. data transfer requirements of a program, SDT will transfer the output of each iteration from GPU to host either immediately or all together after the last GPU kernel call. (3) It supports large data exceeding GPU device memory’s size via data tiling. We propose to extend metadirective’s context selector to obtain similar improvement by enabling target enter/exit data and on-demand data access.
Thursday, November 21

3:30 pm - 5:00 pm

ACM Student Research Semi-Finalists Presentations

Poster 12: Investigating Input Sensitivity in GPU Graph Analytics with Machine Learning
Jacob M. Hope (Texas State University), Trisha Nag (Texas Tech University)

Graph algorithms are at the core of data-intensive applications. As such, efficient graph processing is of great importance. Irregularity in real-world graphs can make performance unpredictable and non-portable across different inputs and architectures. Given a type of graph, the same optimized implementation of an algorithm can produce performance numbers that differ by orders-of-magnitude. We conduct extensive analysis on a set of 1238 graphs to identify input-sensitive performance inefficiencies, including two that have not been studied: (i) register pressure and (ii) CPU-GPU data movement via demand paging. We then build a multiclass decision tree classifier that characterizes the irregular properties of graphs from our data and maps them to an optimal control parameterization at the compiler, system and algorithmic layers, that yield the highest overall algorithmic performance. We then integrate the classifier into a system where it will process a new graph and generate a kernel on the predicted optimal configuration.

Poster 13: Scalable Load-Balancing Concurrent Queues on Many-Core Architectures
Caleb Lehman (Ohio State University)

As the core counts of computing platforms continue to rise, parallel runtime systems with support for very fine-grained tasks become increasingly necessary to fully utilize the available resources. A critical feature of such task-based parallel runtime systems is the ability to balance work evenly and quickly across available cores. We highlight this by studying XTask, a custom parallel runtime system based on XQueue, which is a novel lock-less concurrent queuing system with relaxed ordering semantics that is geared to realizing scalability to hundreds of concurrent threads. We demonstrate the lack of adequate load balancing in the original XQueue design and present several solutions for improving load balancing. We also evaluate the corresponding improvements in
performance on two sample workloads, computation of Fibonacci numbers and computation of Cholesky factorization. Finally, we compare the performance of several versions of XTask along with several implementations of the popular OpenMP runtime system.

Poster 6: Analyzing the Performance of ZFP Compressed Arrays on HPC Kernels
Pavlo D. Triantafyllides (Clemson University)

Per-node memory capacity limits the maximal problem size of HPC applications. Naïve data compression alleviates the memory impact, but requires full decompression before the data is accessed. ZFP compressed arrays reduce the memory footprint, by independently compressing data in fixed sized blocks. Thus, decompressing individual blocks and caching them enables random access and a reduction in decompressions on the critical path. The performance of ZFP compressed arrays is dependent on several key variables: software cache size, cache policy, and compression rate. In this poster, we explore the sensitivity of these ZFP parameters on runtime performance for the matrix-matrix multiplication algorithm. Results show that selection of cache size, policy, and rate yields 8% performance improvement over the default ZFP configuration.

Poster 1: Early Experiences on OpenPOWER Architecture: Analysis of Billion-Scale Atomistic Datasets
Yuya Kawakami (Grinnell College, Oak Ridge National Laboratory)

We present performance analysis on OpenPOWER architecture of an algorithm to generate transversal views of atomistic models. The algorithm was implemented with data parallel primitives in NVIDIA Thrust for architecture portability. We report performance results on IBM Power9 CPUs (OpenMP, Intel Threading Blocks) and NVIDIA Volta GPUs (single and multi GPU). We also evaluate CUDA unified memory performance, exposed by NVIDIA RAPIDS Memory Manager library (RMM).

Poster 2: An Efficient Parallel Algorithm for Dominator Detection
Daniel Giger (University of Massachusetts, Lowell), Hang Liu (Stevens Institute of Technology)

In graph theory, a vertex v dominates a vertex u if every path from the entry vertex to u must go through vertex v. This algorithm is called dominator detection and holds a wide range of applications, such as compiler design, circuit testing, and social network analysis. While the performance of many other graph algorithms soars with respect to the increase of the hardware parallelism, dominator detection algorithm experiences very little advancement due to the hardship of parallelism. This work thus introduces an efficient parallel dominator detection algorithm that is
inspired by Breadth-First Search (BFS), which bests SEMI-NCA on large graphs.

Poster 9: Machine Specific Symbolic Code Generation
Robert King (University of Utah)

New discoveries in science and engineering are primarily driven by numerical simulations of underlying governing equations. Manually written codes to evaluate the main computational kernels lack portability, prone to human errors, ability to perform code optimizations due to the complexity of the underlying equations. In this work we present a symbolic code generation framework, which generates architecture optimized code for different platforms. As the driving application, we primarily use computational relativity where computations of Einstein equations become complicated due to the presence of curvature in spacetime.

The presented framework is based on SymPy with additional modules written to handle complicated partial differential equations (PDEs). The symbolically written equations are converted to a computational graph, which enables to perform architecture (cache, register optimizations) and language specific (SIMD vectorization, CUDA) optimizations. By computing common subexpressions, we reduce the number of compute operations needed, by storing them in temporary variables.

Poster 5: Evaluating Lossy Compressors for Inline Compression
Donald T. Elmore (Clemson University)

HPC applications require massive amounts of data. The data required is growing faster than memory capabilities. An example of this is pySDC, a framework for solving collocation problems iteratively using parallel-in-time methods. pySDC requires storing and exchange of 3D volume data for each parallel point in time. We evaluate several state-of-the-art lossy compressors such as SZ and ZFP for their applicability to inline compression for pySDC. We evaluate the compressors based on compression ratio, compression bandwidth, decompression bandwidth, and overall error introduced.

This poster evaluates state-of-the-art lossy compressors for use in pySDC; shows lossy data compression is an effective tool for reducing memory requirements for pySDC; and highlights current compression/decompression bandwidth is not fast enough for inline lossy compression yet. Results show using SZ with an error bound of 1e-5, we reduce the memory footprint by a factor of 311.99 while maintaining an acceptable level of loss.
Poster 3: A Deep Learning Approach to Noise Prediction and Circuit Optimization for Near-Term Quantum Devices
Alexander Zlokapa (California Institute of Technology), Alexandru Gheorghiu (California Institute of Technology)

Noisy intermediate-scale quantum (NISQ) devices face challenges in achieving high-fidelity computations due to hardware-specific noise. As a basis for noise mitigation, we develop a convolutional neural network noise model to estimate the difference in noise between a given pair of equivalent quantum circuits. On a classically simulated dataset of 1.6 million pairs of quantum circuits with a simplified noise model calibrated to IBM Q hardware, the deep learning approach shows a significant improvement in noise prediction over linear gate count models. A greedy peephole optimization procedure is proposed to minimize noise using the deep learning model as an objective function, showing further improvement in noise mitigation compared to commonly used gate count minimization heuristics.

Poster 10: Learning Page Access Patterns for Algorithms Programmed with GPU UVM
Bennett Cooper (Clemson University), Derek Rodriguez (Clemson University)

We aim to mitigate the performance bottleneck of migrating data between host and device memory in GPU applications by accurately predicting application access patterns using deep neural networks. We model the memory access pattern of any given application by collecting page faults that trigger data migration to the GPU and feed this time series as input to a neural network that outputs the next several page faults. We evaluate these predictions on the basis of what makes a useful prefetch in our context for the GPU. Current work has looked at trivial GPU applications, such as matrix operations, and moving toward real, complex applications. Our work will be presented by bringing attention to the predictive capability of our neural network on the current applications tested.

Poster 8: Mitigating Communication Bottlenecks in MPI-Based Distributed Learning
Abdullah B. Nauman (Ward Melville High School, Brookhaven National Laboratory)

Current commercial and scientific facilities generate and maintain vast amounts of complex data. While machine learning (ML) techniques can provide crucial insight, developing these models is often impractical on a single process. Distributed learning techniques mitigate this problem; however, current models contain significant performance bottlenecks. Here, we conduct a detailed performance analysis of MPI_Learn, a widespread distributed ML framework for high-energy
physics (HEP) applications, on the Summit supercomputer, by training a network to classify simulated collision events from high-energy particle detectors at the CERN Large Hadron Collider (LHC).

We conclude that these bottlenecks occur as a result of increasing communication time between the different processes, and to mitigate the bottlenecks we propose the implementation of a new distributed algorithm for stochastic gradient descent (SGD). We provide a proof of concept by demonstrating better scalability with results on 250 GPUs, and with hyperparameter optimization, show a ten-fold decrease in training time.

**Hearing Single- and Multi-Threaded Program Behavior**

*Mark Wissink (Calvin University), Joel Adams (Calvin University)*

There are many examples of visual program representation, but there is an unexplored area of sonic representation—audialization—of programs. The Thread Safe Audio Library (TSAL) provides the necessary tools to create such program audializations. TSAL is a platform-independent, object-oriented C++ library that provides thread safe classes for synthesizing sound from a program. By adding TSAL calls to a sequential or parallel program, its behavior can be represented with sound. Audializations require a unique interpretation that may greatly benefit auditory learners and other programmers who are looking for a different way to profile a sequential or parallel program.

**Poster 11: Walking the Cost-Accuracy Tightrope: Balancing Trade-Offs in Data-Intensive Genomics**

*Kathryn Leung (Princeton University), Meghan Kimball (DePaul University)*

Scientific applications often exhibit a trade-off between cost and accuracy. However, measuring and predicting cost and accuracy in a way that users can understand these trade-offs is challenging. To address these needs, we present predictive cost and accuracy models for data-intensive genomics applications. We use these models to create a trade-off graph, which researchers can use to selectively trade-off cost and accuracy.

**Poster 31: Designing High-Performance Erasure Coding Schemes for Next-Generation Storage Systems**

*Haiyang Shi (Ohio State University), Xiaoyi Lu (Ohio State University)*

Erasure Coding (EC) NIC offload is a promising technology for designing next-generation distributed
storage systems. To alleviate the EC overhead, various hardware-based Erasure Coding schemes have been proposed to leverage the advanced compute capabilities on modern HPC clusters. However, this poster has identified some critical limitations of current-generation EC NIC offload schemes on modern SmartNICs and the missing of exploiting heterogeneous hardware available on modern HPC clusters in parallel. To address these limitations, this poster first proposes a unified benchmark suite to benchmark, measure and characterize hardware-optimized erasure coders. Then, it proposes a Multi-Rail EC concept which enables upper-layer applications to leverage heterogeneous hardware to perform EC operations simultaneously. Finally, it proposes a new EC NIC offload paradigm based on the tripartite graph model, namely TriEC. TriEC supports both encode-and-send and receive-and-decode offload primitives efficiently.

Poster 27: A Framework for Measuring Hardware Gather-Scatter Support
Patrick Lavin (Georgia Institute of Technology), Jeffrey Young (Georgia Institute of Technology), Richard Vuduc (Georgia Institute of Technology)

In recent years, we have seen the re-addition of vector units to CPUs. While these units easily give speedups for easily vectorized applications with dense memory access, it can be hard to characterize how different access patterns will affect the performance of vectorized code.

We have developed Spatter, a benchmark which allows us to test and investigate the gather-scatter units available on current and upcoming hardware. The information that Spatter reveals to users is of use to everyone from hardware vendors who wish to compare gather-scatter units across platforms, to application developers who wish to test memory access pattern performance in their vectorized code.

Poster 30: A Fast and Efficient Incremental Approach toward Dynamic Community Detection
Neda Zarayeneh (Washington State University)

Community detection is a discovery tool to identify groups of vertices which are tightly connected, called communities. Most of the literature on this subject caters to the static use-case where the underlying network does not change. However, many emerging real-world use-cases give rise to a need to incorporate dynamic graphs as inputs.

We present a fast and efficient incremental approach toward dynamic community detection. The key contribution is a generic technique called Ω-screening, which examines the most recent batch of changes made to an input graph and selects a subset of vertices to reevaluate for potential community (re)assignment. This technique can be incorporated into any of the community detection
methods that use modularity as its objective function for clustering. For demonstration purposes, we incorporated the technique into two well-known community detection tools. Our experiments demonstrate that our approach is able to generate performance speedups without compromising on the output quality.

**Poster 29: Optimal Routing for a Family of Scalable Interconnection Networks**  
Zhipeng Xu (Stony Brook University; Sun Yat-sen University, Guangzhou, China)

We propose a scheme to construct a family of large and high-performance interconnection networks that are scalable, low-radix, minimum diameters. These networks, whose diameters grow linearly as their sizes grow exponentially, are generated by using the Cartesian products of smaller optimal networks of minimum diameters. For the smaller base networks, we design the vertex-balanced routing algorithm by considering the forwarding pressure at each vertex. Comparative benchmarks on a Beowulf cluster show significant improvement in performance after using the new routing algorithm. Each node of the new network generated from base graphs with low-diameter can also sustain balanced forwarding loadings if we apply optimal routing algorithms to the base network. Simulation results for larger networks show that the optimal routing algorithms achieve the gain of communication performance.

**Poster 16: perf-taint: Taint Analysis for Automatic Many-Parameter Performance Modeling**  
Marcin Copik (ETH Zurich)

Performance modeling is a well-known technique for understanding the scaling behavior of an application. Although the modeling process is today often automatic, it still relies on a domain expert selecting program parameters and deciding relevant sampling intervals. Since existing empirical methods attempt blackbox modeling, the decision on which parameters influence a selected part of the program is based on measured data, making empirical modeling sensitive to human errors and instrumentation noise. We introduce a hybrid analysis to mitigate the current limitations of empirical modeling, combining the confidence of static analysis with the ability of dynamic taint analysis to capture the effects of control-flow and memory operations. We construct models of computation and communication volumes that help the modeler to remove effects of noise and improve the correctness of estimated models. Our automatic analysis prunes irrelevant program parameters and brings an understanding of parameter dependencies which helps in designing the experiment.

**Poster 24: Fingerprinting Anomalous Computation with RNN for GPU-Accelerated HPC Machines**
Pengfei Zou (Clemson University)

This paper presents a workload classification framework that discriminates illicit computation from authorized workloads on GPU-accelerated HPC systems. As such systems become more and more powerful, they are exploited by attackers to run malicious and for-profit programs that typically require extremely high computing ability to be successful. Our classification framework leverages the distinctive signatures between illicit and authorized workloads, and explores machine learning methods to learn the workloads and classify them. The framework uses lightweight, non-intrusive workload profiling to collect model input data, and explores multiple machine learning methods, particularly recurrent neural network (RNN) that is suitable for online anomalous workload detection. Evaluation results on four generations of GPU machines demonstrate that the workload classification framework can tell apart the illicit authorized workloads with high accuracy of over 95%. The collected dataset, detection framework, and neural network models will be made available on GitHub.

Poster 33: Enhancing Neural Architecture Search with Speciation and Inter-Epoch Crossover
Matthew E. Baughman (University of Chicago)

As deep learning continues to expand into new areas of application, the demand for efficient use of our HPC resources increase. For new problem domains, new model architectures are developed through a neural architecture search (NAS), which consist of iteratively training many neural networks. To combat the computational waste and maximize compute efficiency for NAS, we demonstrate that the use of genetic algorithms with speciation can be used to both shorten training time and increase accuracy at each iteration.

Poster 14: Data Reuse Analysis for GPU Offloading Using OpenMP
Alok Mishra (Stony Brook University, Lawrence Livermore National Laboratory)

More researchers and developers desire to port their applications to GPU-based clusters, due to their abundant parallelism and energy efficiency. Unfortunately porting or writing an application for accelerators, such as GPUs, requires extensive knowledge of the underlying architectures, the application/algorithm and the interfacing programming model (e.g. OpenMP). Often applications spend a significant portion of their execution time on data transfer. Exploiting data reuse opportunities in an application can reduce its overall execution time. In this research, we present an approach to automatically recognize data reuse opportunities in an application which uses OpenMP for exploiting GPU parallelism, and consequently insert pertinent code to take advantage of data reuse on GPU. Using our approach we were able to retain reused data on the GPU and reduce the overall execution time of multiple benchmark application.
Poster 17: Exploiting Multi-Resource Scheduling for HPC
Yuping Fan (Illinois Institute of Technology)

High-performance computing (HPC) is undergoing significant changes. Next generation HPC systems are equipped with diverse global/local resources. HPC job scheduler plays a crucial role in efficient use of resources. However, traditional job schedulers are single-objective and fail to efficient use of other resources. In our previous work, we present a job scheduling framework named BBSched to schedule CPUs and burst buffers. As we are heading toward exascale computing, a variety of heterogeneous resources are deployed in HPC systems. In this poster, we extend BBSched for managing multiple resources beyond CPUs and burst buffers. We formulate multi-resource scheduling as a general multi-objective optimization (MOO) problem, present a heuristic method to solve the NP-hard MOO problem, and provide a preliminary evaluation for scheduling up to ten resources. The proposed multi-resource scheduling design is intended to enhance Cobalt, a production job scheduler deployed on HPC systems at Argonne Leadership Computing Facility (ALCF).

Poster 15: Cost-Aware Cloudlet Placement in Edge Computing Systems
Dixit Bhatta (University of Delaware)

A well-known challenge in Edge Computing is strategic placement of cloudlets. The fundamental goals of this challenge are to minimize the cloudlet deployment cost and to guarantee minimum latency to the users of edge services. We address this challenge by designing a cost-aware cloudlet placement approach that that ensures user latency requirements while covering all devices in the service region. We first mathematically formulate the problem as a multi-objective integer programming model in a general deployment scenario, which is computationally NP-hard. We then propose a genetic algorithm-based approach, GACP, to find heuristic solutions in significantly reduced time. We investigate the effectiveness of GACP by performing extensive experiments on multiple deployment scenarios based on New York City OpenData. The results presented in the poster show that our approach obtains close to optimal cost solutions with significant time reduction.

Poster 19: Accelerating Real-World Stencil Computations Using Temporal Blocking: Handling Sparse Sources and Receivers
George Bisbas (Imperial College, London)

This paper concerns performance optimization in finite-difference solvers found in seismic imaging.
We tackle a complicating factor that is arising in seismic inversion problems: the addition of waves injected from sources distributed sparsely over the 2D/3D domain, and the need for receivers that interpolate data measurements at a set of points, again distributed sparsely across the domain. In this work we show how to overcome this limitation. We introduce an algorithm for a loop nest transformation policy applied to wave propagation in order to improve data locality and optimize our cache memory use. Our algorithm uses an inspector/executor scheme capable of inspecting measurement operations at sparse locations and then compute the field updates. The ultimate goal of this ongoing work is to automate this scheme for stencil codes generated by Devito.

Poster 23: PERQ: Fair and Efficient Power Management of Power-Constrained Large-Scale Computing Systems
Tirthak Patel (Northeastern University), Devesh Tiwari (Northeastern University)

Large-scale computing systems are becoming increasingly more power-constrained, but these systems employ hardware over-provisioning to achieve higher system throughput because HPC applications often do not consume the peak power capacity of nodes. Unfortunately, focusing on system throughput alone can lead to severe unfairness among multiple concurrently-running applications. This paper introduces PERQ, a new feedback-based principled approach to improve system throughput while achieving fairness among concurrent applications.

Poster 18: Using Non Volatile Memories to Build Energy- and Cost-Efficient Clusters
Onkar Patil (North Carolina State University)

Non-Volatile Memory (NVM) is a byte-addressable, high capacity, high latency, and persistent form of memory that can extend the primary memory hierarchy by another level. It allows clusters to have significantly greater memory capacity per node. Intel’s Optane DC Persistent Memory Module (PMM) is a NVM device that can be used to increase the memory density of high performance computing (HPC) system. With higher memory density, we can run scientific computing applications with larger problem sizes on fewer compute nodes than on current HPC systems. Smaller HPC clusters will reduce the cost of running scientific simulations. We test our hypothesis by running a HPC application with large problem sizes on a single node with NVM and on multiple nodes connected by a high speed interconnect with equivalent amount of volatile memory. We compare the performance characteristics and power consumption of both systems.

Poster 22: Fast Profiling-Based Performance Modeling of Distributed GPU Applications
Jaemin Choi (University of Illinois, Lawrence Livermore National Laboratory)
An increasing number of applications utilize GPUs to accelerate computation, with MPI responsible for communication in distributed environments. Existing performance models only focus on either modeling GPU kernels or MPI communication; few that do model the entire application are often too specialized for a single application and require extensive input from the programmer.

To be able to quickly model different types of distributed GPU applications, we propose a profiling-based methodology for creating performance models. We build upon the roofline performance model for GPU kernels and analytical models for MPI communication, with a significant reduction in profiling time. We also develop a benchmark to model 3D halo exchange that occurs in many scientific applications. Our proposed model for the main iteration loops of MiniFE achieves 6-7% prediction error on LLNL Lassen and 1-2% error on PSC Bridges, with minimal code inspection required to model MPI communication.

**Poster 26: Neural Networks for the Benchmarking of Detection Algorithms**
Silvia Miramontes (Lawrence Berkeley National Laboratory; University of California, Berkeley)

There are several automated methods to detect objects from grayscale images. However, materials scientists still lack basic tools to compare different detection results, particularly when working with microtomography. This poster introduces FibCAM, a convolutional neural network (CNN)-based method using TensorFlow that allows benchmarking fiber detection algorithms. Our contribution is three-fold: (a) the design of a computational framework to compare automated fiber detection models with curated datasets through classification; (b) lossless data reduction by embedding prior knowledge into data-driven models; (c) a scheme to decompose computation into embarrassingly parallel processes for future analysis at scale. Our results show how FibCAM classifies different structures, and how it illustrates the material's composition and frequency distribution of microstructures for improved interpretability of machine learning models. The proposed algorithms support probing the specimen content from gigabyte-sized volumes and enable pinpointing inconsistencies between real structures known a priori and results derived from automated detections.

**Poster 21: Optimization of GPU Kernels for Sparse Matrix Computations in HYPRE**
Chaoyu Zhang (Arkansas State University)

The acceleration of sparse matrix computations on GPUs can significantly enhance the performance of iterative methods for solving linear systems. In this work, we consider the kernels of Sparse Matrix Vector Multiplications (SpMV), Sparse Triangular Matrix Solves (SpTrSv) and Sparse Matrix Matrix
Multiplications (SpMM), which are often demanded by Algebraic Multigrid (AMG) solvers. With the CUDA and the hardware support of the Volta GPUs on Sierra, the existing kernels should be further optimized to fully take the advantage of the new hardware, and the optimizations have shown significant performance improvement. The presented kernels have been put in HYPRE for solving large scale linear systems on HPC equipped with GPUs. These shared-memory kernels for single GPU are the building blocks of distributed matrix operations required by the solver across multiple GPUs and compute nodes. The implementations of these kernels in Hypre and the code optimizations will be discussed.

**Poster 20: From IoT to Cloud: Research Platform for IoT/Cloud Experiments**  
*Jinfeng Lin (University of Notre Dame)*

IoT studies leverage a wide range of lightweight hardware for collecting and processing data in the field. Limited by the resources on the devices, IoT systems have to interoperate with cloud platforms for addressing computation intensive tasks such as image processing, application backend supporting and centralized data storage. Therefore, a testbed for IoT/Cloud experiments should provide infrastructure for IoT to cloud communication, computation deployment, and hardware resource management. With these functionalities, users can focus on research problems without distraction from manually constructing experiment environments. Though cloud providers such as Google, Amazon, and Microsoft all provide IoT to Cloud solutions in general, this commercial model is not entirely compatible with research purpose. We propose a framework named Chameleon IoT testbed (CHIoT) that extends the Chameleon bare-metal cloud resources to support general IoT to Cloud experiments.

**Poster 32: OSU INAM: A Profiling and Visualization Tool for Scalable and In-Depth Analysis of High-Performance GPU-Enabled HPC Clusters**  
*Pouya Kousha (Ohio State University)*

The lack of low-overhead and scalable monitoring tools have prevented a comprehensive study of efficiency and utilization of emerging NVLink-enabled GPU clusters. We address this by proposing and designing an in-depth, real-time analysis, profiling, and visualization tool for high-performance GPU-enabled clusters with NVLinks on the top of the OSU INAM. The proposed tool is capable of presenting a unified and holistic view of MPI-level and fabric level information for emerging NVLink-enabled high-performance GPU clusters. It also provides insights into the efficiency and utilization of underlying interconnects for different communication patterns. We also designed a low overhead and scalable modules to discover the fabric topology and gather fabric metrics by using different levels of threading, bulk insertions and deletions for storage, and using parallel components for
Zheng Miao (Clemson University)

High-Performance Computing systems must simultaneously address both resilience and power. In heterogeneous systems, the trade-offs between resilience and energy-efficiency are more complex for applications using both CPUs and GPUs. A deep understanding of the interplay among energy efficiency, resilience, and performance is required for heterogeneous systems to address them simultaneously.

In this work, we present a new framework for resilient and energy-efficient computing in GPU-accelerated systems. This framework supports partial or full redundancy and checkpointing for resilience, and provides users with flexible hardware resource selection, adjustable precision and power management to improve performance and energy-efficiency. We further perform CUDA-aware MPI to reduce resilience overhead, mainly in message communication between GPUs. Using CG as an example, we show that our framework provides about 40% time and 45% energy savings, comparing to simple extension of RedMPI, a redundancy based resilience framework for homogeneous CPU systems.

Poster 25: Leveraging Smart Data Transfer and Metadirective in Adaptive Computing
Anjia Wang (University of North Carolina, Charlotte; Lawrence Livermore National Laboratory)

In this work, we propose smart data transfer (SDT) along with extensions to metadirective and map constructs in OpenMP 5.0 to improve adaptive computing. The Smith-Waterman algorithm is used as an example, whose naïve implementation does not conduct data transfer efficiently. SDT is used to solve this issue with the following advantages: (1) SDT only transfers necessary data to GPU instead of the whole data set, resulting in 4.5x of speedup in our initial experiments. (2) Depending on computing vs. data transfer requirements of a program, SDT will transfer the output of each iteration from GPU to host either immediately or all together after the last GPU kernel call. (3) It supports large data exceeding GPU device memory’s size via data tiling. We propose to extend metadirective’s context selector to obtain similar improvement by enabling target enter/exit data and on-demand data access.
Awards Presentation

Tuesday, November 19

3:30 pm - 4:15 pm

Test of Time Award Presentation

From Dense Linear Algebra to GPU Occupancy
Vasily Volkov (Nvidia Corporation), James Demmel (University of California, Berkeley)

The CUDA programming model was introduced in 2007 and featured a number of new concepts, such as occupancy and shared memory. In this work, we considered performance implications of these concepts in application to dense matrix factorizations. Our findings were contrarian to the widely accepted recommendations of the day. (i) We found a performance optimization pattern that leads to lower occupancy, whereas it was recommended to maximize occupancy in order to hide memory latencies. (ii) We found that instruction-level parallelism contributes to latency hiding on GPUs, which was believed to be not the case. (iii) We found that performance can be improved by using massive register blocking, whereas it was recommended to minimize register use to maximize occupancy. (iv) We found that shared memory is slower than registers and the use of the latter should be favored over the former, if possible. These novel insights led us to a design of the matrix multiply routine that substantially outperformed the state-of-the-art vendor BLAS library. The optimization pattern we pointed out is found today in many high-performance GPU codes.

Wednesday, November 20

8:30 am - 10:00 am

ACM and IEEE-CS Award Presentations

Session Description: The SC Conference showcases the best and brightest in HPC, both the rising stars and those who have made a lasting impression. A variety of awards from the computing
societies are awarded during the SC Conference, including the ACM/IEEE-CS Ken Kennedy, IEEE-CS Seymour Cray Computer Engineering, and IEEE-CS Sidney Fernbach Memorial Awards. This session will allow the winners of those awards an opportunity to present to SC attendees.

2019 IEEE Seymour Cray Computer Engineering Award
David Kirk (Nvidia Corporation (retired))

Origins of GPU Computing

GPU Computing didn’t just happen. The origin story of GPUs starts with VGA controllers and graphics accelerators. Texture mapping, multi-texture capabilities and frame buffer blending added a very small amount of programmability. Vertex and pixel shaders added real programming – a primitive programming model, but highly parallel processing for many vertices and pixels. The GPGPU movement recognized that GPUs were among the most powerful processors available, but very hard to program. For GPUs to be generally useful, high level programming languages and features were needed. Each step toward increasing capability required ideas, vision, and the hard work of many people who shared that vision. A modern GPU is a mainstream processor that is pervasive and ubiquitous in HPC, data centers, gaming, and many other applications.

2019 ACM/IEEE-CS Ken Kennedy Award
Geoffrey Fox (Indiana University)

Parallel Computing: from Digital Twins to Grids to Deep Learning

We describe a community and personal journey in parallel computing starting (1980-1995) with simulations (perhaps now called digital twins) enabled to go from an at best illustrative two-dimensional model to the realistic three-dimensional case. We had computational science and the third paradigm of scientific study.

Then came the World Wide Web and the programming on it leading the Grid computing initiative (1995-2010). I still remember a caustic referee report from those early days that HTTP was too slow to be serious. We learned in the OGF (Open Grid Forum) how to build communities and eScience was a key theme. Distributed computing was king and pleasing parallel was coined.

Then came a new revolution (2010-?) with Big Data and the fourth paradigm of computing built on data science as the academic underpinning. Distributed computing remained but the Grid
metamorphosed as data processing in the edge and fog. Parallel computing still continues in exascale scientific discovery but also in parallel deep learning.

We muse about this progression noting some common themes -- natural high-level parallel programming remains hard, an area where I had the pleasure of joint research with Ken Kennedy -- and messaging is critical in many different scenarios. We look to a spectacular innovative future where deep learning, simulations and parallel computing synergistically run together; each enhances the other to orders of magnitude better capabilities and performance.

2019 IEEE Sidney Fernbach Award
Alan Edelman (Massachusetts Institute of Technology (MIT))

How the Julia Language Can Fulfill the Promise of Supercomputing

During the first phase of HPC, the emphasis has been on performance above all else. The thinking that carried the day was that you can hack hard enough, bring up performance, and earn your brownie points. The Julia language, through abstractions and multiple dispatch technology, is making it possible to build supercomputing applications more readily in ways that can manage the complexities of modern supercomputers.

We will explain how Julia is transforming HPC in such areas as Scientific Machine Learning, Climate Science, and Medicine. We will conclude with the notion that the one true metric for HPC should not be performance; it should be the number of users. Performance, productivity, scalability, reproducibility, composability and other obvious and non-obvious metrics can then be subsumed by this “prime directive”.

3:30 pm - 5:00 pm

Gordon Bell Prize Finalist Session 1

A Data-Centric Approach to Extreme-Scale Ab Initio Dissipative Quantum Transport Simulations
Alexandros Nikolaos Ziogas (ETH Zurich)

The computational efficiency of a state of the art ab initio quantum transport (QT) solver, capable of revealing the coupled electrothermal properties of atomically-resolved nano-transistors, has been improved by up to two orders of magnitude through a data centric reorganization of the application.
The approach yields coarse- and fine-grained data-movement characteristics that can be used for performance and communication modeling, communication avoidance, and dataflow transformations. The resulting code has been tuned for two top-6 hybrid supercomputers, reaching a sustained performance of 85.45 Pflop/s on 4,560 nodes of Summit (42.55% of the peak) in double precision, and 90.89 Pflop/s in mixed precision. These computational achievements enable the restructured QT simulator to treat realistic nanoelectronic devices made of more than 10,000 atoms within a 14× shorter duration than the original code needs to handle a system with 1,000 atoms, on the same number of CPUs/GPUs and with the same physical accuracy.

Fast, Scalable and Accurate Finite-Element Based Ab Initio Calculations Using Mixed Precision Computing: 46 PFLOPS Simulation of a Metallic Dislocation System
Vikram Gavini (University of Michigan)

Accurate large-scale first principles calculations based on density functional theory (DFT) in metallic systems are prohibitively expensive due to the asymptotic cubic scaling computational complexity with number of electrons. Using algorithmic advances in employing finite-element discretization for DFT (DFT-FE) in conjunction with efficient computational methodologies and mixed precision strategies, we delay the onset of this cubic scaling by significantly reducing the computational prefactor while increasing the arithmetic intensity and lowering the data movement costs. This has enabled fast, accurate, and massively parallel DFT calculations on large-scale metallic systems on both many-core and heterogeneous architectures, with time-to-solution being an order of magnitude faster than state-of-the-art plane-wave DFT codes. We demonstrate an unprecedented sustained performance of 46 PFLOPS (27.8% peak FP64 performance) on a dislocation system in Magnesium containing 105,080 electrons using 3,800 GPU nodes of Summit supercomputer, which is the highest performance to date among DFT codes.

Thursday, November 21
12:45 pm - 1:30 pm
SC19 Awards Ceremony

Session Description: The SC19 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. Everyone with an SC19 badge is welcome to attend.
Tuesday, November 19

12:15 pm - 1:15 pm

MPI 4.0 Is Coming - What Is in It and What Is next?
Martin Schulz (Technical University Munich)

The Message Passing Interface (MPI) API is the most dominant programming approach 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 insights into the current topics discussed in the forum, plans and timelines for upcoming versions of MPI, 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.

12:15 pm - 1:15 pm

The IO-500 and the Virtual Institute of I/O
George Markomanolis (Oak Ridge National Laboratory)

The IO500 is quickly becoming the de facto benchmarking standard for HPC storage. Developed three years ago, the IO500 has released four official lists so far. A BoF highlight is the presentation of the fifth IO-500 list. The general purpose of this BoF is to foster the IO500 and VI4IO communities to ensure forward progress towards the common goals of creating, sharing, and benefiting from a large corpus of shared storage data. We also serve as a repository of detailed information about production storage system architectures over time as a knowledge base for other researchers and system designers to use.
Machine Learning and HPC in Pharma Research and Development

Mohammad Shaikh (Bristol-Myers Squibb Company)

Machine learning and deep learning methods are increasingly being applied in the pharmaceutical industry. The usage is growing exponentially and expanding to more business areas. Discussions will focus on current trends, the many challenges and approaches in applying ML and DL to build insightful models. The ever-increasing scale and velocity of data creates challenges to model accuracies, model refinement and computational scale. Key challenges include network latency and multi-terabyte data sets. HPC is vital to building and running the models in a meaningful timeframe. The interactive panel discussion will focus on experiences and insights by panelists and audience members.

12:15 pm - 1:15 pm

19th Graph500 List

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!). The Graph500 list has grown to over 223 entries and has demonstrated the challenges of even simple analytics. The new SSSP kernel introduced at SC17 has increased the benchmark’s overall difficulty. This BOF will unveil the Graph500 lists, discuss the new streaming kernel, and enhance the new energy metrics the Green Graph500. It will offer a forum for community and provide a rallying point for data intensive supercomputing problems.

12:15 pm - 1:15 pm

SIGHPC Annual Member Meeting

John West (Texas Advanced Computing Center (TACC))

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 will also be available.
12:15 pm - 1:15 pm

Extreme-Scale Scientific Software Stack (E4S)

Sameer Shende (University of Oregon)

The Extreme-Scale Scientific Software Stack (E4S) is a community effort to provide open source software packages for developing, deploying, and running scientific applications on HPC platforms. It provides both source builds as well as container images of a broad collection of HPC software packages. This E4S BoF will provide an update of the current state of the software stack and discuss plans for the future with the wider supercomputing community.

12:15 pm - 1:15 pm

User-Level Threads for Performant and Portable HPC

Stephen Olivier (Sandia National Laboratories)

User-level threading (ULT) libraries serve as flexible frameworks on which to build higher level programming languages and libraries for efficient exploitation of current and future multicore and manycore architectures. This BoF session aims to bring together researchers, developers, vendors and other enthusiasts interested in ULT and tasking models to understand the current state of art and requirements of the broader community. Cross-pollination of ideas among these stakeholders will advance the field, aid in developing common solutions for key emerging technical challenges, and improve the end-user experience.

12:15 pm - 1:15 pm

Ask.CI, the Q&A Platform for Research Computing - Recent Developments

Julie Ma (Massachusetts Green High Performance Computing Center (MGHPCC))

Launched in 2018, Ask.CI aggregates answers to a broad spectrum of questions that are commonly asked by the research computing community, creating a shared, archived, publicly-searchable knowledge base.

Establishing a Q&A site of this nature requires some tenacity. While Ask.CI has gained traction in the year since its launch, attracting nearly 150,000 page views, hundreds of contributors, and
worldwide participation, we are always seeking ways to grow our audience. We recently introduced "locales", institution-specific subcategories where institutions/communities of practice can post FAQs relevant to their constituents. Participants in the Locales pilot will share experiences and future plans at the BoF.

12:15 pm - 1:15 pm

**Quantum Computing at the DOE Laboratories: Status and Future Directions**

*Travis Humble (Oak Ridge National Laboratory)*

Quantum computing is developing rapidly as a key research topic with the potential to influence future generations of high-performance computing as well as other information technology. Quantum computers are also anticipated to enable a variety of new computing applications that address the multidisciplinary missions of the Department of Energy (DOE) for scientific discovery and national security. This BoF will focus on the role of quantum computing in research at the DOE national laboratories and the growing interactions with the broader HPC user community as well as exciting future research directions that arise from these new relationships.

12:15 pm - 1:15 pm

**Energy Efficiency Considerations for HPC Procurements**

*Gert Svensson (KTH Royal Institute of Technology, PDC Center for High Performance Computing)*

The goal when procuring HPC systems is to identify optimal solutions to technical and financial targets. In this BoF, we present some energy efficiency highlights from recent procurements from all over the world which have been analyzed and documented. One of the pre-exaflop sites from Europe will describe their plans to take energy-efficiency into account in an upcoming procurement, a presentation from the United States addresses special considerations related to procuring an energy-efficient container-based system, and a presentation from Japan discusses the procurement of a system with heavy energy-use instrumentation for data analytics. Interaction with the presenters is highly encouraged.

12:15 pm - 1:15 pm

**Unified Communication X (UCX) Community**
Gilad Shainer (Mellanox Technologies)

In order to exploit the capabilities of new HPC systems and to meet their demands in scalability, communication software needs to scale on millions of cores and support applications with adequate functionality. UCX is a collaboration between industry, national labs and academia that consolidates that provides a unified open-source framework.

The UCX project is managed by the UCF consortium (http://www.ucfconsortium.org/) and includes members from LANL, ANL, Ohio State University, AMD, ARM, IBM, Mellanox, NVIDIA and more. The session will serve as the UCX community meeting, and will introduce the latest development to HPC developers and the broader user community.

12:15 pm - 1:15 pm

Power API

Ryan Grant (Sandia National Laboratories)

The HPC community faces considerable constraints on power and energy of HPC installations. A standardized, vendor-neutral API for power measurement and control is needed for portable solutions to these issues at the various layers of the software stack. In this BoF, we discuss the Power API; a community standardized API for measurement and control of power/energy. The BoF will introduce newcomers to these efforts and provide clear differentiation with other ongoing projects. An interactive panel discussion with experts from involved organizations will facilitate discussions on the Power API with ample time for audience questions and comments.

12:15 pm - 1:15 pm

Breaking the Old Rule that HPL Goes with the Pace of the Slowest Node

Dmitry Nemirov (Intel Corporation)

Modern HPC clusters are becoming increasingly heterogeneous both explicitly and implicitly. Top500 race becomes more challenging and heterogeneous tests is a new trend. We will present examples of Intel distribution for HPL use for various real life top500-class machines. For example, a system with two significantly different types of compute nodes with each subset using different interconnect. We will show simpler cases with only compute nodes having configuration differences. In all of these cases, it is possible to achieve combined HPL score. This abstract has a
showcase and interactive parts and will be interesting to all who run HPL frequently.

12:15 pm - 1:15 pm

**Exchanging Best Practices in Supporting Computational and Data-Intensive Research**

*Rudolf Eigenmann (University of Delaware)*

This BOF aims to bring together:

* Researchers developing and using computational and data-intensive (CDI) applications,

* Those who assist these researchers with expertise in CDI technology,

* Developers of tools supporting the development and use of CDI applications.

The BoF will feature brief presentations of a number of panelists who have been involved in creating and providing application support to domain researchers. The audience is then invited to discuss how best practices can be identified and disseminated with the goal of accelerating computational and data-intensive research in the best possible way.

12:15 pm - 1:15 pm

**SAGE2 11th Annual International SC BOF: Scalable Amplified Group Environment for Global Collaboration**

*Jason Leigh (University of Hawaii at Manoa)*

SAGE2 (Scalable Amplified Group Environment) is the de facto operating system for managing Big Data on tiled display walls, providing the scientific community with persistent visualization and collaboration services for global cyberinfrastructure. SAGE2 is an open-source, web-based, user-centered platform for local or distributed groups of researchers, educators and students to communicate information and ideas as if they are in the same room. This year’s BoF provides a forum to provide new and potential users with an overview of new features and enhancements, and to elicit community input on use cases and roadmaps.
12:15 pm - 1:15 pm

**Leading Inclusively - Redefining Effective Leadership in HPC**

Kelly Nolan (Talent Strategy Institute)

Underrepresentation of certain groups of people in HPC leadership is causing acute talent acquisition and retention challenges in the overall pool of HPC professionals. Research shows that future leaders in science, technology, engineering and math (STEM) must master cultural awareness, be aware of the impact of bias, and create a culture of inclusivity. In this interactive BoF we will present the state of leadership in STEM and discuss how HPC can adopt a more inclusive definition of leadership to ensure the sector is able to meet its talent targets and remain competitive.

5:15 pm - 6:45 pm

**The Future of NSF Supported Advanced Cyberinfrastructure**

Manish Parashar (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 (OAC) will update attendees with a blueprint for the national CI ecosystem for science and engineering and discuss the latest funding opportunities across all aspects of the CI ecosystem. Presentations will also discuss Leadership Class Computing, the Research Core program, Advanced Computing Systems and Services, and recent workshops. Substantial time will be devoted to Q&A between attendees and NSF staff.

5:15 pm - 6:45 pm

**Lustre Community BoF: Lustre in HPC, AI, and the Cloud**

Stephen Simms (Indiana University)

Lustre is the leading open-source and open-development file system for HPC. Around two thirds of the top 100 supercomputers use Lustre. It is a community developed technology with contributors from around the world.
Lustre currently supports many HPC infrastructures beyond scientific research, such as financial services, energy, manufacturing and life sciences. Lustre clients are available for a variety of technologies, such as x86, POWER, and Arm. At this BoF, Lustre developers, administrators, and solution providers will gather to discuss recent Lustre developments and challenges, including the role of Lustre in AI and its use in Cloud environments.

5:15 pm - 6:45 pm

Cloud Federation for Large-Scale Collaborations

Martial Michel (Data Machines Corporation)

With the availability for public comments of the "The NIST Cloud Federation Reference Architecture" SP500 draft, we wanted to renew the conversation about the current cloud ecosystem; one in which cloud providers do not interoperate. This has led to a growing recognition that the lack of cloud federation in a landscape of multiple independent cloud providers is a technical and economic challenge. The federation of clouds is essential and necessary enabling technology for the development and translation of innovations in IoT, high-performance computing, distributed big-data analytics, and global scientific collaboration.

5:15 pm - 6:45 pm

Can Arm Take the Lead in HPC?

John Linford (ARM Ltd)

Arm-based processors have gained substantial traction in the HPC community. Sandia’s “Astra” is #156 on the Top500, and projects like the Japanese Post-K “Fugaku”, European Mont-Blanc, U.S. DOE Vanguard, and UK GW4/EPSRC are strong proof points. HPC system integrators like Atos, Cray, Gigabyte, and HPE have Arm commercial offerings, but will Arm become an HPC leader? This BoF brings together experts and luminaries to share their experiences with Arm, discuss the remaining technical and ecosystem challenges, consider the role of codesign in HPC, discuss progress, and lay out a vision for the future state of Arm in the HPC community.

5:15 pm - 6:45 pm

Heterogeneous and Distributed ISO C++ for HPC Status and Directions
Hal Flinkel (Argonne National Laboratory)

After last two year's (SC17, SC18) successful Heterogeneous & Distributed Computing in C++ for HPC BoF, there was popular demand for continuing updates on the progress of adding these capabilities into ISO C++. This includes task dispatch with executors and the property mechanism, data layout, affinity, error handling and Asynchronous execution.

We have also finalized C++20, and this BoF will provide updates on what supports distributed and heterogeneous computing from active participants in the standardization process. We will also look ahead on what is possible for C++20 and for C++23.

5:15 pm - 6:45 pm

HDF5 and Its role in Exascale, Cloud, and Object Stores

Elena Pourmal (HDF Group)

We will provide a forum for the HDF5 user community to learn about HDF5's role in moving science applications to Exascale systems, the Cloud and to Object Stores, and to share initiatives in this area. Elena Pourmal will present HDF5 features that target Exascale systems, the Cloud and Object Stores, HDF5’s role in the DOE’s ECP and EOD projects, and the HDF5 roadmap. Quincey Koziol and Suren Byna will moderate a panel with representatives from research, commercial, and government organizations who will present case studies on how they leverage or plan to leverage HDF technologies for HPC and the Cloud.

5:15 pm - 6:45 pm

Enabling Data Services for HPC

Jerome Soumagne (HDF Group)

Distributed data services can enhance HPC productivity by providing storage, analysis, and visualization capabilities not otherwise present in conventional parallel file systems. Such services are difficult to develop, maintain, and deploy in a scientific workflow, however, due to the complexities of specialized HPC networks, RDMA data transfers, protocol encoding, and fault tolerance.

This BoF will bring together a growing community of researchers, developers, vendors, and facility
operators who are either using or developing HPC data services. Participants will share and discuss practical experiences, implementation examples, new features, and best practices to construct and deploy production-quality, high-performance distributed services.

5:15 pm - 6:45 pm

AI Bench: Toward a Comprehensive AI Benchmark Suite for HPC, Datacenter, Edge, and IoT

Jianfeng Zhan (Institute of Computing Technology, Chinese Academy of Sciences)

As diverse communities pay great attention to innovative AI or machine learning algorithms, architecture, and systems, the pressure of benchmarking rises. However, complexity, diversity, frequently changed workloads, and rapid evolution of AI workloads and systems raise great challenges in AI benchmarking. The aim of this BoF is to discuss how to build a comprehensive AI benchmark suite across different communities with an emphasis on data and workload distributions among HPC, data center, Edge, and IoT.

5:15 pm - 6:45 pm

Designing and Building Next-Generation Computer Systems for Deep Learning

Volodymyr Kindratenko (National Center for Supercomputing Applications (NCSA), University of Illinois)

Deep learning (DL) heavily relies on fast hardware and parallel algorithms to train complex neural networks. This BoF will bring together researchers and developers working on the design of next-generation computer systems for DL and parallel DL algorithms that can exploit the potential of these new systems. Research teams working on major deep learning systems deployed in the field will be invited to discuss latest hardware and software trends and to exchange views on the role Artificial Intelligence (AI) in general and DL in particular will play in the near future for big data analytics and HPC applications.

5:15 pm - 6:45 pm

Ceph in HPC Environments
Benjamin Lynch (University of Minnesota)

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. This BoF session will bring together Ceph implementers to share their deployment experiences, as well as provide feedback to the developer community on needed features and enhancements specific to the HPC community.

5:15 pm - 6:45 pm

Looking Ahead to Tape Technology’s Role in HPC – Future Advancements, Challenges, and Roadmap

Ed Childers (IBM Corporation)

A quick web search could convince someone that cloud has taken over the computing and storage markets, and that tape storage has been utterly abandoned. Tape storage has existed for almost 50 years and was declared “dead” 30 years ago when disk was the leading technology. As an archiving and backup technology, tape remains in prominent use, with every Fortune 100 company still using a tape library. Join us to discuss how tape technology has advanced, as well as how its future roadmap will fuel innovation and enable organizations to access new markets while successfully addressing changing business needs.

5:15 pm - 6:45 pm

Americas HPC Collaboration

Carlos Barrios Hernandez (Advanced Computing Service for Latin America and the Caribbean (SCALAC); Industrial University of Santander, Colombia)

This third BoF Americas HPC Collaboration seeks to showcase collaboration opportunities and experiences between different HPC Networks and Laboratories from countries of the American continent. After the first and second meeting in SC15 and SC17, the goal of this BoF is to show the current state of the art in continental collaboration in HPC, latest developments of regional collaborative networks and, to update the roadmap for the next two years for the Americas HPC partnerships.
5:15 pm - 6:45 pm

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

Jack Wells (Oak Ridge National Laboratory)

OpenACC, a well-established directive-based high-level parallel programming model designed to provide an easy on-ramp to parallel computing on CPUs, GPUs and other devices. The user-friendly programming model has facilitated the acceleration of ~200 applications including FV3, COSMO, GTS, M3D-C1, E3SM, ADSCFD, VASP on multiple platforms and has also seen as an entry-level programming model for the top supercomputers (Top500 list) such as Summit, Sunway Taihulight, and Piz Daint. As in previous years, this BoF invites scientists and programmers to discuss their experiences adopting OpenACC for scientific applications, learn about the roadmaps from implementers and the latest developments in the specification.

5:15 pm - 6:45 pm

Machine-Learning Hardware: Architecture, System Interfaces, and Programming Models

Pete Beckman (Argonne National Laboratory)

Recent years have seen a surge of investment in AI chip companies worldwide. These companies are however mostly targeting applications outside of the scientific computing community. As the use of ML accelerates in the HPC field itself, there is concern that the scientific community should influence the design of this new specialized hardware. In this BoF, we propose to let the community and select vendors engage on questions related to programming models, system interfaces, and architecture trade-offs of these chips.

5:15 pm - 6:45 pm

OpenSHMEM: State of the Union 2019

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 this API across multiple
architectures/implementations. This BoF brings together the OpenSHMEM community to present the latest accomplishments since the release of the 1.4 specification, and discuss future directions for the OpenSHMEM community as we develop version 1.5. 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.

5:15 pm - 6:45 pm

Global Research Platform (GRP): Creating Worldwide Advanced Services and Infrastructure for Science

Joe Mambretti (Northwestern University)

International scientific initiatives collaborate on building, accessing and analyzing data from one-of-a-kind advanced instruments in unique locations around the globe, connected to remote computational, data storage, and visualization systems at speeds of gigabits and terabits per second. The Global Research Platform (GRP) focuses on the design, implementation, and operation strategies for next-generation distributed services and network infrastructure, on a global scale, to facilitate data transfer and accessibility. GRP BoF presentations and discussions address global science drivers and their requirements and describe emerging Research Platforms worldwide that are developing and customizing data fabrics and distributed cyberinfrastructure to support data-intensive scientific workflows.

5:15 pm - 6:45 pm

Community Building for Sustainable and Scalable HPC Outreach

Weronika Filinger (Edinburgh Parallel Computing Centre, University of Edinburgh)

The HPC community has consistently identified public outreach and diversity promotion as vital to the growth of the field. And yet, most people engaged in HPC outreach don’t have specific training in public engagement or education, and do it on a volunteer basis. For HPC Outreach to fulfill its role, it needs to be a community-wide effort. We invite anyone interested in outreach to attend and become a part of this active community. The participants are encouraged to bring their demos, activities, ideas and experiences, which then will be categorized and added to the Outreach repository. Come and get involved!
TOP500 Supercomputers

Erich Strohmaier (Lawrence Berkeley National Laboratory)

The TOP500 list of supercomputers serves as a “Who’s Who” in the field of High Performance Computing (HPC). It started as a list of the most powerful supercomputers in the world and has evolved to a major source of information about trends in HPC. The 54th TOP500 list will be published in November 2019 just in time for SC19.

This BoF will present detailed analyses of the TOP500 and discuss the changes in the HPC marketplace during the past years. The BoF is meant as an open forum for discussion and feedback between the TOP500 authors and the user community.

Wednesday, November 20

12:15 pm - 1:15 pm

Open MPI State of the Union 2019

Jeffrey Squyres (Cisco Systems)

Open MPI continues to drive the start 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.

12:15 pm - 1:15 pm

Impacting Cancer with HPC: Challenges and Opportunities

Patricia Kovatch (Icahn School of Medicine at Mount Sinai)
The rate of advancement in the application of AI is creating tremendous opportunities for the use of HPC in advancing research and clinical applications for cancer. Simultaneously, the importance of data and information necessary for successful AI applications to advance research efforts has also grown dramatically. The BoF will focus on opportunities in cancer research and clinical applications for HPC, emphasizing the data challenges and opportunities with AI in key applications like drug discovery and disease diagnosis. A forward focus will be pursued in data availability, model validation, sharing and adoption of developed models, and opportunities for broader collaborative efforts.

12:15 pm - 1:15 pm

Charting the PMIx Roadmap

Ralph Castain (Intel Corporation)

PMIx facilitates interaction between applications, tools, and middleware and system runtime environments. We have had a highly eventful year with continued adoption, release of new capabilities, and formation of a more formal standards body. We'll recap the activities of the past year and present a proposed roadmap for next year for discussion.

The PMIx community includes viewpoints from across the computing spectrum including HPC, AI, and Data Analytics. 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.

Cookies will be served!

12:15 pm - 1:15 pm

BeeGFS – Architecture, Innovative Implementations, and Development Plans

Frank Herold (ThinkParQ)

The open-source parallel file-system BeeGFS is one of the fastest-growing middleware products for HPC and related environments and is often seen as an easy-deployable alternative to the Lustre file-system. It was awarded Best Storage Product or Technology by HPCwire at SC18 and is now gaining acceptance at high-end sites.

Following on from previous BoFs which attracted well over 100 attendees, the SC19 BoF will be
split into three sections with the key objective to bring developers, administrators, solution providers, and end-users together to connect, interact, discuss development plans and share their user experiences with the product.

12:15 pm - 1:15 pm

OpenHPC Community BoF

Karl Schulz (University of Texas)

Over the last several years, OpenHPC has emerged as a community-driven stack providing a variety of common, pre-built ingredients to deploy and manage an HPC Linux cluster. Formed initially in November 2015 and formalized as a Linux Foundation project in June 2016, OpenHPC has been adding new software components and now supports multiple OSes/architectures. At this BoF, speakers from the OpenHPC Technical Steering Committee will provide technical updates from 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.

12:15 pm - 1:15 pm

Getting Scientific Software Installed

Robert McLay (University of Texas, Texas Advanced Computing Center (TACC))

We intend to provide a platform for presenting and discussing tools to deal with the ubiquitous problems that come forward when building and installing scientific software, which is known to be a tedious and time consuming task.

Several user support tools for allowing scientific software to be installed and used will briefly be presented, for example (but not limited to) EasyBuild, Lmod, Spack, Singularity, etc. We would like to bring various experienced members of HPC user support teams and system administrators as well as users together for an open discussion on tools and best practices.

12:15 pm - 1:15 pm

Sisyphus’ Work: Adapting Applications to New HPC Architectures - a European
Perspective.

Jean-Pierre Panzier (European HPC Technology Platform (ETP4HPC), Atos)

EuroHPC has launched a 1B Euro R&D programme aimed at delivering exascale systems, consolidating Europe's HPC technology provision and strengthening its vibrant application landscape.

However, the ever-changing heterogeneous architectures based on different instruction sets (ARM, RISC-V, Intel, etc.), including GPUs and other accelerators (FPGAs, optical accelerators, quantum, neuromorphic, etc.), render shaping applications for the next generation HPC systems a Sisyphean task.

European and international experts will discuss the approaches being pursued in Europe and worldwide and analyse how various architectures types facilitate sustainable application development.

A summary of European HPC and its projects ('European HPC 2019') is available at www.etp4hpc.eu/euexascale.

12:15 pm - 1:15 pm

Emerging Technologies and HPC Education, Outreach, and Training: Challenges and Opportunities

Nitin Sukhija (Slippery Rock University of Pennsylvania)

The rapid advancement and introduction of new HPC technologies has facilitated the convergence of Artificial Intelligence (AI), Machine Learning (ML), Data Analytics, Big Data, Cybersecurity and the HPC domain platforms. Integrated, flexible and scalable pedagogical and andragogical approaches are needed along with traditional HPC instructional practices to address broad workforce needs necessitated by these new convergent platforms and paradigms. Through brief panel-style discussions, this ACM SIGHPC Education Chapter coordinated BOF aims to elicit a better understanding of the factors that lead to successful HPC teaching and learning in AI, ML and Cybersecurity and other emerging communities using HPC infrastructure.

12:15 pm - 1:15 pm

SOLAR Consortium: Accelerated Ray Tracing for Scientific Simulations
Paul Navratil (Texas Advanced Computing Center (TACC), University of Texas)

Many physical simulations incorporate vector mathematics to model phenomena such as radiative transfer and to compute behavior such as particle advection. Hardware-optimized ray tracing engines, tuned by processor manufacturer engineers, can accelerate simulation critical sections that depend on ray-based traversals and intersections. This BoF will serve as a venue for developers and users of simulations, visual analysis codes, and ray tracers to discuss interfaces, capabilities, and performance. This meeting will continue the conversation toward standardization of ray tracer interfaces to facilitate their expanding role throughout scientific workflows, including data evaluation, insight formulation, discovery communication, and presentation-quality artifact generation.

12:15 pm - 1:15 pm

The National Academies’ Report on Reproducibility and Replicability in Science: Inspirations for the SC Reproducibility Initiative

Lorena Barba (George Washington University)

This BoF will disseminate the findings and recommendations of the NASEM report on Reproducibility and Replicability in Science, making connections with the SC Reproducibility Initiative. It will seed the community with ideas for taking inspiration from the report’s recommendations for the conference to continue leading in the computing community, increasing the transparency of research. The session will also communicate new initiatives stemming from the NASEM report, affecting the computing community, like the National Information Standards Organization (NISO) Badging Scheme Working Group.

12:15 pm - 1:15 pm

With Great Power Comes Great Responsibility: Ethics in HPC

Margaret Lawson (Sandia National Laboratories, University of Illinois)

This BoF hopes to start an ongoing, community-wide discussion about the role of ethics in HPC and to leave participants with a new awareness of the ethical considerations involved in this field. During this BoF, we hope to consider topics such as the role of HPC in supporting morally questionable applications, the ways in which HPC indirectly perpetuates inequality on both an international and individual level, and the environmental impact of HPC. By starting this dialogue,
we can help ensure that we are fulfilling HPC’s promise of making the world a better place.

12:15 pm - 1:15 pm

**Open OnDemand User Group Meeting**

Alan Chalker (Ohio Supercomputer Center)

Open OnDemand (OOD) is an NSF-funded open-source HPC portal whose goal is to provide an easy way for system administrators to provide web access to their HPC resources and is in use at over 50 HPC centers.

This BoF is meant to be an open discussion amongst members of the OOD community to guide the four year roadmap for OOD. Relevant topics include installation experiences, priority of upcoming features such as authentication and customized workflows, training users, integration with other science gateways, and growing the community. The organizers will give a short presentation, which will be followed by audience discussion.

12:15 pm - 1:15 pm

**U.S. Strategic Computing – An Update**

Manish Parashar (National Science Foundation)

Under the auspices of the 2015 National Strategic Computing Initiative, U.S. government agencies have been working to maximize the benefits of high-performance computing for scientific discovery, economic competitiveness and national security. However, the scientific and technological landscapes are rapidly evolving, and it is time to revisit the goals and approaches needed to sustain and enhance leadership in strategic computing. To address this, the OSTP and NSTC’s NITRD Subcommittee have formed a Fast Track Action Committee (FTAC) on Strategic Computing to develop recommendations to advance leadership in strategic computing. This BoF will discuss FTAC’s activities and findings with the community.

12:15 pm - 1:15 pm

**InfiniBand In-Network Computing Technology and Roadmap**
Gilad Shainer (Mellanox Technologies)

Being a standard-based interconnect, InfiniBand enjoys the continuous development of new capabilities.

HDR 200G InfiniBand In-Network Computing technology provides innovative engines offloading and accelerating communication frameworks and application algorithms. The session will discuss the InfiniBand In-Network Computing technology and testing results from DoE systems, Canada’s fastest InfiniBand Dragonfly based supercomputer at the University of Toronto, the world’s first HDR 200G InfiniBand systems and more.

As the needs for faster data speed accelerates, the InfiniBand Trade Association has been working to set the goals for future speeds, and this topic will also be covered at the session.

12:15 pm - 1:15 pm

Training and Education for HPC System Administrators: How Can We Do Better?

Neelofer Banglawala (Edinburgh Parallel Computing Centre)

Skilled and successful HPC System Administrators (SysAdmins) are the bedrock of the HPC community, yet well-established pathways and resources for the necessary skill acquisition remain sparse. There is an increasing desire within the community to change this. This session aims to bring together all those interested in HPC SysAdmin training and education to address key issues, such as: what training currently exists? What support do new versus experienced SysAdmins need? How can we share resources and expertise? When one-size-does-not-fit-all, can we establish formal baseline standards? How can we do better? All are welcome to join the discussion!

12:15 pm - 1:15 pm

HPC PowerStack: Community-Driven Collaboration on Power-Aware System Stack

Martin Schulz (Leibniz Supercomputing Centre, Technical University Munich)

This interactive BoF will bring together vendors, labs, and academia to discuss an ongoing community-wide effort to incorporate power-awareness within system-stacks in upcoming
exascale machines. HPC PowerStack is the first-and-only community-driven vendor-neutral effort to identify what power optimization software actors are critical within the modern-day stack; discuss their interoperability, and work towards gluing together existing open-source projects to engineer cost-effective, but cohesive, portable implementations.

This highly interactive BoF will disseminate key insights acquired in the project, provide prototyping status updates, invite attendee feedback on current directions, brainstorm solutions to open questions and solicit participation addressing the imminent power challenge.

12:15 pm - 1:15 pm

PBS Pro Open Source Project Community BoF

Bill Nitzberg (Altair Engineering)

The PBS Pro software performs job scheduling and workload management 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 drive the future of the PBS Pro.

5:15 pm - 6:45 pm

Achieving Performance on Current and Future Large-Scale Intel-Based Systems

Richard Gerber (National Energy Research Scientific Computing Center (NERSC), Lawrence Berkeley National Laboratory)

This BoF, organized by the Intel eXtreme Performance Computing Users Group (IXPUG), will focus on achieving performance at scale on current and future large Intel-based systems. The IXPUG community has in-depth knowledge and experience preparing codes for Xeon Phi and Xeon Scalable family of processors and coprocessors. With a planned 2021 exascale system that includes Intel GPUs, IXPUG’s optimization targets will be expanding to include the full line of Intel processors, including FPGAs. Through invited talks and open discussion, the BoF will provide a forum to share tips and techniques and look to the future.
5:15 pm - 6:45 pm

OpenMP API Version 5.0 - State of the Union

Jeff Larkin (Nvidia Corporation)

The OpenMP ARB released version 5.0 of the OpenMP API specification in November 2018. This BOF will review the progress made in adopting the new API version. A series of lightning talks on OpenMP in real applications will give the audience insight into how modern OpenMP is used on recent leadership and flagship HPC systems. We will also host ARB representatives of OpenMP vendors to report out on compiler availability for version 5.0. A panel discussion with leading ARB experts will give the audience the opportunity to interact directly with the ARB and to ask questions.

5:15 pm - 6:45 pm

Analyzing Parallel I/O

Shane Snyder (Argonne National Laboratory)

Parallel I/O performance can be a critical bottleneck for applications, yet users are often ill-equipped for identifying and diagnosing I/O performance issues. Increasingly complex hierarchies of storage hardware and software deployed on many systems only compound this problem. Tools that can effectively capture, analyze, and tune I/O behavior for these systems empower users to realize performance gains for many applications.

In this BoF, we form a community around best practices in analyzing parallel I/O and cover recent advances to help address the problem presented above, drawing on the expertise of users, I/O researchers, and administrators in attendance.

5:15 pm - 6:45 pm

Engaging the DoD High Performance Computing Modernization Program (HPCMP)

Kevin Newmeyer (US Department of Defense HPC Modernization Program)

The HPCMP is a Department of Defense (DoD) wide program and national asset providing a
spectrum of high performance computing capabilities and expertise to solve DoD’s most critical mission challenges. This BoF will use lightning talks about strategic directions, current research, technology acquisition plans, and software development needs and interests to illustrate DoD goals and opportunities for engagement. These lightning talks are intended to help external organizations and researchers connect with DoD users and sites to encourage partnerships and help solve problems. External engagement will help DoD users and HPC sites grow expertise and connect to the larger HPC community.

5:15 pm - 6:45 pm

Containers in HPC

Andrew Younge (Sandia National Laboratories)

Container computing has revolutionized how many industries and enterprises develop and deploy software and services. Recently, this model has gained traction in the HPC community through enabling technologies like Charliecloud, Docker, Kubernetes, Podman, Shifter, and Singularity. While containers look to provide greater software flexibility, reliability, ease of deployment, and portability for users, there are still open questions that need to be addressed as we race toward adoption in the Exascale era. In this BoF, we will provide an opportunity for the HPC community to engage with the leaders in the field who can provide real-world experiences to containerization in HPC.

5:15 pm - 6:45 pm

LLVM in HPC: Exciting Developments Again This Year!

Hal Finkel (Argonne National Laboratory, LLVM Foundation)

The LLVM compiler infrastructure is a core open-source project powering compilers and language-processing tools, both open source and proprietary, across the HPC ecosystem. In this BoF, a group of LLVM experts who are driving the development of LLVM (including contributors from ARM, Intel, NVIDIA, and other vendors) will give short presentations of their work, discuss exciting developments, then answer your questions on LLVM and its future in HPC.

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 desire.
5:15 pm - 6:45 pm

The Green 500: Trends in Energy Efficient Supercomputing

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.

5:15 pm - 6:45 pm

Reconfigurable/FPGA Clusters for High Performance Computing

Martin Herbordt (Boston University)

FPGA-Centric Clusters (FCCs) co-locate computation and communication to potentially yield unmatched scalability for HPC applications. FCCs have made inroads in certain domains—cloud, finance, bioinformatics, oil & gas, etc.—and are the object of much academic research with more than a dozen major efforts under way. However, nearly all aspects of building FCC systems are still ad hoc. In this BoF, leading researchers from industry, government, and academia will present the state-of-the-art in applications, architecture, programming productivity, and communication targeting FCCs. Discussion will aim at determining best practices and creating a roadmap towards future FCC-enable supercomputing. Audience participation is highly encouraged.

5:15 pm - 6:45 pm

Cloud and Open Infrastructure Solutions To Run HPC Workloads

Martial Michel (Data Machines Corporation)

Virtualization and containers have grown to see more prominent use within the realm of HPC. Adoption of these tools has enabled IT Organizations to reduce costs all the while making it easier
to manage large pools of compute, storage and networking resources. However, performance
overheads, networking integrations, and system complexity pose daunting architectural challenges.

OpenStack, Containers, and the orchestration thereof, all pose their own set of unique benefits and
challenges. This BoF is aimed at architects, administrators, software engineers, and scientists
interested in designing and deploying cloud infrastructure solutions to run HPC workloads.

5:15 pm - 6:45 pm

**HPC Graph Toolkits and the GraphBLAS Forum**

Antonino Tumeo (Pacific Northwest National Laboratory (PNNL))

Government agencies, industry, and academia are demanding a new generation of tools to
efficiently solve large scale analytics problems in a variety of business, scientific and national
security applications. This BoF aims at gathering the community of people interested in frameworks
and workflows for large scale graph analytics, surveying the current approaches, identifying new
challenges and opportunities, and laying a path toward future interoperable infrastructures. As in
previous editions, we will invite the GraphBLAS community to participate in a discussion of the
current state and evolution of GraphBLAS, with the goal of developing requirements and
recommendations for future tools.

5:15 pm - 6:45 pm

**Multi-Level Memory and Storage for HPC, Data Analytics, and AI**

Hans-Christian Hoppe (Intel Corporation)

This BoF investigates the opportunities arising from the progress in storage class memory (SCM)
technology and the parallel rapid emergence of data-intensive application in the HPC context,
which increasingly combine simulations with data analytics, AI, or graph analytics techniques.
Topics include the benefit of SCM for applications, integration with the system architecture, SW
interfaces, and, of course, results from proof of concept projects.

The session brings together technology providers, application and system SW developers, and
system operators to engage in a discussion with the audience.
MPICH: A High Performance Open-Source MPI Implementation

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.

Big Data and Exascale Computing (BDEC2) Community Roundtable

Peter Beckman (Argonne National Laboratory, Northwestern University)

The emergence of data-intensive machine learning and AI in a wide variety of scientific fields, and the explosive growth of data generated by new instruments and IoT in edge environments, is severely disrupting the landscape for scientific computing. The international Big Data and Extreme-scale Computing (BDEC) initiative has started a second workshop series—BDEC2—with a formulate a coordinated international response to these powerful trends. This BoF will review the results of the three workshops held 2018-2019 and seek input from the SC community into the next three workshops, which aim to deliver a candidate software infrastructure roadmap in 2021.

Challenges and Opportunities in Academic HPC Systems Research in 2030

Paul Carpenter (Barcelona Supercomputing Center)

Radical changes are foreseen in all aspects of high-performance computing systems over the next decade. In this context, in 2017, the European Eurolab-4-HPC project published its long-term vision for academic HPC systems research. This session, which follows a successful BoF at ISC 2019, is part of the process of preparing the 2020 update. In a lively and hopefully controversial discussion,
involving expert presenters and the audience, we will identify, discuss and assess the most important long-term challenges and opportunities in HPC systems research. The overall goal is to build a research community that is high quality and relevant.

5:15 pm - 6:45 pm

**Software Engineering and Reuse in Modeling, Simulation, and Data Analytics for Science and Engineering**

David Bernholdt (Oak Ridge National Laboratory)

Software engineering (SWE) for modeling, simulation, and data analytics for computational science and engineering (CSE) is challenging, with ever-more sophisticated, higher fidelity simulation of ever-larger, more complex problems involving larger data volumes, more domains, and more researchers. Targeting both commodity and custom high-end computers multiplies these challenges. We invest significantly in creating these codes, but rarely talk about that experience; we just focus on the results.

We seek to raise awareness of SWE for CSE on supercomputers as a major challenge, and develop an international “community of practice” to continue these important discussions outside of workshops and other “traditional” venues.

**Thursday, November 21**

12:15 pm - 1:15 pm

**MLPerf: A Benchmark for Machine Learning**

Tom St. John (Tesla Inc)

Machine learning applications are rapidly expanding into scientific domains and challenging the hallmarks of traditional high performance computing workloads. This BoF presents MLPerf, a community-driven system performance benchmark which spans a range of individual machine learning tasks. The speakers at this BoF are experts in the fields of high performance computing, machine learning, and computer architecture, representing academia, government research organizations, and private industry. The goal of this session is to introduce MLPerf to the broader HPC community and solicit input from interested parties to drive the further adoption of this
For a couple of years now, multiple supercomputing sites around the globe have development and implementation projects underway for expanded monitoring frameworks collecting operational parameters of HPC systems and facility support infrastructure into a single unified database, providing a new and more comprehensive overview of all operations. These early adopter sites have already deployed these systems into production and are collecting a valuable repository of performance data. What to do with this wealth of data, how to process and analyze it, and how to feed it back into improved operations, will be the topic of this BoF session.

The panel will consist of representatives from the industry and academia with a background in HPC. The panel will share advice on different career options in HPC, and their experiences in their respective career trajectories. The primary audience for this event is current graduate students and post-doctoral researchers. The format will include a brief introduction by each speaker, followed by a moderated discussion based on a set of previously submitted questions and ending with further questions from the audience. We may also have a brief keynote speaker.

Currently, HPC archiving is a largely hopeful enterprise because the community lacks the
infrastructure needed to help users identify what an archived dataset contains and who generated it. Although metadata management offers a potential solution, the community has yet to realize a system that fully addresses the archiving problem. The goal of this BoF is to bring together the groups that have been working individually on this problem to gather requirements for an effective, exascale solution and to try and generate a common framework for moving forward.

### 12:15 pm - 1:15 pm

**Slurm Community BoF**

*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, cloud bursting, 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 upcoming version 20.02, describe the Slurm roadmap, and solicit user feedback. Everyone interested in Slurm use and/or development is encouraged to attend.

### 12:15 pm - 1:15 pm

**Whither Advanced Scientific Computing After Exascale Is Achieved?**

*Roscoe Giles (Boston University)*

The goal of this BoF is to discuss the landscape for advanced scientific computing after the United States achieves the exascale milestone. What are possible future breakthroughs? What are promising computing and mathematics research directions? How do we organize people and resources to move forward? How do we sustain and broaden the impact of exascale technology?

These questions will be introduced by members of a task force that is studying beyond the Exascale Computing Project. The community discussion will inform both the participants and the task force members.

### 12:15 pm - 1:15 pm

**Khronos SYCL Heterogeneous C++ Status and Directions**
Many HPC programmers have not heard of SYCL, however, with the increasing importance of modern C++ in HPC, or just seeking alternatives to proprietary languages, SYCL is becoming critical as a vendor neutral way to write C++ code that embraces heterogeneous parallelism. SYCL is an open standard. There are multiple implementations available, both open source and commercial.

In this BoF, experts and SYCL implementers will explain its advantages, how the language is governed, where it is going, and why you need to be aware of it if you are intending to write C++ code which targets HPC machines.

12:15 pm - 1:15 pm

**Spack Community BoF**

Adam Stewart (University of Illinois)

Spack is a package manager for scientific computing, with a rapidly growing open-source community. Spack has over 445 contributors from academia, industry, and laboratories across the world, and is used to manage software releases for the U.S. Exascale Computing Project. At this BoF, Spack developers will give updates on the community, new features, and the roadmap for future development. We will poll the audience to gather valuable information on how Spack is being used, and will open the floor for questions. All are invited to provide feedback, request features, and discuss future directions. Help us make installing HPC software simple!

12:15 pm - 1:15 pm

**Quality Assurance and Coding Standards for Parallel Software**

Manuel Arenaz (Appentra Solutions; University of A Coruña, Spain)

The automation of testing is critical in software development to improve quality assurance (QA), but today 80% of testing is manual (Gartner) and $32 billion is spent annually on QA (IDC/Nelson Hall). Coding standards in automotive and cybersecurity (e.g. CWE, MISRA) provide developers with rules and recommendations to prevent faulty code patterns. The ever-increasing complexity of HPC software and hardware pushes the developers to critically reevaluate testing methods, but there is no coding standard for parallel programming yet. Our goal is to form a community interested in quality assurance and best practices for parallel programming.
European HPC Ecosystem Summit

Oriol Pineda (Partnership for Advanced Computing in Europe (PRACE), Barcelona Supercomputing Center)

PRACE has recently engaged in the coordination of European HPC services and activities through a series of events and workshops. This includes access to HPC systems, HPC user support, training in HPC, HPC policy, HPC technology development, HPC operations and dissemination.

The objective of this BoF is to present and discuss the final conclusions of this initiative with the key stakeholders. The conclusions from this BoF will be used to structure the new "HPC in Europe" services portal that will collect the results from this initiative, with a special focus on User Support and Training.

Charm++ and AMPI: Adaptive and Asynchronous Parallel Programming

Kavitha Chandrasekar (University of Illinois)

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.

HPC System Testing: Procedures, Acceptance, Regression Testing, and Automation
Verónica Melesse Vergara (Oak Ridge National Laboratory)

This BoF will briefly highlight acceptance and regression testing procedures at several large-scale HPC centers. The goal of this BoF is to bring together testing efforts from multiple leadership-class supercomputer centers and ideas from the HPC community to discuss the different strategies used and to document lessons learned. BoF attendees will have the opportunity to share their experiences conducting acceptance and regression testing at their institutions and exchange best practices with other HPC centers.

12:15 pm - 1:15 pm

Science and Technology Requirements for Leadership Computing in Open Science

Dan Stanzione (Texas Advanced Computing Center (TACC))

The NSF in the US has begun a planning process for a National-Scale HPC facility providing roughly an 0.5 – 1.0 exaFLOPS in computing capability. This BoF is one of a series of sessions that will guide the design of this facility. The Texas Advanced Computing Center is seeking input from both the science and technology communities regarding the key capabilities this facility must provide to effectively support large scale open science. The session will begin with a brief discussion of the roadmap for the proposed facility, followed by audience discussion and Q&A to stimulate a broad-based discussion of requirements.

12:15 pm - 1:15 pm

HPCG Benchmark Update

Michael Heroux (Sandia National Laboratories, St. John’s University)

The High Performance Conjugate Gradients (HPCG) Benchmark is a TOP500 benchmark for ranking high performance computing systems. First results were released at ISC’14, including optimized results for systems built upon Fujitsu, Intel, Nvidia technologies. Optimized versions of HPCG are available for all major HPC platforms. Industry analyses cite HPCG in combination with LINPACK as indication of system balance.

In this BoF, we first present HPCG updates, then follow with presentations from vendors who have
participated in recent HPCG optimization efforts, including recent AMD and Arm efforts. We end with open discussion about future improvements.

12:15 pm - 1:15 pm

**Benchmarking Machine Learning Ecosystem on HPC Systems**

Murali Emani (Argonne National Laboratory)

High-performance computing is seeing an upsurge in workloads that require data analysis. Machine learning and Deep learning models are used in several science domains such as cosmology, particle physics, biology with data in unprecedented scale from simulations. These applications include tasks such as image detection, segmentation, synthetic data generation and in-situ data analysis. Emerging HPC systems have diverse hardware including many-core, multi-core and heterogeneous accelerators. It is critical to understand the performance of Machine learning/deep learning models on HPC systems at scale. Benchmarking will help to better understand the model-system interactions and help co-design future HPC systems for ML workloads.

12:15 pm - 1:15 pm

**Advanced Architecture Testbeds: Community Resources for Enhanced HPC Research**

Jeff Young (Georgia Institute of Technology)

The recent explosion in architectural diversity in the supercomputing community is supported by the use of user-focused testbeds that allow researchers to test their applications on hardware that may be hard to access at smaller scales. This BoF will bring together panelists from advanced architecture testbed efforts including CSCS’s User lab, PNNL’s CENATE testbed, HAAPS at SNL, the Rogues Gallery at Georgia Tech, ExCL at ORNL, and the Maui HPC Center to discuss increasingly diverse architectures and challenges for using them. Panelists and attendees will debate topics to better meet the growing demand for access to novel architecture systems.

12:15 pm - 1:15 pm

**Python for High Performance and Scientific Computing**
The Python for High Performance and Scientific Computing BoF is intended to provide current and potential Python users and tool providers in the high performance and scientific computing communities a forum to talk about their current projects; ask questions of experts; explore methodologies including interactive supercomputing, data science, and machine learning; delve into issues with the language, modules, tools, and libraries; build community; and discuss the path forward. In the absence of a PyHPC workshop this year, this BoF provides a much needed platform for continued discussion and interaction amongst those in the Python and HPC space.
Computing for Change

Tuesday, November 19

10:00 am - 6:00 pm

SC Theater

Hybrid Quantum-Classical Algorithms for Graph Problems: Forging a Path to Near-Term Applications
Ruslan Shaydulin (Clemson University)

Technology Challenge Introduction and Jury Introduction

Technology Challenge UEN Demo

Technology Challenge RENCI Demo Prep

Technology Challenge RENCI Demo

Break

SCinet Architecture

Technology Challenge ANL Demo Prep

Technology Challenge ANL Demo

Technology Challenge Jury Closing Remarks

Student Cluster Challenge "How We Got Here"

Student Cluster Challenge "How We Got Here"

Student Cluster Challenge "How We Got Here"
Modern HPC approaches to solve time dependent Quantum Mechanical equations

Konstantin Rygol (University of Bergen)

Wednesday, November 20

10:00 am - 6:00 pm

SC Theater

2:30 pm - 4:30 pm

Computing4Change Lightning Talks

Session Description: Computing4Change is an undergraduate student competition sponsored by SIGHCPC and co-located with SC19. Students work in assigned teams on a data intensive, socially relevant topic assigned at SC19. Each student will present their technical approach to the given problem and their results in this lightning talk session. Come support these undergraduate students from across the country and hear their innovative approaches for this competition.
Thursday, November 21

10:00 am - 3:00 pm

SC Theater

1:30 pm - 5:00 pm

Computing4Change Student Presentations

Session Description: Computing4Change is an undergraduate student competition sponsored by SIGHPC and co-located with SC19. Students work in assigned teams on a data intensive, socially relevant topic assigned at SC19. Each student will present their technical approach to the given problem and their results in this session. Come support these undergraduate students from across the country and hear their innovative approaches for this competition.
Doctoral Showcase

Tuesday, November 19

8:30 am - 5:00 pm

Doctoral Showcase Posters Display

Poster 36: Modeling Non-Determinism in HPC Applications
Dylan Chapp (University of Delaware, University of Tennessee)

As HPC applications migrate from the petascale systems of today to the exascale systems of tomorrow, the increasing need to embrace asynchronous, irregular, and dynamic communication patterns will lead to a corresponding decrease in application-level determinism. Two critical challenges emerge from this trend. First, unchecked non-determinism coupled with the non-associativity of floating-point arithmetic undermines numerical reproducibility of scientific applications. Second, the prevalence of non-determinism amplifies the cost of debugging, both in terms of computing resources and human effort. In this thesis, we present a modeling methodology to quantify and characterize communication non-determinism in parallel applications. Our methodology consists of three core components. First, we build graph-structured models of relevant communication events from execution traces. Second, we apply similarity metrics based on graph kernels to quantify run-to-run variability and thus identify the regions of executions where non-determinism manifests most prominently. Third, we leverage our notion of execution similarity to characterize applications via clustering, anomaly detection, and extraction of representative patterns of non-deterministic communication which we dub "non-determinism motifs". Our work will amplify the effectiveness of software tools that target mitigation or control of application-level non-determinism (e.g., record-and-replay tools) by providing them with a common metric for quantifying communication non-determinism in parallel applications and a common language for describing it.

Poster 35: Scaling Up Pipeline Optimization with High Performance Computing
Robert Lim (University of Oregon)
My research focuses on developing a pipeline optimization infrastructure that automates the design and code generation of neural networks through the use of high-performance computing. The problem has the following objectives: unify automated machine learning (AutoML) and compilation, archive profiles for creation of a knowledge base for a data-driven approach toward search, explore various search optimizations for model design and code generation. The field of automated deep learning includes hyperparameter optimization and neural architecture search (NAS), which requires domain expertise in designing a model, in addition to the tuning parameters related to learning and the model itself. The search space is complex and deciding which parameters factor into the overall accuracy of a model is a non-trivial task. Once a model is trained, the next step compiles the model, which maps to the backend of a targeted architecture, whether GPU, embedded mobile phones, or FPGA. The compilation phase also involves choices, in terms of optimizations applied, which can include a set of flags, or direct code transformations. Various efficient search techniques are explored, archiving results along the way for facilitating in code generation options, which could possibly inform the model design process. Our previous work reduced the search space for code generation of various domain kernels by 92%, and this work investigates whether the same approach can be applied in all aspects of AutoML design and code generation.

Poster 38: High-Performance Backpropagation in Scientific Computing
Navjot Kukreja (Imperial College, London)

Devito is a domain-specific language for the automatic generation of high-performance solvers for finite difference equations provided in a high-level symbolic representation. The primary use of Devito is to enable Full-waveform inversion, which is an adjoint-based optimization problem that uses the wave equation as part of its objective function. This doctoral project consists of three elements of Devito that are key to solving adjoint-based optimization problems (i.e. the back propagation). The first is automatic differentiation of stencil loops. This allows the automatic generation of high-performance code implementing the derivative of any provided function. This is essential to enable the use of a wider range of physical equations - in order to use better physical approximations. An essential feature of the generated derivatives is that they show scaling and performance behaviour that is very similar to the forward function. A common issue in solving adjoint-based optimization problems is the prohibitively high memory requirement. The second contribution is a runtime for automatic execution of checkpoint-recompute schedules (called pyRevolve) to alleviate this memory requirement by trading it off for re-computations. These schedules may involve storing some intermediate states on disk, some in memory, and recomputing others. The third contribution is the use of lossy compression (ZFP/SZ), which is a second means of trading off memory and compute, automatically as part of checkpointing in pyRevolve. This is the first time it has been shown that lossy compression combined with checkpoint-recomputation can provide a much better error-memory tradeoff than using checkpoint-recomputation or lossy
compression alone.

**Poster 37: Large Scale Ultrasound Simulations on Accelerated Clusters**

*Filip Vaverka (Brno University of Technology)*

Many emerging non-invasive medical procedures such as high intensity focused ultrasound treatment (HIFU), photoacoustic imaging, or transcranial ultrasound therapy require ultrasound wave propagation simulations. Typically soft tissue can be modeled as fluid, weakly heterogeneous medium with frequency dependent absorption and non-linear effects at high ultrasound intensities. The k-Wave acoustic toolbox, widely used in medical community, uses highly efficient k-space pseudo-spectral time-domain (KSTD) discretization of acoustic equations to solve these problems. However, it is difficult to efficiently map the KSTD method onto modern cluster architectures with accelerators such as GPUs. The present thesis analyzes shortcomings of the KSTD method in respect to modern clusters and proposes local Fourier basis approach to improve scaling of the method. It is shown that the proposed method is able to achieve 5x speedup, while having sufficient accuracy for these medical applications. Behavior of the method is analyzed across variety of GPU and MIC (Intel Xeon Phi) accelerated clusters and results are presented.

**Poster 34: Analysis of Automata Processing Acceleration on Disparate Hardware Technologies**

*Marziyeh Nourian (North Carolina State University)*

Pattern matching is a computation that maps naturally onto finite automata (FA) abstractions. There has been a substantial amount of work on accelerating FA processing on various parallel platforms. However, the advantages and disadvantages of different automata processing accelerators and the innovation space in this area are still unclear. We target this problem and propose a compiler toolchain that automates the deployment of non-deterministic finite automata (NFAs) onto different target platforms. Using this toolchain, we perform an apples-to-apples comparison between AP, GPU- and FPGA-based NFA accelerator designs on large-scale datasets. Specifically, we observe that memory-based designs are limited by memory size and bandwidth. To address this issue, we target fixed-topology NFAs and propose a memory-efficient design that embeds the automata topology in code and stores only the transition symbols in memory. Our solution is suitable for SIMD architectures and is called SIMD_NFA. We design a compiler that automates the deployment of this design on SIMD platforms. We showcase our compiler framework on GPU and Intel platforms. Additionally, we observe that for NFAs with a grid-like fixed-topology (e.g., NFAs for Levenshtein and Hamming distance-based matching), transitions do not need to be encoded within the traversal code but can be inferred from the reference string to be matched and the knowledge of the NFA topology. Lastly, SIMD_NFA is a good fit for FPGA deployment using OpenCL-to-FPGA toolchains.
We investigate the deployment of the OpenCL version of SIMD_NFA, on FPGA and explore a set of optimizations techniques to retarget SIMD_NFA to FPGA.

Poster 42: Power and Thermal Management Runtimes for HPC Applications in the Era of Exascale Computing
Daniele Cesarini (University of Bologna, CINECA)

In the scope of technical and scientific computing, the rush toward larger simulations has been so far assisted by a steady downsizing of micro-processing units, which has allowed to increase the compute capacity of general-purpose architectures at constant power. As side effects of the end of Dennard's scaling, this process is now hitting its ultimate power limits and is just about to come to an end. This implies an increase in the energy cost for computation, performance loss due to a design based on worst-case power consumption, and performance loss due to overheating, and thermal gradients. As result, thermal and power bound supercomputing machines show performance degradation and heterogeneity which limit the peak performance of the system. This doctoral showcase presents software strategies to tackle the main bottlenecks induced by power and thermal issues that affect next-generation supercomputers. To respond to the above challenges, my work shows that propagating workload requirements from application to the runtime and operating system levels is the key to provide efficiency. This is possible only if the proposed software methodologies cause little or no overhead in term of application performance. With this in mind in my work, I have designed application-aware node-level optimal thermal management algorithms and runtimes, lazy node-level power capping, and energy reduction runtime. The experimental results show a significant step forward with respect to the current state-of-the-art solutions in power and thermal control of HPC systems.

Poster 41: Co-Designing Communication Middleware and Deep Learning Frameworks for High-Performance DNN Training on HPC Systems
Ammar Ahmad Awan (Ohio State University)

Recent advances in Machine/Deep Learning techniques have triggered key success stories in many application domains like Computer Vision, Speech Comprehension and Recognition, and Natural Language Processing. Large-scale Deep Neural Networks (DNNs), that are at the core of state-of-the-art AI technologies, have been the primary drivers of this success. Training very complicated and large DNN architectures using a large number of training examples (data) is compute-intensive and can take from weeks to months to achieve state-of-the-art prediction capabilities. To achieve higher accuracy, making the DNN deeper is also a common strategy. These requirements have led to a simple but powerful approach called Data Parallelism to achieve shorter training times. This has
resulted in various research studies and ML/DL software like TensorFlow and PyTorch as well as distributed-training middleware like Horovod. In addition, for DNNs that do not fit the GPU memory, a new DNN workloads are emerging that we call Out-of-Core DNNs and different strategies (out-of-core training and model-parallelism) are needed to train them. Clearly, large-scale DNN training brings forward new requirements for computation runtimes like CUDA and communication middleware like the MVAPICH2 MPI library. In this thesis, we broadly explore three different strategies to train DNNs on modern CPU and GPU architectures: 1) Data Parallelism, 2) Model Parallelism, and 3) Out-of-Core Training. We address the key challenge: How to co-design computation and communication in modern ML/DL frameworks with execution runtimes like CUDA and communication middleware like MVAPICH2 to enable scalable, high-performance, and efficient training of DNNs on large-scale HPC systems?

**Poster 40: Performance, Portability, and Productivity for Data-Parallel Computations on Multi- and Many-Core Architectures**

Ari Rasch (University of Münster)

This thesis presents an approach to performance, portability, and productivity for data-parallel computations on multi- and many-core architectures, e.g., Intel CPU and NVIDIA GPU. We introduce the algebraic formalism of Multi-Dimensional Homomorphisms (MDHs) – a class of functions that cover important data-parallel computations, e.g., linear algebra routines (BLAS) and stencil computations. For our MDHs, we propose a Domain-Specific Language (DSL), based on patterns of parallelism (a.k.a. algorithmic skeletons), to enable conveniently expressing MDH functions. We introduce a code generation approach for our DSL to automatically generate for MDHs optimized program code targeting multi- and many-core architectures. Our code generation approach relies on OpenCL – an emerging de-facto standard for uniformly programming parallel architectures, such as CPU and GPU. A major feature of our generated code is that it is targeted to OpenCL’s abstract device models (rather than a particular architecture) by being parameterized in performance-critical parameters of these abstract models (e.g., the number of threads and size of tiles). With our code generation approach, we enable both high performance and performance portability: we fully automatically optimize our generated code -- for any given combination of an MDH function, architecture, and input size -- by automatically choosing (auto-tuning) optimized values of our code’s performance-critical parameters using our own Auto-Tuning Framework (ATF). Our experimental results on CPU and GPU demonstrate competitive and often significantly better performance of our MDH+ATF approach as compared to the currently best-performing competitors, e.g., Intel MKL/MKL-DNN, NVIDIA cuBLAS/cuDNN, and Facebook’s Tensor Comprehensions framework.
Poster 45: Characterization and Modeling of Error Resilience in HPC Applications
Luanzheng Guo (University of California, Merced)

As high-performance computing systems scale in size and computational power, the occurrence of transient faults grows. Without protection by efficient and effective fault tolerance mechanisms, transient errors can cause incorrect execution outcome and even lead to a catastrophe in safety-critical HPC applications. Previous work attributes error resilience in HPC applications at a high-level to either the probabilistic or iterative nature of the application, whereas the community still lacks the fundamental understanding of the program constructs that result in natural error resilience. We design FlipTracker, a framework to analytically track error propagation and to provide a fine-grained understanding of the propagation and tolerance of errors. After running FlipTracker on representative HPC applications, we summarize six resilience computation patterns that lead to nature error resilience in HPC applications. With a better understanding of natural resilience in HPC applications, we aim to model application resilience on data objects to transient faults. Many common application-level fault tolerance mechanisms focus on data objects. Understanding application resilience on data objects can be helpful to direct those mechanisms. The common practice to understand application resilience (random fault injection) gives us little knowledge of how and where errors are tolerated. Understanding "how" and "where" is necessary to understand how to apply application-level fault tolerance mechanisms effectively and efficiently. We design a practical model (MOARD) to measure application resilience on data objects by analytically quantifying error masking events happening to the data object. Using our model, users can compare application resilience on different data objects with different data types.

Poster 43: Efficient and Scalable Communication Middleware for Emerging Dense-GPU Clusters
Ching-Hsiang Chu (Ohio State University)

In the era of post Moore's law, the traditional CPU is not able to keep the pace up and provide the computing power demanded by the modern compute-intensive and highly parallelizable applications. Under this context, various accelerator architectures such as general-purpose graphics processing unit (GPU), which equipped with the high-bandwidth memory (HBM) and massive parallelizable streaming multiprocessors, has been widely adopted in high-performance computing (HPC) and cloud systems to significantly accelerate numerous scientific and emerging machine/deep learning applications. Message Passing Interface (MPI), the standard programming model for parallel applications, has been widely used for GPU communication. However, the state-of-the-art MPI libraries are only optimizing GPU communication by leveraging advanced technology like Remote Direct Memory Access (RDMA) and not fully utilizing the computational power of GPUs. In this work, we propose GPU-enabled communication schemes to harness GPU
computational resources, and cutting-edge interconnects such as NVIDIA NVLink for communication operations on the emerging heterogeneous systems. In this work, three primary MPI operations are addressed. First, intelligent communication scheduling, efficient packing/unpacking, and packing-free schemes are proposed to accelerate non-contiguous data transfer in scientific HPC applications. Second, scalable broadcast operations are presented to leverage the low-level hardware multicast feature to speed up GPU communication at scale. Finally, we also design topology-aware, link-efficient, and cooperative GPU kernels to significantly accelerate All-reduce operation, which is the primary performance bottleneck in deep learning applications. The proposed designs demonstrate significant performance improvements over the state-of-the-art communication schemes for various HPC and deep learning applications.

Poster 39: Designing Next-Generation Communication Middlewares for Many-Core Architectures
Jahanzeb Maqbool Hashmi (Ohio State University)

Modern multi-petaflop HPC systems are powered by dense multi-/many-core architectures and this trend is expected to grow for the next-generation supercomputing systems. This rapid adoption of high core-density architectures by the current- and next-generation HPC systems is further fueled by the emerging application trends such as Deep Learning. This is putting more emphasis on middleware designers to optimize various communication protocols to meet the diverse needs of the applications. While the novelties in the processor architectures have led to the increased on-chip parallelism, they come at the cost of rendering traditional designs, employed by the communication runtimes such as MPI, to suffer from higher degree of intra-node communication latencies. Tackling the computation and communication challenges that accompany these dense multi-/many-cores garner special design considerations. The proposed work in this thesis tries to address the performance challenges posed by a diverse range of applications and the lacking support in state-of-the-art communication libraries such as MPI to exploit high-concurrency architectures. The author first proposes a "shared-address-spaces"-based communication substrate to derive intra-node communication in MPI. Atop this framework, the author has re-designed various MPI primitives such as point-to-point communication protocols (e.g., user-space zero-copy rendezvous transfer), collective communication (e.g., load/store based collectives, truly zero-copy and partitioning-based reduction algorithms), and efficient MPI derived datatypes processing (e.g., memoization-based "packing-free" communication) to exploit the potential of emerging multi-/many-core architectures and high throughput networks. The proposed designs have demonstrated significant improvement over state-of-the-art for various scientific and deep learning applications.
As HPC applications migrate from the petascale systems of today to the exascale systems of tomorrow, the increasing need to embrace asynchronous, irregular, and dynamic communication patterns will lead to a corresponding decrease in application-level determinism. Two critical challenges emerge from this trend. First, unchecked non-determinism coupled with the non-associativity of floating-point arithmetic undermines numerical reproducibility of scientific applications. Second, the prevalence of non-determinism amplifies the cost of debugging, both in terms of computing resources and human effort. In this thesis, we present a modeling methodology to quantify and characterize communication non-determinism in parallel applications. Our methodology consists of three core components. First, we build graph-structured models of relevant communication events from execution traces. Second, we apply similarity metrics based on graph kernels to quantify run-to-run variability and thus identify the regions of executions where non-determinism manifests most prominently. Third, we leverage our notion of execution similarity to characterize applications via clustering, anomaly detection, and extraction of representative patterns of non-deterministic communication which we dub "non-determinism motifs". Our work will amplify the effectiveness of software tools that target mitigation or control of application-level non-determinism (e.g., record-and-replay tools) by providing them with a common metric for quantifying communication non-determinism in parallel applications and a common language for describing it.
and code generation of neural networks through the use of high-performance computing. The problem has the following objectives: unify automated machine learning (AutoML) and compilation, archive profiles for creation of a knowledge base for a data-driven approach toward search, explore various search optimizations for model design and code generation. The field of automated deep learning includes hyperparameter optimization and neural architecture search (NAS), which requires domain expertise in designing a model, in addition to the tuning parameters related to learning and the model itself. The search space is complex and deciding which parameters factor into the overall accuracy of a model is a non-trivial task. Once a model is trained, the next step compiles the model, which maps to the backend of a targeted architecture, whether GPU, embedded mobile phones, or FPGA. The compilation phase also involves choices, in terms of optimizations applied, which can include a set of flags, or direct code transformations. Various efficient search techniques are explored, archiving results along the way for facilitating in code generation options, which could possibly inform the model design process. Our previous work reduced the search space for code generation of various domain kernels by 92%, and this work investigates whether the same approach can be applied in all aspects of AutoML design and code generation.

**Poster 38: High-Performance Backpropagation in Scientific Computing**  
Navjot Kukreja (Imperial College, London)

Devito is a domain-specific language for the automatic generation of high-performance solvers for finite difference equations provided in a high-level symbolic representation. The primary use of Devito is to enable Full-waveform inversion, which is an adjoint-based optimization problem that uses the wave equation as part of its objective function. This doctoral project consists of three elements of Devito that are key to solving adjoint-based optimization problems (i.e. the back propagation). The first is automatic differentiation of stencil loops. This allows the automatic generation of high-performance code implementing the derivative of any provided function. This is essential to enable the use of a wider range of physical equations - in order to use better physical approximations. An essential feature of the generated derivatives is that they show scaling and performance behaviour that is very similar to the forward function. A common issue in solving adjoint-based optimization problems is the prohibitively high memory requirement. The second contribution is a runtime for automatic execution of checkpoint-recompute schedules (called pyRevolve) to alleviate this memory requirement by trading it off for re-computations. These schedules may involve storing some intermediate states on disk, some in memory, and recomputing others. The third contribution is the use of lossy compression (ZFP/SZ), which is a second means of trading off memory and compute, automatically as part of checkpointing in pyRevolve. This is the first time it has been shown that lossy compression combined with checkpoint-recomputation can provide a much better error-memory tradeoff than using checkpoint-recomputation or lossy compression alone.
Poster 37: Large Scale Ultrasound Simulations on Accelerated Clusters
Filip Vaverka (Brno University of Technology)

Many emerging non-invasive medical procedures such as high intensity focused ultrasound treatment (HIFU), photoacoustic imaging, or transcranial ultrasound therapy require ultrasound wave propagation simulations. Typically soft tissue can be modeled as fluid, weakly heterogeneous medium with frequency dependent absorption and non-linear effects at high ultrasound intensities. The k-Wave acoustic toolbox, widely used in medical community, uses highly efficient k-space pseudo-spectral time-domain (KSTD) discretization of acoustic equations to solve these problems. However, it is difficult to efficiently map the KSTD method onto modern cluster architectures with accelerators such as GPUs. The present thesis analyzes shortcomings of the KSTD method in respect to modern clusters and proposes local Fourier basis approach to improve scaling of the method. It is shown that the proposed method is able to achieve 5x speedup, while having sufficient accuracy for these medical applications. Behavior of the method is analyzed across variety of GPU and MIC (Intel Xeon Phi) accelerated clusters and results are presented.

Poster 34: Analysis of Automata Processing Acceleration on Disparate Hardware Technologies
Marziyeh Nourian (North Carolina State University)

Pattern matching is a computation that maps naturally onto finite automata (FA) abstractions. There has been a substantial amount of work on accelerating FA processing on various parallel platforms. However, the advantages and disadvantages of different automata processing accelerators and the innovation space in this area are still unclear. We target this problem and propose a compiler toolchain that automates the deployment of non-deterministic finite automata (NFAs) onto different target platforms. Using this toolchain, we perform an apples-to-apples comparison between AP, GPU- and FPGA-based NFA accelerator designs on large-scale datasets. Specifically, we observe that memory-based designs are limited by memory size and bandwidth. To address this issue, we target fixed-topology NFAs and propose a memory-efficient design that embeds the automata topology in code and stores only the transition symbols in memory. Our solution is suitable for SIMD architectures and is called SIMD_NFA. We design a compiler that automates the deployment of this design on SIMD platforms. We showcase our compiler framework on GPU and Intel platforms. Additionally, we observe that for NFAs with a grid-like fixed-topology (e.g., NFAs for Levenshtein and Hamming distance-based matching), transitions do not need to be encoded within the traversal code but can be inferred from the reference string to be matched and the knowledge of the NFA topology. Lastly, SIMD_NFA is a good fit for FPGA deployment using OpenCL-to-FPGA toolchains. We investigate the deployment of the OpenCL version of SIMD_NFA, on FPGA and explore a set of
optimizations techniques to retarget SIMD_NFA to FPGA.

**Poster 42: Power and Thermal Management Runtimes for HPC Applications in the Era of Exascale Computing**

Daniele Cesarini (University of Bologna, CINECA)

In the scope of technical and scientific computing, the rush toward larger simulations has been so far assisted by a steady downsizing of micro-processing units, which has allowed to increase the compute capacity of general-purpose architectures at constant power. As side effects of the end of Dennard's scaling, this process is now hitting its ultimate power limits and is just about to come to an end. This implies an increase in the energy cost for computation, performance loss due to a design based on worst-case power consumption, and performance loss due to overheating, and thermal gradients. As result, thermal and power bound supercomputing machines show performance degradation and heterogeneity which limit the peak performance of the system. This doctoral showcase presents software strategies to tackle the main bottlenecks induced by power and thermal issues that affect next-generation supercomputers. To respond to the above challenges, my work shows that propagating workload requirements from application to the runtime and operating system levels is the key to provide efficiency. This is possible only if the proposed software methodologies cause little or no overhead in term of application performance. With this in mind in my work, I have designed application-aware node-level optimal thermal management algorithms and runtimes, lazy node-level power capping, and energy reduction runtime. The experimental results show a significant step forward with respect to the current state-of-the-art solutions in power and thermal control of HPC systems.

**Poster 41: Co-Designing Communication Middleware and Deep Learning Frameworks for High-Performance DNN Training on HPC Systems**

Ammar Ahmad Awan (Ohio State University)

Recent advances in Machine/Deep Learning techniques have triggered key success stories in many application domains like Computer Vision, Speech Comprehension and Recognition, and Natural Language Processing. Large-scale Deep Neural Networks (DNNs), that are at the core of state-of-the-art AI technologies, have been the primary drivers of this success. Training very complicated and large DNN architectures using a large number of training examples (data) is compute-intensive and can take from weeks to months to achieve state-of-the-art prediction capabilities. To achieve higher accuracy, making the DNN deeper is also a common strategy. These requirements have led to a simple but powerful approach called Data Parallelism to achieve shorter training times. This has resulted in various research studies and ML/DL software like TensorFlow and PyTorch as well as
distributed-training middleware like Horovod. In addition, for DNNs that do not fit the GPU memory, a new DNN workloads are emerging that we call Out-of-Core DNNs and different strategies (out-of-core training and model-parallelism) are needed to train them. Clearly, large-scale DNN training brings forward new requirements for computation runtimes like CUDA and communication middleware like the MVAPICH2 MPI library. In this thesis, we broadly explore three different strategies to train DNNs on modern CPU and GPU architectures: 1) Data Parallelism, 2) Model Parallelism, and 3) Out-of-Core Training. We address the key challenge: How to co-design computation and communication in modern ML/DL frameworks with execution runtimes like CUDA and communication middleware like MVAPICH2 to enable scalable, high-performance, and efficient training of DNNs on large-scale HPC systems?

**Poster 40: Performance, Portability, and Productivity for Data-Parallel Computations on Multi- and Many-Core Architectures**

Ari Rasch (University of Münster)

This thesis presents an approach to performance, portability, and productivity for data-parallel computations on multi- and many-core architectures, e.g., Intel CPU and NVIDIA GPU. We introduce the algebraic formalism of Multi-Dimensional Homomorphisms (MDHs) – a class of functions that cover important data-parallel computations, e.g., linear algebra routines (BLAS) and stencil computations. For our MDHs, we propose a Domain-Specific Language (DSL), based on patterns of parallelism (a.k.a. algorithmic skeletons), to enable conveniently expressing MDH functions. We introduce a code generation approach for our DSL to automatically generate for MDHs optimized program code targeting multi- and many-core architectures. Our code generation approach relies on OpenCL – an emerging de-facto standard for uniformly programming parallel architectures, such as CPU and GPU. A major feature of our generated code is that it is targeted to OpenCL’s abstract device models (rather than a particular architecture) by being parameterized in performance-critical parameters of these abstract models (e.g., the number of threads and size of tiles). With our code generation approach, we enable both high performance and performance portability: we fully automatically optimize our generated code -- for any given combination of an MDH function, architecture, and input size -- by automatically choosing (auto-tuning) optimized values of our code’s performance-critical parameters using our own Auto-Tuning Framework (ATF). Our experimental results on CPU and GPU demonstrate competitive and often significantly better performance of our MDH+ATF approach as compared to the currently best-performing competitors, e.g., Intel MKL/MKL-DNN, NVIDIA cuBLAS/cuDNN, and Facebook’s Tensor Comprehensions framework.

**Poster 45: Characterization and Modeling of Error Resilience in HPC Applications**
Luanzheng Guo (University of California, Merced)

As high-performance computing systems scale in size and computational power, the occurrence of transient faults grows. Without protection by efficient and effective fault tolerance mechanisms, transient errors can cause incorrect execution outcome and even lead to a catastrophe in safety-critical HPC applications. Previous work attributes error resilience in HPC applications at a high-level to either the probabilistic or iterative nature of the application, whereas the community still lacks the fundamental understanding of the program constructs that result in natural error resilience. We design FlipTracker, a framework to analytically track error propagation and to provide a fine-grained understanding of the propagation and tolerance of errors. After running FlipTracker on representative HPC applications, we summarize six resilience computation patterns that lead to nature error resilience in HPC applications. With a better understanding of natural resilience in HPC applications, we aim to model application resilience on data objects to transient faults. Many common application-level fault tolerance mechanisms focus on data objects. Understanding application resilience on data objects can be helpful to direct those mechanisms. The common practice to understand application resilience (random fault injection) gives us little knowledge of how and where errors are tolerated. Understanding "how" and "where" is necessary to understand how to apply application-level fault tolerance mechanisms effectively and efficiently. We design a practical model (MOARD) to measure application resilience on data objects by analytically quantifying error masking events happening to the data object. Using our model, users can compare application resilience on different data objects with different data types.

Poster 43: Efficient and Scalable Communication Middleware for Emerging Dense-GPU Clusters
Ching-Hsiang Chu (Ohio State University)

In the era of post Moore’s law, the traditional CPU is not able to keep the pace up and provide the computing power demanded by the modern compute-intensive and highly parallelizable applications. Under this context, various accelerator architectures such as general-purpose graphics processing unit (GPU), which equipped with the high-bandwidth memory (HBM) and massive parallelizable streaming multiprocessors, has been widely adopted in high-performance computing (HPC) and cloud systems to significantly accelerate numerous scientific and emerging machine/deep learning applications. Message Passing Interface (MPI), the standard programming model for parallel applications, has been widely used for GPU communication. However, the state-of-the-art MPI libraries are only optimizing GPU communication by leveraging advanced technology like Remote Direct Memory Access (RDMA) and not fully utilizing the computational power of GPUs. In this work, we propose GPU-enabled communication schemes to harness GPU computational resources, and cutting-edge interconnects such as NVIDIA NVLink for
communication operations on the emerging heterogeneous systems. In this work, three primary MPI operations are addressed. First, intelligent communication scheduling, efficient packing/unpacking, and packing-free schemes are proposed to accelerate non-contiguous data transfer in scientific HPC applications. Second, scalable broadcast operations are presented to leverage the low-level hardware multicast feature to speed up GPU communication at scale. Finally, we also design topology-aware, link-efficient, and cooperative GPU kernels to significantly accelerate All-reduce operation, which is the primary performance bottleneck in deep learning applications. The proposed designs demonstrate significant performance improvements over the state-of-the-art communication schemes for various HPC and deep learning applications.

Poster 39: Designing Next-Generation Communication Middlewares for Many-Core Architectures
Jahanzeb Maqbool Hashmi (Ohio State University)

Modern multi-petaflop HPC systems are powered by dense multi-/many-core architectures and this trend is expected to grow for the next-generation supercomputing systems. This rapid adoption of high core-density architectures by the current- and next-generation HPC systems is further fueled by the emerging application trends such as Deep Learning. This is putting more emphasis on middleware designers to optimize various communication protocols to meet the diverse needs of the applications. While the novelties in the processor architectures have led to the increased on-chip parallelism, they come at the cost of rendering traditional designs, employed by the communication runtimes such as MPI, to suffer from higher degree of intra-node communication latencies. Tackling the computation and communication challenges that accompany these dense multi-/many-cores garner special design considerations. The proposed work in this thesis tries to address the performance challenges posed by a diverse range of applications and the lacking support in state-of-the-art communication libraries such as MPI to exploit high-concurrency architectures. The author first proposes a "shared-address-spaces"-based communication substrate to derive intra-node communication in MPI. Atop this framework, the author has re-designed various MPI primitives such as point-to-point communication protocols (e.g., user-space zero-copy rendezvous transfer), collective communication (e.g., load/store based collectives, truly zero-copy and partitioning-based reduction algorithms), and efficient MPI derived datatypes processing (e.g., memoization-based "packing-free" communication) to exploit the potential of emerging multi-/many-core architectures and high throughput networks. The proposed designs have demonstrated significant improvement over state-of-the-art for various scientific and deep learning applications.
Thursday, November 21

8:30 am - 5:00 pm

Doctoral Showcase Posters Display

**Poster 36: Modeling Non-Determinism in HPC Applications**
Dylan Chapp (University of Delaware, University of Tennessee)

As HPC applications migrate from the petascale systems of today to the exascale systems of tomorrow, the increasing need to embrace asynchronous, irregular, and dynamic communication patterns will lead to a corresponding decrease in application-level determinism. Two critical challenges emerge from this trend. First, unchecked non-determinism coupled with the non-associativity of floating-point arithmetic undermines numerical reproducibility of scientific applications. Second, the prevalence of non-determinism amplifies the cost of debugging, both in terms of computing resources and human effort. In this thesis, we present a modeling methodology to quantify and characterize communication non-determinism in parallel applications. Our methodology consists of three core components. First, we build graph-structured models of relevant communication events from execution traces. Second, we apply similarity metrics based on graph kernels to quantify run-to-run variability and thus identify the regions of executions where non-determinism manifests most prominently. Third, we leverage our notion of execution similarity to characterize applications via clustering, anomaly detection, and extraction of representative patterns of non-deterministic communication which we dub "non-determinism motifs". Our work will amplify the effectiveness of software tools that target mitigation or control of application-level non-determinism (e.g., record-and-replay tools) by providing them with a common metric for quantifying communication non-determinism in parallel applications and a common language for describing it.

**Poster 35: Scaling Up Pipeline Optimization with High Performance Computing**
Robert Lim (University of Oregon)

My research focuses on developing a pipeline optimization infrastructure that automates the design and code generation of neural networks through the use of high-performance computing. The problem has the following objectives: unify automated machine learning (AutoML) and compilation, archive profiles for creation of a knowledge base for a data-driven approach toward search, explore various search optimizations for model design and code generation. The field of automated deep learning includes hyperparameter optimization and neural architecture search (NAS), which requires
domain expertise in designing a model, in addition to the tuning parameters related to learning and the model itself. The search space is complex and deciding which parameters factor into the overall accuracy of a model is a non-trivial task. Once a model is trained, the next step compiles the model, which maps to the backend of a targeted architecture, whether GPU, embedded mobile phones, or FPGA. The compilation phase also involves choices, in terms of optimizations applied, which can include a set of flags, or direct code transformations. Various efficient search techniques are explored, archiving results along the way for facilitating in code generation options, which could possibly inform the model design process. Our previous work reduced the search space for code generation of various domain kernels by 92%, and this work investigates whether the same approach can be applied in all aspects of AutoML design and code generation.

Poster 38: High-Performance Backpropagation in Scientific Computing
Navjot Kukreja (Imperial College, London)

Devito is a domain-specific language for the automatic generation of high-performance solvers for finite difference equations provided in a high-level symbolic representation. The primary use of Devito is to enable Full-waveform inversion, which is an adjoint-based optimization problem that uses the wave equation as part of its objective function. This doctoral project consists of three elements of Devito that are key to solving adjoint-based optimization problems (i.e. the back propagation). The first is automatic differentiation of stencil loops. This allows the automatic generation of high-performance code implementing the derivative of any provided function. This is essential to enable the use of a wider range of physical equations - in order to use better physical approximations. An essential feature of the generated derivatives is that they show scaling and performance behaviour that is very similar to the forward function. A common issue in solving adjoint-based optimization problems is the prohibitively high memory requirement. The second contribution is a runtime for automatic execution of checkpoint-recompute schedules (called pyRevolve) to alleviate this memory requirement by trading it off for re-computations. These schedules may involve storing some intermediate states on disk, some in memory, and recomputing others. The third contribution is the use of lossy compression (ZFP/SZ), which is a second means of trading off memory and compute, automatically as part of checkpointing in pyRevolve. This is the first time it has been shown that lossy compression combined with checkpoint-recomputation can provide a much better error-memory tradeoff than using checkpoint-recomputation or lossy compression alone.

Poster 37: Large Scale Ultrasound Simulations on Accelerated Clusters
Filip Vaverka (Brno University of Technology)
Many emerging non-invasive medical procedures such as high intensity focused ultrasound treatment (HIFU), photoacoustic imaging, or transcranial ultrasound therapy require ultrasound wave propagation simulations. Typically soft tissue can be modeled as fluid, weakly heterogeneous medium with frequency dependent absorption and non-linear effects at high ultrasound intensities. The k-Wave acoustic toolbox, widely used in medical community, uses highly efficient k-space pseudo-spectral time-domain (KSTD) discretization of acoustic equations to solve these problems. However, it is difficult to efficiently map the KSTD method onto modern cluster architectures with accelerators such as GPUs. The present thesis analyzes shortcomings of the KSTD method in respect to modern clusters and proposes local Fourier basis approach to improve scaling of the method. It is shown that the proposed method is able to achieve 5x speedup, while having sufficient accuracy for these medical applications. Behavior of the method is analyzed across variety of GPU and MIC (Intel Xeon Phi) accelerated clusters and results are presented.

**Poster 34: Analysis of Automata Processing Acceleration on Disparate Hardware Technologies**

Marziyeh Nourian (North Carolina State University)

Pattern matching is a computation that maps naturally onto finite automata (FA) abstractions. There has been a substantial amount of work on accelerating FA processing on various parallel platforms. However, the advantages and disadvantages of different automata processing accelerators and the innovation space in this area are still unclear. We target this problem and propose a compiler tool-chain that automates the deployment of non-deterministic finite automata (NFAs) onto different target platforms. Using this toolchain, we perform an apples-to-apples comparison between AP, GPU- and FPGA-based NFA accelerator designs on large-scale datasets. Specifically, we observe that memory-based designs are limited by memory size and bandwidth. To address this issue, we target fixed-topology NFAs and propose a memory-efficient design that embeds the automata topology in code and stores only the transition symbols in memory. Our solution is suitable for SIMD architectures and is called SIMD_NFA. We design a compiler that automates the deployment of this design on SIMD platforms. We showcase our compiler framework on GPU and Intel platforms. Additionally, we observe that for NFAs with a grid-like fixed-topology (e.g., NFAs for Levenshtein and Hamming distance-based matching), transitions do not need to be encoded within the traversal code but can be inferred from the reference string to be matched and the knowledge of the NFA topology. Lastly, SIMD_NFA is a good fit for FPGA deployment using OpenCL-to-FPGA toolchains. We investigate the deployment of the OpenCL version of SIMD_NFA, on FPGA and explore a set of optimizations techniques to retarget SIMD_NFA to FPGA.

**Poster 42: Power and Thermal Management Runtimes for HPC Applications in the Era of Exascale Computing**
In the scope of technical and scientific computing, the rush toward larger simulations has been so far assisted by a steady downsizing of micro-processing units, which has allowed to increase the compute capacity of general-purpose architectures at constant power. As side effects of the end of Dennard’s scaling, this process is now hitting its ultimate power limits and is just about to come to an end. This implies an increase in the energy cost for computation, performance loss due to a design based on worst-case power consumption, and performance loss due to overheating, and thermal gradients. As result, thermal and power bound supercomputing machines show performance degradation and heterogeneity which limit the peak performance of the system. This doctoral showcase presents software strategies to tackle the main bottlenecks induced by power and thermal issues that affect next-generation supercomputers. To respond to the above challenges, my work shows that propagating workload requirements from application to the runtime and operating system levels is the key to provide efficiency. This is possible only if the proposed software methodologies cause little or no overhead in term of application performance. With this in mind in my work, I have designed application-aware node-level optimal thermal management algorithms and runtimes, lazy node-level power capping, and energy reduction runtime. The experimental results show a significant step forward with respect to the current state-of-the-art solutions in power and thermal control of HPC systems.

Poster 41: Co-Designing Communication Middleware and Deep Learning Frameworks for High-Performance DNN Training on HPC Systems
Ammar Ahmad Awan (Ohio State University)

Recent advances in Machine/Deep Learning techniques have triggered key success stories in many application domains like Computer Vision, Speech Comprehension and Recognition, and Natural Language Processing. Large-scale Deep Neural Networks (DNNs), that are at the core of state-of-the-art AI technologies, have been the primary drivers of this success. Training very complicated and large DNN architectures using a large number of training examples (data) is compute-intensive and can take from weeks to months to achieve state-of-the-art prediction capabilities. To achieve higher accuracy, making the DNN deeper is also a common strategy. These requirements have led to a simple but powerful approach called Data Parallelism to achieve shorter training times. This has resulted in various research studies and ML/DL software like TensorFlow and PyTorch as well as distributed-training middleware like Horovod. In addition, for DNNs that do not fit the GPU memory, a new DNN workloads are emerging that we call Out-of-Core DNNs and different strategies (out-of-core training and model-parallelism) are needed to train them. Clearly, large-scale DNN training brings forward new requirements for computation runtimes like CUDA and communication middleware like the MVAPICH2 MPI library. In this thesis, we broadly explore three different
strategies to train DNNs on modern CPU and GPU architectures: 1) Data Parallelism, 2) Model Parallelism, and 3) Out-of-Core Training. We address the key challenge: How to co-design computation and communication in modern ML/DL frameworks with execution runtimes like CUDA and communication middleware like MVAPICH2 to enable scalable, high-performance, and efficient training of DNNs on large-scale HPC systems?

Poster 40: Performance, Portability, and Productivity for Data-Parallel Computations on Multi- and Many-Core Architectures
Ari Rasch (University of Münster)

This thesis presents an approach to performance, portability, and productivity for data-parallel computations on multi- and many-core architectures, e.g., Intel CPU and NVIDIA GPU. We introduce the algebraic formalism of Multi-Dimensional Homomorphisms (MDHs) – a class of functions that cover important data-parallel computations, e.g., linear algebra routines (BLAS) and stencil computations. For our MDHs, we propose a Domain-Specific Language (DSL), based on patterns of parallelism (a.k.a. algorithmic skeletons), to enable conveniently expressing MDH functions. We introduce a code generation approach for our DSL to automatically generate for MDHs optimized program code targeting multi- and many-core architectures. Our code generation approach relies on OpenCL – an emerging de-facto standard for uniformly programming parallel architectures, such as CPU and GPU. A major feature of our generated code is that it is targeted to OpenCL’s abstract device models (rather than a particular architecture) by being parameterized in performance-critical parameters of these abstract models (e.g., the number of threads and size of tiles). With our code generation approach, we enable both high performance and performance portability: we fully automatically optimize our generated code -- for any given combination of an MDH function, architecture, and input size -- by automatically choosing (auto-tuning) optimized values of our code’s performance-critical parameters using our own Auto-Tuning Framework (ATF). Our experimental results on CPU and GPU demonstrate competitive and often significantly better performance of our MDH+ATF approach as compared to the currently best-performing competitors, e.g., Intel MKL/MKL-DNN, NVIDIA cuBLAS/cuDNN, and Facebook’s Tensor Comprehensions framework.

Poster 45: Characterization and Modeling of Error Resilience in HPC Applications
Luanzheng Guo (University of California, Merced)

As high-performance computing systems scale in size and computational power, the occurrence of transient faults grows. Without protection by efficient and effective fault tolerance mechanisms, transient errors can cause incorrect execution outcome and even lead to a catastrophe in safety-
critical HPC applications. Previous work attributes error resilience in HPC applications at a high-
level to either the probabilistic or iterative nature of the application, whereas the community still
lacks the fundamental understanding of the program constructs that result in natural error
resilience. We design FlipTracker, a framework to analytically track error propagation and to provide
a fine-grained understanding of the propagation and tolerance of errors. After running FlipTracker
on representative HPC applications, we summarize six resilience computation patterns that lead to
natural error resilience in HPC applications. With a better understanding of natural resilience in HPC
applications, we aim to model application resilience on data objects to transient faults. Many
common application-level fault tolerance mechanisms focus on data objects. Understanding
application resilience on data objects can be helpful to direct those mechanisms. The common
practice to understand application resilience (random fault injection) gives us little knowledge of
how and where errors are tolerated. Understanding "how" and "where" is necessary to understand
how to apply application-level fault tolerance mechanisms effectively and efficiently. We design a
practical model (MOARD) to measure application resilience on data objects by analytically
quantifying error masking events happening to the data object. Using our model, users can compare
application resilience on different data objects with different data types.

**Poster 43: Efficient and Scalable Communication Middleware for Emerging Dense-GPU
Clusters**

Ching-Hsiang Chu (Ohio State University)

In the era of post Moore’s law, the traditional CPU is not able to keep the pace up and provide the
computing power demanded by the modern compute-intensive and highly parallelizable
applications. Under this context, various accelerator architectures such as general-purpose graphics
processing unit (GPU), which equipped with the high-bandwidth memory (HBM) and massive
parallelizable streaming multiprocessors, has been widely adopted in high-performance computing
(HPC) and cloud systems to significantly accelerate numerous scientific and emerging
machine/deep learning applications. Message Passing Interface (MPI), the standard programming
model for parallel applications, has been widely used for GPU communication. However, the state-
of-the-art MPI libraries are only optimizing GPU communication by leveraging advanced technology
like Remote Direct Memory Access (RDMA) and not fully utilizing the computational power of
GPUs. In this work, we propose GPU-enabled communication schemes to harness GPU
computational resources, and cutting-edge interconnects such as NVIDIA NVLink for
communication operations on the emerging heterogeneous systems. In this work, three primary MPI
operations are addressed. First, intelligent communication scheduling, efficient packing/unpacking,
and packing-free schemes are proposed to accelerate non-contiguous data transfer in scientific
HPC applications. Second, scalable broadcast operations are presented to leverage the low-level
hardware multicast feature to speed up GPU communication at scale. Finally, we also design
designs demonstrate significant performance improvements over the state-of-the-art communication schemes for various HPC and deep learning applications.

Poster 39: Designing Next-Generation Communication Middlewares for Many-Core Architectures
Jahanzeb Maqbool Hashmi (Ohio State University)

Modern multi-petaflop HPC systems are powered by dense multi-/many-core architectures and this trend is expected to grow for the next-generation supercomputing systems. This rapid adoption of high core-density architectures by the current- and next-generation HPC systems is further fueled by the emerging application trends such as Deep Learning. This is putting more emphasis on middleware designers to optimize various communication protocols to meet the diverse needs of the applications. While the novelties in the processor architectures have led to the increased on-chip parallelism, they come at the cost of rendering traditional designs, employed by the communication runtimes such as MPI, to suffer from higher degree of intra-node communication latencies. Tackling the computation and communication challenges that accompany these dense multi-/many-cores garner special design considerations. The proposed work in this thesis tries to address the performance challenges posed by a diverse range of applications and the lacking support in state-of-the-art communication libraries such as MPI to exploit high-concurrency architectures. The author first proposes a "shared-address-spaces"-based communication substrate to derive intra-node communication in MPI. Atop this framework, the author has re-designed various MPI primitives such as point-to-point communication protocols (e.g., user-space zero-copy rendezvous transfer), collective communication (e.g., load/store based collectives, truly zero-copy and partitioning-based reduction algorithms), and efficient MPI derived datatypes processing (e.g., memoization-based "packing-free" communication) to exploit the potential of emerging multi-/many-core architectures and high throughput networks. The proposed designs have demonstrated significant improvement over state-of-the-art for various scientific and deep learning applications.

10:30 am - 12:00 pm

Doctoral Showcase I Presentations

Poster 42: Power and Thermal Management Runtimes for HPC Applications in the Era of
In the scope of technical and scientific computing, the rush toward larger simulations has been so far assisted by a steady downsizing of micro-processing units, which has allowed to increase the compute capacity of general-purpose architectures at constant power. As side effects of the end of Dennard's scaling, this process is now hitting its ultimate power limits and is just about to come to an end. This implies an increase in the energy cost for computation, performance loss due to a design based on worst-case power consumption, and performance loss due to overheating, and thermal gradients. As result, thermal and power bound supercomputing machines show performance degradation and heterogeneity which limit the peak performance of the system. This doctoral showcase presents software strategies to tackle the main bottlenecks induced by power and thermal issues that affect next-generation supercomputers. To respond to the above challenges, my work shows that propagating workload requirements from application to the runtime and operating system levels is the key to provide efficiency. This is possible only if the proposed software methodologies cause little or no overhead in term of application performance. With this in mind in my work, I have designed application-aware node-level optimal thermal management algorithms and runtimes, lazy node-level power capping, and energy reduction runtime. The experimental results show a significant step forward with respect to the current state-of-the-art solutions in power and thermal control of HPC systems.

Poster 45: Characterization and Modeling of Error Resilience in HPC Applications
Luanzheng Guo (University of California, Merced)

As high-performance computing systems scale in size and computational power, the occurrence of transient faults grows. Without protection by efficient and effective fault tolerance mechanisms, transient errors can cause incorrect execution outcome and even lead to a catastrophe in safety-critical HPC applications. Previous work attributes error resilience in HPC applications at a high-level to either the probabilistic or iterative nature of the application, whereas the community still lacks the fundamental understanding of the program constructs that result in natural error resilience. We design FlipTracker, a framework to analytically track error propagation and to provide a fine-grained understanding of the propagation and tolerance of errors. After running FlipTracker on representative HPC applications, we summarize six resilience computation patterns that lead to nature error resilience in HPC applications. With a better understanding of natural resilience in HPC applications, we aim to model application resilience on data objects to transient faults. Many common application-level fault tolerance mechanisms focus on data objects. Understanding application resilience on data objects can be helpful to direct those mechanisms. The common practice to understand application resilience (random fault injection) gives us little knowledge of
how and where errors are tolerated. Understanding "how" and "where" is necessary to understand how to apply application-level fault tolerance mechanisms effectively and efficiently. We design a practical model (MOARD) to measure application resilience on data objects by analytically quantifying error masking events happening to the data object. Using our model, users can compare application resilience on different data objects with different data types.

**Poster 39: Designing Next-Generation Communication Middlewares for Many-Core Architectures**

Jahanzeb Maqbool Hashmi (Ohio State University)

Modern multi-petaflop HPC systems are powered by dense multi-/many-core architectures and this trend is expected to grow for the next-generation supercomputing systems. This rapid adoption of high core-density architectures by the current- and next-generation HPC systems is further fueled by the emerging application trends such as Deep Learning. This is putting more emphasis on middleware designers to optimize various communication protocols to meet the diverse needs of the applications. While the novelties in the processor architectures have led to the increased on-chip parallelism, they come at the cost of rendering traditional designs, employed by the communication runtimes such as MPI, to suffer from higher degree of intra-node communication latencies. Tackling the computation and communication challenges that accompany these dense multi-/many-cores garner special design considerations. The proposed work in this thesis tries to address the performance challenges posed by a diverse range of applications and the lacking support in state-of-the-art communication libraries such as MPI to exploit high-concurrency architectures. The author first proposes a "shared-address-spaces"-based communication substrate to derive intra-node communication in MPI. Atop this framework, the author has re-designed various MPI primitives such as point-to-point communication protocols (e.g., user-space zero-copy rendezvous transfer), collective communication (e.g., load/store based collectives, truly zero-copy and partitioning-based reduction algorithms), and efficient MPI derived datatypes processing (e.g., memoization-based "packing-free" communication) to exploit the potential of emerging multi-/many-core architectures and high throughput networks. The proposed designs have demonstrated significant improvement over state-of-the-art for various scientific and deep learning applications.

1:30 pm - 3:00 pm

**Doctoral Showcase II Presentations**
Poster 41: Co-Designing Communication Middleware and Deep Learning Frameworks for High-Performance DNN Training on HPC Systems
Ammar Ahmad Awan (Ohio State University)

Recent advances in Machine/Deep Learning techniques have triggered key success stories in many application domains like Computer Vision, Speech Comprehension and Recognition, and Natural Language Processing. Large-scale Deep Neural Networks (DNNs), that are at the core of state-of-the-art AI technologies, have been the primary drivers of this success. Training very complicated and large DNN architectures using a large number of training examples (data) is compute-intensive and can take from weeks to months to achieve state-of-the-art prediction capabilities. To achieve higher accuracy, making the DNN deeper is also a common strategy. These requirements have led to a simple but powerful approach called Data Parallelism to achieve shorter training times. This has resulted in various research studies and ML/DL software like TensorFlow and PyTorch as well as distributed-training middleware like Horovod. In addition, for DNNs that do not fit the GPU memory, a new DNN workloads are emerging that we call Out-of-Core DNNs and different strategies (out-of-core training and model-parallelism) are needed to train them. Clearly, large-scale DNN training brings forward new requirements for computation runtimes like CUDA and communication middleware like the MVAPICH2 MPI library. In this thesis, we broadly explore three different strategies to train DNNs on modern CPU and GPU architectures: 1) Data Parallelism, 2) Model Parallelism, and 3) Out-of-Core Training. We address the key challenge: How to co-design computation and communication in modern ML/DL frameworks with execution runtimes like CUDA and communication middleware like MVAPICH2 to enable scalable, high-performance, and efficient training of DNNs on large-scale HPC systems?

Poster 40: Performance, Portability, and Productivity for Data-Parallel Computations on Multi- and Many-Core Architectures
Ari Rasch (University of Münster)

This thesis presents an approach to performance, portability, and productivity for data-parallel computations on multi- and many-core architectures, e.g., Intel CPU and NVIDIA GPU. We introduce the algebraic formalism of Multi-Dimensional Homomorphisms (MDHs) – a class of functions that cover important data-parallel computations, e.g., linear algebra routines (BLAS) and stencil computations. For our MDHs, we propose a Domain-Specific Language (DSL), based on patterns of parallelism (a.k.a. algorithmic skeletons), to enable conveniently expressing MDH functions. We introduce a code generation approach for our DSL to automatically generate MDHs optimized program code targeting multi- and many-core architectures. Our code generation approach relies on OpenCL – an emerging de-facto standard for uniformly programming parallel architectures, such as CPU and GPU. A major feature of our generated code is that it is targeted to
OpenCL’s abstract device models (rather than a particular architecture) by being parameterized in performance-critical parameters of these abstract models (e.g., the number of threads and size of tiles). With our code generation approach, we enable both high performance and performance portability: we fully automatically optimize our generated code -- for any given combination of an MDH function, architecture, and input size -- by automatically choosing (auto-tuning) optimized values of our code’s performance-critical parameters using our own Auto-Tuning Framework (ATF). Our experimental results on CPU and GPU demonstrate competitive and often significantly better performance of our MDH+ATF approach as compared to the currently best-performing competitors, e.g., Intel MKL/MKL-DNN, NVIDIA cuBLAS/cuDNN, and Facebook’s Tensor Comprehensions framework.

**Poster 43: Efficient and Scalable Communication Middleware for Emerging Dense-GPU Clusters**  
Ching-Hsiang Chu (Ohio State University)

In the era of post Moore’s law, the traditional CPU is not able to keep the pace up and provide the computing power demanded by the modern compute-intensive and highly parallelizable applications. Under this context, various accelerator architectures such as general-purpose graphics processing unit (GPU), which equipped with the high-bandwidth memory (HBM) and massive parallelizable streaming multiprocessors, has been widely adopted in high-performance computing (HPC) and cloud systems to significantly accelerate numerous scientific and emerging machine/deep learning applications. Message Passing Interface (MPI), the standard programming model for parallel applications, has been widely used for GPU communication. However, the state-of-the-art MPI libraries are only optimizing GPU communication by leveraging advanced technology like Remote Direct Memory Access (RDMA) and not fully utilizing the computational power of GPUs. In this work, we propose GPU-enabled communication schemes to harness GPU computational resources, and cutting-edge interconnects such as NVIDIA NVLink for communication operations on the emerging heterogeneous systems. In this work, three primary MPI operations are addressed. First, intelligent communication scheduling, efficient packing/unpacking, and packing-free schemes are proposed to accelerate non-contiguous data transfer in scientific HPC applications. Second, scalable broadcast operations are presented to leverage the low-level hardware multicast feature to speed up GPU communication at scale. Finally, we also design topology-aware, link-efficient, and cooperative GPU kernels to significantly accelerate All-reduce operation, which is the primary performance bottleneck in deep learning applications. The proposed designs demonstrate significant performance improvements over the state-of-the-art communication schemes for various HPC and deep learning applications.
Devito is a domain-specific language for the automatic generation of high-performance solvers for finite difference equations provided in a high-level symbolic representation. The primary use of Devito is to enable Full-waveform inversion, which is an adjoint-based optimization problem that uses the wave equation as part of its objective function. This doctoral project consists of three elements of Devito that are key to solving adjoint-based optimization problems (i.e. the back propagation). The first is automatic differentiation of stencil loops. This allows the automatic generation of high-performance code implementing the derivative of any provided function. This is essential to enable the use of a wider range of physical equations - in order to use better physical approximations. An essential feature of the generated derivatives is that they show scaling and performance behaviour that is very similar to the forward function. A common issue in solving adjoint-based optimization problems is the prohibitively high memory requirement. The second contribution is a runtime for automatic execution of checkpoint-recompute schedules (called pyRevolve) to alleviate this memory requirement by trading it off for re-computations. These schedules may involve storing some intermediate states on disk, some in memory, and recomputing others. The third contribution is the use of lossy compression (ZFP/SZ), which is a second means of trading off memory and compute, automatically as part of checkpointing in pyRevolve. This is the first time it has been shown that lossy compression combined with checkpoint-recomputation can provide a much better error-memory tradeoff than using checkpoint-recomputation or lossy compression alone.
the model itself. The search space is complex and deciding which parameters factor into the overall accuracy of a model is a non-trivial task. Once a model is trained, the next step compiles the model, which maps to the backend of a targeted architecture, whether GPU, embedded mobile phones, or FPGA. The compilation phase also involves choices, in terms of optimizations applied, which can include a set of flags, or direct code transformations. Various efficient search techniques are explored, archiving results along the way for facilitating in code generation options, which could possibly inform the model design process. Our previous work reduced the search space for code generation of various domain kernels by 92%, and this work investigates whether the same approach can be applied in all aspects of AutoML design and code generation.

**Poster 34: Analysis of Automata Processing Acceleration on Disparate Hardware Technologies**

*Marziyeh Nourian (North Carolina State University)*

Pattern matching is a computation that maps naturally onto finite automata (FA) abstractions. There has been a substantial amount of work on accelerating FA processing on various parallel platforms. However, the advantages and disadvantages of different automata processing accelerators and the innovation space in this area are still unclear. We target this problem and propose a compiler toolchain that automates the deployment of non-deterministic finite automata (NFAs) onto different target platforms. Using this toolchain, we perform an apples-to-apples comparison between AP, GPU- and FPGA-based NFA accelerator designs on large-scale datasets. Specifically, we observe that memory-based designs are limited by memory size and bandwidth. To address this issue, we target fixed-topology NFAs and propose a memory-efficient design that embeds the automata topology in code and stores only the transition symbols in memory. Our solution is suitable for SIMD architectures and is called SIMD_NFA. We design a compiler that automates the deployment of this design on SIMD platforms. We showcase our compiler framework on GPU and Intel platforms. Additionally, we observe that for NFAs with a grid-like fixed-topology (e.g., NFAs for Levenshtein and Hamming distance-based matching), transitions do not need to be encoded within the traversal code but can be inferred from the reference string to be matched and the knowledge of the NFA topology. Lastly, SIMD_NFA is a good fit for FPGA deployment using OpenCL-to-FPGA toolchains. We investigate the deployment of the OpenCL version of SIMD_NFA, on FPGA and explore a set of optimizations techniques to retarget SIMD_NFA to FPGA.

**Poster 36: Modeling Non-Determinism in HPC Applications**

*Dylan Chapp (University of Delaware, University of Tennessee)*

As HPC applications migrate from the petascale systems of today to the exascale systems of tomorrow, the increasing need to embrace asynchronous, irregular, and dynamic communication
patterns will lead to a corresponding decrease in application-level determinism. Two critical challenges emerge from this trend. First, unchecked non-determinism coupled with the non-associativity of floating-point arithmetic undermines numerical reproducibility of scientific applications. Second, the prevalence of non-determinism amplifies the cost of debugging, both in terms of computing resources and human effort. In this thesis, we present a modeling methodology to quantify and characterize communication non-determinism in parallel applications. Our methodology consists of three core components. First, we build graph-structured models of relevant communication events from execution traces. Second, we apply similarity metrics based on graph kernels to quantify run-to-run variability and thus identify the regions of executions where non-determinism manifests most prominently. Third, we leverage our notion of execution similarity to characterize applications via clustering, anomaly detection, and extraction of representative patterns of non-deterministic communication which we dub "non-determinism motifs". Our work will amplify the effectiveness of software tools that target mitigation or control of application-level non-determinism (e.g., record-and-replay tools) by providing them with a common metric for quantifying communication non-determinism in parallel applications and a common language for describing it.

**Poster 37: Large Scale Ultrasound Simulations on Accelerated Clusters**  
*Filip Vaverka (Brno University of Technology)*

Many emerging non-invasive medical procedures such as high intensity focused ultrasound treatment (HIFU), photoacoustic imaging, or transcranial ultrasound therapy require ultrasound wave propagation simulations. Typically soft tissue can be modeled as fluid, weakly heterogeneous medium with frequency dependent absorption and non-linear effects at high ultrasound intensities. The k-Wave acoustic toolbox, widely used in medical community, uses highly efficient k-space pseudo-spectral time-domain (KSTD) discretization of acoustic equations to solve these problems. However, it is difficult to efficiently map the KSTD method onto modern cluster architectures with accelerators such as GPUs. The present thesis analyzes shortcomings of the KSTD method in respect to modern clusters and proposes local Fourier basis approach to improve scaling of the method. It is shown that the proposed method is able to achieve 5x speedup, while having sufficient accuracy for these medical applications. Behavior of the method is analyzed across variety of GPU and MIC (Intel Xeon Phi) accelerated clusters and results are presented.
Early Career Program

Monday, November 18

8:30 am - 10:00 am

Early Career Program: Welcome and Leadership (8:30-10am)

Early Career Program: Welcome and Leadership Panel/Discussion
Ralph McEldowney (US Department of Defense), Valerie Taylor (Argonne National Laboratory), Satoshi Matsuoka (Tokyo Institute of Technology), Alison Kennedy (Hartree Centre, Science and Technology Facilities Council (STFC)), Daniel S. Katz (University of Illinois), Catherine Jones (Science and Technology Facilities Council (STFC))

This session will include an ice-breaking exercise that helps Early Career Program participants understand leadership skills, followed by a panel of SC and community leaders giving short presentations, followed by questions and discussion with the audience.

10:30 am - 12:00 pm

Early Career Program: Career Development (10:30-12pm)

Early Career Program: Career Development
Mary Hall (University of Utah), Terri Quinn (Lawrence Livermore National Laboratory), Atiq Bajwa (Ampere Systems), Chris J. Newburn (Nvidia Corporation), Carina Haupt (German Aerospace Center (DLR))

This session features insightful stories and kernels of wisdom that are fruits of careers of three warm and welcoming speakers that span industry, academia, and government labs. It offers practical wisdom on being intentional about drawing on your strengths, discerning objectives, and pursuing them in a collaborative way. Finally, participants work through and discuss the results of a “SMART” work that focuses on a specific next step in our career.
1:30 pm - 3:00 pm

Early Career Program: Key Professional Communications (1:30-3pm)

Early Career Program: Key Professional Communications
James Hetherington (Turing Institute), Joanna Leng (University of Leeds), Barney Maccabe (Oak Ridge National Laboratory), Almadena Chtchelkanova (National Science Foundation), David E. Bernholdt (Oak Ridge National Laboratory), Louise Brown (University of Nottingham)
This session will focus on several kinds of communications that are important to Early Career professionals: your CV, interactions with research sponsors (including proposals), and crafting research papers.

Tailoring your CV for the Next Career Step. This practical activity will focus on how to maximize the impact of your CV when you’re applying for a post at the next career level, so that shortlisters can find the important information.

Talking to Research Sponsors. A panel of experienced researchers and representatives of research sponsors will discuss their recommendations for communicating with sponsors, through proposals, highlights, and otherwise.

Panelists: Almadena Chtchelkanova - National Science Foundation, James Hetherington - Alan Turing Institute, Joanna Leng - University of Leeds, Barney Maccabe - Oak Ridge National Laboratory

Is Your Paper Conveying the Right Message? This practical activity will explore how the very start of a paper, the title and abstract, can help or hinder the reader from understanding and appreciating your work.

3:30 pm - 5:00 pm

Mentor Protégé Mixer

How to Build Successful Mentoring Relationships

In this session we discuss strategies to build successful and long lasting mentoring relationships. We will address issues on identifying mentors and mentoring needs, building a mentoring network, and keeping these connections for the long run. The session will start with a mentoring panel and
will culminate with an exercise on networking.

**Speed Mentoring**

The second part of the Mixer includes a speed mentoring session. This is a succession of one on one conversations between a mentee and mentor. The mentor rotates to several mentees during a specified amount of time to provide advice and establish a connection.
Exhibitor Forum

(back to top)

Tuesday, November 19

10:30 am - 12:00 pm

Exhibit Forum: Cloud

Navigating HPC Cloud Migration: Best Practices and Beyond
Fritz Ferstl (Univa Corporation)

To optimize cost and efficiencies, organizations are embracing high-performance computing (HPC) solutions that can scale with their compute-intensive workloads. HPC gives these organizations a competitive advantage within the market as they reap the benefits that come from virtually unlimited HPC capacity and extreme-scale. And yet with all of the advantages of migrating workloads to the cloud, this also brings the potential to escalate preexisting complexities. If companies do not have good visibility into how their existing infrastructure is being used, it will be impossible to tell whether the addition of cloud is yielding the results they are looking to achieve. This presentation will highlight best practices for HPC solutions and discuss ways to combat challenges that many companies face.

A Practical Approach for Projecting HPC Workload Costs on the Cloud
Philippe Bricard (UCit), Benjamin Depardon (UCit), Jean-Frederic Thebault (Storengy SAS)

An increasing number of HPC workloads are eligible for a move in Public Cloud. But beside the technical impact of such migration, it is essential for companies and institutions to understand what would be the associated costs.

UCit will present its approach to this challenge and demonstrate how its Analyze-IT software solution helps companies and institutions to select, budget, and track HPC workloads on AWS. The presentation will then feature how Storengy SAS applied it to move a very demanding HPC workload on AWS.
Tailor-Made HPC in the Cloud
Dean Hildebrand (Google LLC)

HPC in the cloud should simplify and optimize application execution and management, but care needs to be taken to ensure that on-premise issues don’t become cloud issues. Workload I/O contention, scheduling, and efficiency issues can all be reduced with Google Cloud. The key to avoiding lift and shifting issues is to ensure that the cloud infrastructure is tailor-made to your applications. Our goal at Google is to allow users to focus on application needs, and not on the toil of deployment, maintenance, and optimization.

This session focuses on how Google Cloud users can renew their focus on application requirements. In one example, file system deployments can be tailored to the IOPs and bandwidth requirements for individual workloads, which is simply not possible on-premise. By taking advantage of resources unique to the cloud, many of the existing issues simply no longer exist, and users can now focus on their HPC applications.

10:30 am - 12:00 pm
Exhibitor Forum: Managing and Storing Data I

Starfish for Research Data Management: Analytics, Metadata, Data Movement, and Orchestration
Jacob Farmer (Starfish Storage, Cambridge Computer Services Inc), Peter Galvin (Starfish Storage)

Starfish is software for managing HPC-scale file systems and object stores. Starfish associates metadata with files and directories. Metadata are used for reporting, web portals, and for shaping storage management policies. Starfish is the most sophisticated and scalable indexing and reporting platform for big file systems. It supports billions and can execute batch processes based on query results.

We will be showcasing our "Storage Zones" end-user portal that enables users to analyze their own file collections and orchestrate data management (backup, archive, delete, migrate, etc.). Once you give your users the proper tools, you can hold them accountable for cleaning up their storage messes and being the stewards of their own data.

Starfish has been battle tested on some of the largest and most demanding file systems in the world. Our software is running in production in DOE labs, NIH labs, universities, and household
Beyond Discoverability: Metadata to Drive Your Data Management
Terrell Russell (Renaissance Computing Institute (RenCI), iRODS Consortium)

As commercial, governmental, and research organizations continue to move from manual pipelines to automated processing of their vast and growing datasets, they are struggling to find meaning in their repositories.

Many products and approaches now provide data discoverability through indexing and aggregate counts, but few also provide the level of confidence needed for making strong assertions about data provenance. For that, a system needs policy to be enforced; a model for data governance that provides understanding about what is in the system and how it came to be.

With an open, policy-based platform, metadata can be elevated beyond assisting in just search and discoverability. Metadata can associate datasets, help build cohorts for analysis, coordinate data movement and scheduling, and drive the very policy that provides the data governance.

Data management should be data-centric and metadata driven.

Hybrid Computing (HC) – Next-Gen Data Access Technology
Mitsuo Koikawa (tonio Co Ltd), Hiroyoshi Takahashi (Shinonome Inc, tonio Co Ltd)

Today, the world around the Internet is flooded with data, which is scattered all over the place. For example, when performing data analytics, it is common to first upload data to the cloud or to an on-premises server, and then perform tasks. However, this method has problems such as the task of uploading data, time, and cost, and therefore the data analytics may not be performed quickly and efficiently.

Hybrid Computing (HC) under development by Our Company is a framework not only to solve these problems at once, but also to expect performance improvement and cost reduction. Our solution is not data-driven but code-driven, an architecture that uniquely handles a wide variety of clients and servers. (In fact, our tests have shown that it performs about 9 to 50 times better.)
Exhibitor Forum: Hardware for AI

Adding Low Latency Capability to Panasas PanFS for AI and HPC
Curtis Anderson (Panasas Inc)

Neural networks are bringing low-latency requirements to HPC environments, but even traditional HPC workloads are changing. Mixed workloads containing small files are increasing in importance.

Commodity platforms are now available for the underlying hardware layers for cost-effective hybrid (HDD plus SSD) and low-latency NVMeof flash tiers. The challenge is in the software architecture for storing and managing content on, and across, those tiers in a way that will consistently meet the performance, cost, reliability, and ease-of-use needs of both HPC and AI workloads.

In this talk we will explore the drivers and constraints on the architectural approaches we might take to add low-latency capabilities to the Panasas PanFS filesystem.

Building a Wafer-Scale Deep Learning System: Lessons Learned
Jean-Philippe Fricker (Cerebras Systems), Andy Hock (Cerebras Systems)

Deep learning has emerged as one of the most important workloads of our time. While it has applicability to many problems, its computational demands are profound. Compute requirements to train the largest deep learning models increased by 300,000x between 2012-2018 [1]. Traditional processors are not well-suited to meet this demand.

To address this challenge, Cerebras has developed a new computer system optimized for deep learning, the CS-1. This system is powered by the largest chip ever built: the Cerebras Wafer-Scale Engine (WSE). Cerebras’ WSE is a single integrated 46,225 mm^2 silicon chip with >1.2 trillion transistors and 400,000 compute cores. It is >56x larger than today’s largest GPU, with 3,000x more on-chip memory and >10,000x memory bandwidth.

Here, we provide a technical overview of the CS-1 and discuss the unique engineering challenges associated with packaging, powering, cooling, and I/O for a wafer-scale processor.


A New Block Floating Point Arithmetic Unit for AI/ML Workloads
Quinn Jacobson (Achronix Semiconductor Corporation)

Block Floating Point (BFP) is a hybrid of floating-point and fixed-point arithmetic where a block of data is assigned a common exponent. We describe a new arithmetic unit that natively performs Block Floating Point for common matrix arithmetic operations and creates floating-point results. The BFP arithmetic unit supports several data formats with varying precision and range. BFP offers substantial power and area savings over traditional floating-point arithmetic units by trading off some precision. This new arithmetic unit has been implemented in the new family of 7nm FPGAs from Achronix. We cover the architecture and supported operations of the BFP unit. In this presentation, artificial intelligence and machine learning workloads are benchmarked to demonstrate the performance improvement and power savings of BFP as compared to half-precision (FP16) operations.

3:30 pm - 5:00 pm

Exhibitor Forum: AI Applications

Implementing Deep Learning in Prostate Cancer Diagnosis
Ping-Chang Lin (University of Chicago), Hakizumwami Birali Runesha (University of Chicago), Teodora Szasz (University of Chicago), Jaime Puente (Lenovo)

AI-based solutions start impacting our daily life in different aspects including facilitating medical imaging analysis and clinical diagnosis. In the area of diagnosing prostate cancer (PCa), which currently affects one in eight men, there is no well-accepted screening method for this important public health problem. In this presentation, we present an infrastructure that implement high performance computing resource to train and validate a proposed deep learning architecture to predict/detect the loci of prostate cancer by using a multi-parametric MRI (mp-MRI) prostate cancer dataset. The outcomes of the optimized deep learning model are further integrated into the web interface that can assist the radiologists in early prediction of PCa. Our tool has the potential to revolutionize the prostate cancer diagnosis and improve patient quality of life, while decreasing its cost by minimizing unnecessary prostate biopsies and surgeries.

Alops: Bringing Artificial Intelligence to the Data Center
David Sickinger (National Renewable Energy Laboratory (NREL)), Sergey Serebryakov (Hewlett Packard Enterprise), Tahir Cader (Hewlett Packard Enterprise)
Nearly one third of all data centers had an incident in 2018 up from 25 percent the year before, according to a recent survey by the Uptime Institute. The average cost of downtime was $260,000 per hour. Yet 80 percent of the incidents could be prevented through robust detection systems. Today, data center incident detection is typically conducted through threshold-based techniques leading to either too many false alarms or late detection of incidents.

We utilize machine learning and AI to detect incidents early, while reducing false alarm rates. We estimate metric statistical distributions across time and metric subsets, and compare the metric behaviour against statistical baselines. We employ correlation engines coupled with domain expertise to identify metric causality relationships. We further use neural networks to identify data center metric behaviour. Our systems are trained and customized for HPC systems with thousands of nodes and tens of thousands of processors.

**Application of Artificial Intelligence to Heart Disease**
David Ellison (Lenovo), Milan Sonka (University of Iowa)

Artificial Intelligence (AI) has the opportunity to change the prevention and treatment of heart disease. Heart attack frequently strikes without warning and even expert cardiologists cannot tell, which of the potentially many initially small coronary plaques that almost invariably exist in all humans starting at their 2nd decade of life will remain stable. A small number of plaques will rupture, and even cause a heart attack. Optical coherence tomography (OCT) imaging depicts plaque tissue at an unprecedented level of detail. Our approach employs deep learning to localize and quantify existing plaques and predict future plaque vulnerability using OCT image data. For heart disease treatment, we developed a mixture that can be injected to repair the heart after a myocardial infarction. We used a novel device that leverages AI to replicate transplanted cellular secretions. From these replicated secretions, we create a mixture that can be used to repair injured cardiac cells.

**Wednesday, November 20**

10:30 am - 12:00 pm

**Exhibitor Forum: Networking**

**400GE Technology and Research Network**
Tae Hwang (Cisco Systems)
400GE networking is here. Networking devices and optics are readily available, and 400GE is gaining momentum in cloud scale data center. HigherEd/Research institutions are interested in building in non-hindered research network within many campus buildings. The session/forum will discuss the basics of 400GE including market trend and technology (i.e. how many fiber pairs I would need). The session will discuss 400GE campus network design both using Layer 1 optical device such as DWDM and without. The session will discuss ScienceDMZ architecture in 400GE with large buffer placement and FW bypass. The session will also cover 400GE as HPC cluster as it can provide exceptionally low oversubscription ratio among leaf and spine network topology. Lastly, the session will discuss Netflow type of advantage in Ethernet/IP on HPC fabric.

**Super-Connecting the World's Number One Supercomputers**
Scot A. Schultz (Mellanox Technologies)

Mellanox continues our leadership as the best interconnect worldwide for HPC. InfiniBand is leading around 55% of the HPC systems listed on the June 2019 TOP500 supercomputers list, including the top three and accelerates the four of the top five most powerful systems on the list, including the TACC Frontera system, the highest ranked HDR 200Gb/s InfiniBand connected supercomputer. Mellanox has demonstrated exceptional performance and application scalability with adaptive routing, MPI Allreduce latency and higher performance for AI by integrating SHARP capabilities into machine learning frameworks. We will also cover the advancements with InfiniBand HDR such as SHIELD, an autonomous self-healing technology that can instantly re-route data paths from intelligent switches.

By providing the best-in-class network computing, native hardware-based offloads and acceleration engines we will explore the next generation capabilities of these smart-interconnect devices as we prepare to approach the next milestone to exascale computing and extreme distributed machine learning.

**Real-Time Traffic Monitoring at SC19**
Neil H. McKee (InMon Corporation)

How is the SCinet network monitored? The technologies and visualizations behind real-time traffic monitoring in a multivendor high-speed network are described, with live demo examples from SC19. The implications of real time (sub-second) monitoring for dynamic optimization of HPC networks are discussed.
10:30 am - 12:00 pm

Exhibitor Forum: Managing and Storing Data II

Multi-Tiered Storage Solutions – ClusterStor Data Services
Torben Kling Petersen (Cray Inc)

Storage solutions for HPC continue to grow faster than the computational facilities as the amount of
data (both used for analysis and generated by simulations) keep increasing at an exponential rate.
To alleviate bandwidth and IOPS bottlenecks in the I/O path, many turns to tiering with flash-based
solutions handling the most demanding workloads, HDD based capacity storage and tape or cloud
storage systems acting as long-term data repositories. While this approach improved
computational performance and solve the problem with bandwidth and IOPS related I/O, it creates
new problems with data placement and expensive data movement.

Cray has been working on technologies and methodologies trying to solve these problems for a
long time and by utilizing modern features in open source file systems such as Lustre, developed
the means to optimized data placement, minimize data movement while maintaining a coherent
view of the name space. We call this ClusterStor Data Services.

Flexible HPC Storage: The Architectures and Pathways to an Independent Future
James Coomer (DataDirect Networks (DDN))

The importance of scalable data architectures which can be flexibly deployed in multi-cloud
environments is continuing to grow as research institutions and commercial businesses alike seek
to optimize and balance performance, data access, data retention and cost. Emerging workflows in
AI, Deep Learning, Analytics, and HPC are continuing to push I/O infrastructures and creating
additional complexity because of the diversity of data.

DDN is focused on removing the barriers to maximizing the utilization of data and making a
dramatic leap forward in solving complexity for customers struggling to manage data from edge to
core, and over various architectures from flash to hybrid and multi-cloud. Come hear Dr. James
Coomer, senior VP of products for DDN, discuss the newest technologies and DDN's plans to
address evolving communications networks like 5G, data management for IoT, and how storage will
evolve to address Big Data, streaming, and mixed workloads efficiently at scale.
DAOS: Revolutionizing Distributed Storage with Non-Volatile Memory
Kelsey Prantis (Intel Corporation), Johann Lombardi (Intel Corporation)

This session will introduce the architecture and performance of Distributed Asynchronous Object Storage (DAOS), an open-source software-defined multi-tenant scale-out object store designed to take advantage of next generation Non-Volatile Memory (NVM) technology to address the increasing needs of hyperconverged HPC (AI, Data Analytics, and traditional HPC). Storage Class Memory (SCM), such as Intel’s Optane DC Persistent Memory (DCPMM), offers applications byte-granular storage at higher bandwidths, lower latencies, and higher IOPs than solid state block storage devices alone. DAOS stores its metadata and fine-grained application data to Optane DCPMM, and bulk data to NVMe storage (e.g. Optane or 3D-NAND SSDs). In addition, DAOS provides transactional non-blocking I/O, OS bypass, advanced data protection with self-healing, end-to-end data integrity, fine grained data control, and fast data retrieval/indexing. DAOS offers a unified storage model over domain-specific data models, such as HDF5, MPI-IO and Apache Spark, while supporting legacy applications through a POSIX emulation layer.

1:30 pm - 3:00 pm

Exhibitor Forum: Hardware and Architectures I

Innovation in HPC with ThunderX Arm®-Based Server Processors
Craig T. Prunty (Marvell Technology Group Ltd), Larry Wilelius (Marvell Technology Group Ltd)

Marvell Arm®-based Server Processors are the leading Arm® computing solutions to the High-Performance Computing and data center market. Marvell has established a leadership position in Arm®-based compute solutions in High Performance Computing and data center markets. This year at SC19, Marvell will provide exciting updates on our product strategy and, in particular, details and updates on our ThunderX family of Server processors.

Arm Architecture in HPC
Brent Gorda (ARM Ltd)

Arm growing in HPC

You have likely heard of Arm-based systems such as Sandia National Lab’s Astra (by HPE) or the Riken Computer Center’s Fugaku (a.k.a. Post-K by Fujitsu) in Japan, and a number of systems in the
UK and EU (by Atos, HPE & Cray). Arm systems are starting to arrive, and the experience has been quite positive.

In addition, the HPC community is looking for ways to continue growth in performance. With the race to exascale and a focus on efficiency, several Arm systems offer a variety of solutions that will scale both up and down.

Arm is changing up the business model for server (HPC/Cloud/Data Center) resulting in more choice of architectural features, better efficiency, and the ability for new entrants to drive innovation in silicon design.

Come and learn the latest about this exciting new technology.

An Innovative Persistent Memory Solution with Today’s Memory
Damir Jamsek (IBM Corporation), Adam McPadden (IBM Corporation)

IBM continues its innovation around new memory technologies with the recently developed Hybrid Memory Subsystem (HMS) adapter, enabling 3TB per card of persistent memory in today’s POWER9 system. Applications which require large memory footprints can utilize up to 24TB of load/store persistent memory in a 4 socket system, at a lower cost/GB than DRAM. This is IBM’s first storage class memory card leveraging the OpenCAPI 25GB/s high speed serial interface in existing IBM systems. This presentation will cover the HMS architecture, software stack enablement around the standardized SNIA programming model and illustrate the benefits of persistent memory on a high speed serial interface. Additionally, we will be demonstrating HMS in a system during the exhibit, showing real-world benefits not previously achievable.

3:30 pm - 5:00 pm

Exhibitor Forum: Hardware and Architectures II

The Impact of a High Performance 2D NoC on Next-Generation FPGAs
Quinn Jacobson (Achronix Semiconductor Corporation)

With accelerators being used to boost high-performance computing, FPGAs are increasingly popular options offering a reconfigurable datapath. Traditional FPGA architectures connect the external interfaces to the periphery beachfront of the FPGA fabric. But as FPGA compute capacity
and external interface bandwidth continue to grow exponentially, the frequency for the FPGA logic and interconnect does not scale, leading to a mismatch in connecting the FPGA fabric to the outside world. An innovative Network-on-Chip (NoC) architecture, with a bi-sectional bandwidth of 20Tbps, addresses this issue. It allows external interface controllers to connect to each other, and gracefully moves data in, out, and around the FPGA fabric. The NoC overlays a 2D network of high-speed packet-based communication infrastructure providing 160 access points connecting into the FPGA fabric. Learn how 2D NoC improves performance and eases design by reducing routing congestion, eliminating floor-planning challenges, and decoupling clock domains.

**CCIX: Simplifying Development for Seamless Acceleration For Heterogeneous Compute Systems**

*Gaurav Singh (CCIX Consortium, Xilinx Inc)*

The CCIX Consortium continues to move forward as the industry’s leading cache coherent chip-to-chip interconnect standard. CCIX Consortium members including Arm, Huawei, Xilinx, Synopsys and Cadence have delivered a growing list of CCIX-enabled platforms to solve the high-bandwidth, low latency and seamless acceleration needs of heterogeneous compute systems.

But even as the adoption of the CCIX Standard and list of full-production systems continues to grow, the work is just starting. With the CCIX production specification, ecosystem, and hardware now in place, the focus is shifting to software development and workload analysis to optimize and simplify how heterogeneous systems are architected leading to the deployment of production systems.

In this session, CCIX Consortium members will explain how CCIX simplifies development and ultimately allows system designers to seamlessly integrate the right combination of heterogeneous components to address their specific system needs.

**The latest information of NEC Vector Supercomputer, SX-Aurora TSUBASA**

*Masashi Ikuta (NEC Corporation)*

The NEC SX-Aurora TSUBASA is the newest in the line of NEC SX Vector Processors with the world’s highest memory bandwidth. The Processor that is implemented in a PCI-e formfactor can be configured in many flexible configurations together with a standard x86 cluster. No special framework is needed to build the applications, but standard programming paradigms in C/C++ and Fortran, for which the NEC compiler will automatically carry out vectorization and parallelization.
Thursday, November 21

10:30 am - 12:00 pm

Exhibitor Forum: Futures and Exascale

Supercomputer “Fugaku” and Fujitsu Supercomputer PRIMEHPC FX Series
Toshiyuki Shimizu (Fujitsu Ltd)

Fujitsu has provided supercomputers for over forty years; high-end supercomputers and x86 clusters supporting wide application areas and customer requirements.

RIKEN and Fujitsu are now developing the supercomputer “Fugaku,” formerly known as Post-K, as Japan’s new national supercomputer project. Fugaku targets up to 100 times higher application performance than that of K computer, with superior power efficiency. Fugaku employs the newly developed FUJITSU A64FX CPU featuring Armv8-A instruction set architecture and the Scalable Vector Extension (SVE) to widen application opportunities. Fugaku contributes to the Arm ecosystem for HPC applications as well as science and society.

At SC19, Fujitsu will provide updates on Fugaku and unveil Fujitsu’s commercial products based on the technologies developed for Fugaku.

Machine Learning on Near-Term Quantum Computers
Yudong Cao (Zapata Computing Inc)

Quantum computers are known to solve certain problems that are otherwise intractable for the best classical algorithms. However, quantum devices in the near-term are still yet to develop capabilities that can yield practical advantages. A more fruitful direction for near-term quantum devices is then to use them as "sampling devices" which allow for efficient sampling from probability distributions that are arguably hard to simulate using classical computers.

One such hard sampling task exists in the training of Boltzmann Machine, which is a versatile model of probabilistic neural network that finds applications in a variety of areas in machine learning. In particular, accurately computing the gradient during the training of a restricted Boltzmann Machine...
requires extensive sampling from the model distribution, which is costly. We will present two different lines of work that have been pursued by our team, one using quantum annealers and the other using gate-model quantum devices.

**Three Key Recommendations to Master the Transition to Exascale and Beyond**

Jean-Pierre Panzieria (Atos)

HPC is on the verge of a new era: the exascale. Beyond the mere thrill of reaching new heights in terms of computing power and the potential discoveries it entails, this new frontier also brings unprecedented challenges that organizations need to address. While preparing for this new deal, three important areas are worth focusing on. First, invest in accelerated computing: GPU, FPGA, Quantum processors – which are not that far down the road! – but also AI-accelerated HPC and Edge computing. Secondly, optimize data flows through a diverse storage strategy but also through a better knowledge of application IOs. Finally, keep in mind that these transitions will only be successful if users are in a position to nurture their skills set and acquire new ones, be it quantum programming or IO monitoring, for HPC remains a collective adventure.

10:30 am - 12:00 pm

**Exhibitor Forum: Software**

**SUSE HPC - Enabling Discoveries**

Jay Kruemcke (SUSE), Jeff Reser (SUSE)

SUSE offers a highly scalable, high performance open source computing system that enables users to harness the power of the supercomputer and manage parallel computing environments. SUSE accelerates innovation with new solutions developed and delivered with a strong ecosystem of partners. We will discuss the latest exciting initiatives within this ecosystem which improve scale and performance, realize faster time to value, and answer the needs of future HPC systems.

**Dynamic Analysis and Debugging Techniques for HPC Applications**

Bill Burns (Perforce Software Inc)

Debugging and dynamically analyzing HPC applications requires a tool designed to meet the demands of today’s highly complex HPC applications. It must handle applications making extensive
use of C++ templates and the STL, many shared libraries, optimized code, code leveraging GPU
collectors, and applications constructed with multiple languages.

This interactive session highlights the different dynamic analysis and debugging technologies
provided by the TotalView HPC debugger. We’ll show how easy it is to incorporate these
techniques into your daily routine, so you can easily understand complex code and quickly solve
difficult problems. You will learn how to:

* Leverage reverse debugging technology to replay how your program ran.
* See a unified view across applications that utilize Python and C++.
* Debug CUDA applications.
* Find memory leaks in your HPC codes.

Implementing these techniques along with TotalView, will help you find bugs and improve the
quality of your code.

The Next Frontier and New Summit for Precision Medicine

Frank Lee (IBM Corporation)

From the world's fastest genomics pipeline to the largest variant database, we will share our latest
advancements focusing on key capabilities that makes genomics platform faster, easier to use, more
cost-efficient and collaborative: 1) "autonomous pipeline": end-to-end automation of genomics
workflow; 2) "radar on data": metadata tagging and provenance tracking for ocean of data from
biomedical research; 3) hybrid multicloud deployment: orchestration of workloads leveraging
containers and cloud 4) self-service App Hub: graphical user interface for applications and
platforms.

We will also showcase respective prototypes for full GATK pipeline, automated PHI detector and a
cloud-enabled RNAseq pipeline. Join us and learn about the underlying reference architecture for
high performance data and AI which was used to construct Summit - IBM's world-class
supercomputer - and how it can also be implemented on a much smaller system to power any lab,
department or institution to support the next frontier of precision medicine.
Exhibits

(back to top)

Monday, November 18

6:45 pm - 7:00 pm
Exhibit Floor Ribbon Cutting

7:00 pm - 9:00 pm
Gala Opening Reception

Session Description: SC19 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 19

10:00 am - 6:00 pm
Exhibits

Session Description: Come visit with over 300 exhibitors that range from industry to academia and research. There is sure to be something for everyone.

Wednesday, November 20

10:00 am - 6:00 pm
Exhibits

Session Description: Come visit with over 300 exhibitors that range from industry to academia and research. There is sure to be something for everyone.

4:00 pm - 6:00 pm

Family Day

Session Description: On Wednesday, November 20, 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 21

10:00 am - 3:00 pm

Exhibits

Session Description: Come visit with over 300 exhibitors that range from industry to academia and research. There is sure to be something for everyone.
HPC Impact Showcase

Tuesday, November 19
1:30 pm - 3:00 pm

HPC Impact Showcase 1

High Performance Computing for Numerical Weather Prediction at The Weather Company

Todd Hutchinson (IBM Corporation), John Wong (IBM Corporation), William Sheridan (IBM Corporation)

The Weather Company (TWC), an IBM Business, is using HPC to run the world’s first weather prediction system that provides global, rapidly-updating weather forecasts at a very high resolution. The system, referred to as IBM GRAF (Global High-Resolution Atmospheric Forecast System) is driven by MPAS, a global numerical weather prediction system developed at the National Center for Atmospheric Research.

This presentation will describe how a cluster of IBM AC922 servers enables these weather forecasts to be run globally, yet still predict events as local as thunderstorms, with forecasts being updated every hour. In order to exploit the capabilities of the high-performance cluster, MPAS was ported to run very efficiently on hundreds of interconnected CPUs and GPUs. Further, the presentation will show how output derived from the weather prediction system is used to aid decision-making in business areas such as commercial aviation, energy, and agriculture.

LLNL and ExxonMobil's High-Performance Collaboration in Reservoir Simulation

Ruslan Rin (ExxonMobil Corporation)

LLNL and ExxonMobil partnered together to combine their scientific talents to harness the power of HPC for generating insight into cost-effective and sustainable ways to develop oil and gas reservoirs. This partnership lead to significant accomplishments in modeling complex subsurface
flow processes by combining ExxonMobil’s reservoir simulation technology with LLNL’s HPC expertise and resources. As the industry moves to ever more complicated oil and gas fields, the ability to optimize facilities and operations is increasingly critical. Mega-models with fast turn-around for dynamic simulation enable engineers and geoscientists to rapidly evaluate multiple scenarios. ExxonMobil and LLNL scientists worked together to perform a scalability study of three-phase black-oil multi-billion cell models on LLNL’s Vulcan and Sequoia supercomputers. The results of these efforts demonstrated exceptional scaling on the range of 1k to 1.5M cores. This is the largest number of processor counts reported by the oil and gas industry today.

Virtual Designs, Real Results

Raymond Leto (TotalSim LLC)

Providing engineers the ability to easily review, analyze and visualize the copious amounts of data generated from advanced modeling and simulation is crucial to inform design and development. Due to the growing complexity of models, relying on internal HPC alone was not going to meet our requirements for providing CFD consultation to clients.

We will highlight how increasing our HPC capacity through a partnership with the Ohio Supercomputer Center (OSC), has allowed us to scale our consulting capabilities. We’ll highlight working with racing clients’ like NASCAR R&D and Honda Racing’s complex geometry, and show how we’ve been able to run massive simulations and develop accurate models for clients rather than having to scaling down and adversely affect results and recommendations.

Using HPC at this scale does creates other issues like ‘Drowning in Data’

Our partnership with OSC has also enabled us to create a solver agnostic, Web App called ‘TS Results’ that allows us (and other users) to collect data and results from most CFD solvers into a standardized format. This accelerates design by publishing the results, desired images, movies and plots in one place for quick and easy comparison, giving the user more time to analyze results and refine designs by spending less time on collating and plotting data.

We will discuss how adopting expanded HPC has allowed us to provide clients with solutions that enhance stakeholder collaboration amongst multiple agencies and departments, provide more timely analysis, all while increasing overall throughput and deriving more from existing data.

3:30 pm - 5:00 pm
Next Generation Network for Advanced Simulation and Visualization of Large-Scale Modeling

Alan Benjamin (GigaIO), Dawei Mu (University of Illinois)

Nowadays, numerical simulation plays a vital role in analyzing and assessing various physical phenomena. Storage I/O performance and network bandwidth have not kept pace with the growth of computing power; as a result, post-processing has become a bottleneck to end-to-end simulation performance. GigaIO has developed a next generation network technology, FabreX, in order to address this challenge. The presentation will describe how GigaIO partnered with SDSC to deploy and benchmark its solution using earthquake simulation as a representative problem. The presentation will discuss three approaches used to address this performance bottleneck: 1) Reducing the amount of output data by implementing in-situ visualization, which constructs the visualization concurrent with the running simulation; 2) Running the tests on a system utilizing the GigaIO™ FabreX™ next generation network solution; and (3) Using HPC resources provided by SDSC’s Comet supercomputer to benchmark alternative approaches.

The presentation will describe how this project was impactful for both science (earthquake simulation platforms) and commerce (a new interconnect technology). The partnership with SDSC permitted benchmarking the GigaIO technology in a laboratory environment against state-of-the-art production supercomputer resources, providing valuable feedback for GigaIO’s product development efforts and validating a potential use case for the technology in modeling and simulation.

HPC and AI - Accelerating Design of Clean High-Efficiency Engines

Yuanjiang Pei (Aramco Services Company), Sibendu Som (Argonne National Laboratory)

Supercomputers have long been utilized in scientific fields in “capability” fashion, i.e., for performing scale-resolved simulations in a short time. However, very few real-world applications have been demonstrated using a supercomputer in “capacity” computing fashion, i.e., by running many high-fidelity simulations simultaneously. The presentation will demonstrate how an IBM Blue Gene/Q supercomputer at Argonne National Laboratory was utilized in “capacity” fashion to optimize a heavy-duty engine running gasoline for improved efficiency. Combustion engines are extremely challenging to accurately simulate due to the disparate length and time scales, combined with a multitude of physical sub-processes (injection, evaporation, etc.), intertwined with complicated fuel
and emission chemistry. A high-fidelity simulation approach was first developed and incorporated into the CONVERGE CFD code, and validated against experiments. A first-of-its-kind study was performed to guide engine design by exploring a large design space and operating range. Thousands of high-fidelity engine design combinations that would ordinarily require months on a typical cluster were simulated within days on Mira. The accelerated simulation time allowed the evaluation of an unprecedented number of variations within a short time span. Machine Learning was used to develop a surrogate model for the simulations using a “superlearner” approach. This further offered the potential for reducing design time for subsequent iterations. Finally, significant fuel efficiency improvement and lower emissions were demonstrated using the real-world hardware recommended by the simulation. The ability to perform many high-fidelity calculations of a complex device on a supercomputer reduces time-to-science and opens a new frontier in the automotive industry.

Wednesday, November 20
1:30 pm - 2:30 pm
HPC Impact Showcase 3

Using a Supercomputer to Hunt Malware

Chahm An (Webroot Inc), Maurice Schmidtler (Webroot Inc)

The Problem

The internet is a vast, continuously growing resource where, given its dynamic nature, an arbitrarily large amount of new content is introduced every day. Over the past decade, cybercriminals have learned how to leverage the Internet for malicious intent. Modern cybersecurity solutions need to quickly and accurately determine the security risk of visiting a website by classifying the complex information found within long sequences of data.

Effective machine learning solutions handle this problem by learning patterns from arbitrarily long input sequences in a way that captures signals correlated to malicious activity. These sequence modeling approaches are computationally expensive, requiring either large amounts of memory to individually keep track of a huge number of relevant subsequence fragments, or expansive computing power to iteratively evaluate and update networks that model the sequence as a whole.
The Solution

We’ve successfully used the resources at the San Diego Supercomputer Center (SDSC) to address these difficult computational challenges, making significant advances for cybersecurity solutions. We will walk conference attendees through our process and share real-life examples. For instance, the unique resources at SDSC made possible the training of our Real-Time Anti-Phishing model, which requires rapid random access across nearly a terabyte of data. The high-memory computing resources allow us to accelerate the random access bottleneck to speed up the training process several orders of magnitude, from days to minutes.

Combating Traffic Congestion Using Massive HPC Analytics in the Google Cloud Platform

Boyd Wilson (Omnibond Systems LLC), Amy Cannon (Omnibond Systems LLC)

Traffic and congestion is one of the pressing problems facing modern society. Traffic congestion cost in the US is an estimated $166 billion and more than 3.3 billion gallons of fuel is wasted in stalled traffic.

A study performed by TrafficVision and the Clemson DICE Lab, shows how over a million hours of highway video can be processed in the Google Cloud Platform requiring over a million concurrent vCPUs of processing. This talk will discuss how the large scale traffic video analysis using Cloud HPC in the Google Cloud Platform can help agencies better understand congestion and incident hot spots within the camera coverage areas. This HPC process is also used for neural net auto-annotation allowing for AI algorithm improvements helping to provide better insights in the future.

We will also discuss the impact of real-time parallel video analytics on the transportation industry. How real time incident detection through video analytics is leveraged to reduce incident response time for emergency responders saving lives and reducing the duration of congestion periods. This capability also provides for roadway situational awareness across a broad transportation network.

3:30 pm - 4:30 pm

HPC Impact Showcase 4

HPC and the FRB of Dallas: The Impact of BigTex and UT Dallas on Research in The Federal Reserve System
Christopher Simmons (University of Texas, Dallas; OpenHPC), Alexander Richter (Federal Reserve Bank of Dallas)

The University of Texas at Dallas (UTD) and the Federal Reserve Bank of Dallas have partnered to launch BigTex, an OpenHPC-based system for Federal Reserve economists and their co-authors to use for research purposes. BigTex is not only the largest environment in the Federal Reserve system but also the first environment available to every economist in the system. In this talk, we will discuss how HPC is now being used by the Federal Reserve to conduct research as well as a few of the unique challenges the sysadmin team from UTD encountered in deploying the system and supporting the economists. In doing so, we will provide the community an inside look into the process of deploying a new system and porting new software to an HPC system from a non-traditional HPC domain. The lessons from our initiative will provide insight for other institutions looking to move non-traditional HPC workloads / users to HPC systems.

HPC Needs at JPMorgan Chase
Bradford Spiers (JPMorgan Chase)
JPMorgan Chase runs a large, global business that serves tens of millions of households and 80% of the Fortune 500. Historically, Wall Street has turned to HPC solutions for low-latency trading. In this talk, we describe another domain in which JP Morgan has begun to turn to HPC technology---graph pattern match. Due to the nature of graph challenges, typical cluster- based solutions will not scale to meet our needs. We describe our work to look for known-bad cyber patterns as an application of graph pattern matching on a shared-memory supercomputer.

Thursday, November 21
1:30 pm - 3:00 pm

HPC Impact Showcase 5

Beating Bugs with Big Data: Harnessing HPC to Realize the Potential of Genomics in the NHS in Wales

Thomas Richard Connor (Public Health Wales, Cardiff University)

The ability to read the genome of an infectious agent, essentially the blueprint that describes how that organism is 'built' including how it causes disease, is transformative for healthcare. Using
genomics data, one can predict the correct drug to use to treat an individual, identify vaccine targets, or track the spread of an epidemic across the globe.

In 2017 the Welsh Government launched the Genomics for Precision Medicine Strategy, including over £8M of investment to date to develop new clinical genomics services to enable better, more personalised and more precise medical care.

Because genome sequence data is digital, rich and complex, substantial computational resources are required to extract clinically actionable information. Furthermore, within a healthcare setting, diagnostic tests have to be standardised and repeatable, as well as generating results in a short space of time to ensure maximum patient benefit. Bringing together the NHS in Wales, Cardiff University and partners in Dell, a combination of HPC, on-premise Cloud and containerisation have been deployed to power national clinical genomics services for the 3 million people covered by the Welsh NHS.

Here we describe how our computational environment is delivering the first national genomics-based service for HIV drug resistance testing in the UK, supporting outbreak response, and undertaking seasonal influenza surveillance that in the 2018/19 flu season saw Wales contributing surveillance data to global surveillance efforts faster than anyone else in the world.

Genomics is proving to be revolutionary for healthcare, and in Wales that revolution is founded upon our HPC.

Open New Horizons for Oil and Gas with Total's PANGEA III System

Long Qu (Total, France), Elies Bergounioux (Total, France), Diego Klahr (Total, France), Pascal Vezolle (IBM France)

Oil and Gas industry is continuously challenged to increase its hydrocarbon production in response to the growing demand for energy. Finding new oil and gas fields has become more challenging as resources are no longer easily accessible with an increase of geological complexity and environmental constraints. High-resolution and reliable subsurface imaging is getting more and more key for any oil and gas company to reduce the risk of seismic exploration, as well as estimating extractable hydrocarbon volumes. Since seismic imaging of complex geological structures requires extremely compute intensive algorithms and huge amount of data storage, HPC has a critical role in the process of seismic exploration, even more now with the emergence of promising AI algorithms. Therefore, Oil and Gas companies have often been pioneers in giant HPC cluster installation and will continue to dominate the industrial supercomputer market. For the new
generation of systems, disruptive accelerated technology brings the mandatory computing power to be able to run higher resolution algorithms and to process richer data coming from the seismic acquisition, while reducing simulation times and energy costs. To tackle its technical and business challenges, including the integration of AI solutions, Total installed one of the most powerful supercomputers based on IBM Power9 and NVIDIA GPU technologies, called PANGEA III. In this presentation, we outline HPC key role in exploration and explain the choice of this new technology and how the solution matters in term of simulation and businesses improvement, as well as HPC application engineering and innovations.

Supercharging Digital Pathology AI with Unified Memory and HPC

Joe Yeh (AetherAI, Taiwan)

Application of AI in digital pathology has attracted increasing attention in recent years. As the field is in its nascent days, there are numerous challenges in methodology to overcome. One of the main challenges is the extreme spatial dimension of digital whole slide images (WSIs), often larger than billion pixels. Although compute accelerator such as GPU significantly speeds up deep neural network training, its rather limited memory cannot accommodate large image inputs such as WSIs. The most prevalent solution to this problem is patch-based approach that divides WSIs into small image inputs. However, this method has significant drawbacks, most notably the laborious process to perform detailed annotation to provide ground truth for individual patches. Our solution to this problem is to utilize CUDA Unified Memory to allow DNN training on entire WSI inputs. Leveraging the HPC power of TAIWANIA 2, along with our proposed memory optimization techniques, we’re able to train a DNN on 1000 WSIs to reach maximum performance (0.99 AUROC) in 16 hours, with 550x speedup on a total of 128 GPUs. Our experience on the Nasopharyngenl Carcinoma detection case, in collaboration with Chang Gung Memorial Hospital, shows this brand-new training pipeline could significantly shorten the production cycle. By reducing 6 months of annotation efforts, the method only takes 2 weeks to train a DNN model without accuracy loss. Our optimization, combined with HPC, resulted in an easily scalable solution that will greatly facilitate the development of digital pathology AI.

3:30 pm - 5:00 pm

HPC Impact Showcase 6

Modeling Rotary Atomization for Material and Process Improvements
Aaron Zatorsky (PPG Industries Inc)

The U.S. automotive industry spray-applied over 60 million gallons of paint in 2014. Much of this paint is applied by an electrostatic rotary bell atomizer. Commonly used bells are capable of applying paint at 20% higher throughput compared to traditional gun atomizers. In these systems, desirable atomization, which controls paint appearance and the frequency of defects such as solvent pop and paint sag, suffers when fluid delivery is increased. To improve on this performance, new fundamental understanding of the relationships between fluid material properties, the atomization process, and the quality and performance of the resulting sprays is needed.

In a U.S Department of Energy HPC4Mfg collaboration between PPG and LBNL, we are developing computational models to explore paint behavior during rotary bell atomization as a function of paint fluid properties. In this collaborative effort, coating and experimental know-how within PPG are being combined with numerical modelling expertise and supercomputing capabilities at LBNL to advance understanding in this high value space. With this information, new coatings that atomize well at higher flow rates can be developed to increase productivity and ultimately reduce booth size. These improvements can deliver significant energy savings and enhance manufacturing competitiveness.

The Paradigm Shifts in Offshore Engineering that HPC Is Bringing In

Jang Kim (Technip USA Inc, Genesis Oil and Gas Consultants Ltd)

With the advancements in HPC, Computational Fluid Dynamics (CFD) is rapidly widening its territory in offshore engineering applications. There are speculations that the CFD simulation will eventually take over most of the role of the routine physical model tests in the offshore platform design.

HPC also made it possible to simulate the global and structural responses of the offshore platforms during their whole service life, typically longer than 20 years, including extreme events such as hurricane and loop current.

TechnipFMC Genesis has been working closely with operators to develop engineering solutions based on the HPC-based simulation that can be used during offshore projects. The new HPC-based engineering solutions provides more safe and economic design of offshore structures than the conventional design approaches relying on empirical engineering analysis and physical model tests. The technology status and expected timelines of transition from conventional to HPC-based design approach will be reviewed. Also presented will be the areas where HPC-based CFD
simulations can provide solutions beyond the limit of the physical model tests.

**Move It but Don’t Lose It: Re-Creating the Barkla HPC Cluster in the Cloud during a Complete HPC Systems Relocation.**

Cristin Merritt (Alces Flight Limited)

What do you do when you have to move a cluster but cannot afford to lose the resource time? You could rely on a Disaster Recover (DR) plan, but having a complete mirror can be cost-prohibitive in both physical hardware and systems management - especially if your back-up site isn’t really geared-up for HPC. So why not try public cloud? That’s what the team at the University of Liverpool decided to do. They were searching for a means to keep resources available during the move of the Barkla HPC cluster, a 5,000-core Dell EMC cluster containing Xeon Phi and Nvidia Tesla V100 GPUs, into its updated facilities at the Department for Advanced Computing, an on-site data center located on the university campus. This move, and the prospective downtime period of up to two weeks that would come with it, sparked the need to completely overhaul how the team approached HPC resource management when a ‘planned disaster’ was on the way - and their secondary on-site facility not built to handle the HPC workload at the required level for that duration. The solution? Lean on the ephemeral strength of not one, but two public cloud platforms (AWS and Microsoft Azure) to architect a mission critical variant of Barkla that would not only cover off the cluster move, but also ensure users had (nearly) seamless access to cloud resources, lay the foundation for cloud bursting projects, and create a stronger, long-term DR process for their HPC facilities.
HPC is Now Plenary

Monday, November 18

5:30 pm - 6:30 pm

HPC Is Now Plenary
Imaging the Unseen: Taking the First Picture of a Black Hole
Katie Bouman (California Institute of Technology)

This talk presents the methods and procedures used to produce the first image of a black hole from the Event Horizon Telescope. It has been theorized for decades that a black hole will leave a "shadow" on a background of hot gas. Taking a picture of this black hole shadow could help to address a number of important scientific questions, both on the nature of black holes and the validity of general relativity. Unfortunately, due to its small size, traditional imaging approaches require an Earth-sized radio telescope. In this talk, I discuss techniques we have developed to photograph a black hole using the Event Horizon Telescope (EHT), a network of telescopes scattered across the globe. Imaging a black hole’s structure with this computational telescope requires us to reconstruct images from sparse measurements, heavily corrupted by atmospheric error. The resulting image is the distilled product of an observation campaign that collected approximately five petabytes of data over four evenings in 2017. The process for recording, correlating, calibrating, and imaging the data has been built-out by the EHT over the last decade. We will summarize how the data from the 2017 observations were calibrated and imaged, and we will explain some of the challenges that arise with a heterogeneous telescope array like the EHT.

An Embarrassment of Riches: Thanks to HPC, We Now Have Better Topography for the Ice on Earth than the land
Paul Morin (University of Minnesota)

For years, those of us that made maps of the poles apologized. We apologized for the blank spaces on the maps, we apologized for mountains being in the wrong place and out-of-date information. Over the past 10 years the situation improved. An image mosaic of Antarctica was built, and a
constellation of satellites started to stream data at ever higher resolution, at an increasing tempo and even during the long polar winters.

Now a diverse collaboration of US science and intelligence agencies with universities has produced REMA, the Reference Elevation Model of Antarctica, and ArcticDEM, using open source software to extract Digital Elevation Models (DEM) on Blue Waters at a resolution of 2m. The data have an accuracy of a foot and repeat coverage of 90% of the poles an average of 10 times over 6 years. This project was too large for any one agency, university, or company. It required an enormous allocation on Blue Waters, 4 satellites that continuously collected sub-meter optical imagery for 5 years, two satellites that produced ground truth, 100Gbit networking, and petabytes of storage.

We never thought that we would see a time when the science community has better topography for ice than land and better topography for the Transantarctic Mountains than the Rocky Mountains. Even we are having a difficult time understanding what we have made.

We now apologize to the polar science community for a different reason. They have to keep up. And the current DEMs are only the beginning. We now face an avalanche of imagery and derived products in an increasingly complex landscape of small-sats launched by the dozen. It is a complex, exciting time.

1:30 pm - 3:00 pm

Invited Talks - Tuesday Afternoon

US Administration Activities in Artificial Intelligence and HPC
Lynne Parker (White House Office of Science and Technology Policy)

Artificial intelligence (AI) is transforming everything from healthcare to transportation to manufacturing. Recognizing the importance of AI to the United States, in February 2019, the US President announced the American AI Initiative. This Initiative is a whole-of-government strategy to advance AI in collaboration and engagement with the private sector, academia, the public, and international allies. One of the key priorities of this Initiative is AI research and development (R&D), to include not only fundamental AI algorithms, but also the underlying cyberinfrastructure, systems, and data that provide the foundation for complete AI systems. Investments in AI and high performance computing are among the most important areas of emerging technology at work for our nation, both inside and outside government. In this talk, I will discuss the Administration’s activities and priorities in AI, and in high performance computing (HPC), highlighting open R&D
questions at the nexus of AI and HPC.

**Scientific Machine Learning**  
Nathan Baker (Pacific Northwest National Laboratory (PNNL))

Machine learning (ML) is rapidly changing the field of scientific computing. However, there exist key research gaps that limit the impact of current ML methods on scientific problems. These gaps were identified and discussed at a recent Department of Energy Basic Research Needs Workshop on Scientific Machine Learning. This workshop identified six priority research directions to increase the impact of ML on scientific problems. In this talk, I will review these six areas and present recent results from Pacific Northwest National Laboratory in domain-aware machine learning, one of the six priority research directions.

**Wednesday, November 20**

10:30 am - 12:00 pm

**Invited Talks - Wednesday Morning**

**Stellar Collisions at the Limits of HPC**  
Orsola De Marco (Macquarie University, Australia)

Stellar explosions, outbursts, jets, nebulae, and the emission of gravitational waves are all phenomena that happen on very short timescales of seconds to days. Traditionally astronomers have been taking pictures that, beautiful as they were, only observed tiny patches of a vast sky at specific moments in time. This is no longer the case, as astronomers have learned how to make whole sky movies. With these we have discovered a huge range of transient phenomena, and we have realized that we cannot explain them. 3D hydrodynamic simulations are one of the techniques used to determine the underlying mechanisms of this zoo of stellar interactions from supernovae to merging black holes. I will talk about what 3DHD can give us, the techniques that we use (adaptive mesh, smooth particles hydrodynamics or moving mesh), the limitations they all have (lack of scalability), and possible solutions.

**The Role of Cyberinfrastructure in Science: Challenges and Opportunities**  
Ewa Deelman (University of Southern California)
Cyberinfrastructure (CI): instruments, computational platforms, software, and people form a sophisticated platform that powers modern scientific discoveries. As demands for greater collaboration, more data analysis, more complex simulations, and computational power have grown in science, so have the challenges of developing and delivering robust CI to users. Sustainable and dependable CI is not developed in a vacuum; rather, it benefits from advances in Computer Science and provides a unique laboratory for Computer Science research. Grounded in the challenging and ever-increasing needs of a multitude of scientific applications, CI is continuously enhanced and driven to innovate.

This presentation examines selected modern scientific discoveries, such as the detection of gravitational waves from the CI perspective. It aims to answer the following key questions: first, what were the key CI solutions that enabled a particular scientific result? and second, what were the challenges that needed to be overcome? The CI examples will include science automation technologies, such as the Pegasus workflow management system and how it evolved over the last two decades as science applications and broader CI have grown in heterogeneity and scale. It will describe the algorithms and solutions that enable the robust execution of complex scientific workflows on DOE and NSF-funded high-performance and high-throughput computing systems.

This talk will conclude with emerging challenges and opportunities in the CI landscape.

1:30 pm - 3:00 pm
Invited Talks - Wednesday Afternoon

Computational Challenges to Reconstructing Evolution with Large Datasets
Stephen Smith (University of Michigan)

Recent advances in data generation technologies have provided the biological sciences with enormous resources for both DNA sequences and morphological measurements. These data have created entirely new fields and have facilitated our ability to address classic questions in biology that had previously been data-limited. However, despite these advances, new challenges have also emerged. For example, while biologists previously thought that simply increasing the available data would result in more accurate results, most new analyses have uncovered extensive underlying heterogeneity that result from biological processes. Computational challenges have also limited our ability to take full advantage of the new resources. Both the size of the datasets and the underlying complexity have dramatically hindered optimization and inference. I will describe several of the computational challenges facing evolutionary biology along with several solutions. I will also
describe some of the limits to our ability to reconstruct the phylogenetic past, despite our enormous efforts.

HPC Solutions for Geoscience Application on the Sunway Supercomputer

Lin Gan (Tsinghua University, China)

In recent years, many complex and challenging numerical problems, in areas such as climate modeling and earthquake simulation, have been efficiently resolved on Sunway TaihuLight, and have been successfully scaled to over 10 million cores with inspiring good performance. To carefully deal with different issues such as computing efficiency, data locality, and data movement, novel optimizing techniques from different levels are proposed, including some specific ones that fit well with the unique architectural futures of the system and significantly improve the performance. While the architectures for current- and next-generation supercomputers have already diverted, it is important to see the pro and cons of whatever we already have, and to make up the bottlenecks as well as maximize the advantages. This talk contains the summary of the most essential HPC solutions that greatly contribute to the performance boost in our efforts on porting geoscience applications, the discussions and rethinking, and the potential improvements we could undertake.

Thursday, November 21

8:30 am - 10:00 am

Invited Talks - Thursday Morning Plenary

OpenSpace – Visualizing the Universe

Anders Ynnerman (Linköping University, Sweden), Alexander Bock (Linköping University, Sweden)

This talk will present and demonstrate the NASA funded open source initiative, OpenSpace, which is a tool for space and astronomy research and communication, as well as a platform for technical visualization research. OpenSpace is a scalable software platform that paves the path for the next generation of public outreach in immersive environments such as dome theaters and planetariums. It opens up the possibilities to dynamically load and visualize large scale data from HPC simulations, distributed data repositories, and data feeds from space craft instrumentation. The technical focus will be on interactive data visualization on clustered high-end GPUs and in the presentation a number of features will be described and demonstrated such as: globe browsing which enables contextualization of extreme resolution imagery of planetary surfaces, interactive visualization of
large star fields e.g., 1 billion stars from the Gaia DR2 data release, and volumetric rendering of MHD simulations of coronal mass ejections and operational space weather forecasting. The presentation will conclude with a journey to the end of the observable universe and the big bang. The project builds on a collaboration between Linköping University, The American Museum of Natural History, NASA Goddard Space Flight Center, New York University, University of Utah, and University of Vienna.

Predictive Data Science for Physical Systems: From Model Reduction to Scientific Machine Learning
Karen Willcox (University of Texas)

Achieving predictive data science for physical systems requires a synergistic combination of data and physics-based models, as well as a critical need to quantify uncertainties. For many frontier science and engineering challenge problems, a purely data-focused perspective will fall short---these problems are characterized by complex multi-scale multi-physics dynamics, high-dimensional uncertain parameters that cannot be observed directly, and a need to issue predictions that go beyond the specific conditions where data may be available. Learning from data through the lens of models is a way to bring structure to an otherwise intractable problem: it is a way to respect physical constraints, to embed domain knowledge, to bring interpretability to results, and to endow the resulting predictions with quantified uncertainties. This talk highlights how physics-based models and data together unlock predictive modeling approaches through two examples: first, building a Digital Twin for structural health monitoring of an unmanned aerial vehicle, and second, learning low-dimensional models to speed up computational simulations for design of next-generation rocket engines.

10:30 am - 12:00 pm

Invited Talks - Thursday Morning

Next Generation Disaster Intelligence Using the Continuum of Computing and Data Technologies
Ilkay Altintas (San Diego Supercomputer Center)

Modeling of the extent and dynamics of evolving plethora of environmental hazards, and their socio-economic and human impacts, is a data-driven discipline with a vibrant scientific community of computational modeling, remote sensing, data science, technology, and social science experts,
driven by the urgent societal need to mitigate the rising frequency and severity of hazards. However, there are still challenges and opportunities in integration of the scientific discoveries and data-driven methods for hazards with the advances in technology and computing in a way that provides and enables different modalities of sensing and computing. Previously NSF-funded WIFIRE project took the first steps to tackle this problem with a goal to create an integrated system and services for wildfire monitoring, simulation, and response. Today, WIFIRE provides an end-to-end management infrastructure from the data sensing and collection to modeling efforts using a continuum of computing methods that integrate edge, cloud, and high-performance computing. Though this cyberinfrastructure, the WIFIRE project provides data driven knowledge for a wide range of public and private sector users enabling scientific, municipal, and educational use.

This talk will review some of our recent work on building this dynamic data driven cyberinfrastructure and impactful application solution architectures that showcase integration of a variety of existing technologies and collaborative expertise. The lessons learned on use of edge and cloud computing on top of high-speed networks, open data integrity, reproducibility through containerization, and the role of automated workflows and provenance will also be summarized.

“Simulate First” and the Role of HPC – A Caterpillar Perspective
Larry Seitzman (Caterpillar Inc)

“Simulate first” is an imperative spreading across the design and manufacturing communities of multiple commercial companies. To remain competitive in today’s markets, companies must find a way to keep the cost of developing new products as low as possible while simultaneously exploring as much of the design space as possible for qualities the customer expects and values. At Caterpillar, these qualities usually include product performance, robustness, uptime, serviceability, and total cost of ownership. “Simulate first” is a powerful approach to this challenge. We will discuss Caterpillar’s view of what is needed to become a “simulate first” organization and the role that HPC is currently playing and might play in the future. We will focus on how HPC is a central piece of a global compute environment on which product development engineers depend and what challenges that dependence creates for delivering a stable, but ever-modern, HPC environment.
Keynote

(back to top)

Tuesday, November 19

8:30 am - 10:00 am

SC19 Keynote Address

Steven Squyres (Cornell University)

Between 2004 and 2018, mankind spent more than 5,600 cumulative days exploring Mars—yet the Principal Investigator on those missions never needed a spacesuit.

That’s because Spirit and Opportunity, NASA’s two Mars rovers, were guided by the computing power that made it possible for Dr. Steven Squyres to oversee their findings from as much as 250 million miles away.

While he is best known to the public as the “face and voice” who chronicled Spirit and Opportunity’s extended missions—each of which lasted years beyond initial expectations—Dr. Squyres played his most mission-critical role behind the scenes, helping guide the project science as the two rovers examined the terrain of Mars for signs of water or life.

This year NASA celebrates a half-century since Apollo astronauts landed on the moon, using onboard computers that were famously advanced for their time, and just as famously rudimentary compared to the devices most people carry in their pockets today. Unlike the Apollo technology, the Mars rovers had no humans to operate them in person, and even radio commands could take as long as 24 minutes to travel from the Earth. The need for the rovers to carry out complex procedures and navigate unforgiving terrain meant their reliance on computers and software was central to their mission success.
Navigating SC19

(back to top)

Monday, November 18

4:30 pm - 5:15 pm

First Time at the SC Conference

First Time at the SC Conference
Bruce Loftis (University of Colorado (retired)), Diana Dugas (New Mexico State University)
The SC19 Conference extends a warm welcome to all first-time attendees.

A session on Monday afternoon 18 November will provide an overview of the conference, describe what each type of badge permits, and make suggestions about interesting activities in which to participate so you can organize your conference experience.

Attendees will have time to ask questions. Or questions can be submitted before the session to info@info.supercomputing.org

The First-Timer session will be Monday, November 18 at 4:30p to 5:15p in the Four Seasons Ballrooms 1-2-3. Everyone with an SC19 badge is welcome.
Panel

Tuesday, November 19

10:30 am - 12:00 pm

Reconfigurable Computing in HPC: Success Stories Today and Future?

Moderator: Franck Cappello (Argonne National Laboratory)
Panelist: Taisuke Boku (University of Tsukuba), Christian Plessl (Paderborn University), Kentaro Sano (RIKEN Center for Computational Science (R-CCS)), Kazutomo Yoshii (Argonne National Laboratory), Martin Herbordt (Boston University), Jeff Vetter (Oak Ridge National Laboratory)

Reconfigurable computing has been recently introduced in clouds and extreme-scale data centers and shows outstanding performance for specialized applications such as search engine indexing and deep learning. The end of Moore's law, the convergence between HPC and big data, the recent introduction of AI and streaming as execution patterns in HPC, and the availability of devices with exceptional floating-point computing capabilities make reconfigurable computing more attractive than ever for HPC. The SC18 “Reconfigurable computing: will it make it this time?” panel exposed the main challenges to apply reconfigurable computing for HPC and concluded that more progress was needed to demonstrate its relevance and usefulness. In the last 12 months, several HPC platforms with FPGAs have been deployed for production and experimentation. This year’s panel will present success stories about HPC applications that have been ported on these platforms and how they have been programmed, deployed, and optimized.

1:30 pm - 3:00 pm

Developing and Managing Research Software in Universities and National Labs

Moderator: Daniel S. Katz (University of Illinois, National Center for Supercomputing Applications (NCSA))
Panelist: Robert Haines (University of Manchester), Kenton McHenry (University of Illinois, National Center for Supercomputing Applications (NCSA)), Caleb Reinking (University of Notre Dame), Catherine Jones (Science and Technologies Facilities Council, UK), Carina Haupt (German
Modern research in the sciences, engineering, humanities, and other fields depends on software, and specifically, research software. Much of this research software is developed in universities and labs, by academics, researchers, staff engineers, and students. In this panel, we will focus on the role of staff engineers. We will examine a set of different, independently-developed models under which these staff are organized and perform their work, at NCSA, the University of Manchester, the University of Notre Dame, STFC and DLR, and comparatively analyze these models and their consequences on the staff and on the software, considering how the different models support software engineering practices and processes. These insights can be used by software engineering researchers to understand the practices of such organizations and by universities and labs who want to set up similar organizations, and to better produce and maintain research software.

3:30 pm - 5:00 pm

**Within Reach: The Last Mile to Exascale**

*Moderator: Timothy Prickett Morgan (The Next Platform)*
*Panelist: Mark Papermaster (Advanced Micro Devices (AMD) Inc), Steve Scott (Cray Inc), Ivo Bolsens (Xilinx Inc), Maria Girone (European Organization for Nuclear Research (CERN))*

Today, the supercomputing industry is in a global race to exascale computing. Each brigade is vying for the title of ‘first’ or ‘best’ in exascale. This inflection point for computing is creating a new wave of competition within the industry – and internationally. That competition is spurring a wave of innovation. The question is no longer ‘if’ but when and, more importantly shifts to what will be possible in this next great computational era. Regardless of who first breaks the tape, the new technologies and exotic engineering approaches are poised to provide immense industry benefits. The panel will discuss the unique innovations in the coming exascale systems based on an open platform approach, including new technical breakthroughs, software and programming tools, and the practical applications for HPC scenarios.

**Wednesday, November 20**

10:30 am - 12:00 pm

**Edge to Exascale : Computational Innovations in Cancer and the Future for Learning Health Systems**
Moderator: Eric Stahlberg (Frederick National Laboratory for Cancer Research, Leidos Biomedical Research)
Panelist: Gina Tourassi (Oak Ridge National Laboratory), Florence Hudson (FDHint LLC, Indiana University), Roxanne Jensen (National Cancer Institute), Augusto Ochoa (Louisiana State University)

Healthcare is undergoing a revolution. Catalyzed by the emergence of infrastructure to support growing numbers of medical IoT devices, gathering and integrating rapidly growing volumes of information, and use of scalable computing from the edge and at exascale has set the stage for tremendous innovation in how medicine is practiced across the system. The leading edge of such innovations are evident in cancer research and care, where AI, edge computing, and exascale technologies are converging to provide value insights into the future. This panel brings together experts from across the field to inform, to discuss doors opening to new approaches, to frame new challenges and explore new opportunities to envision a dynamic learning environment for improving cancer research, prevention, diagnosis, and ultimately cancer care.

1:30 pm - 3:00 pm

Open Science in a Changing HPC Landscape

Moderator: Beth Plale (National Science Foundation)
Panelist: Manish Parashar (National Science Foundation), Laura Biven (US Department of Energy), Satoshi Sekiguchi (National Institute of Advanced Industrial Science and Technology (AIST)), Frank Wuerthwein (University of California, San Diego; Open Science Grid (OSG)), Anita De Waard (Elsevier), Alex Szalay (Johns Hopkins University)

As scientific research questions have grown larger and increasingly dependent on computational resources, ensuring transparency and rigor of the scientific process has grown urgent. Today both scientific information and the products of research are fundamental to science transparency and rigor. But this attention to rigor is coming at the same time as the high-performance computing landscape changes through big idea science, AI, accelerators, containers, big data, etc. Where does a researcher turn? This multi-stakeholder panel of funders, publishers, researchers, and infrastructure providers will discuss recent encouraging developments in open science through topics such as i) infrastructure and incentives for scientific transparency, ii) emerging approaches for controlled access to restricted data, iii) research results of multi-national teams, iv) AI and data: new demands for transparency, v) clarifying the confusing landscape of supplements, data papers, and appendices; and vi) infrastructure and resources for scholars.
3:30 pm - 5:00 pm

Democratization of HPC through the Use of Web Portals: Different Strategies

Moderator: Patrice Calegari (Atos)
Panelist: Katharine Cahill (Ohio Supercomputer Center), Marlon Pierce (Indiana University), Leo Reiter (Nimbix Inc), Matt Bonyak (Altair Engineering), Kevin Kelly (Rescale), Denis Caromel (ActiveEon, French Institute for Research in Computer Science and Automation (INRIA))

HPC is now used in all domains (scientific, industrial, medical, financial, media, data analytics). More and more diverse user communities access heterogeneous and geographically distributed HPC resources (on-premises, private, and public Cloud), without necessarily being aware of them. This HPC democratization requires user-friendly web portals, science gateways, and web services to expose business related features and hide underlying technical complexity. Web interfaces are now a standard way to fulfill HPC end-users and administrators’ needs from PCs as well as from mobile devices. The goal of this panel is to discuss the diverse user experiences (developers, computational experts, course instructors, unexperienced new comers) and the features they require. Different development and distribution approaches (open source or commercial software, products or services) will also be discussed. The panelists will conclude by sharing their vision on how the evolution of HPC usage will shape the future portals, web services, and related emerging standards.

Thursday, November 21

10:30 am - 12:00 pm

Developing Exascale-Capable Applications: Recent Progress and Lessons Learned from the Exascale Computing Project

Moderator: Erik Draeger (Lawrence Livermore National Laboratory)
Panelist: Paul Kent (Oak Ridge National Laboratory), David McCallen (Lawrence Berkeley National Laboratory), Steven Hamilton (Oak Ridge National Laboratory), Katherine Yelick (Lawrence Berkeley National Laboratory), Tzanio Kolev (Lawrence Livermore National Laboratory)

As the HPC landscape continues to rapidly evolve, developing applications that can effectively utilize the largest available machines is becoming increasingly challenging. The Exascale Computing Project (ECP), launched in 2016 and funded by the US Department of Energy (DOE), is an aggressive research, development, and deployment project focused on delivery of mission-critical applications at exascale. In this panel, we will discuss the strategies ECP application teams are
using to prepare for exascale hardware, showcase recent progress on the Summit supercomputer, and describe the challenges ahead.

1:30 pm - 3:00 pm

New Partners for HPC Centers: Experimental Facilities and International Instrument Collaborations with Exascale Datasets

Moderator: Deborah Bard (National Energy Research Scientific Computing Center (NERSC), Lawrence Berkeley National Laboratory)
Panelist: Sadaf R. Alam (Swiss National Supercomputing Centre (CSCS), ETH Zurich), Lance Wilson (Monash University), Frank Würthwein (University of California, San Diego; Open Science Grid (OSG)), Dirk Pleiter (Forschungszentrum Juelich), Amedeo Perazzo (SLAC National Accelerator Laboratory)

As data sets from experimental user facilities and international instrument collaborations grow in both size and complexity there is an urgent need for new capabilities to transfer, reduce, analyze, store, search, and curate the data in order to facilitate scientific discovery. Supercomputing facilities around the world have begun to expand services and provide new capabilities for both types of communities in support of experiment workflows, from preprocessing to complete analysis. This panel discussion will offer a lively exchange between supercomputing facility staff and the application community, including users of experimental facilities. We will identify:

- What are the biggest pain points for experimental scientists in using HPC?
- What steps are being taken to address these pain points?
- How do experimental facilities partner with HPC centers?
- How do we design a sustainable model for collaboration, incorporating both technology development and business?

3:30 pm - 5:00 pm

Sustainability of HPC Research Computing: Fostering Career Paths for Facilitators, Research Software Engineers, and Gateway Creators

Moderator: Sandra Gesing (University of Notre Dame)
The sustainability of research computing has gained increased attention in academia. This is evident in multiple projects and initiatives such as the Campus Champions, ACI-REF, Research Software Engineer (RSE) communities as well as institutes such as the Science Gateways Community Institute. A major concern in achieving research computing sustainability is improving career paths for facilitators, RSEs, and/or science gateway creators - whether they are staff or faculty. Multiple initiatives and projects are trying to improve the situation, but the interaction between these initiatives is still sparse. While PIs, senior personnel, or organizers of such projects are often involved in at least one other project, there is not a systematic way yet for the organizations to collaborate. This is a good time to join forces and the panel will discuss questions such as "What are good ways to complement actions between these projects and initiatives?"

**Friday, November 22**

8:30 am - 10:00 am

**Quantum Inspired vs. Quantum Computing – What Next?**

Moderator: Matthias Troyer (Microsoft Corporation)

Panelist: Torsten Hoefler (ETH Zurich), Helmut Katzgraber (Microsoft Corporation), Andrew Fursman (1QBit Inc), Daniel Lidar (University of Southern California), Scott Pakin (Los Alamos National Laboratory), Eleanor Rieffel (NASA), Andrew King (D-Wave Systems Inc)

Several independent groups have demonstrated small-scale quantum computers containing several dozens of qubits with limited practical applications. Yet, the promise of quantum computing is huge, and it could lead to several break-throughs in science, technology, and industry. Quantum algorithms are fundamentally different and provide a different way of thinking that can also applied to classical computing systems where we can use them today. We propose to bring the community of industry and academic researchers and users together to discuss recent results and ideas for the way forward.

8:30 am - 10:00 am

**Enabling Machine Learning-Based HPC Performance Diagnostics in Production**
Environments

Moderator: Ann Gentile (Sandia National Laboratories)
Panelist: William Kramer (Pittsburgh Supercomputing Center), Richard Gerber (Lawrence Berkeley National Laboratory), Nick Brown (Edinburgh Parallel Computing Centre), Aaron Saxton (University of Illinois)

With the vast increases in system scales and complexity, a broad range of data is being collected at a scale that is impractical for direct human consumption. Recent advances in Machine Learning techniques and tools show great promise in developing system and application behavioral models that can be utilized to improve operational efficiency and application performance. The panel will discuss their differing perspectives regarding the potential outcomes of this technology, the current state of the art, and paths forward. We have assembled a diverse set of leading experts in the fields of systems management, performance analysis, and real world Machine Learning techniques applied to systems and application data.

10:30 am - 12:00 pm

The Road to Exascale and Beyond Is Paved by Software: How Algorithms, Libraries and Tools Will Make Exascale Performance Real

Moderator: Michael A. Heroux (Sandia National Laboratories, St. John’s University)
Panelist: James P. Ahrens (Los Alamos National Laboratory), Sadaf R. Alam (Swiss National Supercomputing Centre (CSCS), ETH Zurich), Ann Almgren (Lawrence Berkeley National Laboratory), Erin C. Carson (Charles University in Prague), Ada A. Sedova (Oak Ridge National Laboratory)

Designing and building future leadership computer systems is challenging. However, realizing the performance potential of these systems rests almost solely on obtaining additional concurrency from software. Algorithmic innovations and creation of new libraries and tools are essential to obtaining the performance improvements that these platforms promise.

This panel is composed of a diverse group of experts who are leading the creation of new algorithm and software capabilities that will make exascale performance real. Questions addressed by this panel include, How will exascale systems be usable and useful? Can new problem formulations overcome traditional scaling bottlenecks? Can algorithms exploit low precision hardware that is an order of magnitude faster and still obtain robust solutions? Can we raise abstraction layers so applications can leverage higher level common capabilities? What role will tools play in achieving scalability? Can data and visualization workflows adapt to exploit new system capabilities and
10:30 am - 12:00 pm

**HPC Big Data and AI: Computing under Constraints**

Moderator: Daniel Reed (University of Utah)
Panelist: Satoshi Matsuoka (RIKEN Advanced Institute for Computational Science (AICS), National Institute of Advanced Industrial Science and Technology (AIST)), Charles Catlett (Argonne National Laboratory), Rosa M. Badia (Barcelona Supercomputing Center), Greg Koenig (KPMG), Ewa Deelman (University of Southern California)

Big data and AI are today’s memes as we shift from rare and expensive to ubiquitous and inexpensive data. Massive digital data, powerful networks and inexpensive accelerators are bringing new data-driven approaches to technical computing. Extensive instrumentation, monitoring and control of high performance computing systems and their data-centers is the big data behind big data. The emergence of big data as a reality is challenging long held beliefs about technical computing approaches and illuminating old questions in new ways. How do we best meet soft real-time constraints for streaming data, subject to energy, communication, operation and cost constraints? For example, can we use AI for performance and efficiency tuning of both HPC applications and the data-center? How does HPC respond to the rapidly shifting hardware and software vendor ecosystem emphasis on AI hardware and software? This panel will discuss how we got here and where we are likely to go.
Better Data Systems via Better Data Structures

GraphM: An Efficient Storage System for High Throughput of Concurrent Graph Processing
Jin Zhao (Huazhong University of Science and Technology), Yu Zhang (Huazhong University of Science and Technology), Xiaofei Liao (Huazhong University of Science and Technology), Ligang He (University of Warwick), Bingsheng He (National University of Singapore), Hai Jin (Huazhong University of Science and Technology), Haikun Liu (Huazhong University of Science and Technology), Yicheng Chen (Huazhong University of Science and Technology)

Our studies show that the storage engines of existing graph processing systems are inefficient when running concurrent jobs due to redundant data storage and access overhead. We developed a storage system GraphM. It can be integrated into the existing graph processing systems to efficiently support concurrent iterative graph processing jobs for higher throughput by fully exploiting the similarities of the data accesses between these jobs. GraphM regularizes the traversing order of the graph partitions for concurrent graph processing jobs by streaming the partitions into the cache in a common order, and then processes the related jobs concurrently in a novel fine-grained synchronization. Then, the concurrent jobs share the same graph structure data in the cache/memory and also the data accesses to the graph, amortizing the storage consumption and the data access overhead. We plug GraphM into state-of-the-art graph processing systems and show that it improves the throughput by 1.73-13 times.

Best Paper Finalist: no
Best Student Paper Finalist: no

Semantic Query Transformations for Increased Parallelization in Distributed Knowledge Graph Query Processing
HyeongSik Kim (Robert Bosch LLC), Abhisha Bhattacharyya (North Carolina State University)
Ontologies have become an increasingly popular semantic layer for integrating multiple heterogeneous datasets. However, significant challenges remain with supporting efficient and scalable processing of queries with data linked with ontologies (ontological queries). Ontological query processing queries requires explicitly defined query patterns be expanded to capture implicit ones, based on available ontology inference axioms. However, in practice such as in the biomedical domain, the complexity of the ontological axioms results in significantly large query expansions which present day query processing infrastructure cannot support. In particular, it remains unclear how to effectively parallelize such queries.

In this paper, we propose data and query transformations that enable inter-operator parallelism of ontological queries on Hadoop platforms. Our transformation techniques exploit ontological axioms, second order data types, and operator rewritings to eliminate expensive query substructures for increased parallelizability. Comprehensive experiments conducted on benchmark datasets show up to 25x performance improvement over existing approaches.

Best Paper Finalist: no
Best Student Paper Finalist: no

MIQS: Metadata Indexing and Querying Service for Self-Describing File Formats
Wei Zhang (Texas Tech University), Suren Byna (Lawrence Berkeley National Laboratory), Houjun Tang (Lawrence Berkeley National Laboratory), Brody Williams (Texas Tech University), Yong Chen (Texas Tech University)

Scientific applications often store datasets in self-describing data file formats, such as HDF5 and netCDF. Regrettably, to efficiently search the metadata within these files remains challenging due to the sheer size of the datasets. Existing solutions extract the metadata and store it in external database management systems (DBMS) to locate desired data. However, this practice introduces significant overhead and complexity in extraction and querying. In this research, we propose a novel Metadata Indexing and Querying Service (MIQS), which removes the external DBMS and utilizes in-memory index to achieve efficient metadata searching. MIQS follows the self-contained data management paradigm and provides portable and schema-free metadata indexing and querying functionalities for self-describing file formats. We have evaluated MIQS with the state-of-the-art MongoDB-based metadata indexing solution. MIQS achieved up to 99% time reduction in index construction and up to 172k× search performance improvement with up to 75% reduction in memory footprint.
Large-Batch Training for LSTM and Beyond
Yang You (University of California, Berkeley; Google LLC), Jonathan Hseu (Google LLC), Chris Ying (Google LLC), James Demmel (University of California, Berkeley), Kurt Keutzer (University of California, Berkeley), Cho-Jui Hsieh (University of California, Los Angeles (UCLA); Google LLC)

Large-batch training approaches have enabled researchers to utilize distributed processing and greatly accelerate deep neural networks training. However, there are three problems in current large-batch research:

(1) Although RNN approaches like LSTM have been widely used in many applications, current large-batch research is principally focused on CNNs.

(2) Even for CNNs, there is no automated technique for extending the batch size beyond 8K.

(3) To keep the variance in the gradient expectation constant, theory suggests that a Sqrt Scaling scheme should be used in large-batch training.

Unfortunately, there are not many successful applications. In this paper, we propose Dynamic Adaptive-Tuning Engine (DATE) for better large-batch training. DATE achieves a 5.3x average speedup over the baselines for four LSTM-based applications on the same hardware. We finish the ImageNet training with ResNet-50 in two minutes on 1024 v3 TPUs (76.7% top-1 accuracy), which is the fastest version as of June 2019.

Channel and Filter Parallelism for Large-Scale CNN Training
Nikoli Dryden (University of Illinois, Lawrence Livermore National Laboratory), Naoya Maruyama (Lawrence Livermore National Laboratory), Tim Moon (Lawrence Livermore National Laboratory),
Accelerating large-scale CNN training is needed to keep training times reasonable as datasets grow larger and models become more complex. Existing frameworks primarily scale using data-parallelism, but this is limited by the mini-batch size, which cannot grow arbitrarily. We introduce three algorithms that partition channel or filter data to exploit parallelism beyond the sample dimension. Further, they partition the parameters of convolutional layers, replacing global allreduces with segmented allreduces---smaller, concurrent allreduces among disjoint processor sets. These algorithms enable strong scaling, reduced communication overhead, and reduced memory pressure, enabling training of very wide CNNs.

We demonstrate improved strong and weak scaling, including up to 4.1x reductions in training time for residual networks and 4x reductions in allreduce overhead. We also show that wider models provide improved accuracy on ImageNet. We study the current limitations of our algorithms and provide a direction for future optimizations of large-scale deep learning frameworks.

SparCML: High-Performance Sparse Communication for Machine Learning

Applying machine learning techniques to the quickly growing data in science and industry requires highly-scalable algorithms. Large datasets are most commonly processed when distributed across many nodes. Each node's contribution to the overall gradient is summed using a global allreduce. This allreduce is the single communication and thus scalability bottleneck for most machine learning workloads. We observe that frequently, many gradient values are (close to) zero, leading to sparse or sparsifyable communication. To exploit this insight, we analyze, design, and implement a set of communication-efficient protocols for sparse and quantized input data, in conjunction with efficient machine learning algorithms which can leverage these primitives. Our communication protocols generalize standard collective operations, by allowing processes to contribute sparse input data vectors. Our library extends MPI to support features such as non-blocking (asynchronous) operations and low-precision data representation. As such, SparCML provides the basis for future highly-scalable machine learning frameworks.
Computational Fluid Dynamics

Scalable Simulation of Realistic Volume Fraction Red Blood Cell Flows through Vascular Networks
Libin Lu (New York University, Courant Institute of Mathematical Sciences), Matthew J. Morse (New York University, Courant Institute of Mathematical Sciences), Abtin Rahimian (University of Colorado), Georg Stadler (New York University, Courant Institute of Mathematical Sciences), Denis Zorin (New York University, Courant Institute of Mathematical Sciences)

High-fidelity blood flow simulations are a key step toward better understanding biophysical phenomena at the microscale, such as vasodilation, vasoconstriction, and overall vascular resistance. To this end, we present a fast scalable platform for the simulation of red blood cell (RBC) flows through complex capillaries by modeling the physical system as a viscous fluid with immersed deformable particles. We describe a parallel boundary integral equation solver for general elliptic partial differential equations, which we apply to Stokes flow through blood vessels. We also detail a parallel collision avoiding algorithm to ensure RBCs and the blood vessel remain contact-free. We have scaled our code on Stampede2 at the Texas Advanced Computing Center up to 34,816 cores. Our largest simulation enforces a contact-free state between four billion surface elements and solves for three billion degrees of freedom on one million RBCs and a blood vessel composed from two million patches.

Adaptive Neural Network-Based Approximation to Accelerate Eulerian Fluid Simulation
Wenqian Dong (University of California, Merced), Jie Liu (University of California, Merced), Zhen Xie (University of California, Merced), Dong Li (University of California, Merced)

Eulerian Fluid Simulation is an effective formula of modeling fluid in many HPC applications. The high computational complexity of traditional solvers leads to limited deployment, while several neural-network-based methods reduce the labor of computation by learning complicated
relationships and draw wide attention. However, these state-of-the-arts apply single network topology to this input-sensitive application, resulting in inability for achieving required accuracy or inefficiency for leveraging the power of machine learning.

In this work, we propose Smart-fluidnet, which provides various approximate models and enables adaptive approximation to accelerate fluid simulation. We generate multiple network typologies by designed transformation operations and select the most suitable models by offline output-quality controller, then a quality-aware runtime algorithm dynamically singles out the optimal one with ignorable runtime overhead. The results show that Smart-fluidnet is 46% and 590x faster than Tompson's method (the state-of-the-art neural network model) and original fluid simulation respectively on an NVIDIA Pascal GPU.

Best Paper Finalist: no
Best Student Paper Finalist: no

GPU Acceleration of Extreme Scale Pseudo-Spectral Simulations of Turbulence Using Asynchronism
Kiran Ravikumar (Georgia Institute of Technology), David Appelhans (IBM Corporation), P.K. Yeung (Georgia Institute of Technology)

This paper presents new advances in GPU-driven Fourier pseudo-spectral numerical algorithms, which allow the simulation of turbulent fluid flow at problem sizes which exceed the current state of the art. In contrast to several massively parallel petascale systems, the dense nodes of Summit, Sierra, and expected exascale machines can be exploited with coarser MPI decompositions which result in improved MPI all-to-all scaling. An asynchronous batching strategy, combined with the fast hardware connection between the large CPU memory and the fast GPUs allows effective use of the GPUs on problem sizes which are too large to reside in GPU memory. Communication performance is further improved by a hybrid MPI+OpenMP approach. Favorable performance is obtained up to a 18432^3 problem size on 3072 nodes of Summit, with a GPU to CPU speedup of 4.7 for a 12288^3 problem size (the largest problem size previously published in turbulence literature).

Best Paper Finalist: no
Best Student Paper Finalist: yes

1:30 pm - 3:00 pm

High Radix Routing
Practical and Efficient Incremental Adaptive Routing for HyperX Networks
Nic McDonald (Google LLC), Mikhail Isaev (Georgia Institute of Technology), Adriana Flores (Nvidia Corporation), Al Davis (Hewlett Packard Enterprise), John Kim (Korea Advanced Institute of Science and Technology (KAIST))

In efforts to increase performance and reduce cost, modern low-diameter networks are designed for average case traffic and rely on non-minimal adaptive routing for network load-balancing when adversarial traffic patterns are encountered. Source adaptive routing is the predominant method for adaptive routing even though it presents many deficiencies related to making global decisions based solely on local information. In contrast, incremental adaptive routing, which performs an adaptive decision at every hop, is able to increase throughput and reduce latency by overcoming the deficiencies of source adaptive routing. We present two incremental adaptive routing algorithms for HyperX which are the first to be fully implementable in modern high-radix router architectures and interconnection network protocols. Using cycle accurate simulations of a 4096 node network, our evaluation shows these algorithms are able to exceed the performance of prior work by as much as 4x with synthetic traffic and 25% with 27-point stencil traffic.

Best Paper Finalist: no
Best Student Paper Finalist: no

Mitigating Network Noise on Dragonfly Networks through Application-Aware Routing
Daniele De Sensi (University of Pisa, ETH Zurich), Salvatore Di Girolamo (ETH Zurich), Torsten Hoefler (ETH Zurich)

System noise can negatively impact the performance of HPC systems, and the interconnection network is one of the main factors contributing to this problem. To mitigate this effect, adaptive routing sends packets on non-minimal paths if they are less congested. However, while this may mitigate interference caused by congestion, it also generates more traffic since packets traverse additional hops, causing in turn congestion on other applications and on the application itself. In this paper, we first describe how to estimate network noise. By following these guidelines, we show how noise can be reduced by using routing algorithms which select minimal paths with a higher probability. We exploit this knowledge to design an algorithm which changes the probability of selecting minimal paths according to the application characteristics. We validate our solution on microbenchmarks and real-world applications on two systems relying on a Dragonfly interconnection network, showing noise reduction and performance improvement.

Best Paper Finalist: no
Topological-OMC Routing on Dragonfly  
Md Shafayat Rahman (Florida State University), Saptarshi Bhowmik (Florida State University), Yevgeniy Ryasnianskiy (Florida State University), Xin Yuan (Florida State University), Michael Lang (Los Alamos National Laboratory)

The Dragonfly network has been deployed in the current generation supercomputers and will be used in the next generation supercomputers. The Universal Globally Adaptive Load-balance routing (UGAL) is the state-of-the-art routing scheme for Dragonfly. In this work, we show that the performance of the conventional UGAL can be further improved on many practical Dragonfly networks, especially the ones with a small number of groups, by customizing the paths used in UGAL for each topology. We develop a scheme to compute the custom sets of paths for each topology and compare the performance of our topology-custom UGAL routing (T-UGAL) with conventional UGAL. Our evaluation with different UGAL variations and different topologies demonstrates that by customizing the routes, T-UGAL offers significant improvements over UGAL on many practical Dragonfly networks in terms of both latency when the network is under low load and throughput when the network is under high load.

ComDetective: A Lightweight Communication Detection Tool for Threads  
Muhammad Aditya Sasongko (Koc University, Turkey), Milind Chabbi (Scalable Machines Research), Palwisha Akhtar (Koc University, Turkey), Didem Unat (Koc University, Turkey)

Inter-thread communication is a vital performance indicator in shared-memory systems. Prior works on identifying inter-thread communication employed hardware simulators or binary instrumentation and suffered from inaccuracy or high overheads—both space and time—making them impractical for production use. We propose ComDetective, which produces communication matrices that are accurate and introduces low runtime and memory overheads, thus making it practical for production use.
ComDetective employs hardware performance counters to sample memory-access events and uses hardware debug registers to sample communicating pairs of threads. ComDetective can differentiate communication as true or false sharing between threads. Its runtime and memory overheads are only 1.30x and 1.27x, respectively, for the 18 applications studied under 500K sampling period. Using ComDetective, we produce insightful communication matrices for micro-benchmarks, PARSEC benchmark suite, and several CORAL applications and compare the generated matrices against MPI counterparts. Guided by ComDetective we optimize a few codes and achieve up to 13% speedup.

Best Paper Finalist: yes
Best Student Paper Finalist: yes

Pinpointing Performance Inefficiencies via Lightweight Variance Profiling
Pengfei Su (College of William & Mary), Shuyin Jiao (College of William & Mary), Milind Chabbi (Scalable Machines Research), Xu Liu (College of William & Mary)

Execution variance among different invocation instances of the same procedure is often an indicator of performance losses. On the one hand, instrumentation-based tools can insert calipers around procedures and identify execution variance; however, they can introduce high overheads. On the other hand, sampling-based tools insert no instrumentation and have low overheads; however, they cannot synchronize samples with procedure entry and exit.

In this paper, we propose FVSampler, a lightweight, sampling-based variance profiler. FVSampler employs hardware performance monitoring units in conjunction with hardware debug registers to sample and monitor whole procedure instances (invocation till return) and collect hardware metrics in each sampled procedure instance. FVSampler, typically, incurs only 6% runtime overhead and negligible memory overhead making it suitable for HPC-scale production codes. We evaluate FVSampler with several parallel applications and demonstrate its effectiveness in pinpointing execution variance. Guided by FVSampler, we tune data structures and algorithms to obtain significant speedups.

Best Paper Finalist: no
Best Student Paper Finalist: no

Hatchet: Pruning the Overgrowth in Parallel Profiles
Abhinav Bhatel (University of Maryland, Lawrence Livermore National Laboratory), Stephanie Brink (Lawrence Livermore National Laboratory), Todd Gamblin (Lawrence Livermore National...
Performance analysis is critical for eliminating scalability bottlenecks in parallel codes. There are many profiling tools that can instrument codes and gather performance data, but general, easy to use, and programmable analytics and visualization tools are limited. In this paper, we focus on the analytics of structured profiling data, such as that obtained from calling context trees or nested region timers in code. We present a set of techniques and operations that build on the pandas data analysis library to enable parallel profile analysis. We have implemented these techniques in a Python-based library called Hatchet, which allows structured data to be filtered, aggregated, and pruned. Using performance datasets obtained from profiling parallel codes, we demonstrate how common performance analyses can be performed reproducibly with only a few lines of Hatchet code. Hatchet brings the power of modern data science tools to bear on performance analysis.

Cloud Scheduling

Spread-n-Share: Improving Application Performance and Cluster Throughput with Resource-Aware Job Placement
Xiongchao Tang (Tsinghua University, China; Sangfor Technologies Inc.), Haojie Wang (Tsinghua University, China), Xiaosong Ma (Qatar Computing Research Institute), Nosayba El-Sayed (Emory University), Jidong Zhai (Tsinghua University, China), Wenguang Chen (Tsinghua University, China), Ashraf Aboulnaga (Qatar Computing Research Institute)

Traditional batch job schedulers adopt the Compact-and-Exclusive (CE) strategy, packing processes of a parallel job into as few compute nodes as possible. While CE minimizes inter-node network communication, it often leads to self-contention among tasks of a resource-intensive application. Recent studies have used virtual containers to balance CPU utilization and memory capacity across physical nodes, but the unbalanced use of memory bandwidth and the shared last-level cache is still under-investigated.

In this work, we propose Spread-n-Share (SNS), a batch scheduling strategy that automatically scales resource-bound applications out onto more nodes to alleviate their performance bottleneck, and co-locates jobs in a resource compatible manner. We implement Uberun, a prototype scheduler
to validate SNS, considering memory bandwidth and LLC capacity as two types of performance-critical shared resources. Experimental results show that SNS improves the overall system throughput by 19.8% on average over CE, while achieving an average individual job speedup of 1.8%.

Best Paper Finalist: no
Best Student Paper Finalist: no

Swift Machine Learning Model Serving Scheduling: A Region Based Reinforcement Learning Approach
Heyang Qin (University of Nevada, Reno), Syed Zawad (University of Nevada, Reno), Yanqi Zhou (Google Brain), Lei Yang (University of Nevada, Reno), Dongfang Zhao (University of Nevada, Reno), Feng Yan (University of Nevada, Reno)

The success of machine learning has prospered Machine-Learning-as-a-Service (MLaaS) -- deploying trained machine learning (ML) models in cloud to provide low latency inference services at scale. To meet latency Service-Level-Objective (SLO), judicious parallelization at both request and operation levels is utterly important. However, existing ML systems (e.g., Tensorflow) and cloud ML serving platforms (e.g., SageMaker) are SLO-agnostic and rely on users to manually configure the parallelism. To provide low latency ML serving, this paper proposes a swift machine learning serving scheduling framework with a novel Region-based Reinforcement Learning (RRL) approach. RRL can efficiently identify the optimal parallelism configuration under different workloads by estimating performance of similar configurations with that of the known ones. We both theoretically and experimentally show that the RRL approach can outperform state-of-the-art approaches by finding near optimal solutions over 8 times faster while reducing inference latency up to 79.0% and reducing SLO violation up to 49.9%.

Best Paper Finalist: no
Best Student Paper Finalist: no

Slack Squeeze Coded Computing for Adaptive Straggler Mitigation
Krishna Giri Narra (University of Southern California), Zhifeng Lin (University of Southern California), Mehrdad Kiamari (University of Southern California), Salman Avestimehr (University of Southern California), Murali Annavaram (University of Southern California)

While performing distributed computations in today’s cloud-based platforms, execution speed variations among compute nodes can significantly reduce the performance and create bottlenecks
like stragglers. Coded computation techniques leverage coding theory to inject computational redundancy and mitigate stragglers in distributed computations. In this paper, we propose a scheduling strategy for coded computation called SlackSqueeze Coded Computation (S2C2). S2C2 squeezes the compute slack (i.e., overhead) that is built into the coded computing frameworks by efficiently assigning work for all fast and slow nodes according to their speeds and without needing to re-distribute data. We implement an LSTM-based speed prediction algorithm to predict speeds of compute nodes. We evaluate S2C2 on linear algebraic algorithms, gradient descent, graph ranking, and graph filtering algorithms. We demonstrate a 19% to 39% reduction in total computation latency using S2C2 compared to job replication and coded computation. We further show how S2C2 can be applied beyond matrix-vector multiplication.

Best Paper Finalist: yes
Best Student Paper Finalist: yes

3:30 pm - 5:00 pm

State of the Practice

Predicting Faults in High Performance Computing Systems: An In-Depth Survey of the State-of-the-Practice
David Jauk (Technical University Munich), Dai Yang (Technical University Munich), Martin Schulz (Technical University Munich)

As we near exascale, resilience remains a major technical hurdle. Any technique with the goal of achieving resilience suffers from having to be reactive, as failures can appear at any time. A wide body of research therefore aims at predicting failures, i.e., forecasting failures so that evasive actions can be taken while the system is still fully functional and enables a reasoning about its global state.

This research area has grown very diverse with a large number of approaches, yet is currently poorly classified, making it hard to understand impact of existing work. In this paper, we perform an extensive survey of existing literature in failure prediction by analyzing and comparing more than 30 different failure prediction approaches. We develop a taxonomy, which aids in categorizing the approaches, and show how this can help to understand the state-of-the-practice of this field and identify opportunities, gaps as well as future work.

Best Paper Finalist: no
Understanding the state-of-the-practice in MPI usage is paramount for many aspects of supercomputing, including optimizing the communication of HPC applications and informing standardization bodies and HPC systems procurements regarding the most important MPI features. Unfortunately, no previous study has characterized the use of MPI on applications at a significant scale; previous surveys focus either on small data samples or on MPI jobs of specific HPC centers. This paper presents the first comprehensive study of MPI usage in applications. We survey more than one hundred distinct MPI programs covering a significantly large space of the population of MPI applications. We focus on understanding the characteristics of MPI usage with respect to the most used features, code complexity, and programming models and languages. Our study corroborates certain findings previously reported on smaller data samples and presents a number of interesting, previously unreported insights.
Productivity from day one on supercomputers that leverage new technologies requires significant preparation. The institution procuring a novel system architecture often does not have enough people with all the requisite knowledge and skills to prepare for it. Thus, institutions have recently employed the "Center of Excellence" (CoE) concept to prepare for systems such as Summit and Sierra, currently the top two systems in the Top 500.

This paper documents CoE experiences preparing a workload of diverse applications and math libraries for a heterogeneous system. We describe our approach to this preparation, including our management and execution strategies, and detail our experiences with and reasons for employing different programming approaches. Our early science and performance results show the project
enabled significant early seismic science with up to a 14x throughput increase over Cori. In addition to our successes, we discuss our challenges and failures so others may benefit from our experience.

Best Paper Finalist: no
Best Student Paper Finalist: no

3:30 pm - 5:00 pm

Frameworks & Tools

**Diogenes: Looking for an Honest CPU/GPU Performance Measurement Tool**  
Benjamin R. Welton (University of Wisconsin), Barton P. Miller (University of Wisconsin)

GPU accelerators have become common on today’s leadership-class computing platforms. Exploiting the additional parallelism offered by GPUs is fraught with challenges. A key performance challenge faced by developers is how to limit the time consumed by synchronization and memory transfers between the CPU and GPU. We introduce the feed-forward measurement (FFM) performance tool model that automates the identification of unnecessary or inefficient synchronization and memory transfer, providing an estimate of potential benefit if the problem were fixed. FFM uses a new multi-stage/multi-run instrumentation model that adjusts instrumentation based on application behavior from prior runs, guiding FFM to problematic GPU operations that were previously unknown. The collected data feeds a new analysis model that gives an accurate estimate of potential benefit of fixing the problem. We created an implementation of FFM called Diogenes that we have used to identify problems in four real-world scientific applications.

Best Paper Finalist: no
Best Student Paper Finalist: no

**D2P: From Recursive Formulations to Distributed-Memory Codes**  
Nikhil Hegde (Indian Institute of Technology Dharwad, Purdue University), Qifan Chang (Purdue University), Milind Kulkarni (Purdue University)

Recursive formulations of programs are straightforward to reason about and write, often have good locality properties, and readily expose parallelism. We observe that it is easier to automatically generate distributed-memory codes for recursive formulations with certain properties: i) inclusive—a recursive method’s parameters summarize the data access done within the method body. ii)
Intersection—data-set intersection tests among method invocations can be computed efficiently.

In this paper we present D2P, a system that automatically generates distributed-memory codes for recursive divide-conquer algorithms with these properties. D2P produces MPI-based implementations starting from shared-memory specifications of the recursive algorithms. We evaluate D2P with recursive Dynamic Programming (DP) algorithms, since these algorithms have the desired properties and are well known. We show that the generated implementations are scalable and efficient: D2P generated implementations execute faster than implementations generated by recent distributed DP frameworks, and are competitive with (and often faster than) hand-written implementations.

Best Paper Finalist: no
Best Student Paper Finalist: no

**Legate NumPy: Accelerated and Distributed Array Computing**  
*Michael Bauer (Nvidia Corporation), Michael Garland (Nvidia Corporation)*

NumPy is a popular Python library used for performing array-based numerical computations. The canonical implementation of NumPy used by most programmers runs on a single CPU core and is parallelized to use multiple cores for some operations. This restriction to a single-node CPU-only execution limits both the size of data that can be handled and the potential speed of NumPy code. In this work, we introduce Legate, a drop-in replacement for NumPy that requires only a single-line code change and can scale up to an arbitrary number of GPU accelerated nodes. Legate works by translating NumPy programs to the Legion programming model and then leverages the scalability of the Legion runtime system to distribute data and computations across an arbitrary sized machine. Compared to similar programs written in the distributed Dask array library in Python, Legate achieves speed-ups of up to 10x on 1280 CPUs and 100x on 256 GPUs.

Best Paper Finalist: no
Best Student Paper Finalist: no

3:30 pm - 5:00 pm

**Linear Algebra Algorithms**

**Red-Blue Pebbling Revisited: Near Optimal Parallel Matrix Multiplication**
We propose COSMA: a parallel matrix-matrix multiplication algorithm that is near communication-optimal for all combinations of matrix dimensions, processor counts, and memory sizes. The key idea behind COSMA is to derive an optimal (up to a factor of 0.03% for 10MB of fast memory) sequential schedule and then parallelize it, preserving I/O optimality. To achieve this, we use the red-blue pebble game to precisely model MMM dependencies and derive a constructive and tight sequential and parallel I/O lower bound proofs. Compared to 2D or 3D algorithms, which fix processor decomposition upfront and then map it to the matrix dimensions, it reduces communication volume by up to $\sqrt{3}$. COSMA outperforms the established ScaLAPACK, CARMA, and CTF algorithms in all scenarios up to 12.8x (2.2x on average), achieving up to 88% of Piz Daint’s peak performance. Our work does not require any hand tuning and is maintained as an open source implementation.

Best Paper Finalist: yes
Best Student Paper Finalist: yes

AutoFFT: A Template-Based FFT Codes Auto-Generation Framework for ARM and X86 CPUs
Zhihao Li (Institute of Computing Technology, Chinese Academy of Sciences), Haipeng Jia (Institute of Computing Technology, Chinese Academy of Sciences), Yunquan Zhang (Institute of Computing Technology, Chinese Academy of Sciences), Tun Chen (Institute of Computing Technology, Chinese Academy of Sciences), Liang Yuan (Institute of Computing Technology, Chinese Academy of Sciences), Luning Cao (Institute of Computing Technology, Chinese Academy of Sciences), Xiao Wang (Institute of Computing Technology, Chinese Academy of Sciences)

The discrete Fourier transform (DFT) is widely used in scientific and engineering computation. This paper proposes a template-based code generation framework named AutoFFT that can automatically generate high-performance fast Fourier transform (FFT) codes. AutoFFT employs the Cooley-Tukey FFT algorithm, which exploits the symmetric and periodic properties of the DFT matrix as the outer parallelization framework. To further reduce the number of floating-point operations of butterflies, we explore more symmetric and periodic properties of the DFT matrix and formulate two optimized calculation templates for prime and power-of-two radices. To fully exploit hardware resources, we encapsulate a series of optimizations in an assembly template optimizer. Given any DFT problem, AutoFFT automatically generates C FFT kernels using these two templates and transfers them to efficient assembly codes using the template optimizer. Experiments show
that AutoFFT outperforms FFTW, ARMPI, and Intel MKL on average across all FFT types on ARMv8 and Intel x86-64 processors.

Best Paper Finalist: no
Best Student Paper Finalist: no

SLATE: Design of a Modern Distributed and Accelerated Linear Algebra Library
Mark Gates (University of Tennessee), Jakub Kurzak (University of Tennessee), Ali Charara (University of Tennessee), Asim YarKhan (University of Tennessee), Jack Dongarra (University of Tennessee; Oak Ridge National Laboratory, University of Manchester)

The SLATE (Software for Linear Algebra Targeting Exascale) library is being developed to provide fundamental dense linear algebra capabilities for current and upcoming distributed high-performance systems, both accelerated CPU-GPU-based and CPU-based. SLATE will provide coverage of existing ScaLAPACK functionality, including the Parallel BLAS, linear systems using LU and Cholesky, least squares problems, and eigenvalue and singular value problems. In this respect, it will serve as a replacement for ScaLAPACK, which after two decades of operation, cannot adequately be retrofitted for modern accelerated architectures. SLATE uses modern techniques such as communication-avoiding algorithms, lookahead panels to overlap communication and computation, and task-based scheduling, along with a modern C++ framework. Here we present the design of SLATE and initial reports of several of its components.

Best Paper Finalist: no
Best Student Paper Finalist: no

3:30 pm - 5:00 pm

Power and Scale

Uncore Power Scavenger: A Runtime for Uncore Power Conservation on HPC Systems
Neha Gholkar (North Carolina State University), Frank Mueller (North Carolina State University), Barry Rountree (Lawrence Livermore National Laboratory)

The US Department of Energy (DOE) has set a power target of 20-30MW on the first exascale machines. To achieve one exaflop under this power constraint, it is necessary to minimize wasteful consumption of power while striving to improve performance. Toward this end, we investigate
uncore frequency scaling (UFS) as a knob for reducing power footprints of HPC jobs. We propose Uncore Power Scavenger (UPSCavenger), a runtime system that dynamically detects phase changes and automatically sets the best uncore frequency for every phase to save power without significant impact on performance. Our experimental evaluations show that UPSCavenger achieves up to 10% energy savings with under 1% slowdown. It achieves 14% energy savings with the worst case slowdown of 5.5%. We show that UPSCavenger achieves up to 20% speedup and proportional energy savings compared to Intel’s RAPL with equivalent power usage making it a viable solution even for power-constrained computing.

Best Paper Finalist: no
Best Student Paper Finalist: no

PoDD: Power-Capping Dependent Distributed Applications
Huazhe Zhang (University of Chicago), Henry Hoffmann (University of Chicago)

Power budgeting (or capping) has become essential for large-scale computing installations. As these systems scale out, they are able to concurrently execute dependent applications that were serially processed previously. This reduces resource usage and execution time as they communicate at runtime instead of through disk. One challenge for power budgeting systems is how to power cap dependent applications for high performance. Existing approaches, however, have major limitations: (1) poor practicality, due to dependence on application profiles, (2) only optimize for power reallocation between nodes, without considering node-level optimization.

We propose PoDD -- a hierarchical, distributed, dynamic power management system for dependent applications. We implement it on a 49-node cluster and compare it to SLURM, a state-of-the-art power management system, but not considering coupling, and PowerShift, a power capping system for dependent applications without node-level optimization. On average, PoDD increases performance over SLURM by 14-22%, over PowerShift by 11-13%.

Best Paper Finalist: yes
Best Student Paper Finalist: yes

Etalumis: Bringing Probabilistic Programming to Scientific Simulators at Scale
Atilim Gunes Baydin (University of Oxford), Lei Shao (Intel Corporation), Wahid Bhimji (Lawrence Berkeley National Laboratory), Lukas Heinrich (European Organization for Nuclear Research (CERN)), Lawrence F. Meadows (Intel Corporation), Jialin Liu (Lawrence Berkeley National Laboratory), Andreas Munk (University of British Columbia), Saeid Naderipurizi (University of British
Probabilistic programming languages (PPLs) are receiving widespread attention for performing Bayesian inference in complex generative models. However, applications to science remain limited because of the impracticability of rewriting complex scientific simulators in a PPL, the computational cost of inference, and the lack of scalable implementations. To address these, we present a novel PPL framework that couples directly to existing scientific simulators through a cross-platform probabilistic execution protocol and provides Markov chain Monte Carlo (MCMC) and deep-learning-based inference compilation (IC) engines for tractable inference. To guide IC inference, we perform distributed training of a dynamic 3DCNN-LSTM architecture with a PyTorch-MPI-based framework on 1,024 32-core CPU nodes of the Cori supercomputer with a global minibatch size of 128k: achieving a performance of 450 Tflop/s through enhancements to PyTorch. We demonstrate a Large Hadron Collider (LHC) use-case with the C++ Sherpa simulator and achieve the largest-scale posterior inference in a Turing-complete PPL.

Best Paper Finalist: yes
Best Student Paper Finalist: no

Wednesday, November 20

10:30 am - 12:00 pm

Network Evaluation

An Evaluation of the CORAL Interconnects
Christopher Zimmer (Oak Ridge National Laboratory), Scott Atchley (Oak Ridge National Laboratory), Ramesh Pankajaksham (Lawrence Livermore National Laboratory), Brian E. Smith (Oak Ridge National Laboratory), Ian Kariin (Lawrence Livermore National Laboratory), Matt Leininger (Lawrence Livermore National Laboratory), Adam Bertsch (Lawrence Livermore National Laboratory), Brian S. Ryujo (Lawrence Livermore National Laboratory), Jason Burmark (Lawrence Livermore National Laboratory), André Walker-Loud (Lawrence Berkeley National Laboratory), M. A. Clark (Nvidia Corporation), Olga Pearce (Lawrence Livermore National Laboratory)

In 2019, the Department of Energy deployed the Summit and Sierra supercomputers, both
employing the latest interconnect technology. In this paper, we provide an in-depth assessment of the systems' interconnects, that is based on Enhanced Data Rate (EDR) 100 Gb/s Mellanox Infiniband. Both systems use second-generation EDR Host Channel Adapters (HCAs) and switches adding several new features such as Adaptive Routing (AR), switch-based collectives, HCA-based tag matching, and NVMe-over-Fabrics offload. Although based on the same components, Summit's network is "non-blocking" (i.e., fully provisioned) and Sierra's network has a 2:1 taper. We evaluate the two systems' interconnects using traditional communication benchmarks as well as real applications. We find that the new Adaptive Routing dramatically improves performance, but the other new features still need improvement.

Best Paper Finalist: no
Best Student Paper Finalist: no

HyperX Topology: First At-Scale Implementation and Comparison to the Fat-Tree

Jens Domke (RIKEN Center for Computational Science (R-CCS), RIKEN), Satoshi Matsuoka (RIKEN Center for Computational Science (R-CCS), RIKEN), Ivan Radanov Ivanov (Tokyo Institute of Technology), Yuki Tsushima (Tokyo Institute of Technology), Tomoya Yuki (Tokyo Institute of Technology), Akihiro Nomura (Tokyo Institute of Technology), Shin'ichi Miura (Tokyo Institute of Technology), Nic McDonald (Hewlett Packard Enterprise), Dennis Lee Floyd (Hewlett Packard Enterprise), Nicolas Dubé (Hewlett Packard Enterprise)

The de-facto standard topology for modern HPC systems and data-centers are Folded Clos networks, commonly known as Fat-Trees. The number of network endpoints in these systems is steadily increasing. The switch radix increase is not keeping up, forcing an increased path length in these multi-level trees that will limit gains for latency-sensitive applications. Additionally, today's Fat-Trees force the extensive use of active optical cables which carries a prohibitive cost-structure at scale.

To tackle these issues, researchers proposed various low-diameter topologies, such as Dragonfly. Another novel, but only theoretically studied, option is the HyperX. We built the world's first 3 Pflop/s supercomputer with two separate networks, a 3-level Fat-Tree and a 12x8 HyperX. This dual-plane system allows us to perform a side-by-side comparison using a broad set of benchmarks. We show that the HyperX, together with our novel communication pattern-aware routing, can challenge the performance of, or even outperform, traditional Fat-Trees.

Best Paper Finalist: no
Best Student Paper Finalist: no
Bandwidth Steering for HPC Using Silicon Nanophotonics

George Michelogiannakis (Lawrence Berkeley National Laboratory, Stanford University), Yiwen Shen (Columbia University), Min Yee Teh (Columbia University), Xiang Meng (Columbia University), Benjamin Aivazi (Columbia University), Taylor Groves (Lawrence Berkeley National Laboratory), John Shalf (Lawrence Berkeley National Laboratory), Madeleine Glick (Columbia University), Manya Ghobadi (Massachusetts Institute of Technology (MIT)), Larry Dennison (Nvidia Corporation), Keren Bergman (Columbia University)

As bytes-per-FLOP ratios continue to decline, communication is becoming a bottleneck for performance scaling. This paper describes bandwidth steering in HPC using emerging reconfigurable silicon photonic switches. We demonstrate that placing photonics in the lower layers of a hierarchical topology efficiently changes the connectivity and consequently allows operators to recover from system fragmentation that is otherwise hard to mitigate using common task placement strategies. Bandwidth steering enables efficient utilization of the higher layers of the topology and reduces cost with no performance penalties. In our simulations with a few thousand network endpoints, bandwidth steering reduces static power consumption per unit throughput by 36% and dynamic power consumption by 14% compared to a reference fat tree topology. Such improvements magnify as we taper the bandwidth of the upper network layer. In our hardware testbed, bandwidth steering improves total application execution time by 69%, unaffected by bandwidth tapering.

Best Paper Finalist: no
Best Student Paper Finalist: no

10:30 am - 12:00 pm

Machine Learning Optimization

PruneTrain: Fast Neural Network Training by Dynamic Sparse Model Reconfiguration

Sangkug Lym (University of Texas), Esha Choukse (University of Texas), Siavash Zangeneh (University of Texas), Wei Wen (Duke University), Sujay Sanghavi (University of Texas), Mattan Erez (University of Texas)

State-of-the-art convolutional neural networks (CNNs) used in vision applications have large models with numerous weights. Training these models is very compute- and memory-resource intensive. Much research has been done on pruning or compressing these models to reduce the cost of inference, but little work has addressed the costs of training. We focus precisely on
accelerating training. We propose PruneTrain, a cost-efficient mechanism that gradually reduces the training cost during training. PruneTrain uses a structured regularization approach that drives the training optimization toward both high accuracy and small weight values. Small weights can then be periodically removed by reconfiguring the network model to a smaller one. By using a structured-pruning approach and additional reconfiguration techniques we introduce, the pruned model can still be efficiently processed on a GPU accelerator. Overall, PruneTrain achieves a reduction of 39% in the training time of modern CNNs.

Best Paper Finalist: no
Best Student Paper Finalist: yes

Scalable Reinforcement-Learning-Based Neural Architecture Search for Cancer Deep Learning Research
Prasanna Balaprakash (Argonne National Laboratory), Romain Egele (Argonne National Laboratory), Misha Salim (Argonne National Laboratory), Stefan Wild (Argonne National Laboratory), Venkatram Vishwanath (Argonne National Laboratory), Fangfang Xia (Argonne National Laboratory), Tom Brettin (Argonne National Laboratory), Rick Stevens (Argonne National Laboratory)

Cancer is a complex disease. There is a growing need for the design and development of data-driven and, in particular, deep learning methods for various tasks such as cancer diagnosis, detection, prognosis, and prediction. For nonimage and nontext cancer data, designing high-performing deep learning models is a time-consuming, trial-and-error task that requires both cancer domain and deep learning expertise. We develop a reinforcement-learning-based neural architecture search to automate predictive model development for a class of representative cancer data. We develop custom building blocks that allow domain experts to incorporate the cancer-data-specific characteristics. We show that our approach discovers deep neural network architectures that have significantly fewer trainable parameters, shorter training time, and accuracy similar to or higher than those of manually designed architectures. We study and demonstrate the scalability of our approach on up to 1,024 Intel Knights Landing nodes of the Theta supercomputer at the Argonne Leadership Computing Facility.

Best Paper Finalist: no
Best Student Paper Finalist: no

BSTC: A Novel Binarized-Soft-Tensor-Core Design for Accelerating Bit-Based Approximated Neural Nets
We propose binarized-soft-tensor-core as a software-hardware co-design approach to construct the bit-manipulation capability for modern GPUs to effectively harvest the emerging bit-level-parallelism from BNNs and a variety of domains. We propose intra- and inter-layer fusion techniques so that the entire BNN inference process can be realized in one GPU kernel, labeled as Singular-Binarized-Neural-Network. Experiments show that our design can achieve over 1000x speedup for raw inference latency and 10x for inference throughput over state-of-the-art full-precision simulated BNN inference for AlexNet on ImageNet.

Best Paper Finalist: no
Best Student Paper Finalist: no

10:30 am - 12:00 pm

Compression

Significantly Improving Lossy Compression Quality Based on an Optimized Hybrid Prediction Model
Xin Liang (University of California, Riverside), Sheng Di (Argonne National Laboratory), Sihuan Li (University of California, Riverside), Dingwen Tao (University of Alabama), Bogdan Nicolae (Argonne National Laboratory), Zizhong Chen (University of California, Riverside), Franck Cappello (Argonne National Laboratory)

With ever-increasing volumes of data produced by large-scale scientific simulations, error-bounded lossy compression has come into a critical place. In this paper, we design a strategy to improve the compression quality significantly based on an optimized, hybrid prediction framework. The contribution is four-fold. (1) We propose a novel, transform-based predictor and optimize its compression quality. (2) We improve the coefficient-encoding efficiency for the data-fitting predictor. (3) We propose an adaptive framework that can select the bestfit predictor accurately for different datasets. (4) We perform the evaluation by running real-world applications on a supercomputer with 8192 cores. Experiments show that our adaptive compressor can improve the compression ratio by 112~165% over the second-best state-of-the-art lossy compressor. The parallel I/O performance is improved by about 100% because of significantly reduced data size. The total I/O time is reduced by up to 60x with our compressor compared with the original I/O time.
Moment Representation in the Lattice Boltzmann Method on Massively Parallel Hardware
Madhurima Vardhan (Duke University), John Gounley (Oak Ridge National Laboratory), Luiz Hegele (Santa Catarina State University), Erik Draeger (Lawrence Livermore National Laboratory), Amanda Randles (Duke University)

The widely-used lattice Boltzmann method (LBM) for computational fluid dynamics is highly scalable, but also significantly memory bandwidth-bound on current architectures. This paper presents a new regularized LBM implementation that reduces the memory footprint by only storing macroscopic, moment-based data. We show that the amount of data that must be stored in memory during a simulation is reduced by up to 47%. We also present a technique for cache-aware data re-utilization and show that optimizing cache utilization to limit data motion results in a similar improvement in time to solution. These new algorithms are implemented in the hemodynamics solver HARVEY and demonstrated using both idealized and realistic biological geometries. We develop a performance model for the moment representation algorithm and evaluate the performance on Summit.

Slim Graph: Practical Lossy Graph Compression for Approximate Graph Processing, Storage, and Analytics
Maciej Besta (ETH Zurich), Simon Weber (ETH Zurich), Lukas Gianinazzi (ETH Zurich), Robert Gerstenberger (ETH Zurich), Andrey Ivanov (ETH Zurich), Yishai Oltchik (ETH Zurich), Torsten Hoefler (ETH Zurich)

We propose Slim Graph: the first programming model and framework for practical lossy graph compression that facilitates high-performance approximate graph processing, storage, and analytics. Slim Graph enables the developer to express numerous compression schemes using small and programmable compression kernels that can access and modify local parts of input graphs. Such kernels are executed in parallel by the underlying engine, isolating developers from complexities of parallel programming. Our kernels implement novel graph compression schemes that preserve numerous graph properties, for example connected components, minimum spanning trees, or graph spectra. Finally, Slim Graph uses statistical divergences and other metrics to analyze the accuracy of lossy graph compression. We illustrate both theoretically and empirically that Slim
Graph accelerates numerous graph algorithms, reduces storage used by graph datasets, and ensures high accuracy of results. Slim Graph may become the common ground for developing, executing, and analyzing emerging lossy graph compression schemes.

Best Paper Finalist: yes
Best Student Paper Finalist: yes

1:30 pm - 3:00 pm

Sparse Computations

Conflict-Free Symmetric Sparse Matrix-Vector Multiplication on Multicore Architectures
Athena Elafrou (National Technical University of Athens), Georgios Goumas (National Technical University of Athens), Nectarios Koziris (National Technical University of Athens)

Exploiting the numeric symmetry in sparse matrices to reduce their memory footprint is very tempting for optimizing the memory-bound Sparse Matrix-Vector Multiplication (SpMV) kernel. Despite being very beneficial for serial computation, storing the upper or lower triangular part of the matrix introduces race conditions in the updates to the output vector in a parallel execution. Previous work has suggested using local, per-thread vectors to circumvent this problem, introducing a work-inefficient reduction step that limits the scalability of SpMV. In this paper, we address this issue with Conflict-Free Symmetric (CFS) SpMV, an optimization strategy that organizes the parallel computation into phases of conflict-free execution. We identify such phases through graph coloring and propose heuristics to improve the coloring quality for SpMV in terms of load balancing and locality to the input and output vectors. We evaluate our approach on two multicore shared-memory systems and demonstrate improved performance over the state-of-the-art.

Best Paper Finalist: no
Best Student Paper Finalist: no

An Efficient Mixed-Mode Representation of Sparse Tensors
Israt Nisa (Ohio State University), Jiajia Li (Pacific Northwest National Laboratory (PNNL)), Aravind Sukumaran-Rajam (Ohio State University), Prashant Rawat (Ohio State University), Sriram Krishnamoorthy (Pacific Northwest National Laboratory (PNNL)), P. (Saday) Sadayappan (University of Utah)
The Compressed Sparse Fiber (CSF) representation for sparse tensors is a generalization of the Compressed Sparse Row (CSR) format for sparse matrices. For a tensor with $d$ modes, typical tensor methods such as CANDECOMP/PARAFAC decomposition (CPD) require a sequence of $d$ tensor computations, where efficient memory access with respect to different modes is required for each of them. The straightforward solution is to use $d$ distinct representations of the tensor, with each one being efficient for one of the $d$ computations. However, a $d$-fold space overhead is often unacceptable in practice, especially with memory-constrained GPUs.

In this paper, we present a mixed-mode tensor representation that partitions the tensor’s nonzero elements into disjoint sections, each of which is compressed to create fibers along a different mode. Experimental results demonstrate that better performance can be achieved while utilizing only a small fraction of the space required to keep $d$ distinct CSF representations.

**Regularizing Irregularly Sparse Point-to-Point Communications**

Oguz Selvitopi (Lawrence Berkeley National Laboratory), Cevdet Aykanat (Bilkent University, Turkey)

This work tackles the communication challenges posed by the latency-bound applications with irregular communication patterns, i.e., applications with high average and/or maximum message counts. We propose a novel algorithm for reorganizing a given set of irregular point-to-point messages with the objective of reducing total latency cost at the expense of increased volume. We organize processes into a virtual process topology inspired by the $k$-ary $n$-cube networks and regularize irregular messages by imposing regular communication pattern(s) onto them. Exploiting this process topology, we propose a flexible store-and-forward algorithm to control the trade-off between latency and volume. Our approach is able to reduce the communication time of sparse-matrix multiplication with latency-bound instances drastically: up to 22.6x for 16K processes on a 3D Torus network and up to 7.2x for 4K processes on a Dragonfly network, with its performance getting better with increasing number of processes.

1:30 pm - 3:00 pm
Network Congestion and Offload

**GPCNeT: Designing a Benchmark Suite for Inducing and Measuring Contention in HPC Networks**

*Sudheer Chunduri (Argonne National Laboratory), Taylor Groves (Lawrence Berkeley National Laboratory), Peter Mendygral (Cray Inc), Brian Austin (Lawrence Berkeley National Laboratory), Jacob Balma (Cray Inc), Krishna Kandalla (Cray Inc), Kalyan Kumaran (Argonne National Laboratory), Glenn Lockwood (Lawrence Berkeley National Laboratory), Scott Parker (Argonne National Laboratory), Steven Warren (Cray Inc), Nathan Wichmann (Cray Inc), Nicholas Wright (Lawrence Berkeley National Laboratory)*

Network congestion is one of the biggest problems facing HPC systems today, affecting system throughput, performance, user experience, and reproducibility. Congestion manifests as run-to-run variability due to contention for shared resources (like filesystems) or routes between compute endpoints. Despite its significance, current network benchmarks fail to proxy the real-world network utilization seen on congested systems. We propose a new open-source benchmark suite called the Global Performance and Congestion Network Tests (GPCNeT) to advance the state of the practice in this area. The guiding principles used in designing GPCNeT are described, and the methodology employed to maximize its utility is presented. The capabilities of GPCNeT evaluated by analyzing results from several world’s largest HPC systems, including an evaluation of congestion management on a next-generation network. The results show that systems of all technologies and scales are susceptible to congestion, and this work motivates the need for congestion control in next-generation networks.

*Best Paper Finalist: no*

*Best Student Paper Finalist: no*

**Understanding Congestion in High Performance Interconnection Networks Using Sampling**

*Philip A. Taffet (Rice University, Lawrence Livermore National Laboratory), John M. Mellor-Crummey (Rice University)*

Communication cost is an important factor for applications on clusters and supercomputers. To improve communication performance, developers need tools that enable them to understand how their application’s communication patterns interact with the network, especially when those interactions result in congestion. Since communication performance is difficult to reason about analytically and simulation is costly, measurement-based approaches are needed. This paper describes a new sampling-based technique to collect information about the path a packet takes and
congestion it encounters. We describe a variant of this scheme that requires only 5–6 bits of information in a monitored packet, making it practical for use in next-generation networks. Simulations of synthetic benchmarks, miniGhost, and pF3D show that this strategy provides precise application-centric quantitative information about traffic and congestion that can be used to distinguish between problems with an application’s communication patterns, its mapping onto a parallel system, and outside interference.

Best Paper Finalist: no
Best Student Paper Finalist: no

**TriEC: Tripartite Graph Based Erasure Coding NIC Offload**
Haiyang Shi (Ohio State University), Xiaoyi Lu (Ohio State University)

Erasure Coding (EC) NIC offload is a promising technology for designing next-generation distributed storage systems. However, this paper has identified three major limitations of current-generation EC NIC offload schemes on modern SmartNICs. Thus, this paper proposes a new EC NIC offload paradigm based on the tripartite graph model, namely TriEC. TriEC supports both encode-and-send and receive-and-decode operations efficiently. Through theorem-based proofs, co-designs with memcached (i.e., TriEC-Cache), and extensive experiments, we show that TriEC is correct and can deliver better performance than the state-of-the-art EC NIC offload schemes (i.e., BiEC). Benchmark evaluations demonstrate that TriEC outperforms BiEC by up to 1.82x and 2.33x for encoding and recovering, respectively. With extended YCSB workloads, TriEC reduces the average write latency by up to 23.2% and the recovery time by up to 37.8%. TriEC outperforms BiEC by 1.32x for a full-node recovery with 8 million records.

Best Paper Finalist: no
Best Student Paper Finalist: yes

**1:30 pm - 3:00 pm**

**Partitioning & Scheduling**

**A Constraint-Based Approach to Automatic Data Partitioning for Distributed Memory Execution**
Wonchan Lee (Stanford University), Manolis Papadakis (Stanford University), Elliott Slaughter (SLAC National Accelerator Laboratory), Alex Aiken (Stanford University)
Although data partitioning is required to enable parallelism on distributed memory systems, data partitions are not first class objects in most distributed programming models. As a result, automatic parallelizers and application writers encode a particular partitioning strategy in the parallelized program, leading to a program not easily configured or composed with other parallel programs.

We present a constraint-based approach to automatic data partitioning. By introducing abstractions for first-class data partitions, we express a space of correct partitioning strategies. Candidate partitions are characterized by partitioning constraints, which can be automatically inferred from data accesses in parallelizable loops. Constraints can be satisfied by synthesized partitioning code or user-provided partitions. We demonstrate that programs auto-parallelized in our approach are easily composed with manually parallelized parts and have scalability comparable to hand-optimized counterparts.

Best Paper Finalist: yes
Best Student Paper Finalist: yes

Understanding Priority-Based Scheduling of Graph Algorithms on a Shared-Memory Platform
Serif Yesil (University of Illinois), Azin Heidarshenas (University of Illinois), Adam Morrison (Tel Aviv University), Josep Torrellas (University of Illinois)

Many task-based graph algorithms benefit from executing tasks according to some priority order. To support such algorithms, graph frameworks use Concurrent Priority Schedulers (CPSs), which attempt to execute the tasks according to priority order. While CPSs are critical to performance, there is insufficient insight on the relative strengths and weaknesses of the different CPS designs in the literature. Such insights would be invaluable to design better CPSs for graph processing.

This paper performs a detailed empirical performance analysis of several advanced CPS designs in a state-of-the-art graph analytics framework running on a large shared-memory server. Our analysis finds that all CPS designs but one impose major overheads that dominate running time. Only one CPS, the Galois' OBIM, typically imposes negligible overheads, but its performance is input-dependent and can melt down for some inputs. Based on these insights, we develop PMOD that is both robust and delivers the highest performance overall.

Best Paper Finalist: no
Best Student Paper Finalist: no

Almost Deterministic Work Stealing
Shumpei Shiina (University of Tokyo), Kenjiro Taura (University of Tokyo)

With task parallel models, programmers can easily parallelize divide-and-conquer algorithms by using nested fork-join structures. Work stealing, which is a popular scheduling strategy for task parallel programs, can efficiently perform dynamic load balancing; however, it tends to damage data locality and does not scale well with memory-bound applications. This paper introduces Almost Deterministic Work Stealing (ADWS), which addresses the issue of data locality of traditional work stealing by making the scheduling almost deterministic. Specifically, ADWS consists of two parts: (i) deterministic task allocation, which deterministically distributes tasks to workers based on the amount of work for each task, and (ii) hierarchical localized work stealing, which dynamically compensates load imbalance in a locality-aware manner. Experimental results show that ADWS is up to nearly 6 times faster than a traditional work stealing scheduler with memory-bound applications, and that dynamic load balancing works well while maintaining good data locality.

Best Paper Finalist: no
Best Student Paper Finalist: no

3:30 pm - 5:00 pm

Network and Memory Specialization

INCA: In-Network Compute Assistance
Whit Schonbein (Sandia National Laboratories), Ryan E. Grant (Sandia National Laboratories), Matthew G. F. Dosanjh (Sandia National Laboratories), Dorian Arnold (Emory University)

Current proposals for in-network data processing operate on data as it streams through a network switch or endpoint. Since compute resources must be available when data arrives, these approaches provide deadline-based models of execution. This paper introduces a deadline-free general compute model for network endpoints called INCA: In-Network Compute Assistance. INCA builds upon contemporary NIC offload capabilities to provide on-NIC, deadline-free, general-purpose compute capacities that can be utilized when the network is inactive. We demonstrate INCA is Turing complete, and provide a detailed design for extending existing hardware to support this model. We evaluate runtimes for a selection of kernels, including several optimizations, and show INCA can provide up to a 11% speedup for applications with minimal code modifications and between 25% to 37% when applications are optimized for INCA.

Best Paper Finalist: no
Near-Memory Data Transformation for Efficient Sparse Matrix Multi-Vector Multiplication
Daichi Fujiki (University of Michigan), Niladrish Chatterjee (Nvidia Corporation), Donghyuk Lee (Nvidia Corporation), Mike O’Connor (Nvidia Corporation, University of Texas)

Efficient manipulation of sparse matrices is critical to a wide range of HPC applications. We study one common operation, Sparse Matrix Multi-Vector Multiplication (SpMM), and evaluate the impact of the sparsity, distribution of non-zero elements, and tile-traversal strategies on GPU implementations. Using these insights, we determine that operating on these sparse matrices in tiled-DCSR is well-suited to the parallel warp-synchronous execution model of GPU.

Preprocessing or storing the sparse matrix in the tiled-DCSR format, however, often requires significantly more memory storage than conventional CSR or CSC formats. Given that SpMM kernels are often bottlenecked on DRAM bandwidth, the increase in DRAM traffic can result in a slowdown for many matrices.

This work enhances a GPU's last-level cache/memory controller unit to act as a dynamic translator between the compute-optimized representation of data (tiled-DCSR) and its corresponding storage/bandwidth-optimized format (CSC). Our approach achieves 2.26x better performance on average compared to cuSPARSE.

Best Paper Finalist: no
Best Student Paper Finalist: no

Network-Accelerated Non-Contiguous Memory Transfers
Salvatore Di Girolamo (ETH Zurich, Cray Inc), Konstantin Taranov (ETH Zurich), Andreas Kurth (ETH Zurich), Michael Schaffner (ETH Zurich), Timo Schneider (ETH Zurich), Jakub Beranek (IT4Innovations, Czech Republic), Maciej Besta (ETH Zurich), Luca Benini (ETH Zurich), Duncan Roweth (Cray Inc), Torsten Hoefler (ETH Zurich)

Applications often communicate data that is non-contiguous in the send- or the receive-buffer, e.g., when exchanging a column of a matrix stored in row-major order. While non-contiguous transfers are well supported in HPC (e.g., MPI derived datatypes), they can still be up to 5x slower than contiguous transfers of the same size. As we enter the era of network acceleration, we need to investigate which tasks to offload to the NIC: In this work we argue that non-contiguous memory transfers can be transparently network-accelerated, truly achieving zero-copy communications. We
implement and extend sPIN, a packet streaming processor, within a Portals 4 NIC SST model, and evaluate strategies for NIC-offloaded processing of MPI datatypes, ranging from datatype-specific handlers to general solutions for any MPI datatype. We demonstrate up to 8x speedup in the unpack throughput of real applications, demonstrating that non-contiguous memory transfers are a first-class candidate for network acceleration.

Best Paper Finalist: no
Best Student Paper Finalist: no

3:30 pm - 5:00 pm

Software Infrastructures for Applications

A Massively Parallel Infrastructure for Adaptive Multiscale Simulations: Modeling RAS Initiation Pathway for Cancer
Francesco Di Natale (Lawrence Livermore National Laboratory), Harsh Bhatia (Lawrence Livermore National Laboratory), Timothy S. Carpenter (Lawrence Livermore National Laboratory), Chris Neale (Los Alamos National Laboratory), Sara Kokkila Schumacher (IBM Research), Tomas Oppelstrup (Lawrence Livermore National Laboratory), Liam Stanton (San Jose State University), Xiaohua Zhang (Lawrence Livermore National Laboratory), Shiv Sundram (Lawrence Livermore National Laboratory), Thomas R. W. Scogland (Lawrence Livermore National Laboratory), Gautham Dharuman (Lawrence Livermore National Laboratory), Michael P. Surh (Lawrence Livermore National Laboratory), Claudia Misale (IBM Research), Lars Schneidenbach (IBM Corporation), Carlos Costa (IBM Corporation), Changhoan Kim (IBM Corporation), Bruce D'Amora (IBM Corporation), Sandrasegaram Gnanakaran (Los Alamos National Laboratory), Dwight V. Nissley (Frederick National Laboratory for Cancer Research), Fred Streitz (Lawrence Livermore National Laboratory), Felice C. Lightstone (Lawrence Livermore National Laboratory), Peer-Timo Bremer (Lawrence Livermore National Laboratory), James N. Glosli (Lawrence Livermore National Laboratory), Helgi I. Ingolfsson (Lawrence Livermore National Laboratory)

Most biological phenomena have microscopic foundations yet span macroscopic length- and time-scales, necessitating multiscale computational models. Efficient simulation of these complex multiscale models on modern heterogeneous architectures poses significant challenges in scheduling and co-managing resources such as computational power, communication bottlenecks, and filesystem bandwidth. To address these challenges, we present a novel massively parallel Multiscale Machine-Learned Modeling Infrastructure (MuMMI), which combines a large length- and
time-scale macro model with a high-fidelity molecular dynamics (MD) micro model using machine
learning. We describe our infrastructure which is designed for high scalability, efficiency,
robustness, portability, and fault tolerance on heterogeneous resources. We demonstrate MuMMI
conducting the largest-of-its-kind simulation to investigate the dynamics of KRAS proteins in
cancer initiation. Concurrently running up to 36,000 jobs on 16,000 GPUs and 176,000 CPU cores,
we executed 120,000 MD simulations surpassing an aggregate simulation time of 200
milliseconds, orders of magnitude greater than comparable studies.

Best Paper Finalist: yes
Best Student Paper Finalist: no

**CARE: Compiler-Assisted Recovery from Soft Failures**
Chao Chen (Georgia Institute of Technology), Greg Eisenhauer (Georgia Institute of Technology),
Santosh Pande (Georgia Institute of Technology), Qiang Guan (Kent State University)

As new architecture designs continue to boost the system performance with higher circuit density,
shrinking process technology and near-threshold voltage operations, the hardware is projected to
be more vulnerable to transient faults. Even though relatively infrequent, crashes due to transient
faults are incredibly disruptive, and are unpredictable necessitating frequent check-pointing, which
would incurs huge overhead.

In this paper, we present CARE, a light-weight and compiler-assisted technique to continue the
execution of applications upon crash-causing errors. CARE repairs corrupted states by recomputing
the data for the crashed architecture states on-the-fly. We evaluated CARE with 5 scientific
workloads with up to 3072 cores. During the normal execution of applications, CARE incurs near-
to-zero overheads, and can recover on an average 83.5% of crash-causing errors within ten of
milliseconds. Moreover, due to such an effective error-recovery mechanism, frequent check-pointing
can be relaxed into a relatively infrequent one, tremendously reducing the overheads.

Best Paper Finalist: no
Best Student Paper Finalist: yes

**Code Generation for Massively Parallel Phase-Field Simulations**
Martin Bauer (University of Erlangen-Nuremberg), Johannes Hötzer (University of Applied Science
Karlsruhe, Karlsruhe Institute of Technology), Dominik Ernst (University of Erlangen-Nuremberg),
Julian Hammer (University of Erlangen-Nuremberg), Marco Seiz (Karlsruhe Institute of Technology),
Henrik Hierl (Karlsruhe Institute of Technology), Jan Hönig (University of Erlangen-Nuremberg),
This article describes the development of automatic program generation technology to create scalable phase-field methods for material science applications. To simulate the formation of microstructures in metal alloys, we employ an advanced, thermodynamically consistent phase-field method. A state-of-the-art large-scale implementation of this model requires extensive, time-consuming, manual code optimization to achieve unprecedented fine mesh resolution. Our new approach starts with an abstract description based on free-energy functionals which is formally transformed into a continuous PDE and discretized automatically to obtain a stencil-based time-stepping scheme. Subsequently, an automatized performance engineering process generates highly optimized, performance-portable code for CPUs and GPUs. We demonstrate the efficiency for real-world simulations on large-scale GPU-based (PizDaint) and CPU-based (SuperMUC-NG) supercomputers. Our technique simplifies program development and optimization for a wide class of models. We further outperform existing, manually optimized implementations as our code can be generated specifically for each phase-field model and hardware configuration.

Best Paper Finalist: no
Best Student Paper Finalist: no

3:30 pm - 5:00 pm

GPU

Compiler Assisted Hybrid Implicit and Explicit GPU Memory Management Under Unified Address Space
Lingda Li (Brookhaven National Laboratory), Barbara Chapman (Brookhaven National Laboratory, Stony Brook University)

To improve programmability and productivity, recent GPUs adopt a virtual memory address space shared with CPUs (e.g., NVIDIA’s unified memory). Unified memory migrates the data management burden from programmers to system software and hardware, and enables GPUs to address datasets that exceed their memory capacity. Our experiments show that while the implicit data transfer of unified memory may bring better data movement efficiency, page fault overhead and data thrashing can erase its benefits. In this paper, we propose several user-transparent unified memory management schemes to achieve adaptive implicit and explicit data transfer and prevent data
Exploiting Reuse and Vectorization in Blocked Stencil Computations on CPUs and GPUs
Tuowen Zhao (University of Utah), Protonu Basu (Facebook), Samuel Williams (Lawrence Berkeley National Laboratory), Mary Hall (University of Utah), Hans Johansen (Lawrence Berkeley National Laboratory)

Stencil computations in real-world scientific applications may contain multiple interrelated stencils, have multiple input grids, and use higher order discretizations with high arithmetic intensity and complex expression structures. In combination, these properties place immense demands on the memory hierarchy that limit performance. Blocking techniques like tiling are used to exploit reuse in cache. Additional fine-grain data blocking can also reduce TLB, hardware prefetch, and cache pressure.

In this paper, we present a code generation approach designed to further improve tiled stencil performance by exploiting reuse within the block, increasing instruction-level parallelism, and exposing opportunities for the backend compiler to eliminate redundant computation. It also enables efficient vector code generation for CPUs and GPUs. For a wide range of complex stencil computations, we are able to achieve substantial speedups over tiled baselines for Intel KNL and Skylake-X and Nvidia P100 architectures.

A Versatile Software Systolic Execution Model for GPU Memory Bound Kernels
Peng Chen (Tokyo Institute of Technology, National Institute of Advanced Industrial Science and Technology (AIST)), Mohamed Wahib (National Institute of Advanced Industrial Science and Technology (AIST)), Shinichiro Takizawa (National Institute of Advanced Industrial Science and Technology (AIST)), Ryousei Takano (National Institute of Advanced Industrial Science and Technology (AIST)), Satoshi Matsuoka (RIKEN Center for Computational Science (R-CCS), Tokyo Institute of Technology)
This paper proposes a versatile high-performance execution model, inspired by systolic arrays, for memory-bound regular kernels running on CUDA-enabled GPUs. We formulate a systolic model that shifts partial sums by CUDA warp primitives for the computation. We also employ register files as a cache resource in order to operate the entire model efficiently. We demonstrate the effectiveness and versatility of the proposed model for a wide variety of stencil kernels that appear commonly in HPC, and also convolution kernels (increasingly important in deep learning workloads). Our algorithm outperforms the top reported state-of-the-art stencil implementations, including implementations with sophisticated temporal and spatial blocking techniques, on the two latest Nvidia architectures: Tesla V100 and P100. For 2D convolution of general filter sizes and shapes, our algorithm is on average 2.5× faster than Nvidia’s NPP on V100 and P100 GPUs.

Best Paper Finalist: yes
Best Student Paper Finalist: yes

Thursday, November 21

10:30 am - 12:00 pm

Improved Performance through Monitoring and Fine-Tuned Orchestration

End-to-End I/O Portfolio for the Summit Supercomputing Ecosystem
Sarp Oral (Oak Ridge National Laboratory, OpenSFS Inc), Sudharshan S. Vazhkudai (Oak Ridge National Laboratory), Fei yi Wang (Oak Ridge National Laboratory), Christopher Zimmer (Oak Ridge National Laboratory), Christopher Brumgard (Oak Ridge National Laboratory), Jesse Hanley (Oak Ridge National Laboratory), George Markomanolis (Oak Ridge National Laboratory), Ross Miller (Oak Ridge National Laboratory), Dustin Leverman (Oak Ridge National Laboratory), Scott Atchley (Oak Ridge National Laboratory), Veronica Vergara Larrea (Oak Ridge National Laboratory)

The I/O subsystem for the Summit supercomputer, No. 1 on the Top500 list, and its ecosystem of analysis platforms is composed of two distinct layers: the in-system layer and the center-wide parallel file system layer (PFS), Spider 3. The in-system layer uses node-local SSDs and provides 26.7 TB/s for reads, 9.7 TB/s for writes, and 4.6 billion IOPS to Summit. The Spider 3 PFS layer uses IBM’s Spectrum ScaleTM and provides 2.5 TB/s and 2.6 million IOPS to Summit and other systems. While deploying them as two distinct layers was operationally efficient, it also presented usability challenges in terms of multiple mount points and lack of transparency in data movement. To address these challenges, we have developed novel end-to-end I/O solutions for the concerted use
of the two storage layers. We present the I/O subsystem architecture, the end-to-end I/O solution space, their design considerations and our deployment experience.

Best Paper Finalist: no
Best Student Paper Finalist: no

From Facility to Application Sensor Data: Modular, Continuous and Holistic Monitoring with DCDB
Alessio Netti (Leibniz Supercomputing Centre, Technical University Munich), Micha Mueller (Leibniz Supercomputing Centre, Technical University Munich), Axel Auweter (MEGWARE Computer), Carla Guillen (Leibniz Supercomputing Centre), Michael Ott (Leibniz Supercomputing Centre), Daniele Tafani (Leibniz Supercomputing Centre), Martin Schulz (Technical University Munich)

Today's HPC installations are highly-complex systems, and their complexity will only increase as we move to exascale and beyond. At each layer, from facilities to systems, from runtimes to applications, a wide range of tuning decisions must be made in order to achieve efficient operation. This, however, requires systematic and continuous monitoring of system and user data. While many insular solutions exist, a system for holistic and facility-wide monitoring is still lacking in the current HPC ecosystem.

In this paper we introduce DCDB, a comprehensive monitoring system capable of integrating data from all system levels. It is designed as a modular and highly-scalable framework based on a plugin infrastructure. All monitored data is aggregated at a distributed noSQL data store for analysis and cross-system correlation. We demonstrate the performance and scalability of DCDB, and describe two use cases in the area of energy management and characterization.

Best Paper Finalist: no
Best Student Paper Finalist: no

Revisiting I/O Behavior in Large-Scale Storage Systems: The Expected and the Unexpected
Tirthak Patel (Northeastern University), Suren Byna (Lawrence Berkeley National Laboratory), Glenn K. Lockwood (Lawrence Berkeley National Laboratory), Devesh Tiwari (Northeastern University)

Large-scale applications typically spend a significant fraction of their execution time performing I/O to a parallel storage system. However, with rapid progress in compute and storage system stack of large-scale systems, it is critical to investigate and update our understanding of the I/O behavior of large-scale applications. Toward that end, in this work, we monitor, collect, and analyze a year's
worth of storage system data from the NERSC parallel storage system which serves NERSC’s two largest supercomputers, Cori and Edison. We perform temporal, spatial, and correlative analysis of the system as a whole, and of individual I/O and metadata servers, and uncover surprising patterns which defy existing assumptions about HPC I/O and have important implications for future systems.

Best Paper Finalist: no
Best Student Paper Finalist: no

10:30 am - 12:00 pm

Molecular Dynamics

**SW_GROMACS: Accelerate GROMACS on SUNWAY TaihuLight**
Tingjian Zhang (Shandong University; National Supercomputing Center, Wuxi), Yuxuan Li (National Tsing Hua University, Taiwan; National Supercomputing Center, Wuxi), Ping Gao (Shandong University; National Supercomputing Center, Wuxi), Qi Shao (Shandong University; National Supercomputing Center, Wuxi), Mingshan Shao (Shandong University; National Supercomputing Center, Wuxi), Meng Zhang (Shandong University; National Supercomputing Center, Wuxi), Jinxiao Zhang (Shandong University), Xiaohui Duan (Shandong University; National Supercomputing Center, Wuxi), Zhao Liu (Tsinghua University, China; National Supercomputing Center, Wuxi), Lin Gan (Tsinghua University, China; National Supercomputing Center, Wuxi), Haohuan Fu (Tsinghua University, China; National Supercomputing Center, Wuxi), Wei Xue (Tsinghua University, China; National Supercomputing Center, Wuxi), Weiguo Liu (Shandong University; National Supercomputing Center, Wuxi), Guangwen Yang (Tsinghua University, China; National Supercomputing Center, Wuxi)

GROMACS is one of the most popular Molecular Dynamic (MD) applications and is widely used in the field of chemical and biomolecular system study. Similar to other MD applications, it is computationally demanding. Therefore, many high-performance platforms have been employed to accelerate it, such as the KNL, Cell Processor, GPU and so on. As the third fastest supercomputer in the world, the TaihuLight contains 40,960 SW26010 processors, and every processor is a typical many-core processor. To make full use of the superior computation ability of the TaihuLight, we port the GROMACS to the SW26010. Finally, we not only achieved the 60x speedup in the hot spot kernel but also came up with a new useful strategy to accelerate the short-range interaction computation. This strategy could deal with the update conflict in the many-core processor acceleration without much performance hit. Besides SW26010, it could also be used in many other processors.
Fully Integrated FPGA Molecular Dynamics Simulations
Chen Yang (Boston University), Tong Geng (Boston University), Tianqi Wang (Boston University), Rushi Patel (Boston University), Qingqing Xiong (Boston University), Ahmed Sanaullah (Boston University), Chunshu Wu (Boston University), Jiayi Sheng (Falcon Computing Solutions Inc), Charles Lin (Silicon Therapeutics), Vipin Sachdeva (Silicon Therapeutics), Woody Sherman (Silicon Therapeutics), Martin Herbordt (Boston University)

The implementation of Molecular Dynamics (MD) on FPGAs has received substantial attention. Previous work, however, has consisted of either proof-of-concept implementations of components, usually the range-limited force; full systems, but with much of the work shared by the host CPU; or prototype demonstrations, e.g., using OpenCL, that neither implement a whole system nor have competitive performance. In this paper, we present what we believe to be the first full-scale FPGA-based simulation engine, and show that its performance is competitive with a GPU (running Amber in an industrial production environment). The system features on-chip particle data storage and management, short- and long-range force evaluation, as well as bonded forces, motion update, and particle migration. Other contributions of this work include exploring numerous architectural trade-offs and analysis on various mappings schemes among particles/cells and the various on-chip compute units.

OpenKMC: a KMC Design for a Hundred-Billion-Atom Simulation Using Millions of Cores on Sunway Taihulight
Kun Li (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences), Honghui Shang (Institute of Computing Technology, Chinese Academy of Sciences), Yunquan Zhang (Institute of Computing Technology, Chinese Academy of Sciences), Shigang Li (ETH Zurich), Baodong Wu (Institute of Computing Technology, Chinese Academy of Sciences; SenseTime Research), Dong Wang (Dalian Ocean University), Libo Zhang (Wuxi Jiangnan Institute of Computing Technology), Fang Li (Wuxi Jiangnan Institute of Computing Technology), Dexun Chen (National Supercomputing Center, Wuxi), Zhiqiang Wei (Qingdao National Laboratory for Marine Science and Technology)

With more attention attached to nuclear energy, the formation mechanism of the solute clusters
precipitation within complex alloys becomes intriguing research in the embrittlement of nuclear reactor pressure vessel (RPV) steels. Such phenomenon can be simulated with atomic kinetic Monte Carlo (AKMC) software, which evaluates the interactions of solute atoms with point defects in metal alloys. In this paper, we propose OpenKMC to accelerate large-scale KMC simulations on Sunway many-core architecture. To overcome the constraints caused by complex many-core architecture, we employ six levels of optimization in OpenKMC: (1) a new efficient potential computation model; (2) a group reaction strategy for fast event selection; (3) a software cache strategy; (4) combined communication optimizations; (5) a Transcription-Translation-Transmission algorithm for many-core optimization; (6) vectorization acceleration. Experiments illustrate that our OpenKMC has high accuracy and good scalability of applying hundred-billion-atom simulation over 5.2 million cores with a performance of over 80.1% parallel efficiency.

Best Paper Finalist: no
Best Student Paper Finalist: no

10:30 am - 12:00 pm

Algorithmic Techniques for Large-Scale Applications

Local-Global Merge Tree Computation with Local Exchanges
Arnur Nigmetov (Graz University of Technology), Dmitriy Morozov (Lawrence Berkeley National Laboratory)

A merge tree is a topological summary of a real-valued function on a graph. Merge trees can be used to find stable features in the data, report the number of connected components above any threshold, or compute other topological descriptors. A local-global merge tree provides a way of distributing a merge tree among multiple processors so that queries can be performed with minimal communication. While this makes them efficient in massively parallel setting, the only known algorithm for computing a local-global merge tree involves global reduction.

Motivated by applications in cosmological simulations, we consider a restricted version of the problem: we compute a local-global tree down to a threshold fixed by the user. We describe two algorithms for computing such a tree via only local exchanges between processors. We present a number of experiments that show the advantage of our method on different simulations.

Best Paper Finalist: no
Best Student Paper Finalist: no
Solving PDEs in Space-Time: 4D Tree-Based Adaptivity, Mesh-Free and Matrix-Free Approaches
Masado Ishii (University of Utah), Milinda Fernando (University of Utah), Kumar Saurabh (Iowa State University), Biswajit Khara (Iowa State University), Baskar Ganapathysubramanian (Iowa State University), Hari Sundar (University of Utah)

We present a kD tree-based parallel strategy for solving with a spacetime-adaptive approach a general class of partial differential equations. This approach is primarily motivated by the necessity of designing computational methodologies that can scale to leverage the availability of very large computing clusters (exascale and beyond). For evolution problems, the standard approach of decomposing the spatial domain is a powerful paradigm of parallelization. However, for a fixed spatial discretization, the efficiency of purely spatial domain decomposition degrades substantially beyond a threshold. To overcome this, we consider the time domain as an additional dimension and simultaneously solve for blocks of time (spacetime), instead of the standard approach of sequential time-stepping. Spacetime discretization includes natural incorporation a posterior error, full-time solution history, and removal of time-stepping constraints. We present scalable algorithms to perform matrix and matrix-free computations on KD trees, and show scalability across 32K cores in Titan.

Best Paper Finalist: no
Best Student Paper Finalist: no

From Piz Daint to the Stars: Simulation of Stellar Mergers Using High-Level Abstractions
Gregor Daiß (University of Stuttgart), Parsa Amini (Louisiana State University), John Biddiscombe (Swiss National Supercomputing Centre (CSCS)), Patrick Diehl (Louisiana State University), Juhan Frank (Louisiana State University), Kevin Huck (University of Oregon), Hartmut Kaiser (Louisiana State University), Dominic Marcello (Louisiana State University), David Pfander (University of Stuttgart), Dirk Pflüger (University of Stuttgart)

We study the simulation of stellar mergers, which requires complex simulations with high computational demands. We have developed Octo-Tiger, a finite volume grid-based hydrodynamics simulation code with Adaptive Mesh Refinement which is unique in conserving both linear and angular momentum to machine precision. To face the challenge of increasingly complex, diverse, and heterogeneous HPC systems, Octo-Tiger relies on high-level programming abstractions.

We use HPX with its futurization capabilities to ensure scalability both between nodes and within,
and present first results replacing MPI with libfabric achieving up to a 2.8x speedup. We extend Octo-Tiger to heterogeneous GPU-accelerated supercomputers, demonstrating node-level performance and portability. We show scalability up to full system runs on Piz Daint. For the scenario’s maximum resolution, the compute-critical parts (hydrodynamics and gravity) achieve 68.1% parallel efficiency at 2048 nodes.

Best Paper Finalist: no
Best Student Paper Finalist: no

10:30 am - 12:00 pm

Resilience and Fault Injection

BinFI: An Efficient Fault Injector for Safety-Critical Machine Learning Systems
Zitao Chen (University of British Columbia), Guanpeng Li (University of British Columbia), Karthik Pattabiraman (University of British Columbia), Nathan DeBardeleben (Los Alamos National Laboratory)

As machine learning (ML) becomes pervasive in high performance computing, ML has found its way into safety-critical domains (e.g., autonomous vehicles). Thus the reliability of ML has grown in importance. Specifically, failures of ML systems can have catastrophic consequences, and can occur due to soft errors, which are increasing in frequency due to system scaling. Therefore, we need to evaluate ML systems in the presence of soft errors.

In this work, we propose BinFI, an efficient fault injector (FI) for finding the safety-critical bits in ML applications. We find the widely-used ML computations are often monotonic. Thus we can approximate the error propagation behavior of a ML application as a monotonic function. BinFI uses a binary-search like FI technique to pinpoint the safety-critical bits (also measure the overall resilience). BinFI identifies 99.56% of safety-critical bits (with 99.63% precision) in the systems, which significantly outperforms random FI, with much lower costs.

Best Paper Finalist: no
Best Student Paper Finalist: no

Assessing the Impact of Timing Errors on HPC Applications
Chun-Kai Chang (University of Texas), Wenyi Yin (University of Texas), Mattan Erez (University of
Texas)

Timing errors are a growing concern for system resilience as technology continues to scale. It is problematic to use low-fidelity errors such as single-bit flips to model realistic timing errors. We address the lack of holistic methodology and tools for evaluating resilience of applications against timing errors. The proposed technique is able to rapidly inject high-fidelity and configurable timing errors to applications at the instruction level. Our implementation has no runtime dependencies on proprietary tools, enabling full parallelism of error injection campaign. Furthermore, because an injection point may not generate an actual error for a particular application run, we propose an acceleration technique to maximize the likelihood of generating errors that contribute to the overall campaign with speedup up to 7x. With our tool, we show that realistic timing errors lead to distinct error profiles from those of radiation-induced errors at both the instruction level and the application level.

Best Paper Finalist: no
Best Student Paper Finalist: no

**FT-iSort: Efficient Fault Tolerance for Introsort**

Sihuan Li (University of California, Riverside), Hongbo Li (University of California, Riverside), Xin Liang (University of California, Riverside), Jieyang Chen (Oak Ridge National Laboratory), Elisabeth Giem (University of California, Riverside), Kaiming Ouyang (University of California, Riverside), Kai Zhao (University of California, Riverside), Sheng Di (Argonne National Laboratory), Franck Cappello (Argonne National Laboratory), Zizhong Chen (University of California, Riverside)

Introspective sorting is a ubiquitous sorting algorithm which underlies many large scale distributed systems. Hardware-mediated soft errors can result in comparison and memory errors, and thus cause introsort to generate incorrect output, which in turn disrupts systems built upon introsort; hence, it is critical to incorporate fault tolerance capability within introsort. This paper proposes the first theoretically-sound, practical fault tolerant introsort with negligible overhead: FT-iSort. To tolerate comparison errors, we use minimal TMR protection via exploiting the properties of the effects of soft errors on introsort. This algorithm-based selective protection incurs far less overhead than naïve TMR protection designed to protect an entire application. To tolerate memory errors that escape DRAM error correcting code, we propose XOR-based re-execution. We incorporate our fault tolerance method into the well-known parallel sorting implementation HykSort, and we find that fault tolerant HykSort incurs negligible overhead and obtains nearly the same scalability as unprotected HykSort.

Best Paper Finalist: no
Graph and Tensor Computations

Distributed Enhanced Suffix Arrays: Efficient Algorithms for Construction and Querying
Patrick Flick (Georgia Institute of Technology), Srinivas Aluru (Georgia Institute of Technology)

Suffix arrays and trees are important and fundamental string data structures which lie at the foundation of many string algorithms, with important applications in computational biology, text processing, and information retrieval. Recent work enables the efficient parallel construction of suffix arrays and trees requiring at most $O(n/p)$ memory per process in distributed memory.

However, querying these indexes in distributed memory has not been studied extensively. Querying common string indexes such as suffix arrays, enhanced suffix arrays, and FM-Index, all require random accesses into $O(n)$ memory - which in distributed memory algorithms becomes prohibitively expensive.

In this paper, we introduce a novel distributed string index, the Distributed Enhanced Suffix Array (DESA). We present efficient algorithms for the construction and for querying of this distributed data structure, all while requiring only $O(n/p)$ memory per process. We further provide a scalable parallel implementation and demonstrate its performance and scalability.

Scalable Generation of Graphs for Benchmarking HPC Community-Detection Algorithms
George M. Slota (Rensselaer Polytechnic Institute (RPI)), Jonathan W. Berry (Sandia National Laboratories), Simon D. Hammond (Sandia National Laboratories), Stephen L. Olivier (Sandia National Laboratories), Cynthia A. Phillips (Sandia National Laboratories), Sivasankaran Rajamanickam (Sandia National Laboratories)

Community detection in graphs is a canonical social network analysis method. We consider the problem of generating suites of terascale synthetic social networks to compare the solution quality of parallel community-detection methods. The standard method, based on the graph generator of
Lancichinetti, Fortunato, and Radicchi (LFR), has been used extensively for modest-scale graphs, but has inherent scalability limitations.

We provide an alternative, based on the scalable Block Two-Level Erdos-Renyi (BTER) graph generator, that enables HPC-scale evaluation of solution quality in the style of LFR. Our approach varies community coherence, and retains other important properties. Our methods can scale real-world networks, e.g., to create a version of the Friendster network that is 512 times larger. With BTER’s inherent scalability, we can generate a 15-terabyte graph (4.6B vertices, 925B edges) in just over one minute. We demonstrate our capability by showing that label-propagation community-detection algorithm can be strong-scaled with negligible solution-quality loss.

Analytical Cache Modeling and Tilesize Optimization for Tensor Contractions
Rui Li (University of Utah), Aravind Sukumaran-Rajam (Ohio State University), Richard Veras (Louisiana State University), Tze Meng Low (Carnegie Mellon University), Fabrice Rastello (French Institute for Research in Computer Science and Automation (INRIA)), Atanas Rountev (Ohio State University), P. Sadayappan (University of Utah)

Data movement between processor and memory hierarchy is a fundamental bottleneck that limits the performance of many applications on modern computer architectures. Tiling and loop permutation are key techniques for improving data locality. However, selecting effective tile-sizes and loop permutations is particularly challenging for tensor contractions due to the large number of loops. Even state-of-the-art compilers usually produce sub-optimal tile-sizes and loop permutations, as they rely on naive cost models. In this paper, we provide an analytical model based approach to multi-level tile size optimization and permutation selection for tensor contractions. Our experimental results show that this approach achieves comparable or better performance than state-of-the-art frameworks and libraries for tensor contractions.

1:30 pm - 3:00 pm
Quantum Applications
Optimizing the Data Movement in Quantum Transport Simulations via Data-Centric Parallel Programming

Alexandros Nikolaos Ziogas (ETH Zurich), Tal Ben-Nun (ETH Zurich), Guillermo Indalecio Fernandez (ETH Zurich), Timo Schneider (ETH Zurich), Mathieu Luisier (ETH Zurich), Torsten Hoefler (ETH Zurich)

Designing efficient cooling systems for integrated circuits (ICs) relies on a deep understanding of the electro-thermal properties of transistors. To shed light on this issue in currently fabricated FinFETs, a quantum mechanical solver capable of revealing atomically-resolved electron and phonon transport phenomena from first-principles is required.

In this paper, we consider a global, data-centric view of a state-of-the-art quantum transport simulator to optimize its execution on supercomputers. The approach yields coarse- and fine-grained data-movement characteristics, which are used for performance and communication modeling, communication-avoidance, and data-layout transformations. The transformations are tuned for the Piz Daint and Summit supercomputers, where each platform requires different caching and fusion strategies to perform optimally. The presented results make ab initio device simulation enter a new era, where nanostructures composed of over 10,000 atoms can be investigated at an unprecedented level of accuracy, paving the way for better heat management in next-generation ICs.

Best Paper Finalist: no
Best Student Paper Finalist: no

Parallel Transport Time Dependent Density Functional Theory Calculations with Hybrid Functional on Summit

Weile Jia (University of California, Berkeley), Lin-Wang Wang (Lawrence Berkeley National Laboratory), Lin Lin (University of California, Berkeley; Lawrence Berkeley National Laboratory)

Real-time time-dependent density functional theory (rt-TDDFT) with hybrid exchange-correlation functional has wide-ranging applications in chemistry and material science simulations. However, it can be thousands of times more expensive than a conventional ground state DFT simulation, hence is limited to small systems. In this paper, we accelerate hybrid functional rt-TDDFT calculations using the parallel transport gauge formalism, and the GPU implementation on Summit. Our implementation can efficiently scale to 786 GPUs for a large system with 1536 silicon atoms, and the wall clock time is only 1.5 hours per femtosecond. This unprecedented speed enables the simulation of large systems with more than 1000 atoms using rt-TDDFT and hybrid functional.
Full-State Quantum Circuit Simulation by Using Data Compression
Xin-Chuan Wu (University of Chicago), Sheng Di (Argonne National Laboratory), Emma Maitreyee Dasgupta (University of Chicago), Franck Cappello (Argonne National Laboratory), Hal Finkel (Argonne National Laboratory), Yuri Alexeev (Argonne National Laboratory), Frederic T. Chong (University of Chicago)

Quantum circuit simulations are critical for evaluating quantum algorithms and machines. However, the number of state amplitudes required for full simulation increases exponentially with the number of qubits. In this study, we leverage data compression to reduce memory requirements, trading computation time and fidelity for memory space. Specifically, we develop a hybrid solution by combining the lossless compression and our tailored lossy compression method with adaptive error bounds at each timestep of the simulation. Our approach optimizes for compression speed and makes sure that errors due to lossy compression are uncorrelated, an important property for comparing simulation output with physical machines.

Experiments show that our approach reduces the memory requirement of simulating the 61-qubit Grover’s search algorithm from 32 exabytes to 768 terabytes of memory on Argonne’s Theta supercomputer using 4,096 nodes. The results suggest that our techniques can increase the simulation size by 2~16 qubits for general quantum circuits.

Improving Next-Generation Performance and Resilience

SSD Failures in the Field: Symptoms, Causes, and Prediction Models
Jacob Alter (College of William & Mary), Ji Xue (College of William & Mary), Alma Dimnaku (NetApp Inc), Evgenia Smirni (College of William & Mary)

In recent years, solid state drives (SSDs) have become a staple of high-performance data centers for their speed and energy efficiency. In this work, we study the failure characteristics of 30,000
drives from a Google data center spanning six years. We characterize the workload conditions that lead to failures and illustrate that their root causes differ from common expectation but remain difficult to discern. Particularly, we study failure incidents that result in manual intervention from the repair process. We observe high levels of infant mortality and characterize the differences between infant and non-infant failures. We develop several machine learning failure prediction models that are shown to be surprisingly accurate, achieving high recall and low false positive rates. These models are used beyond simple prediction as they aid us to untangle the complex interaction of workload characteristics that lead to failures and identify failure root causes from monitored symptoms.

Best Paper Finalist: no
Best Student Paper Finalist: no

An Early Evaluation of Intel’s Optane DC Persistent Memory Module and Its Impact on High-Performance Scientific Applications
Michèle Weiland (Edinburgh Parallel Computing Centre), Holger Brunst (Technical University Dresden), Tiago Quintino (European Centre for Medium-Range Weather Forecasts), Nick Johnson (Edinburgh Parallel Computing Centre), Olivier Iffrig (European Centre for Medium-Range Weather Forecasts), Simon Smart (European Centre for Medium-Range Weather Forecasts), Christian Herold (Technical University Dresden), Antonino Bonanni (European Centre for Medium-Range Weather Forecasts), Adrian Jackson (Edinburgh Parallel Computing Centre), Mark Parsons (Edinburgh Parallel Computing Centre)

Memory and I/O performance bottlenecks in supercomputing simulations are two key challenges that must be addressed on the road to Exascale. The new byte-addressable persistent non-volatile memory technology from Intel, DCPMM, promises to be an exciting opportunity to break with the status quo, with unprecedented levels of capacity at near-DRAM speeds. Here, we explore the potential of DCPMM in the context of two high-performance scientific applications in terms of outright performance, efficiency and usability for both its Memory and App Direct modes. In Memory mode, we show equivalent performance and better efficiency for a CASTEP simulation that is limited by memory capacity on conventional DRAM-only systems without any changes to the application. For IFS, we demonstrate that a distributed object-store over NVRAM reduces the data contention created in weather forecasting data producer-consumer workflows. In addition, we also present the achievable memory bandwidth performance using STREAM.

Best Paper Finalist: no
Best Student Paper Finalist: no
Performance Optimality or Reproducibility: That Is the Question
Tapasya Patki (Lawrence Livermore National Laboratory), Jayaraman J. Thiagarajan (Lawrence Livermore National Laboratory), Alexis Ayala (Western Washington University), Tanzima Z. Islam (Western Washington University)

The era of extremely heterogeneous supercomputing brings with itself the devil of increased performance variation and reduced reproducibility. There is a lack of understanding in the HPC community on how the simultaneous consideration of network traffic, power limits, concurrency tuning, and interference from other jobs impacts application performance.

In this paper, we design a methodology that allows both HPC users and system administrators to understand the trade-off space between optimal and reproducible performance. We present a first-of-its-kind dataset that simultaneously varies multiple system- and user-level parameters on a production cluster, and introduce a new metric, called the desirability score, which enables comparison across different system configurations. We develop a novel, model-agnostic machine learning methodology based on the graph signal theory for comparing the influence of parameters on application predictability, and using a new visualization technique, make practical suggestions for best practices for multi-objective HPC environments.

Best Paper Finalist: no
Best Student Paper Finalist: no

3:30 pm - 5:00 pm
Image Reconstruction

iFDK: A Scalable Framework for Instant High-Resolution Image Reconstruction
Peng Chen (Tokyo Institute of Technology, National Institute of Advanced Industrial Science and Technology (AIST)), Mohamed Wahib (National Institute of Advanced Industrial Science and Technology (AIST)), Shinichiro Takizawa (National Institute of Advanced Industrial Science and Technology (AIST)), Ryousei Takano (National Institute of Advanced Industrial Science and Technology (AIST)), Satoshi Matsuoka (RIKEN Center for Computational Science (R-CCS), Tokyo Institute of Technology)

Computed Tomography (CT) is a widely used technology that requires compute-intense algorithms for image reconstruction. We propose a novel back-projection algorithm that reduces the projection computation cost to 1/6 of the standard algorithm. We also propose an efficient implementation
that takes advantage of the heterogeneity of GPU-accelerated systems by overlapping the filtering and back-projection stages on CPUs and GPUs, respectively. Finally, we propose a distributed framework for high-resolution image reconstruction on state-of-the-art GPU-accelerated supercomputers. The framework relies on an elaborate interleave of MPI collective communication steps to achieve scalable communication. Evaluation on a single Tesla V100 GPU demonstrates that our back-projection kernel performs up to 1.6 times faster than the standard FDK implementation. We also demonstrate the scalability and instantaneous CT capability of the distributed framework by using up to 2,048 V100 GPUs to solve a 4K and 8K problems within 30 seconds and 2 minutes, respectively (including I/O).

Best Paper Finalist: no
Best Student Paper Finalist: no

MemXCT: Memory-Centric X-Ray CT Reconstruction with Massive Parallelization
Mert Hidayetoglu (University of Illinois), Tekin Bicer (Argonne National Laboratory), Simon Garcia de Gonzalo (University of Illinois), Bin Ren (College of William & Mary), Doga Gursoy (Argonne National Laboratory), Rajkumar Kettimuthu (Argonne National Laboratory), Ian T. Foster (Argonne National Laboratory), Wen-Mei W. Hwu (University of Illinois)

X-ray computed tomography (CT) is used regularly at synchrotrons to study the internal morphology of materials at high resolution. However, experimental constraints, such as radiation sensitivity, can result in noisy or undersampled measurements. Further, depending on the resolution, sample size and data acquisition rates, the resulting noisy dataset can be terabyte-scale. Advanced iterative reconstruction techniques can produce high-quality images from noisy measurements, but their computational requirements have made their use exception rather than the rule. We propose here a novel memory-centric approach that avoids redundant computations at the expense of additional memory complexity. We develop a system, MemXCT, that uses an optimized SpMV implementation with multi-stage buffering and two-level pseudo-Hilbert ordering. We evaluate MemXCT on both KNL and GPUs architectures. MemXCT can reconstruct a large (11Kx11K) mouse brain tomogram in ~10 seconds using 4,096 KNL nodes (256K cores), the largest iterative reconstruction achieved in near-real time.

Best Paper Finalist: no
Best Student Paper Finalist: no

Consensus Equilibrium Framework for Super-Resolution and Extreme-Scale CT Reconstruction
Xiao Wang (Harvard Medical School, Boston Children's Hospital), Venkatesh Sridhar (Purdue
Computed tomography (CT) image reconstruction is a crucial technique for many imaging applications. Among various reconstruction methods, Model-Based Iterative Reconstruction (MBIR) enables super-resolution with superior image quality. MBIR, however, has a high memory requirement that limits the achievable image resolution, and the parallelization for MBIR suffers from limited scalability. In this paper, we propose Asynchronous Consensus MBIR (AC-MBIR) that uses Consensus Equilibrium (CE) to provide a super-resolution algorithm with a small memory footprint, low communication overhead and a high scalability. Super-resolution experiments show that AC-MBIR has a 6.8 times smaller memory footprint and 16 times more scalability, compared with the state-of-the-art MBIR implementation, and maintains a 100% strong scaling efficiency at 146880 cores. In addition, AC-MBIR achieves an average bandwidth of 3.5 petabytes per second at 587,520 cores.

Best Paper Finalist: no
Best Student Paper Finalist: no

3:30 pm - 5:00 pm

The Fewer Tiers, the Fewer Tears

Addressing Data Resiliency for Staging Based Scientific Workflows
Shaohua Duan (Rutgers University, Rutgers Discovery Informatics Institute), Pradeep Subedi (Rutgers University, Rutgers Discovery Informatics Institute), Philip E. Davis (Rutgers University, Rutgers Discovery Informatics Institute), Manish Parashar (Rutgers University, Rutgers Discovery Informatics Institute)

As applications move toward extreme scales, data-related challenges are becoming significant concerns, and in-situ workflows based on data staging and in-situ/in-transit data processing have been proposed to address these challenges. Increasing scales are also expected to result in an increase in the rate of silent data corruption errors, which will impact both the correctness and performance of applications. Furthermore, this impact is amplified in the case of in-situ workflows due to the dataflow between the component applications of the workflow. While existing research
has explored silent error detection at the application level, silent error detection for workflows remains an open challenge.

This paper addresses error detection for extreme scale in-situ workflows. The presented approach leverages the idle computation resource in data staging to enable timely detection and recovery from silent data corruption, effectively reducing the propagation of corrupted data and end-to-end workflow execution time in the presence of silent errors.

Best Paper Finalist: no
Best Student Paper Finalist: no

**LPCC: Hierarchical Persistent Client Caching for Lustre**

Yingjin Qian (DataDirect Networks (DDN)), Xi Li (DataDirect Networks (DDN)), Shuichi Ihara (DataDirect Networks (DDN)), Andreas Dilger (Whamcloud Inc), Carlos Thomaz (DataDirect Networks (DDN)), Shilong Wang (DataDirect Networks (DDN)), Wen Cheng (Huazhong University of Science and Technology), Chunyan Li (Huazhong University of Science and Technology), Lingfang Zeng (Huazhong University of Science and Technology), Fang Wang (Huazhong University of Science and Technology), Dan Feng (Huazhong University of Science and Technology), Tim Suesst (Johannes Gutenberg University Mainz), Andre Brinkmann (Johannes Gutenberg University Mainz)

Most high-performance computing (HPC) clusters today use a global parallel file system to enable high data throughput. The parallel file system is typically centralized, and its storage media are physically separated from the compute cluster. Compute nodes as clients of the parallel file system are often additionally equipped with SSDs. The node internal storage media are rarely well-integrated into the I/O and compute workflows.

In this paper, we propose a hierarchical Persistent Client Caching (LPCC) mechanism for the Lustre file system. LPCC integrates with the Lustre HSM solution and the Lustre layout lock mechanism to provide consistent persistent caching services for I/O applications running on client nodes, meanwhile maintaining a global unified namespace of the entire Lustre file system for distributed persistent client caching. The evaluation results presented in this paper show LPCC's advantages for various workloads, enabling even speed-ups linear in the number of clients for several real-world scenarios.

Best Paper Finalist: no
Best Student Paper Finalist: no
Replication Is More Efficient Than You Think
Anne Benoit (ENS Lyon), Thomas Herault (University of Tennessee), Valentin Le Fèvre (ENS Lyon), Yves Robert (ENS Lyon, University of Tennessee)

This paper revisits replication coupled with checkpointing for fail-stop errors. Previously published works use replication with the no-restart strategy: (i) compute the application Mean Time To Interruption $M$ as a function of the number of processor pairs and the individual processor MTBF (Mean Time To Failures); (ii) use checkpointing period $P= \sqrt{2 M C}$, where $C$ is the checkpoint duration; and (iii) never restart failed processors until the application crashes. We introduce the restart strategy, where failed processors are restarted after each checkpoint, which may introduce an additional overhead but prevents the application configuration from degrading throughout successive checkpointing periods. We show how to compute the optimal checkpointing period $P'$ for this restart strategy, and prove that its length is an order of magnitude higher than $P$. Furthermore, we show through simulations that using $P'$ and the restart strategy significantly decreases the overhead induced by replication.

Best Paper Finalist: no
Best Student Paper Finalist: no

3:30 pm - 5:00 pm

Heterogeneous Systems

Stateful Dataflow Multigraphs: A Data-Centric Model for Performance Portability on Heterogeneous Architectures
Tal Ben-Nun (ETH Zurich), Johannes de Fine Licht (ETH Zurich), Alexandros Nikolaos Ziogas (ETH Zurich), Timo Schneider (ETH Zurich), Torsten Hoefler (ETH Zurich)

The ubiquity of accelerators in high-performance computing has driven programming complexity beyond the skill-set of the average domain scientist. To maintain performance portability in the future, it is imperative to decouple architecture-specific programming paradigms from the underlying scientific computations.

We present the Stateful DataFlow multiGraph (SDFG), a data-centric intermediate representation that enables separating program definition from its optimization. By combining fine-grained data dependencies with high-level control-flow, SDFGs are both expressive and amenable to program transformations, such as tiling and double-buffering. These transformations are applied to the
SDFG in an interactive process, using extensible pattern matching, graph rewriting, and a graphical user interface. We demonstrate SDFGs on CPUs, GPUs, and FPGAs over various motifs --- from fundamental computational kernels to graph analytics. We show that SDFGs deliver competitive performance, allowing domain scientists to develop applications naturally and port them to approach peak hardware performance without modifying the original scientific code.

Best Paper Finalist: no
Best Student Paper Finalist: no

**Streaming Message Interface: High-Performance Distributed Memory Programming on Reconfigurable Hardware**
Tiziano De Matteis (ETH Zurich), Johannes de Fine Licht (ETH Zurich), Jakub Beránek (IT4Innovations, Czech Republic), Torsten Hoefler (ETH Zurich)

Distributed memory programming is the established paradigm used in high-performance computing (HPC) systems, requiring explicit communication between nodes and devices. When FPGAs are deployed in distributed settings, communication is typically handled either by going through the host machine, sacrificing performance, or by streaming across fixed device-to-device connections, sacrificing flexibility.

We present Streaming Message Interface (SMI), a communication model and API that unifies explicit message passing with a hardware-oriented programming model, facilitating minimal-overhead, flexible, and productive inter-FPGA communication. Instead of bulk transmission, messages are streamed across the network during computation, allowing communication to be seamlessly integrated into pipelined designs. We present a high-level synthesis implementation of SMI targeting a dedicated FPGA interconnect, exposing runtime-configurable routing with support for arbitrary network topologies, and implement a set of distributed memory benchmarks. Using SMI, programmers can implement distributed, scalable HPC programs on reconfigurable hardware, without deviating from best practices for hardware design.

Best Paper Finalist: no
Best Student Paper Finalist: no

**High Performance Monte Carlo Simulation of Ising Model on TPU Clusters**
Kun Yang (Google LLC), Yi-Fan Chen (Google LLC), Georgios Roumpos (Google LLC), Chris Colby (Google LLC), John Anderson (Google LLC)
Large-scale deep learning benefits from an emerging class of AI accelerators. Some of these accelerators’ designs are general enough for compute-intensive applications beyond AI and Cloud TPU is one such example. In this paper, we demonstrate a novel approach using TensorFlow on Cloud TPU to simulate the two-dimensional Ising Model. TensorFlow and Cloud TPU framework enable the simple and readable code to express the complicated distributed algorithm without compromising the performance. Our code implementation fits into a small Jupyter Notebook and fully utilizes Cloud TPU’s efficient matrix operation and dedicated high speed inter-chip connection. The performance is highly competitive: it outperforms the best published benchmarks to our knowledge by 60% in single-core and 250% in multi-core with good linear scaling. When compared to Tesla V100 GPU, the single-core performance maintains a ~10% gain. We also demonstrate that using low precision arithmetic---bfloat16---does not compromise the correctness of the simulation results.

Best Paper Finalist: no
Best Student Paper Finalist: no
**Posters**

**(back to top)**

**Tuesday, November 19**

8:30 am - 5:00 pm

**ACM Student Research Competition Posters Display**

**Poster 12: Investigating Input Sensitivity in GPU Graph Analytics with Machine Learning**

*Jacob M. Hope (Texas State University), Trisha Nag (Texas Tech University)*

Graph algorithms are at the core of data-intensive applications. As such, efficient graph processing is of great importance. Irregularity in real-world graphs can make performance unpredictable and non-portable across different inputs and architectures. Given a type of graph, the same optimized implementation of an algorithm can produce performance numbers that differ by orders-of-magnitude. We conduct extensive analysis on a set of 1238 graphs to identify input-sensitive performance inefficiencies, including two that have not been studied: (i) register pressure and (ii) CPU-GPU data movement via demand paging. We then build a multiclass decision tree classifier that characterizes the irregular properties of graphs from our data and maps them to an optimal control parameterization at the compiler, system and algorithmic layers, that yield the highest overall algorithmic performance. We then integrate the classifier into a system where it will process a new graph and generate a kernel on the predicted optimal configuration.

**Poster 13: Scalable Load-Balancing Concurrent Queues on Many-Core Architectures**

*Caleb Lehman (Ohio State University)*

As the core counts of computing platforms continue to rise, parallel runtime systems with support for very fine-grained tasks become increasingly necessary to fully utilize the available resources. A critical feature of such task-based parallel runtime systems is the ability to balance work evenly and quickly across available cores. We highlight this by studying XTask, a custom parallel runtime system based on XQueue, which is a novel lock-less concurrent queuing system with relaxed ordering semantics that is geared to realizing scalability to hundreds of concurrent threads. We demonstrate the lack of adequate load balancing in the original XQueue design and present several
solutions for improving load balancing. We also evaluate the corresponding improvements in performance on two sample workloads, computation of Fibonacci numbers and computation of Cholesky factorization. Finally, we compare the performance of several versions of XTask along with several implementations of the popular OpenMP runtime system.

**Poster 6: Analyzing the Performance of ZFP Compressed Arrays on HPC Kernels**  
Pavlo D. Triantafyllides (Clemson University)

Per-node memory capacity limits the maximal problem size of HPC applications. Naive data compression alleviates the memory impact, but requires full decompression before the data is accessed. ZFP compressed arrays reduce the memory footprint, by independently compressing data in fixed sized blocks. Thus, decompressing individual blocks and caching them enables random access and a reduction in decompressions on the critical path. The performance of ZFP compressed arrays is dependent on several key variables: software cache size, cache policy, and compression rate. In this poster, we explore the sensitivity of these ZFP parameters on runtime performance for the matrix-matrix multiplication algorithm. Results show that selection of cache size, policy, and rate yields 8% performance improvement over the default ZFP configuration.

**Poster 1: Early Experiences on OpenPOWER Architecture: Analysis of Billion-Scale Atomistic Datasets**  
Yuya Kawakami (Grinnell College, Oak Ridge National Laboratory)

We present performance analysis on OpenPOWER architecture of an algorithm to generate transversal views of atomistic models. The algorithm was implemented with data parallel primitives in NVIDIA Thrust for architecture portability. We report performance results on IBM Power9 CPUs (OpenMP, Intel Threading Blocks) and NVIDIA Volta GPUs (single and multi GPU). We also evaluate CUDA unified memory performance, exposed by NVIDIA RAPIDS Memory Manager library (RMM).

**Poster 2: An Efficient Parallel Algorithm for Dominator Detection**  
Daniel Giger (University of Massachusetts, Lowell), Hang Liu (Stevens Institute of Technology)

In graph theory, a vertex v dominates a vertex u if every path from the entry vertex to u must go through vertex v. This algorithm is called dominator detection and holds a wide range of applications, such as compiler design, circuit testing, and social network analysis. While the performance of many other graph algorithms soars with respect to the increase of the hardware parallelism, dominator detection algorithm experiences very little advancement due to the hardship
of parallelism. This work thus introduces an efficient parallel dominator detection algorithm that is inspired by Breadth-First Search (BFS), which bests SEMI-NCA on large graphs.

**Poster 9: Machine Specific Symbolic Code Generation**
Robert King (University of Utah)

New discoveries in science and engineering are primarily driven by numerical simulations of underlying governing equations. Manually written codes to evaluate the main computational kernels lack portability, prone to human errors, ability to perform code optimizations due to the complexity of the underlying equations. In this work we present a symbolic code generation framework, which generates architecture optimized code for different platforms. As the driving application, we primarily use computational relativity where computations of Einstein equations become complicated due to the presence of curvature in spacetime.

The presented framework is based on SymPy with additional modules written to handle complicated partial differential equations (PDEs). The symbolically written equations are converted to a computational graph, which enables to perform architecture (cache, register optimizations) and language specific (SIMD vectorization, CUDA) optimizations. By computing common subexpressions, we reduce the number of compute operations needed, by storing them in temporary variables.

**Poster 5: Evaluating Lossy Compressors for Inline Compression**
Donald T. Elmore (Clemson University)

HPC applications require massive amounts of data. The data required is growing faster than memory capabilities. An example of this is pySDC, a framework for solving collocation problems iteratively using parallel-in-time methods. pySDC requires storing and exchange of 3D volume data for each parallel point in time. We evaluate several state-of-the-art lossy compressors such as SZ and ZFP for their applicability to inline compression for pySDC. We evaluate the compressors based on compression ratio, compression bandwidth, decompression bandwidth, and overall error introduced.

This poster evaluates state-of-the-art lossy compressors for use in pySDC; shows lossy data compression is an effective tool for reducing memory requirements for pySDC; and highlights current compression/decompression bandwidth is not fast enough for inline lossy compression yet. Results show using SZ with an error bound of 1e-5, we reduce the memory footprint by a factor of 311.99 while maintaining an acceptable level of loss.
Poster 3: A Deep Learning Approach to Noise Prediction and Circuit Optimization for Near-Term Quantum Devices  
Alexander Zlokapa (California Institute of Technology), Alexandru Gheorghiu (California Institute of Technology)  

Noisy intermediate-scale quantum (NISQ) devices face challenges in achieving high-fidelity computations due to hardware-specific noise. As a basis for noise mitigation, we develop a convolutional neural network noise model to estimate the difference in noise between a given pair of equivalent quantum circuits. On a classically simulated dataset of 1.6 million pairs of quantum circuits with a simplified noise model calibrated to IBM Q hardware, the deep learning approach shows a significant improvement in noise prediction over linear gate count models. A greedy peephole optimization procedure is proposed to minimize noise using the deep learning model as an objective function, showing further improvement in noise mitigation compared to commonly used gate count minimization heuristics.

Poster 10: Learning Page Access Patterns for Algorithms Programmed with GPU UVM  
Bennett Cooper (Clemson University), Derek Rodriguez (Clemson University)  

We aim to mitigate the performance bottleneck of migrating data between host and device memory in GPU applications by accurately predicting application access patterns using deep neural networks. We model the memory access pattern of any given application by collecting page faults that trigger data migration to the GPU and feed this time series as input to a neural network that outputs the next several page faults. We evaluate these predictions on the basis of what makes a useful prefetch in our context for the GPU. Current work has looked at trivial GPU applications, such as matrix operations, and moving toward real, complex applications. Our work will be presented by bringing attention to the predictive capability of our neural network on the current applications tested.

Poster 8: Mitigating Communication Bottlenecks in MPI-Based Distributed Learning  
Abdullah B. Nauman (Ward Melville High School, Brookhaven National Laboratory)  

Current commercial and scientific facilities generate and maintain vast amounts of complex data. While machine learning (ML) techniques can provide crucial insight, developing these models is often impractical on a single process. Distributed learning techniques mitigate this problem; however, current models contain significant performance bottlenecks. Here, we conduct a detailed
performance analysis of MPI_Learn, a widespread distributed ML framework for high-energy physics (HEP) applications, on the Summit supercomputer, by training a network to classify simulated collision events from high-energy particle detectors at the CERN Large Hadron Collider (LHC).

We conclude that these bottlenecks occur as a result of increasing communication time between the different processes, and to mitigate the bottlenecks we propose the implementation of a new distributed algorithm for stochastic gradient descent (SGD). We provide a proof of concept by demonstrating better scalability with results on 250 GPUs, and with hyperparameter optimization, show a ten-fold decrease in training time.

**Hearing Single- and Multi-Threaded Program Behavior**
*Mark Wissink (Calvin University), Joel Adams (Calvin University)*

There are many examples of visual program representation, but there is an unexplored area of sonic representation—audialization—of programs. The Thread Safe Audio Library (TSAL) provides the necessary tools to create such program audializations. TSAL is a platform-independent, object-oriented C++ library that provides thread safe classes for synthesizing sound from a program. By adding TSAL calls to a sequential or parallel program, its behavior can be represented with sound. Audializations require a unique interpretation that may greatly benefit auditory learners and other programmers who are looking for a different way to profile a sequential or parallel program.

**Poster 11: Walking the Cost-Accuracy Tightrope: Balancing Trade-Offs in Data-Intensive Genomics**
*Kathryn Leung (Princeton University), Meghan Kimball (DePaul University)*

Scientific applications often exhibit a trade-off between cost and accuracy. However, measuring and predicting cost and accuracy in a way that users can understand these trade-offs is challenging. To address these needs, we present predictive cost and accuracy models for data-intensive genomics applications. We use these models to create a trade-off graph, which researchers can use to selectively trade-off cost and accuracy.

**Poster 31: Designing High-Performance Erasure Coding Schemes for Next-Generation Storage Systems**
*Haiyang Shi (Ohio State University), Xiaoyi Lu (Ohio State University)*
Erasure Coding (EC) NIC offload is a promising technology for designing next-generation distributed storage systems. To alleviate the EC overhead, various hardware-based Erasure Coding schemes have been proposed to leverage the advanced compute capabilities on modern HPC clusters. However, this poster has identified some critical limitations of current-generation EC NIC offload schemes on modern SmartNICs and the missing of exploiting heterogeneous hardware available on modern HPC clusters in parallel. To address these limitations, this poster first proposes a unified benchmark suite to benchmark, measure and characterize hardware-optimized erasure coders. Then, it proposes a Multi-Rail EC concept which enables upper-layer applications to leverage heterogeneous hardware to perform EC operations simultaneously. Finally, it proposes a new EC NIC offload paradigm based on the tripartite graph model, namely TriEC. TriEC supports both encode-and-send and receive-and-decode offload primitives efficiently.

Poster 27: A Framework for Measuring Hardware Gather-Scatter Support
Patrick Lavin (Georgia Institute of Technology), Jeffrey Young (Georgia Institute of Technology), Richard Vuduc (Georgia Institute of Technology)

In recent years, we have seen the re-addition of vector units to CPUs. While these units easily give speedups for easily vectorized applications with dense memory access, it can be hard to characterize how different access patterns will affect the performance of vectorized code.

We have developed Spatter, a benchmark which allows us to test and investigate the gather-scatter units available on current and upcoming hardware. The information that Spatter reveals to users is of use to everyone from hardware vendors who wish to compare gather-scatter units across platforms, to application developers who wish to test memory access pattern performance in their vectorized code.

Poster 30: A Fast and Efficient Incremental Approach toward Dynamic Community Detection
Neda Zarayeneh (Washington State University)

Community detection is a discovery tool to identify groups of vertices which are tightly connected, called communities. Most of the literature on this subject caters to the static use-case where the underlying network does not change. However, many emerging real-world use-cases give rise to a need to incorporate dynamic graphs as inputs.

We present a fast and efficient incremental approach toward dynamic community detection. The key contribution is a generic technique called Ω-screening, which examines the most recent batch of changes made to an input graph and selects a subset of vertices to reevaluate for potential
community (re)assignment. This technique can be incorporated into any of the community detection methods that use modularity as its objective function for clustering. For demonstration purposes, we incorporated the technique into two well-known community detection tools. Our experiments demonstrate that our approach is able to generate performance speedups without compromising on the output quality.

Poster 29: Optimal Routing for a Family of Scalable Interconnection Networks
Zhipeng Xu (Stony Brook University; Sun Yat-sen University, Guangzhou, China)

We propose a scheme to construct a family of large and high-performance interconnection networks that are scalable, low-radix, minimum diameters. These networks, whose diameters grow linearly as their sizes grow exponentially, are generated by using the Cartesian products of smaller optimal networks of minimum diameters. For the smaller base networks, we design the vertex-balanced routing algorithm by considering the forwarding pressure at each vertex. Comparative benchmarks on a Beowulf cluster show significant improvement in performance after using the new routing algorithm. Each node of the new network generated from base graphs with low-diameter can also sustain balanced forwarding loadings if we apply optimal routing algorithms to the base network. Simulation results for larger networks show that the optimal routing algorithms achieve the gain of communication performance.

Poster 16: perf-taint: Taint Analysis for Automatic Many-Parameter Performance Modeling
Marcin Copik (ETH Zurich)

Performance modeling is a well-known technique for understanding the scaling behavior of an application. Although the modeling process is today often automatic, it still relies on a domain expert selecting program parameters and deciding relevant sampling intervals. Since existing empirical methods attempt blackbox modeling, the decision on which parameters influence a selected part of the program is based on measured data, making empirical modeling sensitive to human errors and instrumentation noise. We introduce a hybrid analysis to mitigate the current limitations of empirical modeling, combining the confidence of static analysis with the ability of dynamic taint analysis to capture the effects of control-flow and memory operations. We construct models of computation and communication volumes that help the modeler to remove effects of noise and improve the correctness of estimated models. Our automatic analysis prunes irrelevant program parameters and brings an understanding of parameter dependencies which helps in designing the experiment.

Poster 24: Fingerprinting Anomalous Computation with RNN for GPU-Accelerated HPC
Machines
Pengfei Zou (Clemson University)

This paper presents a workload classification framework that discriminates illicit computation from authorized workloads on GPU-accelerated HPC systems. As such systems become more and more powerful, they are exploited by attackers to run malicious and for-profit programs that typically require extremely high computing ability to be successful. Our classification framework leverages the distinctive signatures between illicit and authorized workloads, and explore machine learning methods to learn the workloads and classify them. The framework uses lightweight, non-intrusive workload profiling to collect model input data, and explores multiple machine learning methods, particularly recurrent neural network (RNN) that is suitable for online anomalous workload detection. Evaluation results on four generations of GPU machines demonstrate that the workload classification framework can tell apart the illicit authorized workloads with high accuracy of over 95%. The collected dataset, detection framework, and neural network models will be made available on GitHub.

Poster 33: Enhancing Neural Architecture Search with Speciation and Inter-Epoch Crossover
Matthew E. Baughman (University of Chicago)

As deep learning continues to expand into new areas of application, the demand for efficient use of our HPC resources increase. For new problem domains, new model architectures are developed through a neural architecture search (NAS), which consist of iteratively training many neural networks. To combat the computational waste and maximize compute efficiency for NAS, we demonstrate that the use of genetic algorithms with speciation can be used to both shorten training time and increase accuracy at each iteration.

Poster 14: Data Reuse Analysis for GPU Offloading Using OpenMP
Alok Mishra (Stony Brook University, Lawrence Livermore National Laboratory)

More researchers and developers desire to port their applications to GPU-based clusters, due to their abundant parallelism and energy efficiency. Unfortunately porting or writing an application for accelerators, such as GPUs, requires extensive knowledge of the underlying architectures, the application/algorithm and the interfacing programming model (e.g. OpenMP). Often applications spend a significant portion of their execution time on data transfer. Exploiting data reuse opportunities in an application can reduce its overall execution time. In this research, we present an approach to automatically recognize data reuse opportunities in an application which uses OpenMP for exploiting GPU parallelism, and consequently insert pertinent code to take advantage of data reuse on GPU. Using our approach we were able to retain reused data on the GPU and reduce the overall execution
Poster 17: Exploiting Multi-Resource Scheduling for HPC
Yuping Fan (Illinois Institute of Technology)

High-performance computing (HPC) is undergoing significant changes. Next generation HPC systems are equipped with diverse global/local resources. HPC job scheduler plays a crucial role in efficient use of resources. However, traditional job schedulers are single-objective and fail to efficient use of other resources. In our previous work, we present a job scheduling framework named BBSched to schedule CPUs and burst buffers. As we are heading toward exascale computing, a variety of heterogeneous resources are deployed in HPC systems. In this poster, we extend BBSched for managing multiple resources beyond CPUs and burst buffers. We formulate multi-resource scheduling as a general multi-objective optimization (MOO) problem, present a heuristic method to solve the NP-hard MOO problem, and provide a preliminary evaluation for scheduling up to ten resources. The proposed multi-resource scheduling design is intended to enhance Cobalt, a production job scheduler deployed on HPC systems at Argonne Leadership Computing Facility (ALCF).

Poster 15: Cost-Aware Cloudlet Placement in Edge Computing Systems
Dixit Bhatta (University of Delaware)

A well-known challenge in Edge Computing is strategic placement of cloudlets. The fundamental goals of this challenge are to minimize the cloudlet deployment cost and to guarantee minimum latency to the users of edge services. We address this challenge by designing a cost-aware cloudlet placement approach that that ensures user latency requirements while covering all devices in the service region. We first mathematically formulate the problem as a multi-objective integer programming model in a general deployment scenario, which is computationally NP-hard. We then propose a genetic algorithm-based approach, GACP, to find heuristic solutions in significantly reduced time. We investigate the effectiveness of GACP by performing extensive experiments on multiple deployment scenarios based on New York City OpenData. The results presented in the poster show that our approach obtains close to optimal cost solutions with significant time reduction.

Poster 19: Accelerating Real-World Stencil Computations Using Temporal Blocking: Handling Sparse Sources and Receivers
George Bisbas (Imperial College, London)
This paper concerns performance optimization in finite-difference solvers found in seismic imaging. We tackle a complicating factor that is arising in seismic inversion problems: the addition of waves injected from sources distributed sparsely over the 2D/3D domain, and the need for receivers that interpolate data measurements at a set of points, again distributed sparsely across the domain. In this work we show how to overcome this limitation. We introduce an algorithm for a loop nest transformation policy applied to wave propagation in order to improve data locality and optimize our cache memory use. Our algorithm uses an inspector/executor scheme capable of inspecting measurement operations at sparse locations and then compute the field updates. The ultimate goal of this ongoing work is to automate this scheme for stencil codes generated by Devito.

Poster 23: PERQ: Fair and Efficient Power Management of Power-Constrained Large-Scale Computing Systems
Tirthak Patel (Northeastern University), Devesh Tiwari (Northeastern University)

Large-scale computing systems are becoming increasingly more power-constrained, but these systems employ hardware over-provisioning to achieve higher system throughput because HPC applications often do not consume the peak power capacity of nodes. Unfortunately, focusing on system throughput alone can lead to severe unfairness among multiple concurrently-running applications. This paper introduces PERQ, a new feedback-based principled approach to improve system throughput while achieving fairness among concurrent applications.

Poster 18: Using Non Volatile Memories to Build Energy- and Cost-Efficient Clusters
Onkar Patil (North Carolina State University)

Non-Volatile Memory (NVM) is a byte-addressable, high capacity, high latency, and persistent form of memory that can extend the primary memory hierarchy by another level. It allows clusters to have significantly greater memory capacity per node. Intel’s Optane DC Persistent Memory Module (PMM) is a NVM device that can be used to increase the memory density of high performance computing (HPC) system. With higher memory density, we can run scientific computing applications with larger problem sizes on fewer compute nodes than on current HPC systems. Smaller HPC clusters will reduce the cost of running scientific simulations. We test our hypothesis by running a HPC application with large problem sizes on a single node with NVM and on multiple nodes connected by a high speed interconnect with equivalent amount of volatile memory. We compare the performance characteristics and power consumption of both systems.

Poster 22: Fast Profiling-Based Performance Modeling of Distributed GPU Applications
An increasing number of applications utilize GPUs to accelerate computation, with MPI responsible for communication in distributed environments. Existing performance models only focus on either modeling GPU kernels or MPI communication; few that do model the entire application are often too specialized for a single application and require extensive input from the programmer.

To be able to quickly model different types of distributed GPU applications, we propose a profiling-based methodology for creating performance models. We build upon the roofline performance model for GPU kernels and analytical models for MPI communication, with a significant reduction in profiling time. We also develop a benchmark to model 3D halo exchange that occurs in many scientific applications. Our proposed model for the main iteration loops of MiniFE achieves 6-7% prediction error on LLNL Lassen and 1-2% error on PSC Bridges, with minimal code inspection required to model MPI communication.

**Poster 26: Neural Networks for the Benchmarking of Detection Algorithms**

Silvia Miramontes (Lawrence Berkeley National Laboratory; University of California, Berkeley)

There are several automated methods to detect objects from grayscale images. However, materials scientists still lack basic tools to compare different detection results, particularly when working with microtomography. This poster introduces FibCAM, a convolutional neural network (CNN)-based method using TensorFlow that allows benchmarking fiber detection algorithms. Our contribution is three-fold: (a) the design of a computational framework to compare automated fiber detection models with curated datasets through classification; (b) lossless data reduction by embedding prior knowledge into data-driven models; (c) a scheme to decompose computation into embarrassingly parallel processes for future analysis at scale. Our results show how FibCAM classifies different structures, and how it illustrates the material's composition and frequency distribution of microstructures for improved interpretability of machine learning models. The proposed algorithms support probing the specimen content from gigabyte-sized volumes and enable pinpointing inconsistencies between real structures known a priori and results derived from automated detections.

**Poster 21: Optimization of GPU Kernels for Sparse Matrix Computations in Hypre**

Chaoyu Zhang (Arkansas State University)

The acceleration of sparse matrix computations on GPUs can significantly enhance the performance of iterative methods for solving linear systems. In this work, we consider the kernels of Sparse Matrix...
Vector Multiplications (SpMV), Sparse Triangular Matrix Solves (SpTrSv) and Sparse Matrix Matrix Multiplications (SpMM), which are often demanded by Algebraic Multigrid (AMG) solvers. With the CUDA and the hardware support of the Volta GPUs on Sierra, the existing kernels should be further optimized to fully take the advantage of the new hardware, and the optimizations have shown significant performance improvement. The presented kernels have been put in HYPRE for solving large scale linear systems on HPC equipped with GPUs. These shared-memory kernels for single GPU are the building blocks of distributed matrix operations required by the solver across multiple GPUs and compute nodes. The implementations of these kernels in Hypre and the code optimizations will be discussed.

**Poster 20: From IoT to Cloud: Research Platform for IoT/Cloud Experiments**

*Jinfeng Lin (University of Notre Dame)*

IoT studies leverage a wide range of lightweight hardware for collecting and processing data in the field. Limited by the resources on the devices, IoT systems have to interoperate with cloud platforms for addressing computation intensive tasks such as image processing, application backend supporting and centralized data storage. Therefore, a testbed for IoT/Cloud experiments should provide infrastructure for IoT to cloud communication, computation deployment, and hardware resource management. With these functionalities, users can focus on research problems without distraction from manually constructing experiment environments. Though cloud providers such as Google, Amazon, and Microsoft all provide IoT to Cloud solutions in general, this commercial model is not entirely compatible with research purpose. We propose a framework named Chameleon IoT testbed (CHIoT) that extends the Chameleon bare-metal cloud resources to support general IoT to Cloud experiments.

**Poster 32: OSU INAM: A Profiling and Visualization Tool for Scalable and In-Depth Analysis of High-Performance GPU-Enabled HPC Clusters**

*Pouya Kousha (Ohio State University)*

The lack of low-overhead and scalable monitoring tools have prevented a comprehensive study of efficiency and utilization of emerging NVLink-enabled GPU clusters. We address this by proposing and designing an in-depth, real-time analysis, profiling, and visualization tool for high-performance GPU-enabled clusters with NVLinks on the top of the OSU INAM. The proposed tool is capable of presenting a unified and holistic view of MPI-level and fabric level information for emerging NVLink-enabled high-performance GPU clusters. It also provides insights into the efficiency and utilization of underlying interconnects for different communication patterns. We also designed a low overhead and scalable modules to discover the fabric topology and gather fabric metrics by using different
levels of threading, bulk insertions and deletions for storage, and using parallel components for fabric discovery and port metric inquiry.

Zheng Miao (Clemson University)

High-Performance Computing systems must simultaneously address both resilience and power. In heterogeneous systems, the trade-offs between resilience and energy-efficiency are more complex for applications using both CPUs and GPUs. A deep understanding of the interplay among energy efficiency, resilience, and performance is required for heterogeneous systems to address them simultaneously.

In this work, we present a new framework for resilient and energy-efficient computing in GPU-accelerated systems. This framework supports partial or full redundancy and checkpointing for resilience, and provides users with flexible hardware resource selection, adjustable precision and power management to improve performance and energy-efficiency. We further perform CUDA-aware MPI to reduce resilience overhead, mainly in message communication between GPUs. Using CG as an example, we show that our framework provides about 40% time and 45% energy savings, comparing to simple extension of RedMPI, a redundancy based resilience framework for homogeneous CPU systems.

**Poster 25: Leveraging Smart Data Transfer and Metadirective in Adaptive Computing**
Anjia Wang (University of North Carolina, Charlotte; Lawrence Livermore National Laboratory)

In this work, we propose smart data transfer (SDT) along with extensions to metadirective and map constructs in OpenMP 5.0 to improve adaptive computing. The Smith-Waterman algorithm is used as an example, whose naïve implementation does not conduct data transfer efficiently. SDT is used to solve this issue with the following advantages: (1) SDT only transfers necessary data to GPU instead of the whole data set, resulting in 4.5x of speedup in our initial experiments. (2) Depending on computing vs. data transfer requirements of a program, SDT will transfer the output of each iteration from GPU to host either immediately or all together after the last GPU kernel call. (3) It supports large data exceeding GPU device memory’s size via data tiling. We propose to extend metadirective’s context selector to obtain similar improvement by enabling target enter/exit data and on-demand data access.
8:30 am - 5:00 pm

Doctoral Showcase Posters Display

**Poster 36: Modeling Non-Determinism in HPC Applications**
Dylan Chapp *(University of Delaware, University of Tennessee)*

As HPC applications migrate from the petascale systems of today to the exascale systems of tomorrow, the increasing need to embrace asynchronous, irregular, and dynamic communication patterns will lead to a corresponding decrease in application-level determinism. Two critical challenges emerge from this trend. First, unchecked non-determinism coupled with the non-associativity of floating-point arithmetic undermines numerical reproducibility of scientific applications. Second, the prevalence of non-determinism amplifies the cost of debugging, both in terms of computing resources and human effort. In this thesis, we present a modeling methodology to quantify and characterize communication non-determinism in parallel applications. Our methodology consists of three core components. First, we build graph-structured models of relevant communication events from execution traces. Second, we apply similarity metrics based on graph kernels to quantify run-to-run variability and thus identify the regions of executions where non-determinism manifests most prominently. Third, we leverage our notion of execution similarity to characterize applications via clustering, anomaly detection, and extraction of representative patterns of non-deterministic communication which we dub "non-determinism motifs". Our work will amplify the effectiveness of software tools that target mitigation or control of application-level non-determinism (e.g., record-and-replay tools) by providing them with a common metric for quantifying communication non-determinism in parallel applications and a common language for describing it.

**Poster 35: Scaling Up Pipeline Optimization with High Performance Computing**
Robert Lim *(University of Oregon)*

My research focuses on developing a pipeline optimization infrastructure that automates the design and code generation of neural networks through the use of high-performance computing. The problem has the following objectives: unify automated machine learning (AutoML) and compilation, archive profiles for creation of a knowledge base for a data-driven approach toward search, explore various search optimizations for model design and code generation. The field of automated deep learning includes hyperparameter optimization and neural architecture search (NAS), which requires domain expertise in designing a model, in addition to the tuning parameters related to learning and the model itself. The search space is complex and deciding which parameters factor into the overall
accuracy of a model is a non-trivial task. Once a model is trained, the next step compiles the model, which maps to the backend of a targeted architecture, whether GPU, embedded mobile phones, or FPGA. The compilation phase also involves choices, in terms of optimizations applied, which can include a set of flags, or direct code transformations. Various efficient search techniques are explored, archiving results along the way for facilitating in code generation options, which could possibly inform the model design process. Our previous work reduced the search space for code generation of various domain kernels by 92%, and this work investigates whether the same approach can be applied in all aspects of AutoML design and code generation.

**Poster 38: High-Performance Backpropagation in Scientific Computing**

Navjot Kukreja (Imperial College, London)

Devito is a domain-specific language for the automatic generation of high-performance solvers for finite difference equations provided in a high-level symbolic representation. The primary use of Devito is to enable Full-waveform inversion, which is an adjoint-based optimization problem that uses the wave equation as part of its objective function. This doctoral project consists of three elements of Devito that are key to solving adjoint-based optimization problems (i.e. the back propagation). The first is automatic differentiation of stencil loops. This allows the automatic generation of high-performance code implementing the derivative of any provided function. This is essential to enable the use of a wider range of physical equations - in order to use better physical approximations. An essential feature of the generated derivatives is that they show scaling and performance behaviour that is very similar to the forward function. A common issue in solving adjoint-based optimization problems is the prohibitively high memory requirement. The second contribution is a runtime for automatic execution of checkpoint-recompute schedules (called pyRevolve) to alleviate this memory requirement by trading it off for re-computations. These schedules may involve storing some intermediate states on disk, some in memory, and recomputing others. The third contribution is the use of lossy compression (ZFP/SZ), which is a second means of trading off memory and compute, automatically as part of checkpointing in pyRevolve. This is the first time it has been shown that lossy compression combined with checkpoint-recomputation can provide a much better error-memory tradeoff than using checkpoint-recomputation or lossy compression alone.

**Poster 37: Large Scale Ultrasound Simulations on Accelerated Clusters**

Filip Vaverka (Brno University of Technology)

Many emerging non-invasive medical procedures such as high intensity focused ultrasound treatment (HIFU), photoacoustic imaging, or transcranial ultrasound therapy require ultrasound
wave propagation simulations. Typically soft tissue can be modeled as fluid, weakly heterogeneous medium with frequency dependent absorption and non-linear effects at high ultrasound intensities. The k-Wave acoustic toolbox, widely used in medical community, uses highly efficient k-space pseudo-spectral time-domain (KSTD) discretization of acoustic equations to solve these problems. However, it is difficult to efficiently map the KSTD method onto modern cluster architectures with accelerators such as GPUs. The present thesis analyzes shortcomings of the KSTD method in respect to modern clusters and proposes local Fourier basis approach to improve scaling of the method. It is shown that the proposed method is able to achieve 5x speedup, while having sufficient accuracy for these medical applications. Behavior of the method is analyzed across variety of GPU and MIC (Intel Xeon Phi) accelerated clusters and results are presented.

Poster 34: Analysis of Automata Processing Acceleration on Disparate Hardware Technologies
Marziyeh Nourian (North Carolina State University)

Pattern matching is a computation that maps naturally onto finite automata (FA) abstractions. There has been a substantial amount of work on accelerating FA processing on various parallel platforms. However, the advantages and disadvantages of different automata processing accelerators and the innovation space in this area are still unclear. We target this problem and propose a compiler toolchain that automates the deployment of non-deterministic finite automata (NFAs) onto different target platforms. Using this toolchain, we perform an apples-to-apples comparison between AP, GPU- and FPGA-based NFA accelerator designs on large-scale datasets. Specifically, we observe that memory-based designs are limited by memory size and bandwidth. To address this issue, we target fixed-topology NFAs and propose a memory-efficient design that embeds the automata topology in code and stores only the transition symbols in memory. Our solution is suitable for SIMD architectures and is called SIMD_NFA. We design a compiler that automates the deployment of this design on SIMD platforms. We showcase our compiler framework on GPU and Intel platforms. Additionally, we observe that for NFAs with a grid-like fixed-topology (e.g., NFAs for Levenshtein and Hamming distance-based matching), transitions do not need to be encoded within the traversal code but can be inferred from the reference string to be matched and the knowledge of the NFA topology. Lastly, SIMD_NFA is a good fit for FPGA deployment using OpenCL-to-FPGA toolchains. We investigate the deployment of the OpenCL version of SIMD_NFA, on FPGA and explore a set of optimizations techniques to retarget SIMD_NFA to FPGA.

Poster 42: Power and Thermal Management Runtimes for HPC Applications in the Era of Exascale Computing
Daniele Cesarini (University of Bologna, CINECA)
In the scope of technical and scientific computing, the rush toward larger simulations has been so far assisted by a steady downsizing of micro-processing units, which has allowed to increase the compute capacity of general-purpose architectures at constant power. As side effects of the end of Dennard’s scaling, this process is now hitting its ultimate power limits and is just about to come to an end. This implies an increase in the energy cost for computation, performance loss due to a design based on worst-case power consumption, and performance loss due to overheating, and thermal gradients. As result, thermal and power bound supercomputing machines show performance degradation and heterogeneity which limit the peak performance of the system. This doctoral showcase presents software strategies to tackle the main bottlenecks induced by power and thermal issues that affect next-generation supercomputers. To respond to the above challenges, my work shows that propagating workload requirements from application to the runtime and operating system levels is the key to provide efficiency. This is possible only if the proposed software methodologies cause little or no overhead in term of application performance. With this in mind in my work, I have designed application-aware node-level optimal thermal management algorithms and runtimes, lazy node-level power capping, and energy reduction runtime. The experimental results show a significant step forward with respect to the current state-of-the-art solutions in power and thermal control of HPC systems.

Poster 41: Co-Designing Communication Middleware and Deep Learning Frameworks for High-Performance DNN Training on HPC Systems
Ammar Ahmad Awan (Ohio State University)

Recent advances in Machine/Deep Learning techniques have triggered key success stories in many application domains like Computer Vision, Speech Comprehension and Recognition, and Natural Language Processing. Large-scale Deep Neural Networks (DNNs), that are at the core of state-of-the-art AI technologies, have been the primary drivers of this success. Training very complicated and large DNN architectures using a large number of training examples (data) is compute-intensive and can take from weeks to months to achieve state-of-the-art prediction capabilities. To achieve higher accuracy, making the DNN deeper is also a common strategy. These requirements have led to a simple but powerful approach called Data Parallelism to achieve shorter training times. This has resulted in various research studies and ML/DL software like TensorFlow and PyTorch as well as distributed-training middleware like Horovod. In addition, for DNNs that do not fit the GPU memory, new DNN workloads are emerging that we call Out-of-Core DNNs and different strategies (out-of-core training and model-parallelism) are needed to train them. Clearly, large-scale DNN training brings forward new requirements for computation runtimes like CUDA and communication middleware like the MVAPICH2 MPI library. In this thesis, we broadly explore three different strategies to train DNNs on modern CPU and GPU architectures: 1) Data Parallelism, 2) Model Parallelism, and 3) Out-of-Core Training. We address the key challenge: How to co-design
computation and communication in modern ML/DL frameworks with execution runtimes like CUDA and communication middleware like MVAPICH2 to enable scalable, high-performance, and efficient training of DNNs on large-scale HPC systems?

**Poster 40: Performance, Portability, and Productivity for Data-Parallel Computations on Multi- and Many-Core Architectures**  
Ari Rasch (University of Münster)

This thesis presents an approach to performance, portability, and productivity for data-parallel computations on multi- and many-core architectures, e.g., Intel CPU and NVIDIA GPU. We introduce the algebraic formalism of Multi-Dimensional Homomorphisms (MDHs) – a class of functions that cover important data-parallel computations, e.g., linear algebra routines (BLAS) and stencil computations. For our MDHs, we propose a Domain-Specific Language (DSL), based on patterns of parallelism (a.k.a. algorithmic skeletons), to enable conveniently expressing MDH functions. We introduce a code generation approach for our DSL to automatically generate for MDHs optimized program code targeting multi- and many-core architectures. Our code generation approach relies on OpenCL – an emerging de-facto standard for uniformly programming parallel architectures, such as CPU and GPU. A major feature of our generated code is that it is targeted to OpenCL’s abstract device models (rather than a particular architecture) by being parameterized in performance-critical parameters of these abstract models (e.g., the number of threads and size of tiles). With our code generation approach, we enable both high performance and performance portability: we fully automatically optimize our generated code -- for any given combination of an MDH function, architecture, and input size -- by automatically choosing (auto-tuning) optimized values of our code’s performance-critical parameters using our own Auto-Tuning Framework (ATF). Our experimental results on CPU and GPU demonstrate competitive and often significantly better performance of our MDH+ATF approach as compared to the currently best-performing competitors, e.g., Intel MKL/MKL-DNN, NVIDIA cuBLAS/cuDNN, and Facebook’s Tensor Comprehensions framework.

**Poster 45: Characterization and Modeling of Error Resilience in HPC Applications**  
Luanzheng Guo (University of California, Merced)

As high-performance computing systems scale in size and computational power, the occurrence of transient faults grows. Without protection by efficient and effective fault tolerance mechanisms, transient errors can cause incorrect execution outcome and even lead to a catastrophe in safety-critical HPC applications. Previous work attributes error resilience in HPC applications at a high-level to either the probabilistic or iterative nature of the application, whereas the community still
lacks the fundamental understanding of the program constructs that result in natural error resilience. We design FlipTracker, a framework to analytically track error propagation and to provide a fine-grained understanding of the propagation and tolerance of errors. After running FlipTracker on representative HPC applications, we summarize six resilience computation patterns that lead to nature error resilience in HPC applications. With a better understanding of natural resilience in HPC applications, we aim to model application resilience on data objects to transient faults. Many common application-level fault tolerance mechanisms focus on data objects. Understanding application resilience on data objects can be helpful to direct those mechanisms. The common practice to understand application resilience (random fault injection) gives us little knowledge of how and where errors are tolerated. Understanding "how" and "where" is necessary to understand how to apply application-level fault tolerance mechanisms effectively and efficiently. We design a practical model (MOARD) to measure application resilience on data objects by analytically quantifying error masking events happening to the data object. Using our model, users can compare application resilience on different data objects with different data types.

Poster 43: Efficient and Scalable Communication Middleware for Emerging Dense-GPU Clusters
Ching-Hsiang Chu (Ohio State University)

In the era of post Moore’s law, the traditional CPU is not able to keep the pace up and provide the computing power demanded by the modern compute-intensive and highly parallelizable applications. Under this context, various accelerator architectures such as general-purpose graphics processing unit (GPU), which equipped with the high-bandwidth memory (HBM) and massive parallelizable streaming multiprocessors, has been widely adopted in high-performance computing (HPC) and cloud systems to significantly accelerate numerous scientific and emerging machine/deep learning applications. Message Passing Interface (MPI), the standard programming model for parallel applications, has been widely used for GPU communication. However, the state-of-the-art MPI libraries are only optimizing GPU communication by leveraging advanced technology like Remote Direct Memory Access (RDMA) and not fully utilizing the computational power of GPUs. In this work, we propose GPU-enabled communication schemes to harness GPU computational resources, and cutting-edge interconnects such as NVIDIA NVLink for communication operations on the emerging heterogeneous systems. In this work, three primary MPI operations are addressed. First, intelligent communication scheduling, efficient packing/unpacking, and packing-free schemes are proposed to accelerate non-contiguous data transfer in scientific HPC applications. Second, scalable broadcast operations are presented to leverage the low-level hardware multicast feature to speed up GPU communication at scale. Finally, we also design topology-aware, link-efficient, and cooperative GPU kernels to significantly accelerate All-reduce operation, which is the primary performance bottleneck in deep learning applications. The proposed
designs demonstrate significant performance improvements over the state-of-the-art communication schemes for various HPC and deep learning applications.

Poster 39: Designing Next-Generation Communication Middlewares for Many-Core Architectures
Jahanzeb Maqbool Hashmi (Ohio State University)

Modern multi-petaflop HPC systems are powered by dense multi-/many-core architectures and this trend is expected to grow for the next-generation supercomputing systems. This rapid adoption of high core-density architectures by the current- and next-generation HPC systems is further fueled by the emerging application trends such as Deep Learning. This is putting more emphasis on middleware designers to optimize various communication protocols to meet the diverse needs of the applications. While the novelties in the processor architectures have led to the increased on-chip parallelism, they come at the cost of rendering traditional designs, employed by the communication runtimes such as MPI, to suffer from higher degree of intra-node communication latencies. Tackling the computation and communication challenges that accompany these dense multi-/many-cores garner special design considerations. The proposed work in this thesis tries to address the performance challenges posed by a diverse range of applications and the lacking support in state-of-the-art communication libraries such as MPI to exploit high-concurrency architectures. The author first proposes a "shared-address-spaces"-based communication substrate to derive intra-node communication in MPI. Atop this framework, the author has re-designed various MPI primitives such as point-to-point communication protocols (e.g., user-space zero-copy rendezvous transfer), collective communication (e.g., load/store based collectives, truly zero-copy and partitioning-based reduction algorithms), and efficient MPI derived datatypes processing (e.g., memoization-based "packing-free" communication) to exploit the potential of emerging multi-/many-core architectures and high throughput networks. The proposed designs have demonstrated significant improvement over state-of-the-art for various scientific and deep learning applications.

8:30 am - 5:00 pm
Research Posters Display

Poster 74: Enabling Code Portability of a Parallel and Distributed Smooth-Particle Hydrodynamics Application, FleCSPH
Suyash Tandon (University of Michigan), Nicholas Stegmeier (University of Illinois), Vasu Jaganath
Core-collapse supernovae (CCSNe) are integral to the formation and distribution of heavy elements across the universe. However, CCSNe are highly complex and inherently non-linear phenomena. Large-scale simulations of these cosmic events can provide us a glimpse of their hydrodynamic and nucleosynthetic processes which are difficult to observe. To enable these massive numerical simulations on high-performance computing (HPC) centers, this study uses FleCSPH, a parallel and distributed code, based on the smooth-particle hydrodynamics (SPH) formulation. In the recent years, the HPC architecture has evolved and the next generation of exascale computers are expected to feature heterogenous architecture. Therefore, it is important to maintain code portability across platforms. This work demonstrates code portability of FleCSPH through the incorporation of Kokkos C++ library and containers using Charliecloud.

**Poster 73: Accelerating Large-Scale GW Calculations on Hybrid CPU-GPU Architectures**

Mauro Del Ben (Lawrence Berkeley National Laboratory), Charlene Yang (National Energy Research Scientific Computing Center (NERSC)), Felipe Jornada (University of California, Berkeley; Lawrence Berkeley National Laboratory), Steven G. Louie (University of California, Berkeley; Lawrence Berkeley National Laboratory), Jack Deslippe (National Energy Research Scientific Computing Center (NERSC))

In this poster, we present the strategy, progress, and performance while GPU porting one of the major modules, epsilon, of the electronic structure code BerkeleyGW. Epsilon represents the most time-consuming routines in the BerkeleyGW workflow for large-scale material science simulations. Some of the porting/optimization strategies include, changing our original data layout to efficiently use libraries such as cuBLAS and cuFFT, implementation of specific CUDA kernels to minimize data copies between host/device and keeping data on device, efficient use of data streams to leverage high concurrency on the device, asynchronous memory copies and overlapping (MPI) communication on the host and computation on the device. Preliminary results are presented in terms of the speedup compared to the CPU-only implementation, strong/weak scaling, and power efficiency. Excellent acceleration is demonstrated: up to 30x for specific kernels. Our port also exhibits good scalability and about 16x higher FLOPs/watt efficiency compared to the CPU-only implementation.
Poster 89: BeeCWL: A CWL Compliant Workflow Management System
Betis Baheri (Kent State University), Steven Anaya (New Mexico Institute of Mining and Technology), Patricia Grubel (Los Alamos National Laboratory), Qiang Guan (Kent State University), Timothy Randles (Los Alamos National Laboratory)

Scientific workflows are used widely to carry out complex and hierarchical experiments. Although there are many trends to extend the functionality of workflow management systems to cover all possible requirements that may arise from a user community, one unified standard over cloud and HPC systems is still missing. In this paper, we propose a Common Workflow Language (CWL) compliant workflow management system. BeeCWL is a parser to derive meaningful information such as requirements, steps, relationships, etc. from CWL files and to create a graph database from those components. Generated graphs can be passed to an arbitrary scheduler and management system to decide whether there are enough resources to optimize and execute the workflow. Lastly, the user can have control over workflow execution, collecting logs, and restart or rerun some part of a complex workflow.

Poster 150: A Machine Learning Approach to Understanding HPC Application Performance Variation
Burak Aksar (Boston University, Sandia National Laboratories), Benjamin Schwaller (Sandia National Laboratories), Omar Aaziz (Sandia National Laboratories), Emre Ates (Boston University), Jim Brandt (Sandia National Laboratories), Ayse K. Coskun (Boston University), Manuel Egele (Boston University), Vitus Leung (Sandia National Laboratories)

Performance anomalies are difficult to detect because often a “healthy system” is vaguely defined, and the ground truth for how a system should be operating is evasive. As we move to exascale, however, detection of performance anomalies will become increasingly important with the increase in size and complexity of systems. There are very few accepted ways of detecting anomalies in the literature, and there are no published and labeled sets of anomalous HPC behavior. In this research, we develop a suite of applications that represent HPC workloads and use data from a lightweight metric collection service to train machine learning models to predict the future behavior of metrics. In the future, this work will be used to predict anomalous runs in compute nodes and determine some root causes of performance issues to help improve the efficiency of HPC system administrators and users.
Poster 81: Performance of Devito on HPC-Optimised ARM Processors
Hermes Senger (Federal University of São Carlos, Brazil; University of São Paulo), Jaime Freire de Souza (Federal University of São Carlos, Brazil), Edson Satoshi Gomi (University of São Paulo), Fabio Luporini (Imperial College, London), Gerard Gorman (Imperial College, London)

We evaluate the performance of Devito, a domain specific language (DSL) for finite differences on Arm ThunderX2 processors. Experiments with two common seismic computational kernels demonstrate that Devito can apply automatic code generation and optimization across Arm and Intel platforms. The code transformations include: parallelism, and SIMD vectorization (OpenMP >=4); loop tiling (with best block shape obtained via auto-tuning); domain-specific symbolic optimisation such as common sub-expression elimination and factorisation for Flop reduction, polynomial approximations for trigonometry terms, and heuristic hoisting of time-invariant expressions. Results show that Devito can achieve performance on Arm processors which is competitive to other Intel Xeon processors.

Poster 103: LIKWID 5: Lightweight Performance Tools
Thomas Gruber (Erlangen Regional Computing Center), Jan Eitzinger (Erlangen Regional Computing Center), Georg Hager (Erlangen Regional Computing Center), Gerhard Wellein (Erlangen Regional Computing Center)

LIKWID is a tool suite for performance oriented programmers with a worldwide user group. It is developed by the HPC group of the University Erlangen-Nuremberg since 2009 to support them in their daily research and performance engineering of user codes. The HPC landscape has become more and more diverse over the last years with clusters using non-x86 architectures and being equipped with accelerators. With the new major version, the architectural support of LIKWID is extended to ARM and POWER CPUs with the same functionality and features as for x86 architectures. Besides the CPU monitoring, the new version provides access the hardware counting facilities of Nvidia GPUs. This poster introduces the new features and shows the successes of applying LIKWID to identify performance bottlenecks and to test optimizations. Furthermore, the poster gives an overview of how users can integrate the LIKWID tools in their application using a lightweight add-once-and-reuse instrumentation API.
Poster 149: Solving Phase-Field Equations in Space-Time: Adaptive Space-Time Meshes and Stabilized Variational Formulations
Kumar Saurabh (Iowa State University), Biswajit Khara (Iowa State University), Milinda Fernando (University of Utah), Masado Ishii (University of Utah), Hari Sundar (University of Utah), Baskar Ganapathysubramaniam (Iowa State University)

We seek to efficiently solve a generalized class of partial differential equations called the phase-field equations. These non-linear PDE’s model phase transition (solidification, melting, phase-separation) phenomena which exhibit spatially and temporally localized regions of steep gradients. We consider time as an additional dimension and simultaneously solve for the unknown in large blocks of time (i.e. in space-time), instead of the standard approach of sequential time-stepping. We use variational multiscale (VMS) based finite element approach to solve the ensuing space-time equations. This allows us to (a) exploit parallelism not only in space but also in time, (b) gain high order accuracy in time, and (c) exploit adaptive refinement approaches to locally refine region of interest in both space and time. We illustrate this approach with several canonical problems including melting and solidification of complex snow flake structures.

Poster 84: ESTEE: A Simulation Toolkit for Distributed Workflow Execution
Vojtěch Cima (IT4Innovations, Czech Republic), Jakub Beránek (IT4Innovations, Czech Republic), Stanislav Böhm (IT4Innovations, Czech Republic)

Task graphs provide a simple way to describe scientific workflows (sets of tasks with dependencies) that can be executed on both HPC clusters and in the cloud. An important aspect of executing such graphs is the used scheduling algorithm. Many scheduling heuristics have been proposed in existing works; nevertheless, they are often tested in oversimplified environments. We introduce a simulation environment designed for prototyping and benchmarking task schedulers. Our simulation environment, scheduler source codes, and graph datasets are open in order to be fully reproducible. To demonstrate usage of Estee, as an example, we compare the performance of various workflow schedulers in an environment using two different network models.
Poster 93: Robust Data-Driven Power Simulator for Fast Cooling Control Optimization of a Large-Scale Computing System

Takashi Shiraishi (Fujitsu Laboratories Ltd), Hiroshi Endo (Fujitsu Laboratories Ltd), Takaaki Hineno (Fujitsu Laboratories Ltd), Hiroyuki Fukuda (Fujitsu Laboratories Ltd)

Power of large-scale systems such as an HPC or a datacenter is a significant issue. Cooling units consume 30% of the total power. General control policies for cooling units are local and static (manual overall optimization nearly once a week). However, free cooling and IT-load fluctuation may change hourly optimum control variables of the cooling units. In this work, we present a deep neural network (DNN) power simulator that can learn from actual operating logs and can quickly identify the optimum control variables. We demonstrated the power simulator of an actual large-scale system with 4.7-MW-power IT load. Our robust simulator predicted the total power with error of 4.8% without retraining during one year. We achieved optimization by the simulator within 80 seconds that was drastically faster than previous works. The dynamic control optimization each hour showed a 15% power reduction compared to that of conventional policy in the actual system.

Best Poster Finalist: no

Poster 115: sDNA: Software-Defined Network Accelerator Based on Optical Interconnection Architecture

En Shao (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences), Guangming Tan (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences), Zhan Wang (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences), Guojun Yuan (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences), Ninghui Sun (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences)

Software-Defined Network Accelerator (sDNA) is a new accelerated system for the exascale computer. Inspired by the edge forwarding index (EFI), the main contribution of our work is that it presents an extended EFI-based optical interconnection method with slow switching optical device. In our work, we found that sDNA based on extended EFI evaluation is not only able to offload the traffic from an electrical link to an optical link but is also able to avoid congestion inherent to electrical link.

Best Poster Finalist: no

Poster 49: WarpX: Toward Exascale Modeling of Plasma Particle Accelerators on GPU
Particle accelerators are a vital part of the DOE-supported infrastructure of discovery science and applications, but we need game-changing improvements in the size and cost for future accelerators. Plasma-based particle accelerators stand apart in their potential for these improvements. Turning this from a promising technology into mainstream scientific tools depends critically on high-performance, high-fidelity modeling of complex processes.

WarpX is an open-source particle-in-cell (PIC) code supported by the Exascale Computing Project (ECP) that is combining advanced algorithms with adaptive mesh refinement to allow challenging simulations of a multi-stage plasma-based TeV acceleration relevant for future high-energy physics discoveries. WarpX relies on the ECP co-design center for mesh refinement AMReX, and runs on CPU and GPU-accelerated computers. Production simulation have run on Cori KNL at NERSC and Summit at OLCF. In this poster, recent results and strategies on GPU will be presented, along with recent performance results.

Best Poster Finalist: no

Poster 50: Implementing an Adaptive Sparse Grid Discretization (ASGarD) for High Dimensional Advection-Diffusion Problems on Exascale Architectures

M. Graham Lopez (Oak Ridge National Laboratory), David L. Green (Oak Ridge National Laboratory), Lin Mu (University of Georgia), Ed D’Azevedo (Oak Ridge National Laboratory), Wael Elwasif (Oak Ridge National Laboratory), Tyler McDaniel (University of Tennessee), Timothy Younkin (University of Tennessee), Adam McDaniel (Oak Ridge National Laboratory), Diego Del-Castillo-Negrete (Oak Ridge National Laboratory)

Many scientific domains require the solution of high dimensional PDEs. Traditional grid- or mesh-based methods for solving such systems in a noise-free manner quickly become intractable due to the scaling of the degrees of freedom going as $O(N^d)$ sometimes called "the curse of
dimensionality." We are developing an arbitrarily high-order discontinuous-Galerkin finite-element solver that leverages an adaptive sparse-grid discretization whose degrees of freedom scale as \(O(N^{1+\log 2 N^{D-1}})\). This method and its subsequent reduction in the required resources is being applied to several PDEs including time-domain Maxwell’s equations (3D), the Vlasov equation (in up to 6D) and a Fokker-Planck-like problem in ongoing related efforts. Here we present our implementation which is designed to run on multiple accelerated architectures, including distributed systems. Our implementation takes advantage of a system matrix decomposed as the Kronecker product of many smaller matrices which is implemented as batched operations.

Best Poster Finalist: no

**Poster 91: FreeCompilerCamp: Online Training for Extending Compilers**
Alok Mishra (Stony Brook University, Lawrence Livermore National Laboratory), Anjia Wang (University of North Carolina, Charlotte; Lawrence Livermore National Laboratory), Chunhua Liao (Lawrence Livermore National Laboratory), Yonghong Yan (University of North Carolina, Charlotte), Barbara Chapman (Stony Brook University)

In this presentation, we introduce an ongoing effort of an online training platform aimed to automate the training of developers to quickly extend compilers - FreeCompilerCamp.org. Our free and open platform allows anyone who is interested in developing compilers to learn the necessary skills. A live training website, built on top of Play-With-Docker, is set up so that anyone with internet access and a web browser will be able to take this training. The entire training system is open-source and developers with relevant skills can contribute new tutorials and deploy it on a private server, workstation or even laptop. We have created some initial tutorials on how to extend the Clang/LLVM or ROSE compilers to support new OpenMP features. Using a web interface consisting of two side-by-side panels, users can follow the tutorials on one side and immediately practice what they learned in a terminal sandbox embedded on the other.

Best Poster Finalist: yes

**Poster 51: SmartK: Efficient, Scalable, and Winning Parallel MCTS**
Michael S. Davinroy (Swarthmore College), Shawn Pan (Swarthmore College), Bryce Wiedenbeck (Swarthmore College, Davidson College), Tia Newhall (Swarthmore College)

SmartK is our efficient and scalable parallel algorithm for Monte Carlo Tree Search (MCTS), an approximation technique for game searches. MCTS is also used to solve problems as diverse as planning under uncertainty, combinatorial optimization, and high-energy physics. In these problems,
the solution search space is significantly large, necessitating parallel solutions. Shared memory parallel approaches do not scale well beyond the size of a single node's RAM. SmartK is a distributed memory parallelization that takes advantage of both inter-node and intra-node parallelism and a large cumulative RAM found in clusters. SmartK's novel selection algorithm combined with its ability to efficiently search the solution space, results in better solutions than other MCTS parallel approaches. Results of an MPI implementation of SmartK for the game of Hex, show SmartK yields a better win percentage than other parallel algorithms, and that its performance scales to larger search spaces and high degrees of parallelism.

Best Poster Finalist: no

**Poster 70: Numerical Method and Parallelization for the Computation of Coherent Synchrotron Radiation**  
Boqian Shen (Rice University, Los Alamos National Laboratory)

The purpose of this work is to develop and parallelize an accurate and efficient numerical method for the computation of synchrotron radiation from relativistic electrons in the near field. The high-brilliance electron beam and coherent short-wavelength light source provide a powerful method to understand the microscopic structure and dynamics of materials. Such a method supports a wide range of applications including matter physics, structural biology, and medicine development. To understand the interaction between the beam and synchrotron radiation, an accurate and efficient numerical simulation is needed. With millions of electrons, the computational cost of the field would be large. Thus, multilevel parallelism and performance portability are desired since modern supercomputers are getting more complex and heterogeneous. The performance model and performance analysis are presented.

Best Poster Finalist: no

**Poster 146: AI Matrix: A Deep Learning Benchmark for Alibaba Data Centers**  
Wei Zhang (Alibaba Inc), Wei Wei (Alibaba Inc), Lingjie Xu (Alibaba Inc), Lingling Jin (Alibaba Inc), Cheng Li (University of Illinois)

This work introduces AI Matrix, an in-house Deep Learning (DL) benchmark suite developed specifically for Alibaba's e-commerce environment. AI Matrix results from a full investigation of the DL applications used inside Alibaba and aims to cover the typical DL applications that account for more than 90% of the GPU usage in Alibaba data centers. This benchmark suite collects DL models that are either directly used or closely resemble the models used in the company's real e-commerce
applications. It also collects the real e-commerce applications if no similar DL models are not available. Through the high coverage and close resemblance to real applications, AI Matrix fully represents the DL workloads on Alibaba data centers. The collected benchmarks mainly fall into three categories: computer vision, recommendation, and language processing, which consist of the most majority of DL applications in Alibaba. AI Matrix is made open source, hoping it can benefit the public.

Best Poster Finalist: no

Poster 54: Massively Parallel Eigensolvers Based on Unconstrained Energy Functionals Methods
Osni Marques (Lawrence Berkeley National Laboratory), Mauro Del Ben (Lawrence Berkeley National Laboratory), Andrew Canning (Lawrence Berkeley National Laboratory)

This poster focuses on a preconditioned conjugate gradient based iterative eigensolver using an unconstrained energy functional minimization scheme. This scheme avoids an explicit reorthogonalization of the trial eigenvectors and becomes an attractive alternative for the solution of very large problems. The unconstrained formulation is implemented in the first-principles materials and chemistry CP2K code, which performs electronic structure calculations based on a density functional theory approximation to the solution of the many-body Schrödinger equation. The systems we use in our studies have a number of atoms ranging from 2,247 to 12,288. We study the convergence of the unconstrained formulation and its scaling on a Cray XC40 (a partition with 9,688 Intel KNL nodes). We show that there is a trade-off between the preconditioner that leads to fast convergence and lower cost preconditioners that lead to best time to solution.

Best Poster Finalist: yes

Poster 100: Comparison of Array Management Library Performance - A Neuroscience Use Case
Donghe Kang (Ohio State University), Oliver Rübel (Lawrence Berkeley National Laboratory), Suren Byna (Lawrence Berkeley National Laboratory), Spyros Blanas (Ohio State University)

Array management libraries, such as HDF5, Zarr, etc., depend on a complex software stack that consists of parallel I/O middleware (MPI-IO), POSIX-IO, and file systems. Components in the stack are interdependent, such that effort in tuning the parameters in these software libraries for optimal performance is non-trivial. On the other hand, it is challenging to choose an array management library based on the array configuration and access patterns. In this poster, we investigate the performance aspect of two array management libraries, i.e., HDF5 and Zarr, in the context of a neuroscience use case. We highlight the performance variability of HDF5 and Zarr in our preliminary results and discuss potential optimization strategies.
**Poster 118: Self-Driving Reconfigurable Silicon Photonic Interconnects (Flex-LIONS) with Deep Reinforcement Learning**

Roberto Proietti (University of California, Davis), Yu Shang (University of California, Davis), Xian Xiao (University of California, Davis), Xiaoliang Chen (University of California, Davis), Yu Zhang (University of California, Davis), SJ Ben Yoo (University of California, Davis)

We propose a self-driving reconfigurable optical interconnect architecture for HPC systems exploiting a deep reinforcement learning (DRL) algorithm and a reconfigurable silicon photonic (SiPh) switching fabric to adapt the interconnect topology to different traffic demands. Preliminary simulation results show that after training, the DRL-based SiPh fabric provides the lowest average end-to-end latency for time-varying traffic patterns.

Best Poster Finalist: no

**Poster 147: Extremely Accelerated Deep Learning: ResNet-50 Training in 70.4 Seconds**

Akihiro Tabuchi (Fujitsu Laboratories Ltd), Akihiko Kasagi (Fujitsu Laboratories Ltd), Masafumi Yamazaki (Fujitsu Laboratories Ltd), Takumi Honda (Fujitsu Laboratories Ltd), Masahiro Miwa (Fujitsu Laboratories Ltd), Takashi Shiraishi (Fujitsu Laboratories Ltd), Motohiro Kosaki (Fujitsu Laboratories Ltd), Naoto Fukumoto (Fujitsu Laboratories Ltd), Tsuguchika Tabaru (Fujitsu Laboratories Ltd), Atsushi Ike (Fujitsu Laboratories Ltd), Kohta Nakashima (Fujitsu Laboratories Ltd)

Distributed deep learning using a large mini-batch is a key technology to accelerate training in deep learning. However, it is difficult to achieve a high scalability and maintain validation accuracy in distributed learning on large clusters. We introduce two optimizations, reducing the computation time and overlapping the communication with the computation. By applying the techniques and using 2,048 GPUs, we achieved the world's fastest ResNet-50 training in MLPerf, which is a de facto standard DNN benchmark (as of July 2019).

Best Poster Finalist: no

**Poster 111: Multiple HPC Environments-Aware Container Image Configuration for Bioinformatics Application**

Kento Aoyama (Tokyo Institute of Technology, National Institute of Advanced Industrial Science and
Containers have a considerable advantage for application portability in different environments by isolating process with a small performance overhead; thus it has been rapidly getting popular in a wide range of science fields. However, there are problems in container image configuration when run in multiple HPC environments, and it requires users to have knowledge of systems, container runtimes, container image format, and library compatibilities in HPC environments.

In this study, we introduce our HPC container workflow in multiple supercomputing environments that have different system/library specifications (ABCI, TSUBAME3.0). Our workflow provides custom container image configurations for HPC environments by taking into account differences in container runtime, container image, and library compatibility between the host and inside of the container. We also show the parallel performance of our application in each HPC environment.

Best Poster Finalist: no
**Poster 112: Building Complex Software Applications Inside Containers**
*Calvin D. Seamons (Los Alamos National Laboratory)*

High performance computing (HPC) scientific applications require complex dependencies to operate. As user demand for HPC systems increases, it becomes unrealistic to support every unique dependency request. Containers can offer the ability to satisfy the users’ dependency request while simultaneously offering HPC portability across systems. By "containerizing" Model for Prediction Across Scales (MPAS, a large atmospheric simulation suite), we show that it is possible to containerize and run complex software. Furthermore, the container can be run across different HPC systems with nearly identical results (21 bytes difference over 2.1 gigabytes). Containers have the possibility to bring flexibility to code teams in HPC by helping to meet the demand for user defined software stacks (UDSS), and giving teams the ability to choose their software, independently of what is offered by the HPC system.

Best Poster Finalist: no

**Poster 101: Job Performance Overview of Apache Flink and Apache Spark Applications**
*Jan Frenzel (Technical University Dresden), René Jäkel (Technical University Dresden)*

Apache Spark and Apache Flink are two Big Data frameworks used for fast data exploration and analysis. Both frameworks provide the runtime of program sections and performance metrics, such as the number of bytes read or written, via an integrated dashboard. Performance metrics available in the dashboard lack timely information and are only shown aggregated in a separate part of the dashboard. However, performance investigations and optimizations would benefit from an integrated view with detailed performance metric events. Thus, we propose a system that samples metrics at runtime and collects information about the program sections after the execution finishes. The performance data is stored in an established format independent from Spark and Flink versions and can be viewed with state-of-the-art performance tools, i.e. Vampir. The overhead depends on the sampling interval and was below 10% in our experiments.

Best Poster Finalist: no

**Poster 113: Improvements Toward the Release of the Pavilion 2.0 Test Harness**
*Kody J. Everson (Los Alamos National Laboratory, Dakota State University), Maria Francine Lapid (Los Alamos National Laboratory)*


High-performance computing production support entails thorough testing in order to evaluate the efficacy of a system for production-grade workloads. There are various phases of a system’s life-cycle to assess, requiring different methods to accomplish effective evaluation of performance and correctness. Due to the unique and distributed nature of an HPC-system, the necessity for sophisticated tools to automatically harness and assess test results, all while interacting with schedulers and programming environment software, requires a customizable, extensible, and lightweight system to manage concurrent testing. Beginning with the recently refactored codebase of Pavilion 1.0, we assisted with the finishing touches on readying this software for open-source release and production usage. Pavilion 2.0 is a Python 3-based testing framework for HPC clusters that facilitates the building, running, and analysis of tests through an easy-to-use, flexible, YAML-based configuration system. This enables users to write their own tests by simply wrapping everything in Pavilion’s well-defined format.

Best Poster Finalist: no

**Poster 119: Toward Lattice QCD on Fugaku: SVE Compiler Studies and Micro-Benchmarks in the RIKEN Fugaku Processor Simulator**
Nils Meyer (University of Regensburg, Bavaria), Tilo Wettig (University of Regensburg, Bavaria), Yuetsu Kodama (RIKEN Center for Computational Science (R-CCS)), Mitsuhisa Sato (RIKEN Center for Computational Science (R-CCS))

The Fugaku supercomputer, successor to the Japanese flagship K-Computer, will start operation in 2021. Fugaku incorporates the Fujitsu A64FX processor, which is the first hardware implementation supporting the Arm SVE instruction set, in this case a 512-bit version. Real hardware is not accessible today, but RIKEN has designed a simulator of the A64FX. We present micro-benchmarks relevant for Lattice QCD obtained in the RIKEN Fugaku processor simulator and compare three different SVE compilers.

Best Poster Finalist: no

**Poster 62: Emulating Multi-Pattern Quantum Grover’s Search on a High-Performance Reconfigurable Computer**
Naveed Mahmud (University of Kansas), Bennett Haase-Divine (University of Kansas), Bailey K. Srimoungchanh (University of Kansas), Nolan Blankenau (University of Kansas), Annika Kuhnke (University of Kansas), Esam El-Araby (University of Kansas)

Grover’s search(GS) is a widely studied quantum algorithm that can be employed for both single and
multi-pattern search problems and potentially provides quadratic speedup over existing classical search algorithms. In this paper, we propose a multi-pattern quantum search methodology based on a modified GS quantum circuit. The proposed method combines classical post-processing permutations with a modified Grover's circuit to efficiently search for given single/multiple input patterns. Our proposed methodology reduces quantum circuit complexity, realizes space-efficient emulation hardware and improves overall system configurability for dynamic, multi-pattern search. We use a high-performance reconfigurable computer to emulate multi-pattern GS(MGS) and present scalable emulation architectures of a complete multi-pattern search system. We validate the system and provide analysis of experimental results in terms of FPGA resource utilization and emulation time. Our results include a successful hardware architecture that is capable of emulating MGS algorithm up to 32 fully-entangled quantum bits on a single FPGA.

Best Poster Finalist: no

Poster 107: Exploring Interprocess Work Stealing for Balanced MPI Communication
Kaiming Ouyang (University of California, Riverside), Min Si (Argonne National Laboratory), Zizhong Chen (University of California, Riverside)

Workload balance among MPI processes is a critical consideration during the development of HPC applications. However, because of many factors such as complex network interconnections and irregularity of HPC applications, fully achieving workload balance in practice is nearly impossible. Although interprocess job stealing is a promising solution, existing shared-memory techniques that lack necessary flexibility or cause inefficiency during data access cannot provide an applicable job-stealing implementation. To solve this problem, we propose a new process-in-process (PiP) interprocess job-stealing method to balance communication workload among processes on MPI layers. Our initial experimental results show PiP-based job stealing can efficiently help amortize workload, reduce imbalance, and greatly improve intra- and intersocket ping-pong performance compared with original MPI.

Best Poster Finalist: no

Poster 127: sFlow Monitoring for Security and Reliability
Xava A. Grooms (Los Alamos National Laboratory, University of Kentucky), Robert V. Rollins (Los Alamos National Laboratory, Michigan Technological University), Collin T. Rumpca (Los Alamos National Laboratory, Dakota State University)

In the past ten years, High Performance Computing (HPC) has moved far beyond the terascale
performance, making petascale systems the new standard. The drastic improvement in performance has been largely unmatched with insignificant improvements in system monitoring. Thus, there is an immediate need for practical and scalable monitoring solutions to ensure the effectiveness of costly compute clusters. This project aims to explore the viability and impact of sFlow enabled switches in cluster network monitoring for security and reliability. A series of tests and exploits were performed to target specific network abnormalities on a nine-node HPC cluster. The results present web-based dashboards that can aid network administrators in improving a cluster’s security and reliability.

Best Poster Finalist: no

**Poster 77: Extreme Scale Phase-Field Simulations of Sintering Processes**  
*Johannes Hötzer (Karlsruhe University of Applied Sciences), Henrik Hierl (Karlsruhe University of Applied Sciences), Marco Seiz (Karlsruhe Institute of Technology), Andreas Reiter (Karlsruhe Institute of Technology), Britta Nestler (Karlsruhe Institute of Technology)*

The sintering process, which turns loose powders into dense materials, is naturally found in the formation of glaciers, but is also the indispensable process to manufacture ceramic materials. This process is described by a dynamically evolving microstructure, which largely influences the resulting material properties.

To investigate this complex three-dimensional, scale-bridging evolution in realistic domain sizes, a highly optimized and parallelized multiphysics phase-field solver is developed. The solver is optimized in a holistic way, from the application level over the time integration and parallelization, down to the hardware. Optimizations include communication hiding, explicit vectorization, implicit schemes, and local reduction of degrees of freedom.

With this, we are able to investigate large-scale, three-dimensional domains, and long integration times. We have achieved a single-core peak performance of 32.5%, scaled up to 98304 cores on Hazel Hen and SuperMUC-NG, and simulated a multimillion particle system.

Best Poster Finalist: no

**Poster 140: Toward Automatic Function Call Generation for Deep Learning**  
*Shizhi Tang (Tsinghua University, China), Jidong Zhai (Tsinghua University, China)*

Mainstream deep learning frameworks are commonly implemented by invoking underlying high performance tensor libraries on various architectures. However, as these libraries provide increasingly
complex semantics including operator fusions, in-place operations, and various memory layouts, the gap between mathematical deep learning models and the underlying libraries becomes larger. In this paper, inspired by the classic problem of Instruction Selection, we design a theorem solver guided exhausted search algorithm to select functions for complex tensor computations. Preliminary results with some micro-benchmarks and a real model show that our approach can outperform both Tensorflow and Tensor Comprehensions at run time.

Best Poster Finalist: no

**Poster 83: ETL: Elastic Training Layer for Deep Learning**
Lei Xie (Tsinghua University, China), Jidong Zhai (Tsinghua University, China)

Due to the rising of deep learning, clusters for deep learning training are widely deployed in production. However, static task configuration and resource fragmentation problems in existing clusters result in low efficiency and poor quality of service. We propose ETL, an elastic training layer for deep learning, to help address them once for all. ETL adopts many novel mechanisms, such as lightweight and configurable report primitive and asynchronous, parallel and IO-free state replication, to achieve both high elasticity and efficiency. The evaluation demonstrates the low overhead and high efficiency of these mechanisms and reveals the advantages of elastic deep learning supported by ETL.

Best Poster Finalist: no

**Poster 130: Deep Learning-Based Feature-Aware Data Modeling for Complex Physics Simulations**
Qun Liu (Louisiana State University), Subhashis Hazarika (Ohio State University), John M. Patchett (Los Alamos National Laboratory), James P. Ahrens (Los Alamos National Laboratory), Ayan Biswas (Los Alamos National Laboratory)

Data modeling and reduction for in situ is important. Feature-driven methods for in situ data analysis and reduction are a priority for future exascale machines as there are currently very few such methods. We investigate a deep-learning-based workflow that targets in situ data processing using autoencoders. We employ integrated skip connections to obtain higher performance compared to the existing autoencoders. Our experiments demonstrate the initial success of the proposed framework and create optimism for the in situ use case.

Best Poster Finalist: no
Poster 125: Physics Informed Generative Adversarial Networks for Virtual Mechanical Testing
Julian Cuevas (NASA, University of Puerto Rico at Mayaguez), Patrick Leser (NASA), James Warner (NASA), Geoffrey Bomarito (NASA), William Leser (NASA)

Physics-informed generative adversarial networks (PI-GANs) are used to learn the underlying probability distributions of spatially-varying material properties (e.g., microstructure variability in a polycrystalline material). While standard GANs rely solely on data for training, PI-GANs encode physics in the form of stochastic differential equations using automatic differentiation. The goal here is to show that experimental data from a limited number of material tests can be used with PI-GANs to enable unlimited virtual testing for aerospace applications. Preliminary results using synthetically generated data are provided to demonstrate the proposed framework. Deep learning and automatic differentiation capabilities in Tensorflow were implemented on Nvidia Tesla V100 GPUs.

Best Poster Finalist: no

Poster 141: ExaGeoStatR: Harnessing HPC Capabilities for Large Scale Geospatial Modeling Using R
Sameh Abdulah (King Abdullah University of Science and Technology (KAUST)), Yuxiao Li (King Abdullah University of Science and Technology (KAUST)), Jian Cao (King Abdullah University of Science and Technology (KAUST)), Hatem Ltaief (King Abdullah University of Science and Technology (KAUST)), David Keyes (King Abdullah University of Science and Technology (KAUST)), Marc Genton (King Abdullah University of Science and Technology (KAUST)), Ying Sun (King Abdullah University of Science and Technology (KAUST))

Large-scale simulations and parallel computing techniques are becoming essential in Gaussian process calculations to lessen the complexity of geostatistics applications. The log-likelihood function is used in such applications to evaluate the model associated with a given set of measurements in existing n geographic locations. The evaluation of such a function requires O(n^2) memory and O(n^3) computation, which is infeasible for large datasets with existing software tools.

We present ExaGeoStatR, a package for large-scale geostatistics in R that computes the log-likelihood function on shared and distributed-memory, possibly equipped with GPU, using advanced linear algebra techniques. The package provides a high-level abstraction of the underlying architecture while enhancing the R developers' productivity. We demonstrate ExaGeoStatR package by illustrating its implementation details, analyzing its performance on various parallel architectures, and assessing its accuracy using synthetic datasets and a sea surface temperature dataset. The
Hierarchical low-rank approximation can reduce both the storage and computation costs of dense matrices, but its implementation is challenging. In this research, we tackle one of the most difficult problems of GPU parallelization of the factorization of these hierarchical matrices. To this end, we are developing a new runtime system for GPUs that can schedule all tasks into one GPU kernel. Other existing runtime systems, like cuGraph and Stanford Legion, can only manage streams and kernel-level parallelism. Even without too much tuning, we achieved 4x better performance in H-LU factorization with a single GPU when comparing with a well-tuned CPU-based hierarchical matrix library, HLIBpro, on moderately sized matrices. Additionally, we have significantly less runtime overheads exposed when processing smaller matrices.

Best Poster Finalist: no

Poster 145: Improving Data Compression with Deep Predictive Neural Network for Time Evolutional Data
Rupak Roy (Florida State University), Kento Sato (RIKEN Center for Computational Science (R-CCS)), Jian Guo (RIKEN Center for Computational Science (R-CCS)), Jens Domke (RIKEN Center for Computational Science (R-CCS)), Weikuan Yu (Florida State University), Takaki Hatsui (RIKEN SPring-8 Center), Yasumasa Joti (Japan Synchrotron Radiation Research Institute)

Scientific applications/simulations periodically generate huge intermediate data. Storing or transferring such a large scale of data is critical. Fast I/O is important for making this process faster. One of the approaches to achieve fast I/O is data compression. Our goal is to achieve a delta technique that can improve the performance of existing data compression algorithms for time evolutional intermediate data.

In our approach, we compute the delta values from original data and data predicted by the deep predictive neural network. We pass these delta values through three phases which are preprocessing phase, partitioned entropy coding phase, and density-based spatial delta encoding phase.
In our poster, we present how our predictive delta technique can leverage the time evolutional data to produce highly concentrated small values. We show the improvement in compression ratio when our technique, combined with existing compression algorithms, are applied on the intermediate data for different datasets.

Best Poster Finalist: no

**Poster 144: Optimizing Asynchronous Multi-Level Checkpoint/Restart Configurations with Machine Learning**

Tonmoy Dey (Florida State University), Kento Sato (RIKEN Center for Computational Science (R-CCS)), Jian Guo (RIKEN Center for Computational Science (R-CCS)), Bogdan Nicolae (Argonne National Laboratory), Jens Domke (RIKEN Center for Computational Science (R-CCS)), Weikuan Yu (Florida State University), Franck Cappello (Argonne National Laboratory), Kathryn Mohror (Lawrence Livermore National Laboratory)

With the emergence of fast local storage, multi-level checkpointing (MLC) has become a common approach for efficient checkpointing. To utilize MLC efficiently, it is important to determine the optimal configuration for the checkpoint/restart (CR). There are mainly two approaches for determining the optimal configuration for CR, namely modeling and simulation approach. However, with MLC, CR becomes more complicated making the modeling approach inaccurate and the simulation approach though accurate, very slow. In this poster, we focus on optimizing the performance of CR by predicting the optimized checkpoint count and interval. This was achieved by combining the simulation approach with machine learning and neural network to leverage its accuracy without spending time on simulating different CR parameters. We demonstrate that our models can predict the optimized parameter values with minimal error when compared to the simulation approach.

Best Poster Finalist: no

**Poster 132: Optimizing Performance at Runtime Using Binary Rewriting**

Alexis Engelke (Technical University Munich), David Hildenbrand (Technical University Munich), Martin Schulz (Technical University Munich)

In addition to scalability, performance of sequential code in applications is an important factor in HPC. Typically, programs are compiled once, at which time optimizations are applied, and are then run several times. However, not all information relevant for performance optimizations are available at compile-time, restricting optimization possibilities. The generation of specialized code at runtime allows for further optimizations. Performing such specialization on binary code allows for initial code
to be generated at compile-time with only the relevant parts being rewritten at runtime, reducing the optimization overhead. For targeted optimizations and effective use of known runtime information, the rewriting process needs to be guided by the application itself, exploiting information only known to the developer.

We describe three approaches for self-guided binary rewriting explicitly guided by the running application and evaluate the performance of the optimized code as well as the performance of the rewriting process itself.

Best Poster Finalist: no

Poster 53: Unstructured Mesh Technologies for Fusion Simulations
Cameron Smith (Rensselaer Polytechnic Institute (RPI)), Gerrett Diamond (Rensselaer Polytechnic Institute (RPI)), Gopan Perumpilly (Rensselaer Polytechnic Institute (RPI)), Chonglin Zhang (Rensselaer Polytechnic Institute (RPI)), Agnieszka Truszkowska (Rensselaer Polytechnic Institute (RPI)), Morteza Hakimi (Rensselaer Polytechnic Institute (RPI)), Onkar Sahni (Rensselaer Polytechnic Institute (RPI)), Mark Shephard (Rensselaer Polytechnic Institute (RPI)), Eisung Yoon (Ulsan National Institute of Science and Technology, South Korea), Daniel Ibanez (Sandia National Laboratories)

Multiple unstructured mesh technologies are needed to define and execute plasma physics simulations. The domains of interest combine model features defined from physical fields within 3D CAD of the tokamak vessel with an antenna assembly, and 2D cross sections of the tokamak vessel. Mesh generation technologies must satisfy these geometric constraints and additional constraints imposed by the numerical models. Likewise, fusion simulations over these domains study a range of timescales and physical phenomena within a tokamak.

XGCm studies the development of plasma turbulence in the reactor vessel, GITRm studies impurity transport, and PetraM simulations model RF wave propagation in scrape off layer plasmas. GITRm and XGCm developments are using the PUMIpic infrastructure to manage the storage and access of non-uniform particle distributions in unstructured meshes on GPUs. PetraM combines PUMI adaptive unstructured mesh control with MFEM using CAD models and meshes defined with Simmetrix tools.

Best Poster Finalist: no

Poster 99: Eithne: A Framework for Benchmarking Micro-Core Accelerators
Maurice C. Jamieson (Edinburgh Parallel Computing Centre, University of Edinburgh), Nick Brown (Edinburgh Parallel Computing Centre, University of Edinburgh)
Running existing HPC benchmarks as-is on micro-core architectures is at best difficult and most often impossible as they have a number of architectural features that makes them significantly different from traditional CPUs: tiny amounts on-chip RAM (c. 32KB), low-level knowledge specific to each device (including the host / device communications interface), limited communications bandwidth and complex or no device debugging environment. In order to compare and contrast different the micro-core architectures, a benchmark framework is required to abstract much of this complexity.

The modular Eithne framework supports the comparison of a number of micro-core architectures. The framework separates the actual benchmark from the details of how this is executed on the different technologies. The framework was evaluated by running the LINPACK benchmark on the Adapteva Epiphany, PicoRV32 and VectorBlox Orca RISC-V soft-cores, NXP RV32M1, ARM Cortex-A9, and Xilinx MicroBlaze soft-core, and comparing resulting performance and power consumption.

Best Poster Finalist: no

**Poster 133: Portable Resilience with Kokkos**

*Jeffery Miles (Sandia National Laboratories), Nicolas Morales (Sandia National Laboratories), Carson Mould (Sandia National Laboratories), Keita Teranishi (Sandia National Laboratories)*

The Kokkos ecosystem is a programming environment that provides performance and portability to many scientific applications that run on DOE supercomputers as well as other smaller scale systems. Leveraging software abstraction concepts within Kokkos, software resilience for end user code is made portable with abstractions and concepts while implementing the most efficient resilience algorithms internally. This addition enables an application to manage hardware failures reducing the cost of interruption without drastically increasing the software maintenance cost. Two main resilience methodologies have been added to the Kokkos ecosystem to validate the resilience abstractions: 1. Checkpointing includes an automatic mode supporting other checkpointing libraries and a manual mode which leverages the data abstraction and memory space concepts. 2. The redundant execution model anticipates failures by replicating data and execution paths. The design and implementation of these additions are illustrated, and appropriate examples are included to demonstrate the simplicity of use.

Best Poster Finalist: no

**Poster 79: The HPC PowerStack: A Community-Wide Collaboration Toward an Energy Efficient**
Software Stack
Siddhartha Jana (Energy Efficient HPC Working Group, Intel Corporation), Stephanie Brink (Lawrence Livermore National Laboratory), Christopher Cantalupo (Intel Corporation), Jonathan Eastep (Intel Corporation), Masaaki Kondo (RIKEN Advanced Institute for Computational Science (AICS), University of Tokyo), Matthias Maiterth (Intel Corporation), Aniruddha Marathe (Lawrence Livermore National Laboratory), Tapasya Patki (Lawrence Livermore National Laboratory), Barry Rountree (Lawrence Livermore National Laboratory), Ryuichi Sakamoto (University of Tokyo), Martin Schulz (Technical University Munich, Leibniz Supercomputing Centre), Carsten Trinitis (Technical University Munich), Josef Weidendorfer (Technical University Munich, Leibniz Supercomputing Centre)

This poster highlights an ongoing community-wide effort among vendors, labs, and academia, to incorporate power-awareness within system-stacks in upcoming exascale machines. HPC PowerStack is the first-and-only community-driven vendor-neutral effort to identify what power optimization software actors are key within the modern-day stack; discuss their interoperability, and work toward gluing together existing open source projects to engineer cost-effective, but cohesive, portable implementations.

This poster disseminates key insights acquired in the project, provides prototyping status updates, highlights open questions, and solicits participation addressing the imminent exascale power challenge.

Best Poster Finalist: no

Poster 87: Parallelizing Simulations of Large Quantum Circuits
Michael A. Perlin (University of Colorado, National Institute of Standards and Technology (NIST)), Teague Tomesh (Princeton University), Bradley Pearlman (University of Colorado, National Institute of Standards and Technology (NIST)), Wei Tang (Princeton University), Yuri Alexeev (Argonne National Laboratory), Martin Suchara (Argonne National Laboratory)

We present a parallelization scheme for classical simulations of quantum circuits. Our scheme is based on a recent method to "cut" large quantum circuits into smaller sub-circuits that can be simulated independently, and whose simulation results can in turn be re-combined to infer the output of the original circuit. The exponentially smaller classical computing resources needed to simulate smaller circuits are counterbalanced by exponential overhead in terms of classical post-processing costs. We discuss how this overhead can be massively parallelized to reduce classical computing costs.

Best Poster Finalist: no
Poster 120: ILP-Based Scheduling for Linear-Tape Model Trapped-Ion Quantum Computers
Xin-Chuan Wu (University of Chicago), Yongshan Ding (University of Chicago), Yunong Shi (University of Chicago), Yuri Alexeev (Argonne National Laboratory), Hal Finkel (Argonne National Laboratory), Kibaek Kim (Argonne National Laboratory), Frederic T. Chong (University of Chicago)

Quantum computing (QC) is emerging as a potential post-Moore high-performance computing (HPC) technology. Trapped-ion quantum bits (qubits) are among the most leading technologies to reach scalable quantum computers that would solve certain problems beyond the capabilities of even the largest classical supercomputers. In trapped-ion QC, qubits can physically move on the ion trap. The state-of-the-art architecture, linear-tape model, only requires a few laser beams to interact with the entire qubits by physically moving the interacting ions to the execution zone. Since the laser beams are limited resources, the ion chain movement and quantum gate scheduling are critical for the circuit latency. To harness the emerging architecture, we present our mathematical model for scheduling the qubit movements and quantum gates in order to minimize the circuit latency. In our experiment, our scheduling reduces 29.47% circuit latency on average. The results suggest classical HPC would further improve the quantum circuit optimization.

Best Poster Finalist: no

Poster 55: MPI+OpenMP Parallelization of DFT Method in GAMESS
Vladimir Mironov (Lomonosov Moscow State University), Yuri Alexeev (Argonne National Laboratory), Dmitry Fedorov (National Institute of Advanced Industrial Science and Technology (AIST))

In this work, the Density Functional Theory (DFT) method is parallelized with MPI-OpenMP in the quantum chemistry package GAMESS. It has been implemented in both regular and Fragment Molecular Orbital (FMO) based DFT codes. The scalability of the FMO-DFT code was demonstrated on Cray XC40 Theta supercomputer. We demonstrated excellent scalability of the code up 2,048 Intel Xeon Phi nodes (131,072 cores). Moreover, the developed DFT code is about twice as fast as the original code because of our new grid integration algorithm.

Best Poster Finalist: no

Poster 71: AI-Solver: Uncertainty in Prediction and Error Estimation for AI in Engineering
Ahmed Al-Jarro (Fujitsu Laboratories of Europe Ltd), Loic Beheshti (Fujitsu Laboratories of Europe Ltd), Serban Georgescu (Fujitsu Laboratories of Europe Ltd), Koichi Shirahata (Fujitsu Laboratories
The AI-Solver is a deep learning platform that learns from simulation data to extract general behavior based on physical parameters. The AI-Solver can handle a wide variety of classes of problems including those commonly identified in FEA, CFD and CEM, to name a few, with speedups of up to 250,000X and extremely low error rate of 2-3%. In this work, we build on this recent effort. We first integrate uncertainty quantification, via exploiting the approximation of Bayesian Deep Learning. Second, we develop bespoke error estimation mechanisms capable of processing this uncertainty to provide instant feedback on the confidence in predictions without relying on the availability of ground truth data. To our knowledge, the ability to estimate the discrepancy in predictions without labels is a first in the field of AI for Engineering.

Best Poster Finalist: no

Aaron Scheinberg (Princeton Plasma Physics Laboratory), Guangye Chen (Los Alamos National Laboratory), Stephane Ethier (Princeton Plasma Physics Laboratory), Stuart Slattery (Oak Ridge National Laboratory), Robert Bird (Los Alamos National Laboratory), Pat Worley (PHWorley Consulting), Choong-Seock Chang (Princeton Plasma Physics Laboratory)

Numerical plasma physics models such as the particle-in-cell XGC code are important tools to understand phenomena encountered in experimental fusion devices. Adequately resolved simulations are computationally expensive, so optimization is essential. To address the need for consistent high performance by cutting-edge scientific software applications, frameworks such as Kokkos have been developed to enable portability as new architectures require hardware-specific coding implementation for best performance. Cabana, a recent extension to Kokkos developed with the ECP-CoPA project, is a library of common kernels and operations typically necessary for particle-based codes. The Kokkos/Cabana framework enables intuitive construction of particle-based codes, while maintaining portability between architectures. Here, we summarize the adoption by XGC of the execution and data layout patterns offered by this framework. We demonstrate a method for Fortran codes to adopt Kokkos and show that it can provide a single, portable code base that performs well on both GPUs and multicore machines.

Best Poster Finalist: no

Poster 106: Optimizing Hybrid Access Virtual Memory System Using SCM/DRAM Unified Memory Management Unit
Yusuke Shirota (Toshiba Corporation), Shiyo Yoshimura (Toshiba Corporation), Satoshi Shirai (Toshiba Corporation), Tatsunori Kanai (Toshiba Corporation)

In HPC systems, expectations for storage-class memory (SCM) are increasing in large-scale in-memory processing. While SCM can deliver higher capacity and lower standby power than DRAM, it is slower and the dynamic power is higher. Therefore, in order to realize high-speed, low-power and scalable main memory, it is necessary to build an SCM/DRAM unified memory, and dynamically optimize data placement between the two memories according to the memory access pattern.

In this poster, we describe a new hybrid access type virtual memory method using TLB-extended unified memory management unit which enables collecting and extracting fine-grained memory access locality characteristics. We show that with the proposed method, Hybrid Access control, which is a memory hierarchy control that selectively uses Direct Access to bus attached byte-addressable SCM and low power Aggressive Paging using small DRAM as cache, can be made more accurate, and the efficiency of memory access can be significantly improved.

Best Poster Finalist: no

Holistic Measurement Driven System Assessment
Saurabh Jha (University of Illinois), Mike Showerman (National Center for Supercomputing Applications (NCSA), University of Illinois), Aaron Saxton (National Center for Supercomputing Applications (NCSA), University of Illinois), Jeremy Enos (National Center for Supercomputing Applications (NCSA), University of Illinois), Greg Bauer (National Center for Supercomputing Applications (NCSA), University of Illinois), Zbigniew Kalbarczyk (University of Illinois), Ann Gentile (Sandia National Laboratories), Jim Brandt (Sandia National Laboratories), Ravi Iyer (University of Illinois), William T. Kramer (University of Illinois, National Center for Supercomputing Applications (NCSA))

HPC users deploy a suite of monitors to observe patterns of failures and performance anomalies to improve operational efficiency, achieve higher application performance and inform the design of future systems. However, the promises and the potential of monitoring data have largely been not realized due to various challenges such as inadequacy in monitoring, limited availability of data, lack of methods for fusing monitoring data at time-scales necessary for enabling human-in-the-loop or machine-in-the-loop feedback. To address above challenges, in this work we developed a monitoring fabric Holistic Measurement Driven System Assessment (HMDSA) for large-scale HPC facilities, independent of major component vendor, and within budget constraints of money, space, and power. We accomplish this through development and deployment of scalable, platform-independent, open-source tools and techniques for monitoring, coupled with statistical and machine-learning based
runtime analysis and feedback, which enables highly efficient HPC system operation and usage and also informs future system improvements.

Best Poster Finalist: no

**Poster 139: Model Identification of Pressure Drop in Membrane Channels with Multilayer Artificial Neural Networks**

Jiang-hang Gu (Sun Yat-sen University, Zhuhai, School of Chemical Engineering and Technology), Jiu Luo (Sun Yat-sen University, Guangzhou, School of Materials Science and Engineering), Ming-heng Li (California State Polytechnic University, Pomona), Yi Heng (Sun Yat-sen University, Guangzhou, School of Data and Computer Science; Sun Yat-sen University, Guangzhou, China)

This poster presents the work of identifying a data-driven model of pressure drop in spacer-filled reverse osmosis membrane channels and conducting CFD simulations. The established model correlates the pressure drop with a wide range of design objectives, which enables a quantitative description of the geometric structures and operation conditions for improvement. This way, it aims at optimizing the spacer geometry with minimal effort. Furthermore, a high-performance computing strategy is employed to tackle the resulted intractable computational task in the identification procedure and CFD simulations.

Best Poster Finalist: no

**Poster 61: Fast 3D Diffeomorphic Image Registration on GPUs**

Malte Brunn (University of Stuttgart), Naveen Himthani (University of Texas), George Biros (University of Texas), Miriam Mehl (University of Stuttgart), Andreas Mang (University of Houston)

3D image registration is one of the most fundamental and computationally expensive operations in medical image analysis. Here, we present a mixed-precision, Gauss-Newton-Krylov solver for diffeomorphic registration. Our work extends the publicly available CLAIRE library to GPU architectures. Despite the importance of image registration, only a few implementations of large deformation diffeomorphic registration packages support GPUs. Our contributions are new algorithms and dedicated computational kernels to significantly reduce the runtime of the main computational kernels in CLAIRE: derivatives and interpolation. We deploy (i) highly-optimized, mixed-precision GPU-kernels for the evaluation of scattered-data interpolation, (ii) replace FFT-based first-order derivatives with optimized 8th-order finite differences, and (iii) compare with state-of-the-art CPU and GPU implementations. As a highlight, we demonstrate that we can register 256^3 clinical images in less than 6 seconds on a single NVIDIA Tesla V100. This amounts to over 20x
speed-up over CLAIRE and over 30x speed-up over existing GPU implementations.

Best Poster Finalist: no

**Poster 138: Across-Stack Profiling and Characterization of State-of-the-Art Machine Learning Models on GPUs**

Cheng Li (University of Illinois), Abdul Dakkak (University of Illinois), Wei Wei (Alibaba Inc), Jinjun Xiong (IBM Research), Lingjie Xu (Alibaba Inc), Wei Zhang (Alibaba Inc), Wen-mei Hwu (University of Illinois)

The past few years have seen a surge of using Machine Learning (ML) and Deep Learning (DL) algorithms for traditional HPC tasks such as feature detection, numerical analysis, and graph analytics. While ML and DL enable solving HPC tasks, their adoption has been hampered due to the lack of understanding of how they utilize systems. Optimizing these algorithms requires characterizing their performance across the hardware/software (HW/SW) stack, but the lack of simple tools to automate the process and the reliance on researchers to perform manual characterization is a bottleneck. To alleviate this, we propose an across-stack profiling scheme and integrate it within MLModelScope — a hardware and software agnostic tool for evaluating and benchmarking ML/DL at scale. We demonstrate MLModelScope’s ability to characterize state-of-art ML/DL models and give insights that are only possible obtained by performing across-stack profiling.

Best Poster Finalist: no

**Poster 60: Massively Parallel Large-Scale Multi-Model Simulation of Tumor Development**

Marco Berghoff (Karlsruhe Institute of Technology), Jakob Rosenbauer (Forschungszentrum Juelich), Alexander Schug (Forschungszentrum Juelich)

The temporal and spatial resolution in the microscopy of tissues has increased significantly within the last years, yielding new insights into the dynamics of tissue development and the role of the single-cell within it. A thorough theoretical description of the connection of single-cell processes to macroscopic tissue reorganizations is still lacking. Especially in tumor development, single cells play a crucial role in advance of tumor properties.

We developed a simulation framework that can model tissue development up to the centimeter scale with micrometer resolution of single cells. Through a full parallelization, it enables the efficient use of HPC systems, therefore enabling detailed simulations on a large scale. We developed a generalized tumor model that respects adhesion driven cell migration, cell-to-cell signaling, and mutation-driven
tumor heterogeneity. We scan the response of the tumor development depending on division inhibiting substances such as cytostatic agents.

Best Poster Finalist: no

**Poster 114: Optimizing Recommendation System Inference Performance Based on GPU**
Xiaowei Shen (Alibaba Inc), Junrui Zhou (Alibaba Inc), Kan Liu (Alibaba Inc), Lingling Jin (Alibaba Inc), Pengfei Fan (Alibaba Inc), Wei Zhang (Alibaba Inc), Jun Yang (University of Pittsburgh)

Neural network-based recommendation models have been widely applied on tracking personalization and recommendation tasks at large Internet companies such as e-commerce companies and social media companies. Alibaba recommendation system deploys WDL (wide and deep learning) models for product recommendation tasks. The WDL model consists of two main parts: embedding lookup and neural network-based feature ranking model that ranks different products for different users. As more and more products and users the model need to rank, the feature length and batch size of the models are increased. The computation of models is also increased so that traditional model inference implementation on CPU cannot meet the requirement of QPS (query per second) and latency of recommendation tasks. In this poster, we develop a GPU based system to speedup recommendation system inference performance. By model quantization and graph transformation, we can achieve 3.9x performance speedup when compared with a baseline GPU implementation.

Best Poster Finalist: no

**Poster 59: Accelerating BFS and SSSP on a NUMA Machine for the Graph500 Challenge**
Tanuj K. Aasawat (RIKEN), Kazuki Yoshizoe (RIKEN), Tahsin Reza (University of British Columbia), Matei Ripeanu (University of British Columbia)

The NUMA architecture is the design choice for modern multi-CPU shared memory systems. In many ways, a NUMA system resembles a shared-nothing distributed system: memory accesses to remote NUMA domains are more expensive than local accesses.

In this work, we explore how improved data locality and reduced expensive remote communication can be achieved by exploiting "distributed" shared-memory of NUMA machines to develop shared-memory graph processing solutions optimized for NUMA systems. We introduce a novel hybrid design for memory accesses that handles the burst mode in traversal based algorithms, like BFS and SSSP, and reduces the number of remote accesses and updates. We demonstrate that our designs offer up to 84% speedup over our NUMA-oblivious framework Totem and 2.86x over shared-nothing...
distributed design, for BFS and SSSP algorithms.

Best Poster Finalist: no

**Poster 47: Decomposition Algorithms for Scalable Quantum Annealing**

Elijah Pelofske (Los Alamos National Laboratory), Georg Hahn (Harvard University), Hristo Djidjev (Los Alamos National Laboratory)

Commercial adiabatic quantum annealers such as D-Wave 2000Q have the potential to solve NP-complete optimization problems efficiently. One of the primary constraints of such devices is the limited number and connectivity of their qubits. This research presents two exact decomposition methods (for the Maximum Clique and the Minimum Vertex Cover problem) that allow us to solve problems of arbitrarily large sizes by splitting them up recursively into a series of arbitrarily small subproblems. Those subproblems are then solved exactly or approximately using a quantum annealer. Whereas some previous approaches are based on heuristics that do not guarantee optimality of their solutions, our decomposition algorithms have the property that the optimal solution of the input problem can be reconstructed given all generated subproblems are solved optimally as well. We investigate various heuristic and exact bounds as well as reduction methods that help to increase the scalability of our approaches.

Best Poster Finalist: no

**Poster 92: Nanoporous Flow Simulations on the Summit Supercomputer**

Yidong Xia (Idaho National Laboratory), Lixiang Luo (IBM - TJ Watson Research Center), Ansel Blumers (Brown University), Joshua Kane (Idaho National Laboratory), Jan Goral (University of Utah), Yu-Hang Tang (Lawrence Berkeley National Laboratory), Zhen Li (Clemson University, Brown University), Hai Huang (Idaho National Laboratory), Milind Deo (University of Utah)

Fluid flow behaviors in nanoporous materials are distinct from those following the continuum physics. Numerical simulations can be a complement to laboratory experiments. This work presents a dissipative particle dynamics (DPD) package for GPU-accelerated mesoscale flow simulations in nanoporous materials. In an ideal benchmark that minimizes load imbalance, the package delivered nearly perfect strong- and weak-scaling (with up to 4 billion DPD particles) on up to 1,536 V100 GPUs on Oak Ridge National Laboratory's Summit supercomputer. More remarkably, in a benchmark to measure its usefulness with realistic nanopores in SBA-15 silica, the package exhibited more than 20x speedup over its LAMMPS-based CPU counterpart with the same number nodes (e.g., 384 V100 GPUs vs. 2,688 POWER9 cores). It is worth highlighting that the NVLink2 Host-to-Device
interconnects kept the cost of CPU-GPU memory copy as low as only 10% of GPU activity time per rank: 4 times less than their PCIe counterparts.

Best Poster Finalist: yes

**Poster 90: You Have to Break It to Make It: How On-Demand, Ephemeral Public Cloud Projects with Alces Flight Compute Resulted in the Open-Source OpenFlightHPC Project**
Cristin Merritt (Alces Flight Limited; Alces Software Ltd, UK), Wil Mayers (Alces Flight Limited), Stu Franks (Alces Flight Limited)

Over three years ago the Alces Flight team made a decision to explore on-demand public cloud consumption for High Performance Computing (HPC). Our premise was simple, create a fully-featured, scalable HPC environment for research and scientific computing and provide it as a free subscription service to researchers in order to study how they would approach and consume this then new technology. This tool, Alces Flight Compute, would set out to chart how far away from the traditional bare-metal platforms our subscribers were willing to go. What we didn’t expect was that to get to their destination, our users would proceed to take our tool apart. This deconstruction has resulted in a new approach to HPC environment creation (the open-source OpenFlightHPC project), helped us better understand cloud adoption strategies, and handed over a set of guidelines to help those looking to bring public cloud into their HPC solution.

Best Poster Finalist: no

**Poster 126: Enforcing Crash Consistency of Scientific Applications in Non-Volatile Main Memory Systems**
Tyler Coy (Washington State University, Vancouver), Xuechen Zhang (Washington State University, Vancouver)

This poster presents a compiler-assistant technique, NVPath, to automatically generates NVMM-aware persistent data structures which provide the same level of guarantee of crash consistency compared to the baseline code. Compiler-assistant code annotation and transformation is general and can be applied to applications using various data structures. Our experimental results with real-world scientific applications show that the performance of the annotated programs is commensurate with the version using the manual code transformation on the Titan supercomputer.

Best Poster Finalist: no
**Poster 137: Warwick Data Store: A HPC Library for Flexible Data Storage in Multi-Physics Applications**
Richard O. Kirk (University of Warwick), Timothy R. Law (Atomic Weapons Establishment (AWE), UK), Satheesh Maheswaran (Atomic Weapons Establishment (AWE), UK), Stephen A. Jarvis (University of Warwick)

With the increasing complexity of memory architectures and multi-physics applications, developing data structures that are performant, portable, scalable, and support developer productivity, is difficult. In order to manage these complexities and allow rapid prototyping of different approaches we are building a lightweight and extensible C++ template library called the Warwick Data Store (WDS). WDS is designed to abstract details of the data structure away from the user, thus easing application development and optimisation. We show that WDS generates minimal performance overhead, via a variety of different scientific benchmarks and proxy-applications.

Best Poster Finalist: no

**Poster 76: HPChain: An MPI-Based Blockchain Framework for High Performance Computing Systems**
Abdullah Al-Mamun (University of Nevada, Reno; Lawrence Berkeley National Laboratory), Tonglin Li (Lawrence Berkeley National Laboratory), Mohammad Sadoghi (University of California, Davis), Linhua Jiang (Fudan University, Shanghai), Haoting Shen (University of Nevada, Reno), Dongfang Zhao (University of Nevada, Reno; University of California, Davis)

Data fidelity is of prominent importance for scientific experiments and simulations. The state-of-the-art mechanism to ensure data fidelity is through data provenance. However, the provenance data itself may as well exhibit unintentional human errors and malicious data manipulation. To enable a trustworthy and reliable data fidelity service, we advocate achieving the immutability and decentralization of scientific data provenance through blockchains. Specifically, we propose HPChain, a new blockchain framework specially designed for HPC systems. HPChain employs a new consensus protocol compatible with and optimized for HPC systems. Furthermore, HPChain was implemented with MPI and integrated with an off-chain distributed provenance service to tolerate the failures caused by faulty MPI ranks. The HPChain prototype system has been deployed to 500 cores at the University of Nevada's HPC center and demonstrated strong resilience and scalability while outperforming state-of-the-art blockchains by orders of magnitude.

Best Poster Finalist: no
Poster 104: An Adaptive Checkpoint Model For Large-Scale HPC Systems
Subhendu S. Behera (North Carolina State University), Lipeng Wan (Oak Ridge National Laboratory), Frank Mueller (North Carolina State University), Matthew Wolf (Oak Ridge National Laboratory), Scott Klasky (Oak Ridge National Laboratory)

Checkpoint/Restart is a widely used Fault Tolerance technique for application resilience. However, failures and the overhead of saving application state for future recovery upon failure reduces the application efficiency significantly. This work contributes a failure analysis and prediction model making decisions for checkpoint data placement, recovery, and techniques for reducing checkpoint frequency. We also demonstrate a reduction in application overhead by taking proactive measures guided by failure prediction.

Best Poster Finalist: no

Poster 123: Cloud-Native SmartX Intelligence Cluster for AI-Inspired HPC/HPDA Workloads
Jungsu Han (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science), Jun-Sik Shin (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science), JinCheol Kwon (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science), JongWon Kim (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science)

In this poster, we introduce Cloud-native SmartX Intelligence Cluster for flexibly supporting AI-inspired HPC (high performance computing) / HPDA (high performance data analytics) workloads. This work has been continuously refined from 2013 with a futuristic vision for operating 100 petascale data center. Then, we discuss issues and approaches that come with building a Cloud-native SmartX Intelligence Cluster.

Best Poster Finalist: no

Poster 86: High-Performance Custom Computing with FPGA Cluster as an Off-Loading Engine
Takaaki Miyajima (RIKEN Center for Computational Science (R-CCS)), Tomohiro Ueno (RIKEN Center for Computational Science (R-CCS)), Jens Huthmann (RIKEN Center for Computational Science (R-CCS)), Atsushi Koshiba (RIKEN Center for Computational Science (R-CCS)), Kentaro Sano (RIKEN Center for Computational Science (R-CCS)), Mitsuhsa Sato (RIKEN Center for Computational Science (R-CCS))

A heterogeneous system with Field Programmable Gate Arrays (FPGAs) is gathering attention in
High-Performance Computing (HPC) area. We have been researching and developing a high-performance custom computing with FPGA cluster as an off-loading engine for an existing HPC system. In this research poster, we describe the motivation of our research and present research topics on a software bridge between the FPGA cluster and existing HPC servers, and dedicated inter-FPGA networks.

Best Poster Finalist: no

Poster 69: Optimization for Quantum Computer Simulation
Naoki Yoshioka (RIKEN Center for Computational Science (R-CCS)), Hajime Inaoka (RIKEN Center for Computational Science (R-CCS)), Nobuyasu Ito (RIKEN Center for Computational Science (R-CCS)), Fengping Jin (Forschungszentrum Juelich), Kristel Michielsen (Forschungszentrum Juelich), Hans De Raedt (University of Groningen)

Simulator of quantum circuits is developed for massively parallel classical computers, and it is tested on the K computer in RIKEN R-CCS up to 45 qubits. Two optimization techniques are proposed in order to improve performance of the simulator. The "page method" reduces unnecessary copies in each node. It is found that this method makes approximately 17% speed-up maximum. Initial permutation of qubits is also studied how it affects performance of the simulator. It is found that a simple permutation in ascending order of the number of operations for each qubit is sufficient in the case of simulations of quantum adder circuits.

Best Poster Finalist: no

Poster 110: Hierarchical Data Prefetching in Multi-Tiered Storage Environments
Hariharan Devarajan (Illinois Institute of Technology), Anthony Kougkas (Illinois Institute of Technology), Xian-He Sun (Illinois Institute of Technology)

In the era of data-intensive computing, accessing data with a high-throughput and low-latency is very imperative. Data prefetching is used for hiding read latency by requesting data before it is needed to move it from a high-latency medium to a low-latency one. However, existing solutions do not consider the multi-tiered storage and also suffer from under-utilization of prefetching resources and unnecessary evictions. Additionally, existing approaches implement a client-pull model where understanding the application's I/O behavior drives prefetching decisions. Moving toward exascale, where machines run multiple applications concurrently by accessing files in a workflow, a more data-centric approach resolves challenges such as cache pollution and redundancy. In this study, we present HFetch, a truly hierarchical data prefetcher that adopts a server-push approach to data
prefetching. We demonstrate the benefits of such an approach. Results show 10-35% performance gains over existing prefetchers and over 50% when compared to systems with no prefetching.

Best Poster Finalist: yes

**Poster 52: Design and Specification of Large-Scale Simulations for GPUs Using FTX**

Anuva Kulkarni (Carnegie Mellon University), Daniele Spampinato (Carnegie Mellon University), Franz Franchetti (Carnegie Mellon University)

Large-scale scientific simulations can be ported to heterogeneous environments with GPUs using domain decomposition. However, Fast Fourier Transform (FFT) based simulations require all-to-all communication and large memory, which is beyond the capacity of on-chip GPU memory. To overcome this, domain decomposition solutions are combined with adaptive sampling or pruning around the domain to reduce storage. Expression of such operations is a challenge in existing FFT libraries like FFTW, and thus it is difficult to get a high performance implementation of such methods. We demonstrate algorithm specification for one such simulation (Hooke's law) using FTX, an emerging API with a SPIRAL-based code generation back-end, and suggest future extensions useful for GPU-based scientific computing.

Best Poster Finalist: no

**Poster 136: CHAMELEON: Reactive Load Balancing and Migratable Tasks for Hybrid MPI+OpenMP Applications**

Jannis Klinkenberg (RWTH Aachen University), Philipp Samfaß (Technical University Munich), Michael Bader (Technical University Munich), Karl Fürlinger (Ludwig Maximilian University of Munich), Christian Terboven (RWTH Aachen University), Matthias S. Müller (RWTH Aachen University)

Many HPC applications are designed based on underlying performance and execution models. These models could successfully be employed in the past for balancing load within and between compute nodes. However, the increasing complexity of modern software and hardware makes performance predictability and load balancing much more difficult. Tackling these challenges in search for a generic solution, we present a novel library for fine-granular task-based reactive load balancing in distributed memory based on MPI and OpenMP. Our concept allows creating individual migratable tasks that can be executed on any MPI rank. Migration decisions are performed at run time based on online performance or load data. Two fundamental approaches to balance load and at the same time overlap computation and communication are compared. We evaluate our concept under enforced power caps and clock frequency changes using a synthetic benchmark and demonstrate robustness.
Poster 124: Porting Finite State Automata Traversal from GPU to FPGA: Exploring the Implementation Space

Marziyeh Nourian (North Carolina State University), Mostafa Eghbali Zarch (North Carolina State University), Michela Becchi (North Carolina State University)

While FPGAs are traditionally considered hard to program, recently there are efforts to allow using high-level programming models intended for multi-core CPUs and GPUs to program FPGAs. For example, both Intel and Xilinx are now providing OpenCL-to-FPGA toolchains. However, since GPU and FPGA devices offer different parallelism models, OpenCL code optimized for GPU can prove inefficient on FPGA, in terms of both performance and hardware resource utilization.

In this poster, we explore this problem on an emerging workload: finite state automata traversal. Specifically, we explore a set of structural code changes, custom, and best-practice optimizations to retarget an OpenCL NFA engine designed for GPU to FPGA. Our evaluation, which covers traversal throughput and resource utilization, shows that our optimizations lead, on a single execution pipeline, to speedups up to 4x over an already optimized baseline that uses one of the proposed code changes to fit the original code on FPGA.

Poster 68: Linking a Next-Gen Remap Library into a Long-Lived Production Code

Charles R. Ferenbaugh (Los Alamos National Laboratory), Brendan K. Krueger (Los Alamos National Laboratory)

LANL’s long-lived production application xRage contains a remapper capability that maps mesh fields from its native AMR mesh to the GEM mesh format used by some third-party libraries. The current remapper was implemented in a short timeframe and is challenging to maintain. Meanwhile, our next-generation code project has developed a modern remapping library Portage, and the xRage team wanted to link in Portage as an alternate mapper option. But the two codes are very different from each other, and connecting the two required us to deal with a number of challenges. This poster describes the codes, the challenges we worked through, current status, and some initial performance statistics.
Poster 131: Efficiency of Algorithmic Structures
Julian Miller (RWTH Aachen University), Lukas Trümper (RWTH Aachen University), Christian Terboven (RWTH Aachen University), Matthias S. Müller (RWTH Aachen University)

The implementation of high-performance parallel software is challenging and raises issues not seen in serial programs before. It requires a strategy of parallel execution which preserves correctness but maximizes scalability. Efficiently deriving well-scaling solutions remains an unsolved problem especially with the quickly-evolving hardware landscape of high-performance computing (HPC).

This work proposes a framework for classifying the efficiency of parallel programs. It bases on a strict separation between the algorithmic structure of a program and its executed functions. By decomposing parallel programs into a hierarchical structure of parallel patterns, a high-level abstraction is provided which leads to equivalence classes over parallel programs. Each equivalence class possesses efficiency properties, mainly communication and synchronization, dataflow and architecture efficiency. This classification allows for wide application areas and a workflow for structural optimization of parallel algorithms is proposed.

Poster 98: INSPECT Intranode Stencil Performance Evaluation Collection
Julian Hammer (University of Erlangen-Nuremberg), Julian Hornich (University of Erlangen-Nuremberg), Georg Hager (University of Erlangen-Nuremberg), Thomas Gruber (University of Erlangen-Nuremberg), Gerhard Wellein (University of Erlangen-Nuremberg)

Modeling and presenting performance data---even for simple kernels such as stencils---is not trivial. We therefore present an overview on how to interpret and what to learn from an INSPECT report, as well as highlighting best practices for performance data reporting.

INSPECT is the "Intranode Stencil Performance Evaluation Collection", which compiles performance benchmarks and reports of various stencil and streaming kernels on a variety of architectures. The goal is to aid performance-aware developers with reference material and a methodology to analyze their own codes.

INSPECT set out to cover these topics and compile a summary of all necessary information to allow reproduction of the performance results, their interpretation and discussion.
**Poster 97: Optimizing Multigrid Poisson Solver of Cartesian CFD Code CUBE**

Kazuto Ando (RIKEN Center for Computational Science (R-CCS)), Rahul Bale (RIKEN), Keiji Onishi (RIKEN Center for Computational Science (R-CCS)), Kiyoshi Kumahata (RIKEN Center for Computational Science (R-CCS)), Kazuo Minami (RIKEN Center for Computational Science (R-CCS)), Makoto Tsubokura (Kobe University, RIKEN Center for Computational Science (R-CCS))

We demonstrate an optimization of multigrid Poisson solver of Cartesian CFD code "CUBE (Complex Unified Building cubE method)". CUBE is a simulation framework for complex industrial flow problem, such as aerodynamics of vehicles, based on hierarchical Cartesian mesh. In incompressible CFD simulation, solving pressure Poisson equation is the most time-consuming part. In this study, we use a cavity flow simulation as a benchmark problem. With this problem, multigrid Poisson solver dominates 91% of execution time of the time-step loop. Specifically, we evaluate the performance of Gauss-Seidel loop as a computational kernel based on “Byte per Flop” approach. With optimization of the kernel, we achieved 9.8x speedup and peak floating point performance ratio increased from 0.4% to 4.0%. We also measured parallel performance up to 8,192 nodes (65,536 cores) on the K computer. With optimization of the parallel performance, we achieved 2.9x–3.9x sustainable speedup in the time-step loop.

Best Poster Finalist: no

**Poster 85: Hybrid Computing Platform for Combinatorial Optimization with the Coherent Ising Machine**

Junya Arai (Nippon Telegraph and Telephone Corporation), Yagi Satoshi (Nippon Telegraph and Telephone Corporation), Hiroyuki Uchiyama (Nippon Telegraph and Telephone Corporation), Toshimori Honjo (Nippon Telegraph and Telephone Corporation), Takahiro Inagaki (Nippon Telegraph and Telephone Corporation), Kensuke Inaba (Nippon Telegraph and Telephone Corporation), Takuya Ikuta (Nippon Telegraph and Telephone Corporation), Hiroki Takesue (Nippon Telegraph and Telephone Corporation), Keitaro Horikawa (Nippon Telegraph and Telephone Corporation)

Several institutes are operating cloud platforms that offer Web API access to Ising computers such as quantum annealing machines. Platform users can solve complex combinatorial optimization problems by using hybrid algorithms that utilize both users' conventional digital computers and remote Ising computers. However, communication via the Internet takes an order of magnitude longer time than optimization on Ising computers. This overheads seriously degrade the performance of hybrid
algorithms since they involve frequent communication. In this poster, we first state issues in the design of Ising computing platforms, including communication overheads. Then, we answer the issues by introducing the computing platform for the coherent Ising machine (CIM), an Ising computer based on photonics technologies. Our platform offers efficient CIM-digital communication by allowing users to execute their program on digital computers co-located with the CIM. We have released the platform to our research collaborators in this autumn and started the evaluation.

Best Poster Finalist: no

**Poster 117: A New Polymorphic Computing Architecture Based on Fine-Grained Instruction Mobility**

David Hentrich (Illinois Institute of Technology), Erdal Oruklu (Illinois Institute of Technology), Jafar Saniie (Illinois Institute of Technology)

This is a summary of the base concepts behind David Hentrich’s May 2018 Ph.D. dissertation in Polymorphic Computing. Polymorphic Computing is the emerging field of changing the computer architecture around the software, rather than vice versa. The main contribution is a new polymorphic computing architecture. The key idea behind the architecture is to create an array of processors where a program’s instructions can be individually and arbitrarily assigned/mobilized to any processor, even during runtime. The key enablers of this architecture are a dataflow instruction set that is conducive to instruction migration, a microarchitectural block called an “operation cell” (“op-cell”), a processor built around the instruction set and the “op-cells”, and arrays of these processors.

Best Poster Finalist: no

**Poster 67: Genie: an MPEG-G Conformant Software to Compress Genomic Data.**

Brian E. Bliss (University of Illinois), Joshua M. Allen (University of Illinois), Saurabh Baheti (Mayo Clinic), Matthew A. Bockol (Mayo Clinic), Shubham Chandak (Stanford University), Jaime Delgado (Polytechnic University of Catalonia), Jan Fostier (Ghent University), Josep L. Gelpi (University of Barcelona), Steven N. Hart (Mayo Clinic), Michael T. Kalmbach (Mayo Clinic), Eric W. Klee (Mayo Clinic), Liudmila S. Mainzer (University of Illinois), Fabian Müntefering (Leibniz University), Daniel Naro (Barcelona Supercomputing Center), Idoia Ochoa-Alvarez (University of Illinois), Jörn Ostermann (Leibniz University), Tom Paridaens (Ghent University), Christian A. Ross (Mayo Clinic), Jan Voges (Leibniz University), Eric D. Wieben (Mayo Clinic), Mingyu Yang (University of Illinois), Tsachy Weissman (Stanford University), Mathieu Wiepert (Mayo Clinic)
Precision medicine has unprecedented potential for accurate diagnosis and effective treatment. It is supported by an explosion of genomic data, which continues to accumulate at accelerated pace. Yet storage and analysis of petascale genomic data is expensive, and that cost will ultimately be borne by the patients and citizens. The Moving Picture Experts Group (MPEG) has developed MPEG-G, a new open standard to compress, store, transmit and process genomic sequencing data that provides an evolved and superior alternative to currently used genomic file formats. Our poster will showcase software package GENIE, the first open source implementation of an encoder-decoder pair that is compliant with the MPEG-G specifications and delivers all its benefits: efficient compression, selective access, transport and analysis, guarantee of long-term support, and embedded mechanisms for annotation and encryption of compressed information. GENIE will create a step-change in medical genomics by reducing the cost of data storage and analysis.

Best Poster Finalist: no

Poster 82: A View from the Facility Operations Side on the Water/Air Cooling System of the K Computer
Jorji Nonaka (RIKEN Center for Computational Science (R-CCS)), Keiji Yamamoto (RIKEN Center for Computational Science (R-CCS)), Akiyoshi Kuroda (RIKEN Center for Computational Science (R-CCS)), Toshiyuki Tsukamoto (RIKEN Center for Computational Science (R-CCS)), Kazuki Koiso (Kobe University, RIKEN Center for Computational Science (R-CCS)), Naohisa Sakamoto (Kobe University, RIKEN Center for Computational Science (R-CCS))

The Operations and Computer Technologies Division at the RIKEN R-CCS is responsible for the operations of the entire K computer facility, which includes the auxiliary subsystems such as the power supply and water/air cooling systems. It is worth noting that part of these subsystems will be reused in the next supercomputer (Fugaku), thus a better understanding of the operational behavior as well as the potential impacts especially on the hardware failure and energy consumption would be greatly beneficial. In this poster, we will present some preliminary impressions of the impact of the water/air cooling system on the K computer system, focusing on the potential benefits of the use of low water/air temperature respectively for the CPU and DRAM memory modules produced by the cooling system. We expect that the obtained knowledge will be helpful for the decision support and/or operation planning of the next supercomputer.

Best Poster Finalist: no

Poster 135: High-Performance Deep Learning via a Single Building Block
Evangelos Georganas (Intel Corporation), Kunal Banerjee (Intel Corporation), Dhiraj Kalamkar (Intel
Deep learning (DL) is one of the most prominent branches of machine learning. Due to the immense computational cost of DL workloads, industry and academia have developed DL libraries with highly-specialized kernels for each workload/architecture, leading to numerous, complex code-bases that strive for performance, yet they are hard to maintain and do not generalize. In this work, we introduce the batch-reduce-GEMM kernel and show how the most popular DL algorithms can be formulated with this kernel as basic building-block. Consequently, the DL library-development degenerates to mere (potentially automatic) tuning of loops around this sole optimized kernel. By exploiting our kernel we implement Recurrent Neural Networks, Convolution Neural Networks and Multilayer Perceptron training and inference primitives in just 3K lines of high-level-code. Our primitives outperform vendor-optimized libraries on multi-node CPU-Clusters. We also provide CNN kernels targeting GPUs. Finally, we demonstrate that batch-reduce-GEMM kernel within a tensor compiler yields high-performance CNN primitives.

Best Poster Finalist: no

**Poster 56: Reinforcement Learning for Quantum Approximate Optimization**
Sami Khairy (Illinois Institute of Technology), Ruslan Shaydulin (Clemson University), Lukasz Cincio (Los Alamos National Laboratory), Yuri Alexeev (Argonne National Laboratory), Prasanna Balaprakash (Argonne National Laboratory)

The Quantum Approximate Optimization Algorithm (QAOA) is one of the leading candidates for demonstrating quantum advantage. The quality of the solution obtained by QAOA depends on the performance of the classical optimization routine used to optimize the variational parameters. In this work, we propose a Reinforcement Learning (RL) based approach to drastically reduce the number of evaluations needed to find high-quality variational parameters. We train an RL agent on small 8-qubit Max-Cut problem instances on an Intel Xeon Phi supercomputer Bebop, and use (transfer) the learned optimization policy to quickly find high-quality solutions for other larger problem instances coming from different distributions and graph classes. The preliminary results show that our RL based approach is able to improve the quality of the obtained solution by up to 10% within a fixed budget of function evaluations and demonstrate learned optimization policy transferability between different graph classes and sizes.

Best Poster Finalist: no

John Shalf (Lawrence Berkeley National Laboratory), Dilip Vasudevan (Lawrence Berkeley National Laboratory), David Donofrio (Lawrence Berkeley National Laboratory), Anastasia Butko (Lawrence Berkeley National Laboratory), Andrew Chien (University of Chicago), Yuanwei Fang (University of Chicago), Arjun Rawal (University of Chicago), Chen Zou (University of Chicago), Raymond Bair (Argonne National Laboratory), Kristopher Keipert (Argonne National Laboratory), Arun Rodriguez (Sandia National Laboratories), Maya Gokhale (Lawrence Livermore National Laboratory), Scott Lloyd (Lawrence Livermore National Laboratory), Xiaochen Guo (Lehigh University), Yuan Zeng (Lehigh University)

Accelerating technology disruptions and architectural change create growing opportunities and urgency to reduce the latency in for new architectural innovations to be deployed in extreme scale systems. We are exploring new architectural features that improve memory system performance including word-wise scratchpad memory, a flexible Recode engine, hardware message queues, and the data rearrangement engine (DRE). Performance results are promising yielding as much as 20x benefit. Project 38 is a cross-agency effort undertaken by the US Department of Energy (DOE) and Department of Defense (DoD).

Best Poster Finalist: no

**Poster 128: Identifying Time Series Similarity in Large-Scale Earth System Datasets**

Payton Linton (Youngstown State University), William Melodia (Youngstown State University), Alina Lazar (Youngstown State University), Deborah Agarwal (Lawrence Berkeley National Laboratory), Ludovico Bianchi (Lawrence Berkeley National Laboratory), Devarshi Ghoshal (Lawrence Berkeley National Laboratory), Kesheng Wu (Lawrence Berkeley National Laboratory), Gilberto Pastorello (Lawrence Berkeley National Laboratory), Lavanya Ramakrishnan (Lawrence Berkeley National Laboratory)

Scientific data volumes are growing every day and instrument configurations, quality control and software updates result in changes to the data. This study focuses on developing algorithms that detect changes in time series datasets in the context of the Deduce project. We propose a combination of methods that include dimensionality reduction and clustering to evaluate similarity measuring algorithms. This methodology can be used to discover existing patterns and correlations within a dataset. The current results indicate that the Euclidean Distance metric provides the best results in terms of internal cluster validity measures for multi-variable analyses of large-scale earth system datasets. The poster will include details on our methodology, results, and future work.
**Poster 151: Three-Dimensional Characterization on Edge AI Processors with Object Detection Workloads**  
Yujie Hui (Ohio State University), Jeffrey Lien (NovuMind Inc), Xiaoyi Lu (Ohio State University)

The Deep Learning inference applications are moving to the edge side, as edge-side AI platforms are cheap and energy-efficient. Different edge AI processors are diversified, since these processors are designed with different approaches. However, it is hard for customers to select an edge AI processor without an overall evaluation of these processors. We propose a three-dimensional characterization (i.e., accuracy, latency, and energy efficiency) approach on three different kinds of edge AI processors (i.e., Edge TPU, NVIDIA Xavier, and NovuTensor). We deploy YOLOv2 and Tiny-YOLO, which are two YOLO-based object detection systems, on these edge AI platforms with Microsoft COCO dataset. I will present our work starting from the problem statement. And then I'll introduce our experiments setup and hardware configuration. Lastly, I'll conclude our experimental results and current work status, as well as the future work.

**Poster 148: Unsupervised Clustering of Golden Eagle Telemetry Data**  

We use a recurrent autoencoder neural network to encode sequential California golden eagle telemetry data. The encoding is followed by an unsupervised clustering technique, Deep Embedded Clustering (DEC), to iteratively cluster the data into a chosen number of behavior classes. We apply the method to simulated movement data sets and telemetry data for a Golden Eagle. The DEC achieves better unsupervised clustering accuracy scores for the simulated data sets as compared to the baseline K-means clustering result.

**Poster 66: Hybrid CPU/GPU FE2 Multi-Scale Implementation Coupling Alya and Microp**  
Guido Giuntoli (Barcelona Supercomputing Center), Judicaël Grasset (Science and Technology Facilities Council (STFC)), Alejandro Figueroa (George Mason University), Charles Moulinec (Science and Technology Facilities Council (STFC)), Mariano Vázquez (Barcelona Supercomputing Center),
Guillaume Houzeaux (Barcelona Supercomputing Center), Stephen Longshaw (Science and Technology Facilities Council (STFC)), Sergio Oller (Polytechnic University of Catalonia)

This poster exposes the results of a new implementation of the FE2 multi-scale algorithm that is achieved by coupling the multi-physics and massively parallel code Alya with the GPU-based code micropp. The coupled code is mainly designed to solve large scale and realistic composite material problems for the aircraft industry. Alya is responsible of solving the macro-scale equations and micropp for solving the representation of fibres at the microscopic level. The poster shows computational performance results that demonstrate that the technique is scalable for real size industrial problems and also how the execution time is dramatically reduced using GPU-based clusters.

Best Poster Finalist: no

Poster 129: Understanding I/O Behavior in Scientific Workflows on High Performance Computing Systems

Fahim Tahmid Chowdhury (Florida State University, Lawrence Livermore National Laboratory), Francesco Di Natale (Lawrence Livermore National Laboratory), Adam Moody (Lawrence Livermore National Laboratory), Elsa Gonsiorowski (Lawrence Livermore National Laboratory), Kathryn Mohror (Lawrence Livermore National Laboratory), Weikuan Yu (Florida State University)

Leadership high performance computing (HPC) systems have the capability to execute workflows of scientific, research or industry applications. Complex HPC workflows can have significant data transfer and I/O requirements. Heterogeneous storage systems in supercomputers equipped with bleeding-edge non-volatile persistent storage devices can be leveraged to handle these data transfer and I/O requirements efficiently.

In this poster, we describe our efforts to extract the I/O characteristics of various HPC workflows and develop strategies to improve I/O performance by leveraging heterogeneous storage systems. We have implemented an emulator to mimic different types of I/O requirements posed by HPC application workflows. We have analyzed the workflow of Cancer Moonshot Pilot 2 (CMP2) project to determine possible I/O inefficiencies. To date, we have performed a systematic characterization and evaluation on the workloads generated by the workflow emulator and a small scale adaptation of the CMP2 workflow.

Best Poster Finalist: no
Poster 116: Advancements in Ultrasound Simulations Enabled by High-Bandwidth GPU Interconnects
Filip Vaverka (Brno University of Technology), Bradley E. Treeby (University College London), Jiri Jaros (Brno University of Technology)

Realistic ultrasound simulations are becoming integral part of many novel medical procedures such as photoacoustic screening and non-invasive treatment planning. The common denominator of all these applications is the need for cheap and relatively large-scale ultrasound simulations with sufficient accuracy. Typical medical applications require full-wave simulations which take frequency-dependent absorption and non-linearity into account.

This poster investigates the benefits of high-bandwidth low-latency interconnects to k-Wave acoustic toolbox in dense multi-GPU environment. The k-Wave multi-GPU code is based on a variant of the local Fourier basis domain decomposition. The poster compares the behavior of the code on a typical PCI-E 3.0 machine with 8 Nvidia Tesla P40 GPUs and a Nvidia DGX-2 server. The performance constraints of PCI-E platforms built around multiple socket servers on multi-GPU applications are deeply explored. Finally, it is shown the k-Wave toolbox can efficiently utilize NVlink 2.0 and achieve over 4x speedup compared to PCI-E systems.

Best Poster Finalist: no

Poster 65: Comparing Granular Dynamics vs. Fluid Dynamics via Large DOF-Count Parallel Simulation on the GPU
Milad Rakhsha (University of Wisconsin), Conlain Kelly (Georgia Institute of Technology), Nicholas Olsen (University of Wisconsin), Lijing Yang (University of Wisconsin), Radu Serban (University of Wisconsin), Dan Negrut (University of Wisconsin)

In understanding granular dynamics, the commonly-used discrete modeling approach that tracks the motion of all particles is computationally demanding, especially with large system size. In such cases, one can contemplate switching to continuum models that are computationally less expensive. In order to assess when such a discrete to continuum switch is justified, we compare granular and fluid dynamics that scales to handle more than 1 billion degrees of freedom (DOFs); i.e., two orders of magnitude higher than the state-of-the-art. On the granular side, we solve the Newton-Euler equations of motion; on the fluid side, we solve the Navier-Stokes equations. Both solvers leverage parallel computing on the GPU, and are publicly available on GitHub as part of an open-source code called Chrono. We report similarities and differences between the dynamics of the discrete, fully-resolved system and the continuum model via numerical experiments including both static and highly transient scenarios.
Poster 78: Understanding HPC Application I/O Behavior Using System Level Statistics
Arnab K. Paul (Virginia Tech), Olaf Faaland (Lawrence Livermore National Laboratory), Adam Moody (Lawrence Livermore National Laboratory), Elsa Gonsiorowski (Lawrence Livermore National Laboratory), Kathryn Mohror (Lawrence Livermore National Laboratory), Ali R. Butt (Virginia Tech)

The processor performance of high performance computing (HPC) systems is increasing at a much higher rate than storage performance. Storage and file system designers therefore require a deep understanding of how HPC application I/O behavior affects current storage system installations in order to improve storage performance. In this work, we contribute to this understanding using application-agnostic file system statistics gathered on compute nodes as well as metadata and object storage file system servers. We analyze file system statistics of more than 4 million jobs over a period of three years on two systems at Lawrence Livermore National Laboratory that include a 15 PiB Lustre file system for storage. Some key observations in our study show that more than 65% HPC users perform significant I/O which are mostly writes; and less than 22% of HPC users who submit write-intensive jobs perform efficient writes to the file system.

Poster 109: A Runtime Approach for Dynamic Load Balancing of OpenMP Parallel Loops in LLVM
Jonas H. Müller Korndörfer (University of Basel, Switzerland), Florina M. Ciorba (University of Basel, Switzerland), Akan Yilmaz (University of Basel, Switzerland), Christian Iwainsky (Technical University Darmstadt), Johannes Doerfert (Argonne National Laboratory), Hal Finkel (Argonne National Laboratory), Vivek Kale (Brookhaven National Laboratory), Michael Klemm (Intel Corporation)

Load imbalance is the major source of performance degradation in computationally-intensive applications that frequently consist of parallel loops. Efficient scheduling can improve the performance of such programs. OpenMP is the de-facto standard for parallel programming on shared-memory systems. The current OpenMP specification provides only three choices for loop scheduling which are insufficient in scenarios with irregular loops, system-induced interference, or both. Therefore, this work augments the LLVM OpenMP runtime library implementation with eleven ready to use scheduling techniques. We tested existing and added scheduling strategies on several applications from NAS, SPEC OMP 2012, and CORAL2 benchmark suites. Experiments show that implemented scheduling techniques outperform others in certain application and system configurations. We
measured performance gains of up to 6% compared to the fastest standard scheduling technique. This work aims to be a convincing step toward beyond-standard scheduling options in OpenMP for the benefit of evolving applications executing on multicore architectures.

Best Poster Finalist: no

**Poster 143: Quantum Natural Language Processing**

Lee James O'Riordan (Irish Centre for High-End Computing (ICHEC), National University of Ireland Galway (NUIG)), Myles Doyle (Irish Centre for High-End Computing (ICHEC), National University of Ireland Galway (NUIG)), Venkatesh Kannan (Irish Centre for High-End Computing (ICHEC), National University of Ireland Galway (NUIG)), Fabio Baruffa (Intel Corporation)

Natural language processing (NLP) algorithms that operate over strings of words are limited since they analyse meanings of the component words in a corpus without information about grammatical rules of the language. Consequently, they often produce unsatisfactory results with increase in problem complexity.

The "distributed compositional semantics" (DisCo) model incorporates grammatical structure of sentences into the algorithms, and offers significant improvements to the quality of results. However, their main challenge is the need for large classical computational resources. The DisCo model presents two quantum algorithms which lower storage and compute requirements compared to a classic HPC implementation.

In this project, we implement the two DisCo model quantum algorithms on the Intel Quantum Simulator deployed on the Irish national supercomputer. We target corpuses with ~1000 most-common words using up to 36 qubits simulation. The solution will be able to compute the meanings of two sentences and decide if their meanings match.

Best Poster Finalist: no

**Poster 152: Deep Domain Adaptation for Runtime Prediction in Dynamic Workload Scheduler**

Hoang H. Nguyen (National Center for Atmospheric Research (NCAR); University of Illinois, Chicago), Ben Matthews (National Center for Atmospheric Research (NCAR)), Irfan Elahi (National Center for Atmospheric Research (NCAR))

In HPC systems, users' requested runtime for submitted jobs plays a crucial role in efficiency. While underestimation of job runtime could terminate jobs before completion, overestimation could result
in long queuing of other jobs in HPC systems. In reality, runtime prediction in HPC is challenging due to the complexity and dynamics of running workloads. Most of the current predictive runtime models are trained on static workloads. This poses a risk of over-fitting the predictions with bias from the learned workload distribution. In this work, we propose an adaptation of Correlation Alignment method in our deep neural network architecture (DCORAL) to alleviate the domain shift between workloads for better runtime predictions. Experiments on both standard benchmark workloads and NCAR real-time production workloads reveal that our proposed method results in a more stable training model across different workloads with low accuracy variance as compared to the other state-of-the-art methods.

Best Poster Finalist: no

Poster 75: libCEED - Lightweight High-Order Finite Elements Library with Performance Portability and Extensibility
Jeremy Thompson (University of Colorado), Valeria Barra (University of Colorado), Yohann Dudouit (Lawrence Livermore National Laboratory), Oana Marin (Argonne National Laboratory), Jed Brown (University of Colorado)

High-order numerical methods are widely used in PDE solvers, but software packages that have provided high-performance implementations have often been special-purpose and intrusive. libCEED is a new library that offers a purely algebraic interface for matrix-free operator representation and supports run-time selection of implementations tuned for a variety of computational device types, including CPUs and GPUs. We introduce the libCEED API and demonstrate how it can be used in standalone code or integrated with other packages (e.g., PETSc, MFEM, Nek5000) to solve examples of problems that often arise in the scientific computing community, ranging from fast solvers via geometric multigrid methods to Computational Fluid Dynamics (CFD) applications.

Best Poster Finalist: no

Progress on the Exascale Transition of the VSim Multiphysics PIC code
Benjamin M. Cowan (Tech-X Corporation), Sergey N. Averkin (Tech-X Corporation), John R. Cary (Tech-X Corporation), Jarrod Leddy (Tech-X Corporation), Scott W. Sides (Tech-X Corporation), Ilya A. Zilberter (Tech-X Corporation)

The highly performant, flexible plasma simulation code VSim was designed nearly 20 years ago (originally as Vorpal), with its first applications roughly four years later. Using object oriented
methods, VSIm was designed to allow runtime selection from multiple field solvers, particle dynamics, and reactions. It has been successful in modeling for many areas of physics, including fusion plasmas, particle accelerators, microwave devices, and RF and dielectric structures. Now it is critical to move to exascale systems, with their compute accelerator architectures, massive threading, and advanced instruction sets. Here we discuss how we are moving this complex, multiphysics computational application to the new computing paradigm, and how it is done in a way that kept the application producing physics during the move. We present performance results showing significant speedups in all parts of the PIC loop, including field updates, particle pushes, and reactions.

Best Poster Finalist: no

**Poster 58: Lock-Free van Emde Boas Array**
Ziyuan Guo (University of Tokyo)

Lock-based data structures have some potential issues such as deadlock, livelock, and priority inversion, and the progress can be delayed indefinitely if the thread that is holding locks cannot acquire a timeslice from the scheduler. Lock-free data structures, which guarantees the progress of some method call, can be used to avoid these problems. This poster introduces the first lock-free concurrent van Emde Boas Array which is a variant of van Emde Boas Tree. It is linearizable, and the benchmark shows significant performance improvement comparing to other lock-free search trees when the date set is large and dense enough.

Best Poster Finalist: no

**Poster 63: Adaptive Execution Planning in Biomedical Workflow Management Systems**
Marta Jaros (Brno University of Technology), Bradley E. Treeby (University College London), Jiri Jaros (Brno University of Technology)

Biomedical simulations require very powerful computers. Their execution is described by a workflow consisting of a number of different cooperating tasks. The manual execution of individual tasks may be tedious for expert users, but prohibiting for most inexperienced clinicians. k-Dispatch offers a ‘run and forget’ approach where the users are completely screened out from the complexity of HPC systems. k-Dispatch provides task scheduling, execution, monitoring, and fault tolerance. Since the task execution configuration strongly affects the final tasks mapping on the computational resources, the execution planning is of the highest priority. Unlike other tools, k-Dispatch considers a variable amount of computational resources per individual tasks. Since the scaling of the individual
HPC codes is never perfect, k-Dispatch may find such a good mapping even an experienced user would miss. The proposed adaptive execution planning is based on collected performance data and the current cluster utilization monitoring.

Best Poster Finalist: no

**Poster 142: Training Deep Neural Networks Directly on Hundred-Million-Pixel Histopathology Images on a Large-Scale GPU Cluster**

Chi-Chung Chen (AetherAI, Taiwan), Wen-Yu Chuang (Chang-Gung Memorial Hospital, Taiwan), Wei-Hsiang Yu (AetherAI, Taiwan), Hsi-Ching Lin (National Center for High-Performance Computing (NCHC), Taiwan), Shuen-Tai Wang (National Center for High-Performance Computing (NCHC), Taiwan), Fang-An Kuo (National Center for High-Performance Computing (NCHC), Taiwan), Chao-Chun Chuang (National Center for High-Performance Computing (NCHC), Taiwan), Chao-Yuan Yeh (AetherAI, Taiwan)

Deep learning for digital pathology is challenging because the resolution of whole-slide-images (WSI) is extremely high, often in billions. The most common approach is patch-based method, where WSIs are divided into small patches to train convolutional neural networks (CNN). This approach has significant drawbacks. To have ground truth for individual patches, detailed annotations by pathologists are required. This laborious process has become the major impediment to the development of digital pathology AI. End-to-end WSI training, however, faces the difficulties of fitting the task into limited GPU memory. In this work, we improved the efficiency of using system memory for GPU compute by 411% through memory optimization and deployed the training pipeline on 8 nodes, totally 32 GPUs distributed system, achieving 147.28x speedup. We demonstrated that CNN is capable of learning features without detailed annotations. The trained CNN can correctly classify cancerous specimen, with performance level closely matching the patch-based methods.

Best Poster Finalist: no

**Poster 96: TSQR on TensorCores**

Hiroyuki Ootomo (Tokyo Institute of Technology), Rio Yokota (Tokyo Institute of Technology)

Tall-Skinny QR (TSQR) is an efficient algorithm for calculating the QR decomposition of m x n matrices where m << n, which is done by recursively performing QR decomposition on subdivided blocks of the tall and skinny matrix. Such operations are useful for low-rank approximation methods, which are replacing more and more dense linear algebra in both scientific computing and machine
learning fields. The present work focuses on the implementation of this important algorithm on Tensor Cores, which are available on the latest NVIDIA GPUs. We evaluate the speed, accuracy, and stability of TSQR on TensorCores.

Best Poster Finalist: yes

**Poster 95: A Heterogeneous HEVC Video Encoder Based on OpenPOWER Acceleration Platform**

Chenhao Gu (Fudan University, Shanghai, State Key Laboratory of ASIC and System; IBM Corporation), Yang Chen (Fudan University, Shanghai, State Key Laboratory of ASIC and System; IBM Corporation), Yanheng Lu (IBM Corporation), Pengfei Gou (IBM Corporation), Yong Lu (IBM Corporation), Yang Dai (IBM Corporation), Yue Xu (IBM Corporation), Yang Liu (IBM Corporation), Yibo Fan (Fudan University, Shanghai, State Key Laboratory of ASIC and System)

This poster describes a heterogeneous HEVC video encoder system based on the OpenPOWER platform. Our design leverages the Coherent Accelerator Processor Interface (CAPI) on the OpenPOWER, which provides cache-coherent access for FPGA. This technology highly improves CPU-FPGA data communication bandwidth and programming efficiency. X265 is optimized on the OpenPOWER platform to improve its performance with both architecture specific methods and hardware-acceleration methods. For hardware acceleration, frame-level acceleration and functional-unit-level acceleration are introduced and evaluated in this work.

Best Poster Finalist: no

**Poster 102: Fast Training of an AI Radiologist: Leveraging Data Pipelining to Efficiently Utilize GPUs**

Rakshith Vasudev (Dell EMC), John A. Lockman III (Dell EMC), Lucas A. Wilson (Dell EMC), Srinivas Varadharajan (Dell EMC), Frank Han (Dell EMC), Rengan Xu (Dell EMC), Quy Ta (Dell EMC)

In a distributed deep learning training setting, using accelerators such as GPUs can be challenging to develop a high throughput model. If the accelerators are not utilized effectively, this could mean more time to solution, and thus the model's throughput is low. To use accelerators effectively across multiple nodes, we need to utilize an effective data pipelining mechanism that handles scaling gracefully so GPUs can be exploited of their parallelism. We study the effect of using the correct pipelining mechanism that is followed by tensorflow official models vs a naive pipelining mechanism that doesn't scale well, on two image classification models. Both the models using the optimized data pipeline demonstrate effective linear scaling when GPUs are added. We also show
that converting to TF Records is not always necessary.

Best Poster Finalist: no

**Poster 94: Multi-GPU Optimization of a Non-Hydrostatic Numerical Ocean Model with Multigrid Preconditioned Conjugate Gradient Method**
Takateru Yamagishi (Research Organization for Information Science and Technology, Japan), Yoshimasa Matsumura (University of Tokyo, Atmosphere and Ocean Research Institute), Hiroyasu Hasumi (University of Tokyo, Atmosphere and Ocean Research Institute)

The conjugate gradient method with multigrid preconditioners (MGCG) is used in scientific applications because of its high performance and scalability with many computational nodes. GPUs are thought to be good candidates for accelerating such applications with many meshes where an MGCG solver could show high performance. No previous studies have evaluated and discussed the numerical character of an MGCG solver on GPUs. Consequently, we have implemented and optimized our “kinaco” numerical ocean model with an MGCG solver on GPUs. We evaluated its performance and discussed inter-GPU communications on a coarse grid on which GPUs could be intrinsically problematic. We achieved 3.9 times speedup compared to CPUs and learned how inter-GPU communications depended on the number of GPUs and the aggregation level of information in a multigrid method.

Best Poster Finalist: no

**Poster 108: Power Prediction for High-Performance Computing**
Shigeto Suzuki (Fujitsu Laboratories Ltd), Michiko Hiraoka (Fujitsu Ltd), Takashi Shiraishi (Fujitsu Laboratories Ltd), Enxhi Kreshpa (Fujitsu Laboratories Ltd), Takuji Yamamoto (Fujitsu Laboratories Ltd), Hiroyuki Fukuda (Fujitsu Laboratories Ltd), Shuji Matsui (Fujitsu Ltd), Masahide Fujisaki (Fujitsu Ltd), Atsuya Uno (RIKEN Center for Computational Science (R-CCS))

Exascale computers consume large amounts of power both for computing and cooling-units. As power of the computer varies dynamically corresponding to the load change, cooling-units are desirable to follow it for effective energy management. Because of time lags in cooling-unit operations, advance control is inevitable and an accurate prediction is a key for it. Conventional prediction methods make use of the similarity between job information while in queue. The prediction fails if there is no previously similar job. We developed two models to correct the prediction after queued jobs start running. By taking power histories into account, power-correlated topic model reselects more suitable candidate and recurrent-neural-network model considering
variable network sizes predicts power variation from shape features of it. We integrated these into a single algorithm and demonstrated high-precision prediction with an average relative error of 5.7% in K computer as compared to the 18.0% obtained using the conventional method.

Best Poster Finalist: no

**Poster 80: Sharing and Replicability of Notebook-Based Research on Open Testbeds**
Maxine V. King (University of Chicago), Jason Anderson (University of Chicago), Kate Keahey (Argonne National Laboratory, University of Chicago)

We seek to facilitate replicability by creating a way to share experiments easily in and out of notebook-based, open testbed environments and a sharing platform for such experiments in order to allow researchers to combine shareability, consistency of code environment, and well-documented process.

Best Poster Finalist: no

**Poster 121: HFlush: Realtime Flushing for Modern Storage Environments**
Jaime Cernuda (Illinois Institute of Technology), Hugo Trivino (Illinois Institute of Technology), Hariharan Devarajan (Illinois Institute of Technology), Anthony Kougkas (Illinois Institute of Technology), Xian-He Sun (Illinois Institute of Technology)

Due to the unparalleled magnitude of data movement in extreme scale computing, I/O has become a central challenge. Modern storage environments have proposed the use of multiple layers between applications and the PFS. Nonetheless, the difference in capacities and speeds between storage layers makes it extremely challenging to evict data from upper layers to lower layers efficiently. However, current solutions are executed in batches, compromising latency; are also hash-based implementations, compromising resource utilization. Hence, we propose HFlush, a continuous data eviction mechanism built on a streaming architecture that is pull-based and in which each component is decoupled and executed in parallel. Initial results have shown RFlush to obtain a 7X latency reduction and a 2X bandwidth improvement over a baseline batch-based system. Therefore, RFlush is a promising solution to the growing challenges of extreme scale data generation and eviction shortcomings when archiving data across multiple tiers of storage.

Best Poster Finalist: no
Linux containers are an increasingly popular method used by HPC centers to meet increasing demand for greater software flexibility. A common concern is that containers may introduce application performance overhead. Prior work has not tested a broad set of HPC container technologies on a broad set of benchmarks. This poster addresses the gap by comparing performance of the three HPC container implementations (Charliecloud, Shifter, and Singularity) and bare metal on multiple dimensions using industry-standard benchmarks.

We found no meaningful performance differences between the four environments with the possible exception of modest variation in memory usage, which is broadly consistent with prior results. This result suggests that HPC users should feel free to containerize their applications without concern about performance degradation, regardless of the container technology used. It is an encouraging development on the path towards greater adoption of user-defined software stacks to increase the flexibility of HPC.

Best Poster Finalist: no

We propose herein an approach for reformulating an equation-based modeling algorithm to an algorithm similar to that of training artificial intelligence (AI) and accelerate this algorithm using high-performance accelerators to reduce the huge computational costs encountered for physics equation-based modeling in earthquake disaster mitigation. A fast scalable equation-based implicit solver on unstructured finite elements is accelerated with a Tensor Core-enabled matrix-vector product kernel. The developed kernel attains 1.10 ExaFLOPS, leading to 416 PFLOPS for the whole solver on full Summit. This corresponds to a 75-fold speedup from a previous state-of-the-art solver running on full Piz Daint. This result could lead to breakthroughs in earthquake disaster
mitigation. Our new idea in the HPC algorithm design of combining equation-based modeling with AI is expected to have broad impacts in other earth science and industrial problems.

Best Poster Finalist: no

8:30 am - 5:00 pm

Scientific Visualization & Data Analytics Showcase Posters Display

Visualization of Entrainment and Mixing Phenomena at Cloud Edges
Bipin Kumar (Indian Institute of Tropical Meteorology), Matt Rehme (National Center for Atmospheric Research (NCAR)), Neethi Suresh (Indian Institute of Tropical Meteorology), Nihanth Cherukuru (National Center for Atmospheric Research (NCAR)), Stanislaw Jaroszynski (National Center for Atmospheric Research (NCAR)), Shaomeng Li (National Center for Atmospheric Research (NCAR)), Scott Pearse (National Center for Atmospheric Research (NCAR)), Tim Scheitlin (National Center for Atmospheric Research (NCAR))

Entrainment and mixing processes occur during the entire life of a cloud. These processes change the droplet size distribution, which determines rain formation and radiative properties. Since it is a microphysical process, it cannot be resolved in large scale weather forecasting models. Small scale simulations such as Direct Numerical Simulations (DNS) are required to resolve the smallest scale of these processes. However, it has been a challenge to visualize these processes in a 3D domain as it generates petabytes of data. Visualization plays a vital role in analyzing and understanding these huge data outputs. Here, we present different techniques for 3D visualization of data obtained from DNS carried out at Indian Institute of Tropical Meteorology (IITM) to understand cloud microphysical properties more closely.

Multiple tools were used to conduct a visual analysis of this data. Two of these tools are well established and tested technologies: ParaView and VAPOR. The others are emergent technologies created at the National Center for Atmospheric Research (NCAR) and are in the development phase. This data simulation, in addition to exploring DNS as mentioned above, provided an opportunity to test and improve development of several tools and methods.

Visualizing Deep Learning at Scale for the Construction of Galaxy Catalogs in the Dark Energy Survey
Janet Y. K. Knowles (Argonne National Laboratory), Joseph A. Insley (Argonne National Laboratory,
The advent of machine and deep learning algorithms on petascale supercomputers is accelerating the pace of discovery in astrophysics and poses significant challenges to the interpretability of these deep neural networks. We present a novel visualization of a deep neural network output during training as it is learning to classify galaxy images as either spiral or elliptical. The network is trained using labeled datasets from the citizen science campaign, Galaxy Zoo, adopted by the Sloan Digital Sky Survey. These trained neural network models can then be used to classify galaxies in the Dark Energy Survey that overlap the footprint of both surveys. Visualizing a reduced representation of the network output, projected into 3-D parameter space, reveals how the network has discovered two distinct clusters of features which allows it to classify galaxies into two groups. These visualizations of the neural network during training aid in the interpretability of the black box of deep learning and reveal how the network responds to the input images at various stages of training. Finally, it allows a wider net to be cast to a general audience, thereby generating interest in and visibility to an otherwise highly specialized field.

NVIDIA IndeX Accelerated Computing for Visualizing Cholla's Galactic Winds
Christopher Lux (Nvidia Corporation), Alexander Kuhn (Nvidia Corporation), Evan Schneider (Princeton University), Brant Robertson (University of California, Santa Cruz), Marc Nienhaus (Nvidia Corporation)

Galactic winds – outflows of gas driven out of galaxies by the combined effects of thousands of supernovae – are a crucial feature of galaxy evolution. Despite their importance, a complete theoretical picture of these winds has been elusive. Simulating the complicated interaction between the hot, high pressure gas created by supernovae and the cooler, high density gas in the galaxy disk requires massive computational resources and sophisticated software. For this purpose, Computational Hydrodynamics On Parallel Architectures (Cholla) has been demonstrated to be a scalable and efficient tool that operates in large, distributed multi-GPU environments at high levels of performance. This additional performance results in unprecedented resolution for this type of simulation and massive amounts of generated data. This raises the requirements for analysis tools that can cope with scale and complexity of the simulated physical processes. To address those challenges, we utilize NVIDIA IndeX as a scalable framework to visualize the simulation output: NVIDIA IndeX features a streaming-based architecture to interactively explore simulation results in large-scale, multi-GPU environments. We utilize customized sampling programs for multi-volume and surface rendering to address analysis questions of galactic wind simulations. This combination
of massively parallel simulation and analysis allows us to utilize recent supercomputer capabilities and to speed up the exploration of galactic wind simulations.

**Visualizing the World's Largest Turbulence Simulation**
Salvatore Cielo (Leibniz Supercomputing Centre), Luigi Iapichino (Leibniz Supercomputing Centre), Johannes Günther (Intel Corporation), Christoph Federrath (Australian National University, Research School of Astronomy and Astrophysics), Elisabeth Mayer (Leibniz Supercomputing Centre), Markus Wiedemann (Leibniz Supercomputing Centre)

We present the visualization of the largest interstellar turbulence simulations ever performed, unravelling key astrophysical processes concerning the formation of stars and the relative role of magnetic fields. The simulations, including pure hydrodynamical (HD) and magneto-hydrodynamical (MHD) runs, up to a size of \(10048^3\) grid elements, were produced on the supercomputers of the Leibniz Supercomputing Centre and visualized using the hybrid parallel (MPI + TBB) ray-tracing engine OSPRay associated with VisIt. Besides revealing features of turbulence with an unprecedented resolution, the visualizations brilliantly showcase the stretching-and-folding mechanisms through which astrophysical processes such as supernova explosions drive turbulence and amplify the magnetic field in the interstellar gas, and how the first structures, the seeds of newborn stars are shaped by this process.

**An Accessible Visual Narrative for the Primary Energy Source of Life from the Fulldome Show Birth of Planet Earth**
Melih Sener (University of Illinois), Stuart Levy (University of Illinois), AJ Christensen (University of Illinois), Robert Patterson (University of Illinois), Kalina Borkiewicz (University of Illinois), John E. Stone (University of Illinois), Barry Isralewitz (University of Illinois), Jeffrey Carpenter (University of Illinois), Donna Cox (University of Illinois)

Conversion of sunlight into chemical energy, namely photosynthesis, is the primary energy source of life on Earth. An explanatory visualization depicting this process is presented in the form of an excerpt from the fulldome show Birth of Planet Earth. This accessible visual narrative shows a lay audience, especially children, how the energy of sunlight is captured, converted, and stored through a chain of proteins to power living cells. The visualization is the result of a multi-year collaboration among biophysicists, visualization scientists, and artists, which, in turn, is based on a decade-long experimental-computational collaboration on structural and functional modeling that produced an atomic detail description of a bacterial bioenergetic organelle, the chromatophore. The energy conversion steps depicted feature an integration of function from electronic to cell levels, spanning nearly 12 orders of magnitude in time scales modeled with multi-scale computational approaches.
This atomic detail description uniquely enables a modern retelling of one of humanity’s earliest stories---the interplay between light and life.

**Visualizing Supersonic Retropropulsion for Mars: The Summit Early Science Results**
Patrick Moran (NASA Ames Research Center), Timothy Sandstrom (NASA Ames Research Center), Ashley Korzun (NASA Langley Research Center), Eric Nielsen (NASA Langley Research Center), Aaron Walden (NASA Langley Research Center)

As part of the Oak Ridge Leadership Computing Facility (OLCF) Early Science program, NASA ran a series of simulations on the new, GPU-accelerated system Summit, using the flow solver FUN3D to produce solutions for a set of representative descent conditions. The raw output from these simulations was over a petabyte of data. We produced visualizations at the NASA Advanced Supercomputing Division at Ames Research Center using parallel techniques where feasible to accelerate the process. Included with this summary is an animation with some of the highlights of our visualizations so far. We discuss some of our visualization design choices, and plans for future work.

5:15 pm - 7:00 pm

Poster Reception

**Wednesday, November 20**

8:30 am - 5:00 pm

ACM Student Research Competition Posters Display

**Poster 12: Investigating Input Sensitivity in GPU Graph Analytics with Machine Learning**
Jacob M. Hope (Texas State University), Trisha Nag (Texas Tech University)

Graph algorithms are at the core of data-intensive applications. As such, efficient graph processing is of great importance. Irregularity in real-world graphs can make performance unpredictable and non-portable across different inputs and architectures. Given a type of graph, the same optimized implementation of an algorithm can produce performance numbers that differ by orders-of-magnitude. We conduct extensive analysis on a set of 1238 graphs to identify input-sensitive
performance inefficiencies, including two that have not been studied: (i) register pressure and (ii) CPU-GPU data movement via demand paging. We then build a multiclass decision tree classifier that characterizes the irregular properties of graphs from our data and maps them to an optimal control parameterization at the compiler, system and algorithmic layers, that yield the highest overall algorithmic performance. We then integrate the classifier into a system where it will process a new graph and generate a kernel on the predicted optimal configuration.

**Poster 13: Scalable Load-Balancing Concurrent Queues on Many-Core Architectures**

*Caleb Lehman (Ohio State University)*

As the core counts of computing platforms continue to rise, parallel runtime systems with support for very fine-grained tasks become increasingly necessary to fully utilize the available resources. A critical feature of such task-based parallel runtime systems is the ability to balance work evenly and quickly across available cores. We highlight this by studying XTask, a custom parallel runtime system based on XQueue, which is a novel lock-less concurrent queuing system with relaxed ordering semantics that is geared to realizing scalability to hundreds of concurrent threads. We demonstrate the lack of adequate load balancing in the original XQueue design and present several solutions for improving load balancing. We also evaluate the corresponding improvements in performance on two sample workloads, computation of Fibonacci numbers and computation of Cholesky factorization. Finally, we compare the performance of several versions of XTask along with several implementations of the popular OpenMP runtime system.

**Poster 6: Analyzing the Performance of ZFP Compressed Arrays on HPC Kernels**

*Pavlo D. Triantafyllides (Clemson University)*

Per-node memory capacity limits the maximal problem size of HPC applications. Naïve data compression alleviates the memory impact, but requires full decompression before the data is accessed. ZFP compressed arrays reduce the memory footprint, by independently compressing data in fixed sized blocks. Thus, decompressing individual blocks and caching them enables random access and a reduction in decompressions on the critical path. The performance of ZFP compressed arrays is dependent on several key variables: software cache size, cache policy, and compression rate. In this poster, we explore the sensitivity of these ZFP parameters on runtime performance for the matrix-matrix multiplication algorithm. Results show that selection of cache size, policy, and rate yields 8% performance improvement over the default ZFP configuration.

**Poster 1: Early Experiences on OpenPOWER Architecture: Analysis of Billion-Scale Atomistic**
Datasets
Yuya Kawakami (Grinnell College, Oak Ridge National Laboratory)

We present performance analysis on OpenPOWER architecture of an algorithm to generate transversal views of atomistic models. The algorithm was implemented with data parallel primitives in NVIDIA Thrust for architecture portability. We report performance results on IBM Power9 CPUs (OpenMP, Intel Threading Blocks) and NVIDIA Volta GPUs (single and multi GPU). We also evaluate CUDA unified memory performance, exposed by NVIDIA RAPIDS Memory Manager library (RMM).

Poster 2: An Efficient Parallel Algorithm for Dominator Detection
Daniel Giger (University of Massachusetts, Lowell), Hang Liu (Stevens Institute of Technology)

In graph theory, a vertex v dominates a vertex u if every path from the entry vertex to u must go through vertex v. This algorithm is called dominator detection and holds a wide range of applications, such as compiler design, circuit testing, and social network analysis. While the performance of many other graph algorithms soars with respect to the increase of the hardware parallelism, dominator detection algorithm experiences very little advancement due to the hardship of parallelism. This work thus introduces an efficient parallel dominator detection algorithm that is inspired by Breadth-First Search (BFS), which bests SEMI-NCA on large graphs.

Poster 9: Machine Specific Symbolic Code Generation
Robert King (University of Utah)

New discoveries in science and engineering are primarily driven by numerical simulations of underlying governing equations. Manually written codes to evaluate the main computational kernels lack portability, prone to human errors, ability to perform code optimizations due to the complexity of the underlying equations. In this work we present a symbolic code generation framework, which generates architecture optimized code for different platforms. As the driving application, we primarily use computational relativity where computations of Einstein equations become complicated due to the presence of curvature in spacetime.

The presented framework is based on SymPy with additional modules written to handle complicated partial differential equations (PDEs). The symbolically written equations are converted to a computational graph, which enables to perform architecture (cache, register optimizations) and language specific (SIMD vectorization, CUDA) optimizations. By computing common subexpressions, we reduce the number of compute operations needed, by storing them in temporary variables.
Poster 5: Evaluating Lossy Compressors for Inline Compression
Donald T. Elmore (Clemson University)

HPC applications require massive amounts of data. The data required is growing faster than memory capabilities. An example of this is pySDC, a framework for solving collocation problems iteratively using parallel-in-time methods. pySDC requires storing and exchange of 3D volume data for each parallel point in time. We evaluate several state-of-the-art lossy compressors such as SZ and ZFP for their applicability to inline compression for pySDC. We evaluate the compressors based on compression ratio, compression bandwidth, decompression bandwidth, and overall error introduced.

This poster evaluates state-of-the-art lossy compressors for use in pySDC; shows lossy data compression is an effective tool for reducing memory requirements for pySDC; and highlights current compression/decompression bandwidth is not fast enough for inline lossy compression yet. Results show using SZ with an error bound of $1e^{-5}$, we reduce the memory footprint by a factor of $311.99$ while maintaining an acceptable level of loss.

Poster 3: A Deep Learning Approach to Noise Prediction and Circuit Optimization for Near-Term Quantum Devices
Alexander Zlokapa (California Institute of Technology), Alexandru Gheorghiu (California Institute of Technology)

Noisy intermediate-scale quantum (NISQ) devices face challenges in achieving high-fidelity computations due to hardware-specific noise. As a basis for noise mitigation, we develop a convolutional neural network noise model to estimate the difference in noise between a given pair of equivalent quantum circuits. On a classically simulated dataset of $1.6$ million pairs of quantum circuits with a simplified noise model calibrated to IBM Q hardware, the deep learning approach shows a significant improvement in noise prediction over linear gate count models. A greedy peephole optimization procedure is proposed to minimize noise using the deep learning model as an objective function, showing further improvement in noise mitigation compared to commonly used gate count minimization heuristics.

Poster 10: Learning Page Access Patterns for Algorithms Programmed with GPU UVM
Bennett Cooper (Clemson University), Derek Rodriguez (Clemson University)
We aim to mitigate the performance bottleneck of migrating data between host and device memory in GPU applications by accurately predicting application access patterns using deep neural networks. We model the memory access pattern of any given application by collecting page faults that trigger data migration to the GPU and feed this time series as input to a neural network that outputs the next several page faults. We evaluate these predictions on the basis of what makes a useful prefetch in our context for the GPU. Current work has looked at trivial GPU applications, such as matrix operations, and moving toward real, complex applications. Our work will be presented by bringing attention to the predictive capability of our neural network on the current applications tested.

**Poster 8: Mitigating Communication Bottlenecks in MPI-Based Distributed Learning**

Abdullah B. Nauman (Ward Melville High School, Brookhaven National Laboratory)

Current commercial and scientific facilities generate and maintain vast amounts of complex data. While machine learning (ML) techniques can provide crucial insight, developing these models is often impractical on a single process. Distributed learning techniques mitigate this problem; however, current models contain significant performance bottlenecks. Here, we conduct a detailed performance analysis of MPI_Learn, a widespread distributed ML framework for high-energy physics (HEP) applications, on the Summit supercomputer, by training a network to classify simulated collision events from high-energy particle detectors at the CERN Large Hadron Collider (LHC).

We conclude that these bottlenecks occur as a result of increasing communication time between the different processes, and to mitigate the bottlenecks we propose the implementation of a new distributed algorithm for stochastic gradient descent (SGD). We provide a proof of concept by demonstrating better scalability with results on 250 GPUs, and with hyperparameter optimization, show a ten-fold decrease in training time.

**Hearing Single- and Multi-Threaded Program Behavior**

Mark Wissink (Calvin University), Joel Adams (Calvin University)

There are many examples of visual program representation, but there is an unexplored area of sonic representation—audialization—of programs. The Thread Safe Audio Library (TSAL) provides the necessary tools to create such program audializations. TSAL is a platform-independent, object-oriented C++ library that provides thread safe classes for synthesizing sound from a program. By adding TSAL calls to a sequential or parallel program, its behavior can be represented with sound. Audializations require a unique interpretation that may greatly benefit auditory learners and other
Poster 11: Walking the Cost-Accuracy Tightrope: Balancing Trade-Offs in Data-Intensive Genomics
Kathryn Leung (Princeton University), Meghan Kimball (DePaul University)

Scientific applications often exhibit a trade-off between cost and accuracy. However, measuring and predicting cost and accuracy in a way that users can understand these trade-offs is challenging. To address these needs, we present predictive cost and accuracy models for data-intensive genomics applications. We use these models to create a trade-off graph, which researchers can use to selectively trade-off cost and accuracy.

Poster 31: Designing High-Performance Erasure Coding Schemes for Next-Generation Storage Systems
Haiyang Shi (Ohio State University), Xiaoyi Lu (Ohio State University)

Erasure Coding (EC) NIC offload is a promising technology for designing next-generation distributed storage systems. To alleviate the EC overhead, various hardware-based Erasure Coding schemes have been proposed to leverage the advanced compute capabilities on modern HPC clusters. However, this poster has identified some critical limitations of current-generation EC NIC offload schemes on modern SmartNICs and the missing of exploiting heterogeneous hardware available on modern HPC clusters in parallel. To address these limitations, this poster first proposes a unified benchmark suite to benchmark, measure and characterize hardware-optimized erasure coders. Then, it proposes a Multi-Rail EC concept which enables upper-layer applications to leverage heterogeneous hardware to perform EC operations simultaneously. Finally, it proposes a new EC NIC offload paradigm based on the tripartite graph model, namely TriEC. TriEC supports both encode-and-send and receive-and-decode offload primitives efficiently.

Poster 27: A Framework for Measuring Hardware Gather-Scatter Support
Patrick Lavin (Georgia Institute of Technology), Jeffrey Young (Georgia Institute of Technology), Richard Vuduc (Georgia Institute of Technology)

In recent years, we have seen the re-addition of vector units to CPUs. While these units easily give speedups for easily vectorized applications with dense memory access, it can be hard to characterize how different access patterns will affect the performance of vectorized code.
We have developed Spatter, a benchmark which allows us to test and investigate the gather-scatter units available on current and upcoming hardware. The information that Spatter reveals to users is of use to everyone from hardware vendors who wish to compare gather-scatter units across platforms, to application developers who wish to test memory access pattern performance in their vectorized code.

**Poster 30: A Fast and Efficient Incremental Approach toward Dynamic Community Detection**

Neda Zarayeneh (Washington State University)

Community detection is a discovery tool to identify groups of vertices which are tightly connected, called communities. Most of the literature on this subject caters to the static use-case where the underlying network does not change. However, many emerging real-world use-cases give rise to a need to incorporate dynamic graphs as inputs.

We present a fast and efficient incremental approach toward dynamic community detection. The key contribution is a generic technique called $\Delta$-screening, which examines the most recent batch of changes made to an input graph and selects a subset of vertices to reevaluate for potential community (re)assignment. This technique can be incorporated into any of the community detection methods that use modularity as its objective function for clustering. For demonstration purposes, we incorporated the technique into two well-known community detection tools. Our experiments demonstrate that our approach is able to generate performance speedups without compromising on the output quality.

**Poster 29: Optimal Routing for a Family of Scalable Interconnection Networks**

Zhipeng Xu (Stony Brook University; Sun Yat-sen University, Guangzhou, China)

We propose a scheme to construct a family of large and high-performance interconnection networks that are scalable, low-radix, minimum diameters. These networks, whose diameters grow linearly as their sizes grow exponentially, are generated by using the Cartesian products of smaller optimal networks of minimum diameters. For the smaller base networks, we design the vertex-balanced routing algorithm by considering the forwarding pressure at each vertex. Comparative benchmarks on a Beowulf cluster show significant improvement in performance after using the new routing algorithm. Each node of the new network generated from base graphs with low-diameter can also sustain balanced forwarding loadings if we apply optimal routing algorithms to the base network. Simulation results for larger networks show that the optimal routing algorithms achieve the gain of communication performance.
Poster 16: perf-taint: Taint Analysis for Automatic Many-Parameter Performance Modeling
Marcin Copik (ETH Zurich)

Performance modeling is a well-known technique for understanding the scaling behavior of an application. Although the modeling process is today often automatic, it still relies on a domain expert selecting program parameters and deciding relevant sampling intervals. Since existing empirical methods attempt blackbox modeling, the decision on which parameters influence a selected part of the program is based on measured data, making empirical modeling sensitive to human errors and instrumentation noise. We introduce a hybrid analysis to mitigate the current limitations of empirical modeling, combining the confidence of static analysis with the ability of dynamic taint analysis to capture the effects of control-flow and memory operations. We construct models of computation and communication volumes that help the modeler to remove effects of noise and improve the correctness of estimated models. Our automatic analysis prunes irrelevant program parameters and brings an understanding of parameter dependencies which helps in designing the experiment.

Poster 24: Fingerprinting Anomalous Computation with RNN for GPU-Accelerated HPC Machines
Pengfei Zou (Clemson University)

This paper presents a workload classification framework that discriminates illicit computation from authorized workloads on GPU-accelerated HPC systems. As such systems become more and more powerful, they are exploited by attackers to run malicious and for-profit programs that typically require extremely high computing ability to be successful. Our classification framework leverages the distinctive signatures between illicit and authorized workloads, and explore machine learning methods to learn the workloads and classify them. The framework uses lightweight, non-intrusive workload profiling to collect model input data, and explores multiple machine learning methods, particularly recurrent neural network (RNN) that is suitable for online anomalous workload detection. Evaluation results on four generations of GPU machines demonstrate that the workload classification framework can tell apart the illicit authorized workloads with high accuracy of over 95%. The collected dataset, detection framework, and neural network models will be made available on GitHub.

Poster 33: Enhancing Neural Architecture Search with Speciation and Inter-Epoch Crossover
Matthew E. Baughman (University of Chicago)

As deep learning continues to expand into new areas of application, the demand for efficient use of our HPC resources increase. For new problem domains, new model architectures are developed
through a neural architecture search (NAS), which consist of iteratively training many neural networks. To combat the computational waste and maximize compute efficiency for NAS, we demonstrate that the use of genetic algorithms with speciation can be used to both shorten training time and increase accuracy at each iteration.

**Poster 14: Data Reuse Analysis for GPU Offloading Using OpenMP**
Alok Mishra (Stony Brook University, Lawrence Livermore National Laboratory)

More researchers and developers desire to port their applications to GPU-based clusters, due to their abundant parallelism and energy efficiency. Unfortunately porting or writing an application for accelerators, such as GPUs, requires extensive knowledge of the underlying architectures, the application/algorithm and the interfacing programming model (e.g. OpenMP). Often applications spend a significant portion of their execution time on data transfer. Exploiting data reuse opportunities in an application can reduce its overall execution time. In this research, we present an approach to automatically recognize data reuse opportunities in an application which uses OpenMP for exploiting GPU parallelism, and consequently insert pertinent code to take advantage of data reuse on GPU. Using our approach we were able to retain reused data on the GPU and reduce the overall execution time of multiple benchmark application.

**Poster 17: Exploiting Multi-Resource Scheduling for HPC**
Yuping Fan (Illinois Institute of Technology)

High-performance computing (HPC) is undergoing significant changes. Next generation HPC systems are equipped with diverse global/local resources. HPC job scheduler plays a crucial role in efficient use of resources. However, traditional job schedulers are single-objective and fail to efficient use of other resources. In our previous work, we present a job scheduling framework named BBsched to schedule CPUs and burst buffers. As we are heading toward exascale computing, a variety of heterogeneous resources are deployed in HPC systems. In this poster, we extend BBsched for managing multiple resources beyond CPUs and burst buffers. We formulate multi-resource scheduling as a general multi-objective optimization (MOO) problem, present a heuristic method to solve the NP-hard MOO problem, and provide a preliminary evaluation for scheduling up to ten resources. The proposed multi-resource scheduling design is intended to enhance Cobalt, a production job scheduler deployed on HPC systems at Argonne Leadership Computing Facility (ALCF).

**Poster 15: Cost-Aware Cloudlet Placement in Edge Computing Systems**
Dixit Bhatta (University of Delaware)

A well-known challenge in Edge Computing is strategic placement of cloudlets. The fundamental goals of this challenge are to minimize the cloudlet deployment cost and to guarantee minimum latency to the users of edge services. We address this challenge by designing a cost-aware cloudlet placement approach that ensures user latency requirements while covering all devices in the service region. We first mathematically formulate the problem as a multi-objective integer programming model in a general deployment scenario, which is computationally NP-hard. We then propose a genetic algorithm-based approach, GACP, to find heuristic solutions in significantly reduced time. We investigate the effectiveness of GACP by performing extensive experiments on multiple deployment scenarios based on New York City OpenData. The results presented in the poster show that our approach obtains close to optimal cost solutions with significant time reduction.

Poster 19: Accelerating Real-World Stencil Computations Using Temporal Blocking: Handling Sparse Sources and Receivers
George Bisbas (Imperial College, London)

This paper concerns performance optimization in finite-difference solvers found in seismic imaging. We tackle a complicating factor that is arising in seismic inversion problems: the addition of waves injected from sources distributed sparsely over the 2D/3D domain, and the need for receivers that interpolate data measurements at a set of points, again distributed sparsely across the domain. In this work we show how to overcome this limitation. We introduce an algorithm for a loop nest transformation policy applied to wave propagation in order to improve data locality and optimize our cache memory use. Our algorithm uses an inspector/executor scheme capable of inspecting measurement operations at sparse locations and then compute the field updates. The ultimate goal of this ongoing work is to automate this scheme for stencil codes generated by Devito.

Poster 23: PERQ: Fair and Efficient Power Management of Power-Constrained Large-Scale Computing Systems
Tirthak Patel (Northeastern University), Devesh Tiwari (Northeastern University)

Large-scale computing systems are becoming increasingly more power-constrained, but these systems employ hardware over-provisioning to achieve higher system throughput because HPC applications often do not consume the peak power capacity of nodes. Unfortunately, focusing on system throughput alone can lead to severe unfairness among multiple concurrently-running applications. This paper introduces PERQ, a new feedback-based principled approach to improve system throughput while achieving fairness among concurrent applications.
**Poster 18: Using Non Volatile Memories to Build Energy- and Cost-Efficient Clusters**  
Onkar Patil (North Carolina State University)

Non-Volatile Memory (NVM) is a byte-addressable, high capacity, high latency, and persistent form of memory that can extend the primary memory hierarchy by another level. It allows clusters to have significantly greater memory capacity per node. Intel’s Optane DC Persistent Memory Module (PMM) is a NVM device that can be used to increase the memory density of high performance computing (HPC) system. With higher memory density, we can run scientific computing applications with larger problem sizes on fewer compute nodes than on current HPC systems. Smaller HPC clusters will reduce the cost of running scientific simulations. We test our hypothesis by running a HPC application with large problem sizes on a single node with NVM and on multiple nodes connected by a high speed interconnect with equivalent amount of volatile memory. We compare the performance characteristics and power consumption of both systems.

**Poster 22: Fast Profiling-Based Performance Modeling of Distributed GPU Applications**  
Jaemin Choi (University of Illinois, Lawrence Livermore National Laboratory)

An increasing number of applications utilize GPUs to accelerate computation, with MPI responsible for communication in distributed environments. Existing performance models only focus on either modeling GPU kernels or MPI communication; few that do model the entire application are often too specialized for a single application and require extensive input from the programmer.

To be able to quickly model different types of distributed GPU applications, we propose a profiling-based methodology for creating performance models. We build upon the roofline performance model for GPU kernels and analytical models for MPI communication, with a significant reduction in profiling time. We also develop a benchmark to model 3D halo exchange that occurs in many scientific applications. Our proposed model for the main iteration loops of MiniFE achieves 6-7% prediction error on LLNL Lassen and 1-2% error on PSC Bridges, with minimal code inspection required to model MPI communication.

**Poster 26: Neural Networks for the Benchmarking of Detection Algorithms**  
Silvia Miramontes (Lawrence Berkeley National Laboratory; University of California, Berkeley)

There are several automated methods to detect objects from grayscale images. However, materials scientists still lack basic tools to compare different detection results, particularly when working with
microtomography. This poster introduces FibCAM, a convolutional neural network (CNN)-based method using TensorFlow that allows benchmarking fiber detection algorithms. Our contribution is three-fold: (a) the design of a computational framework to compare automated fiber detection models with curated datasets through classification; (b) lossless data reduction by embedding prior knowledge into data-driven models; (c) a scheme to decompose computation into embarrassingly parallel processes for future analysis at scale. Our results show how FibCAM classifies different structures, and how it illustrates the material's composition and frequency distribution of microstructures for improved interpretability of machine learning models. The proposed algorithms support probing the specimen content from gigabyte-sized volumes and enable pinpointing inconsistencies between real structures known a priori and results derived from automated detections.

**Poster 21: Optimization of GPU Kernels for Sparse Matrix Computations in Hypre**  
Chaoyu Zhang (Arkansas State University)

The acceleration of sparse matrix computations on GPUs can significantly enhance the performance of iterative methods for solving linear systems. In this work, we consider the kernels of Sparse Matrix Vector Multiplications (SpMV), Sparse Triangular Matrix Solves (SpTrSv) and Sparse Matrix Matrix Multiplications (SpMM), which are often demanded by Algebraic Multigrid (AMG) solvers. With the CUDA and the hardware support of the Volta GPUs on Sierra, the existing kernels should be further optimized to fully take the advantage of the new hardware, and the optimizations have shown significant performance improvement. The presented kernels have been put in HYPRE for solving large scale linear systems on HPC equipped with GPUs. These shared-memory kernels for single GPU are the building blocks of distributed matrix operations required by the solver across multiple GPUs and compute nodes. The implementations of these kernels in Hypre and the code optimizations will be discussed.

**Poster 20: From IoT to Cloud: Research Platform for IoT/Cloud Experiments**  
Jinfeng Lin (University of Notre Dame)

IoT studies leverage a wide range of lightweight hardware for collecting and processing data in the field. Limited by the resources on the devices, IoT systems have to interoperate with cloud platforms for addressing computation intensive tasks such as image processing, application backend supporting and centralized data storage. Therefore, a testbed for IoT/Cloud experiments should provide infrastructure for IoT to cloud communication, computation deployment, and hardware resource management. With these functionalities, users can focus on research problems without distraction from manually constructing experiment environments. Though cloud providers such as
Google, Amazon, and Microsoft all provide IoT to Cloud solutions in general, this commercial model is not entirely compatible with research purpose. We propose a framework named Chameleon IoT testbed (CHIoT) that extends the Chameleon bare-metal cloud resources to support general IoT to Cloud experiments.

**Poster 32: OSU INAM: A Profiling and Visualization Tool for Scalable and In-Depth Analysis of High-Performance GPU-Enabled HPC Clusters**

Pouya Kousha (Ohio State University)

The lack of low-overhead and scalable monitoring tools have prevented a comprehensive study of efficiency and utilization of emerging NVLink-enabled GPU clusters. We address this by proposing and designing an in-depth, real-time analysis, profiling, and visualization tool for high-performance GPU-enabled clusters with NVLinks on the top of the OSU INAM. The proposed tool is capable of presenting a unified and holistic view of MPI-level and fabric level information for emerging NVLink-enabled high-performance GPU clusters. It also provides insights into the efficiency and utilization of underlying interconnects for different communication patterns. We also designed a low overhead and scalable modules to discover the fabric topology and gather fabric metrics by using different levels of threading, bulk insertions and deletions for storage, and using parallel components for fabric discovery and port metric inquiry.


Zheng Miao (Clemson University)

High-Performance Computing systems must simultaneously address both resilience and power. In heterogeneous systems, the trade-offs between resilience and energy-efficiency are more complex for applications using both CPUs and GPUs. A deep understanding of the interplay among energy efficiency, resilience, and performance is required for heterogeneous systems to address them simultaneously.

In this work, we present a new framework for resilient and energy-efficient computing in GPU-accelerated systems. This framework supports partial or full redundancy and checkpointing for resilience, and provides users with flexible hardware resource selection, adjustable precision and power management to improve performance and energy-efficiency. We further perform CUDA-aware MPI to reduce resilience overhead, mainly in message communication between GPUs. Using CG as an example, we show that our framework provides about 40% time and 45% energy savings, comparing to simple extension of RedMPI, a redundancy based resilience framework for
homogeneous CPU systems.

**Poster 25: Leveraging Smart Data Transfer and Metadirective in Adaptive Computing**  
Anjia Wang (University of North Carolina, Charlotte; Lawrence Livermore National Laboratory)

In this work, we propose smart data transfer (SDT) along with extensions to metadirective and map constructs in OpenMP 5.0 to improve adaptive computing. The Smith-Waterman algorithm is used as an example, whose naïve implementation does not conduct data transfer efficiently. SDT is used to solve this issue with the following advantages: (1) SDT only transfers necessary data to GPU instead of the whole data set, resulting in 4.5x of speedup in our initial experiments. (2) Depending on computing vs. data transfer requirements of a program, SDT will transfer the output of each iteration from GPU to host either immediately or all together after the last GPU kernel call. (3) It supports large data exceeding GPU device memory's size via data tiling. We propose to extend metadirective's context selector to obtain similar improvement by enabling target enter/exit data and on-demand data access.

8:30 am - 5:00 pm

**Doctoral Showcase Posters Display**

**Poster 36: Modeling Non-Determinism in HPC Applications**  
Dylan Chapp (University of Delaware, University of Tennessee)

As HPC applications migrate from the petascale systems of today to the exascale systems of tomorrow, the increasing need to embrace asynchronous, irregular, and dynamic communication patterns will lead to a corresponding decrease in application-level determinism. Two critical challenges emerge from this trend. First, unchecked non-determinism coupled with the non-associativity of floating-point arithmetic undermines numerical reproducibility of scientific applications. Second, the prevalence of non-determinism amplifies the cost of debugging, both in terms of computing resources and human effort. In this thesis, we present a modeling methodology to quantify and characterize communication non-determinism in parallel applications. Our methodology consists of three core components. First, we build graph-structured models of relevant communication events from execution traces. Second, we apply similarity metrics based on graph kernels to quantify run-to-run variability and thus identify the regions of executions where non-determinism manifests most prominently. Third, we leverage our notion of execution similarity to characterize applications via clustering, anomaly detection, and extraction of representative
patterns of non-deterministic communication which we dub "non-determinism motifs". Our work will amplify the effectiveness of software tools that target mitigation or control of application-level non-determinism (e.g., record-and-replay tools) by providing them with a common metric for quantifying communication non-determinism in parallel applications and a common language for describing it.

**Poster 35: Scaling Up Pipeline Optimization with High Performance Computing**
Robert Lim (University of Oregon)

My research focuses on developing a pipeline optimization infrastructure that automates the design and code generation of neural networks through the use of high-performance computing. The problem has the following objectives: unify automated machine learning (AutoML) and compilation, archive profiles for creation of a knowledge base for a data-driven approach toward search, explore various search optimizations for model design and code generation. The field of automated deep learning includes hyperparameter optimization and neural architecture search (NAS), which requires domain expertise in designing a model, in addition to the tuning parameters related to learning and the model itself. The search space is complex and deciding which parameters factor into the overall accuracy of a model is a non-trivial task. Once a model is trained, the next step compiles the model, which maps to the backend of a targeted architecture, whether GPU, embedded mobile phones, or FPGA. The compilation phase also involves choices, in terms of optimizations applied, which can include a set of flags, or direct code transformations. Various efficient search techniques are explored, archiving results along the way for facilitating in code generation options, which could possibly inform the model design process. Our previous work reduced the search space for code generation of various domain kernels by 92%, and this work investigates whether the same approach can be applied in all aspects of AutoML design and code generation.

**Poster 38: High-Performance Backpropagation in Scientific Computing**
Navjot Kukreja (Imperial College, London)

Devito is a domain-specific language for the automatic generation of high-performance solvers for finite difference equations provided in a high-level symbolic representation. The primary use of Devito is to enable Full-waveform inversion, which is an adjoint-based optimization problem that uses the wave equation as part of its objective function. This doctoral project consists of three elements of Devito that are key to solving adjoint-based optimization problems (i.e. the back propagation). The first is automatic differentiation of stencil loops. This allows the automatic generation of high-performance code implementing the derivative of any provided function. This is essential to enable the use of a wider range of physical equations - in order to use better physical
approximations. An essential feature of the generated derivatives is that they show scaling and performance behaviour that is very similar to the forward function. A common issue in solving adjoint-based optimization problems is the prohibitively high memory requirement. The second contribution is a runtime for automatic execution of checkpoint-recompute schedules (called pyRevolve) to alleviate this memory requirement by trading it off for re-computations. These schedules may involve storing some intermediate states on disk, some in memory, and recomputing others. The third contribution is the use of lossy compression (ZFP/SZ), which is a second means of trading off memory and compute, automatically as part of checkpointing in pyRevolve. This is the first time it has been shown that lossy compression combined with checkpoint-recomputation can provide a much better error-memory tradeoff than using checkpoint-recomputation or lossy compression alone.

**Poster 37: Large Scale Ultrasound Simulations on Accelerated Clusters**  
*Filip Vaverka (Brno University of Technology)*

Many emerging non-invasive medical procedures such as high intensity focused ultrasound treatment (HIFU), photoacoustic imaging, or transcranial ultrasound therapy require ultrasound wave propagation simulations. Typically soft tissue can be modeled as fluid, weakly heterogeneous medium with frequency dependent absorption and non-linear effects at high ultrasound intensities. The k-Wave acoustic toolbox, widely used in medical community, uses highly efficient k-space pseudo-spectral time-domain (KSTD) discretization of acoustic equations to solve these problems. However, it is difficult to efficiently map the KSTD method onto modern cluster architectures with accelerators such as GPUs. The present thesis analyzes shortcomings of the KSTD method in respect to modern clusters and proposes local Fourier basis approach to improve scaling of the method. It is shown that the proposed method is able to achieve 5x speedup, while having sufficient accuracy for these medical applications. Behavior of the method is analyzed across variety of GPU and MIC (Intel Xeon Phi) accelerated clusters and results are presented.

**Poster 34: Analysis of Automata Processing Acceleration on Disparate Hardware Technologies**  
*Marziyeh Nourian (North Carolina State University)*

Pattern matching is a computation that maps naturally onto finite automata (FA) abstractions. There has been a substantial amount of work on accelerating FA processing on various parallel platforms. However, the advantages and disadvantages of different automata processing accelerators and the innovation space in this area are still unclear. We target this problem and propose a compiler toolchain that automates the deployment of non-deterministic finite automata (NFAs) onto different target platforms. Using this toolchain, we perform an apples-to-apples comparison between AP,
GPU- and FPGA-based NFA accelerator designs on large-scale datasets. Specifically, we observe that memory-based designs are limited by memory size and bandwidth. To address this issue, we target fixed-topology NFAs and propose a memory-efficient design that embeds the automata topology in code and stores only the transition symbols in memory. Our solution is suitable for SIMD architectures and is called SIMD_NFA. We design a compiler that automates the deployment of this design on SIMD platforms. We showcase our compiler framework on GPU and Intel platforms. Additionally, we observe that for NFAs with a grid-like fixed-topology (e.g., NFAs for Levenshtein and Hamming distance-based matching), transitions do not need to be encoded within the traversal code but can be inferred from the reference string to be matched and the knowledge of the NFA topology. Lastly, SIMD_NFA is a good fit for FPGA deployment using OpenCL-to-FPGA toolchains. We investigate the deployment of the OpenCL version of SIMD_NFA, on FPGA and explore a set of optimizations techniques to retarget SIMD_NFA to FPGA.

Poster 42: Power and Thermal Management Runtimes for HPC Applications in the Era of Exascale Computing
Daniele Cesarini (University of Bologna, CINECA)

In the scope of technical and scientific computing, the rush toward larger simulations has been so far assisted by a steady downsizing of micro-processing units, which has allowed to increase the compute capacity of general-purpose architectures at constant power. As side effects of the end of Dennard's scaling, this process is now hitting its ultimate power limits and is just about to come to an end. This implies an increase in the energy cost for computation, performance loss due to a design based on worst-case power consumption, and performance loss due to overheating, and thermal gradients. As result, thermal and power bound supercomputing machines show performance degradation and heterogeneity which limit the peak performance of the system. This doctoral showcase presents software strategies to tackle the main bottlenecks induced by power and thermal issues that affect next-generation supercomputers. To respond to the above challenges, my work shows that propagating workload requirements from application to the runtime and operating system levels is the key to provide efficiency. This is possible only if the proposed software methodologies cause little or no overhead in term of application performance. With this in mind in my work, I have designed application-aware node-level optimal thermal management algorithms and runtimes, lazy node-level power capping, and energy reduction runtime. The experimental results show a significant step forward with respect to the current state-of-the-art solutions in power and thermal control of HPC systems.

Poster 41: Co-Designing Communication Middleware and Deep Learning Frameworks for High-Performance DNN Training on HPC Systems
Recent advances in Machine/Deep Learning techniques have triggered key success stories in many application domains like Computer Vision, Speech Comprehension and Recognition, and Natural Language Processing. Large-scale Deep Neural Networks (DNNs), that are at the core of state-of-the-art AI technologies, have been the primary drivers of this success. Training very complicated and large DNN architectures using a large number of training examples (data) is compute-intensive and can take from weeks to months to achieve state-of-the-art prediction capabilities. To achieve higher accuracy, making the DNN deeper is also a common strategy. These requirements have led to a simple but powerful approach called Data Parallelism to achieve shorter training times. This has resulted in various research studies and ML/DL software like TensorFlow and PyTorch as well as distributed-training middleware like Horovod. In addition, for DNNs that do not fit the GPU memory, a new DNN workloads are emerging that we call Out-of-Core DNNs and different strategies (out-of-core training and model-parallelism) are needed to train them. Clearly, large-scale DNN training brings forward new requirements for computation runtimes like CUDA and communication middleware like the MVAPICCH2 MPI library. In this thesis, we broadly explore three different strategies to train DNNs on modern CPU and GPU architectures: 1) Data Parallelism, 2) Model Parallelism, and 3) Out-of-Core Training. We address the key challenge: How to co-design computation and communication in modern ML/DL frameworks with execution runtimes like CUDA and communication middleware like MVAPICCH2 to enable scalable, high-performance, and efficient training of DNNs on large-scale HPC systems?

Poster 40: Performance, Portability, and Productivity for Data-Parallel Computations on Multi- and Many-Core Architectures
Ari Rasch (University of Münster)

This thesis presents an approach to performance, portability, and productivity for data-parallel computations on multi- and many-core architectures, e.g., Intel CPU and NVIDIA GPU. We introduce the algebraic formalism of Multi-Dimensional Homomorphisms (MDHs) – a class of functions that cover important data-parallel computations, e.g., linear algebra routines (BLAS) and stencil computations. For our MDHs, we propose a Domain-Specific Language (DSL), based on patterns of parallelism (a.k.a. algorithmic skeletons), to enable conveniently expressing MDH functions. We introduce a code generation approach for our DSL to automatically generate for MDHs optimized program code targeting multi- and many-core architectures. Our code generation approach relies on OpenCL – an emerging de-facto standard for uniformly programming parallel architectures, such as CPU and GPU. A major feature of our generated code is that it is targeted to OpenCL’s abstract device models (rather than a particular architecture) by being parameterized in performance-critical parameters of these abstract models (e.g., the number of threads and size of
tiles). With our code generation approach, we enable both high performance and performance portability: we fully automatically optimize our generated code -- for any given combination of an MDH function, architecture, and input size -- by automatically choosing (auto-tuning) optimized values of our code’s performance-critical parameters using our own Auto-Tuning Framework (ATF). Our experimental results on CPU and GPU demonstrate competitive and often significantly better performance of our MDH+ATF approach as compared to the currently best-performing competitors, e.g., Intel MKL/MKL-DNN, NVIDIA cuBLAS/cuDNN, and Facebook’s Tensor Comprehensions framework.

**Poster 45: Characterization and Modeling of Error Resilience in HPC Applications**  
_Luanzheng Guo (University of California, Merced)_

As high-performance computing systems scale in size and computational power, the occurrence of transient faults grows. Without protection by efficient and effective fault tolerance mechanisms, transient errors can cause incorrect execution outcome and even lead to a catastrophe in safety-critical HPC applications. Previous work attributes error resilience in HPC applications at a high-level to either the probabilistic or iterative nature of the application, whereas the community still lacks the fundamental understanding of the program constructs that result in natural error resilience. We design FlipTracker, a framework to analytically track error propagation and to provide a fine-grained understanding of the propagation and tolerance of errors. After running FlipTracker on representative HPC applications, we summarize six resilience computation patterns that lead to nature error resilience in HPC applications. With a better understanding of natural resilience in HPC applications, we aim to model application resilience on data objects to transient faults. Many common application-level fault tolerance mechanisms focus on data objects. Understanding application resilience on data objects can be helpful to direct those mechanisms. The common practice to understand application resilience (random fault injection) gives us little knowledge of how and where errors are tolerated. Understanding "how" and "where" is necessary to understand how to apply application-level fault tolerance mechanisms effectively and efficiently. We design a practical model (MOARD) to measure application resilience on data objects by analytically quantifying error masking events happening to the data object. Using our model, users can compare application resilience on different data objects with different data types.

**Poster 43: Efficient and Scalable Communication Middleware for Emerging Dense-GPU Clusters**  
_Ching-Hsiang Chu (Ohio State University)_

In the era of post Moore’s law, the traditional CPU is not able to keep the pace up and provide the
computing power demanded by the modern compute-intensive and highly parallelizable applications. Under this context, various accelerator architectures such as general-purpose graphics processing unit (GPU), which equipped with the high-bandwidth memory (HBM) and massive parallelizable streaming multiprocessors, has been widely adopted in high-performance computing (HPC) and cloud systems to significantly accelerate numerous scientific and emerging machine/deep learning applications. Message Passing Interface (MPI), the standard programming model for parallel applications, has been widely used for GPU communication. However, the state-of-the-art MPI libraries are only optimizing GPU communication by leveraging advanced technology like Remote Direct Memory Access (RDMA) and not fully utilizing the computational power of GPUs. In this work, we propose GPU-enabled communication schemes to harness GPU computational resources, and cutting-edge interconnects such as NVIDIA NVLink for communication operations on the emerging heterogeneous systems. In this work, three primary MPI operations are addressed. First, intelligent communication scheduling, efficient packing/unpacking, and packing-free schemes are proposed to accelerate non-contiguous data transfer in scientific HPC applications. Second, scalable broadcast operations are presented to leverage the low-level hardware multicast feature to speed up GPU communication at scale. Finally, we also design topology-aware, link-efficient, and cooperative GPU kernels to significantly accelerate All-reduce operation, which is the primary performance bottleneck in deep learning applications. The proposed designs demonstrate significant performance improvements over the state-of-the-art communication schemes for various HPC and deep learning applications.

Poster 39: Designing Next-Generation Communication Middlewares for Many-Core Architectures
Jahanzeb Maqbool Hashmi (Ohio State University)

Modern multi-petaflop HPC systems are powered by dense multi-/many-core architectures and this trend is expected to grow for the next-generation supercomputing systems. This rapid adoption of high core-density architectures by the current- and next-generation HPC systems is further fueled by the emerging application trends such as Deep Learning. This is putting more emphasis on middleware designers to optimize various communication protocols to meet the diverse needs of the applications. While the novelties in the processor architectures have led to the increased on-chip parallelism, they come at the cost of rendering traditional designs, employed by the communication runtimes such as MPI, to suffer from higher degree of intra-node communication latencies. Tackling the computation and communication challenges that accompany these dense multi-/many-cores garner special design considerations. The proposed work in this thesis tries to address the performance challenges posed by a diverse range of applications and the lacking support in state-of-the-art communication libraries such as MPI to exploit high-concurrency architectures. The author first proposes a "shared-address-spaces"-based communication
substrate to derive intra-node communication in MPI. Atop this framework, the author has re-designed various MPI primitives such as point-to-point communication protocols (e.g., user-space zero-copy rendezvous transfer), collective communication (e.g., load/store based collectives, truly zero-copy and partitioning-based reduction algorithms), and efficient MPI derived datatypes processing (e.g., memoization-based "packing-free" communication) to exploit the potential of emerging multi-/many-core architectures and high throughput networks. The proposed designs have demonstrated significant improvement over state-of-the-art for various scientific and deep learning applications.

8:30 am - 5:00 pm

Research Posters Display

Poster 74: Enabling Code Portability of a Parallel and Distributed Smooth-Particle Hydrodynamics Application, FleCSPH
Suyash Tandon (University of Michigan), Nicholas Stegmeier (University of Illinois), Vasu Jaganath (University of Wyoming), Jennifer Ranta (Michigan State University), Rathish Ratnasingam (Newcastle University), Elizabeth Carlson (University of Nebraska), Julien Loiseau (Los Alamos National Laboratory), Vinay Ramakrishnaiah (Los Alamos National Laboratory), Robert Pavel (Los Alamos National Laboratory)

Core-collapse supernovae (CCSNe) are integral to the formation and distribution of heavy elements across the universe. However, CCSNe are highly complex and inherently non-linear phenomena. Large-scale simulations of these cosmic events can provide us a glimpse of their hydrodynamic and nucleosynthetic processes which are difficult to observe. To enable these massive numerical simulations on high-performance computing (HPC) centers, this study uses FleCSPH, a parallel and distributed code, based on the smooth-particle hydrodynamics (SPH) formulation. In the recent years, the HPC architecture has evolved and the next generation of exascale computers are expected to feature heterogenous architecture. Therefore, it is important to maintain code portability across platforms. This work demonstrates code portability of FleCSPH through the incorporation of Kokkos C++ library and containers using Charliecloud.

Best Poster Finalist: no

Poster 73: Accelerating Large-Scale GW Calculations on Hybrid CPU-GPU Architectures
Mauro Del Ben (Lawrence Berkeley National Laboratory), Charlene Yang (National Energy Research...
In this poster, we present the strategy, progress, and performance while GPU porting one of the major modules, epsilon, of the electronic structure code BerkeleyGW. Epsilon represents the most time-consuming routines in the BerkeleyGW workflow for large-scale material science simulations. Some of the porting/optimization strategies include, changing our original data layout to efficiently use libraries such as cuBLAS and cuFFT, implementation of specific CUDA kernels to minimize data copies between host/device and keeping data on device, efficient use of data streams to leverage high concurrency on the device, asynchronous memory copies and overlapping (MPI) communication on the host and computation on the device. Preliminary results are presented in terms of the speedup compare to the CPU-only implementation, strong/weak scaling, and power efficiency. Excellent acceleration is demonstrated: up to 30x for specific kernels. Our port also exhibits good scalability and about 16x higher FLOPs/watt efficiency compared to the CPU-only implementation.

Best Poster Finalist: no

**Poster 89: BeeCWL: A CWL Compliant Workflow Management System**

Betis Baheri (Kent State University), Steven Anaya (New Mexico Institute of Mining and Technology), Patricia Grubel (Los Alamos National Laboratory), Qiang Guan (Kent State University), Timothy Randles (Los Alamos National Laboratory)

Scientific workflows are used widely to carry out complex and hierarchical experiments. Although there are many trends to extend the functionality of workflow management systems to cover all possible requirements that may arise from a user community, one unified standard over cloud and HPC systems is still missing. In this paper, we propose a Common Workflow Language (CWL) compliant workflow management system. BeeCWL is a parser to derive meaningful information such as requirements, steps, relationships, etc. from CWL files and to create a graph database from those components. Generated graphs can be passed to an arbitrary scheduler and management system to decide whether there are enough resources to optimize and execute the workflow. Lastly, the user can have control over workflow execution, collecting logs, and restart or rerun some part of a complex workflow.

Best Poster Finalist: no
Poster 150: A Machine Learning Approach to Understanding HPC Application Performance Variation
Burak Aksar (Boston University, Sandia National Laboratories), Benjamin Schwaller (Sandia National Laboratories), Omar Aaziz (Sandia National Laboratories), Emre Ates (Boston University), Jim Brandt (Sandia National Laboratories), Ayse K. Coskun (Boston University), Manuel Egele (Boston University), Vitus Leung (Sandia National Laboratories)

Performance anomalies are difficult to detect because often a “healthy system” is vaguely defined, and the ground truth for how a system should be operating is evasive. As we move to exascale, however, detection of performance anomalies will become increasingly important with the increase in size and complexity of systems. There are very few accepted ways of detecting anomalies in the literature, and there are no published and labeled sets of anomalous HPC behavior. In this research, we develop a suite of applications that represent HPC workloads and use data from a lightweight metric collection service to train machine learning models to predict the future behavior of metrics. In the future, this work will be used to predict anomalous runs in compute nodes and determine some root causes of performance issues to help improve the efficiency of HPC system administrators and users.

Best Poster Finalist: no

Poster 81: Performance of Devito on HPC-Optimised ARM Processors
Hermes Senger (Federal University of São Carlos, Brazil; University of São Paulo), Jaime Freire de Souza (Federal University of São Carlos, Brazil), Edson Satoshi Gomi (University of São Paulo), Fabio Luporini (Imperial College, London), Gerard Gorman (Imperial College, London)

We evaluate the performance of Devito, a domain specific language (DSL) for finite differences on Arm ThunderX2 processors. Experiments with two common seismic computational kernels demonstrate that Devito can apply automatic code generation and optimization across Arm and Intel platforms. The code transformations include: parallelism, and SIMD vectorization (OpenMP >=4); loop tiling (with best block shape obtained via auto-tuning); domain-specific symbolic optimisation such as common sub-expression elimination and factorisation for Flop reduction, polynomial approximations for trigonometry terms, and heuristic hoisting of time-invariant expressions. Results show that Devito can achieve performance on Arm processors which is competitive to other Intel Xeon processors.

Best Poster Finalist: no
Poster 103: LIKWID 5: Lightweight Performance Tools
Thomas Gruber (Erlangen Regional Computing Center), Jan Eitzinger (Erlangen Regional Computing Center), Georg Hager (Erlangen Regional Computing Center), Gerhard Wellein (Erlangen Regional Computing Center)

LIKWID is a tool suite for performance oriented programmers with a worldwide user group. It is developed by the HPC group of the University Erlangen-Nuremberg since 2009 to support them in their daily research and performance engineering of user codes. The HPC landscape has become more and more diverse over the last years with clusters using non-x86 architectures and being equipped with accelerators. With the new major version, the architectural support of LIKWID is extended to ARM and POWER CPUs with the same functionality and features as for x86 architectures. Besides the CPU monitoring, the new version provides access the hardware counting facilities of Nvidia GPUs. This poster introduces the new features and shows the successes of applying LIKWID to identify performance bottlenecks and to test optimizations. Furthermore, the poster gives an overview of how users can integrate the LIKWID tools in their application using a lightweight add-once-and-reuse instrumentation API.

Best Poster Finalist: no

Poster 149: Solving Phase-Field Equations in Space-Time: Adaptive Space-Time Meshes and Stabilized Variational Formulations
Kumar Saurabh (Iowa State University), Biswajit Khara (Iowa State University), Milinda Fernando (University of Utah), Masado Ishii (University of Utah), Hari Sundar (University of Utah), Baskar Ganapathysubramanian (Iowa State University)

We seek to efficiently solve a generalized class of partial differential equations called the phase-field equations. These non-linear PDE’s model phase transition (solidification, melting, phase-separation) phenomena which exhibit spatially and temporally localized regions of steep gradients. We consider time as an additional dimension and simultaneously solve for the unknown in large blocks of time (i.e. in space-time), instead of the standard approach of sequential time-stepping. We use variational multiscale (VMS) based finite element approach to solve the ensuing space-time equations. This allows us to (a) exploit parallelism not only in space but also in time, (b) gain high order accuracy in time, and (c) exploit adaptive refinement approaches to locally refine region of interest in both space and time. We illustrate this approach with several canonical problems including melting and solidification of complex snow flake structures.

Best Poster Finalist: no
**Poster 84: ESTEE: A Simulation Toolkit for Distributed Workflow Execution**

Vojtěch Cima (IT4Innovations, Czech Republic), Jakub Beránek (IT4Innovations, Czech Republic), Stanislav Böhm (IT4Innovations, Czech Republic)

Task graphs provide a simple way to describe scientific workflows (sets of tasks with dependencies) that can be executed on both HPC clusters and in the cloud. An important aspect of executing such graphs is the used scheduling algorithm. Many scheduling heuristics have been proposed in existing works; nevertheless, they are often tested in oversimplified environments. We introduce a simulation environment designed for prototyping and benchmarking task schedulers. Our simulation environment, scheduler source codes, and graph datasets are open in order to be fully reproducible. To demonstrate usage of Estee, as an example, we compare the performance of various workflow schedulers in an environment using two different network models.

Best Poster Finalist: no

**Poster 93: Robust Data-Driven Power Simulator for Fast Cooling Control Optimization of a Large-Scale Computing System**

Takashi Shiraishi (Fujitsu Laboratories Ltd), Hiroshi Endo (Fujitsu Laboratories Ltd), Takaaki Hineno (Fujitsu Laboratories Ltd), Hiroyuki Fukuda (Fujitsu Laboratories Ltd)

Power of large-scale systems such as an HPC or a datacenter is a significant issue. Cooling units consume 30% of the total power. General control policies for cooling units are local and static (manual overall optimization nearly once a week). However, free cooling and IT-load fluctuation may change hourly optimum control variables of the cooling units. In this work, we present a deep neural network (DNN) power simulator that can learn from actual operating logs and can quickly identify the optimum control variables. We demonstrated the power simulator of an actual large-scale system with 4.7-MW-power IT load. Our robust simulator predicted the total power with error of 4.8% without retraining during one year. We achieved optimization by the simulator within 80 seconds that was drastically faster than previous works. The dynamic control optimization each hour showed a 15% power reduction compared to that of conventional policy in the actual system.

Best Poster Finalist: no

**Poster 115: sDNA: Software-Defined Network Accelerator Based on Optical Interconnection Architecture**

En Shao (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of
Software-Defined Network Accelerator (sDNA) is a new accelerated system for the exascale computer. Inspired by the edge forwarding index (EFI), the main contribution of our work is that it presents an extended EFI-based optical interconnection method with slow switching optical device. In our work, we found that sDNA based on extended EFI evaluation is not only able to offload the traffic from an electrical link to an optical link but is also able to avoid congestion inherent to electrical link.

Best Poster Finalist: no

**Poster 49: WarpX: Toward Exascale Modeling of Plasma Particle Accelerators on GPU**

Maxence Thevenet (Lawrence Berkeley National Laboratory), Jean-Luc Vay (Lawrence Berkeley National Laboratory), Ann Almgren (Lawrence Berkeley National Laboratory), Diana Amorim (Lawrence Berkeley National Laboratory), John Bell (Lawrence Berkeley National Laboratory), Axel Huebl (Lawrence Berkeley National Laboratory), Revathi Jambunathan (Lawrence Berkeley National Laboratory), Rémi Lehe (Lawrence Berkeley National Laboratory), Andrew Myers (Lawrence Berkeley National Laboratory), Jaehong Park (Lawrence Berkeley National Laboratory), Olga Shapoval (Lawrence Berkeley National Laboratory), Weiqun Zhang (Lawrence Berkeley National Laboratory), Lixin Ge (SLAC National Accelerator Laboratory), Mark Hogan (SLAC National Accelerator Laboratory), Cho Ng (SLAC National Accelerator Laboratory), David Grote (Lawrence Livermore National Laboratory)

Particle accelerators are a vital part of the DOE-supported infrastructure of discovery science and applications, but we need game-changing improvements in the size and cost for future accelerators. Plasma-based particle accelerators stand apart in their potential for these improvements. Turning this from a promising technology into mainstream scientific tools depends critically on high-performance, high-fidelity modeling of complex processes.

WarpX is an open-source particle-in-cell (PIC) code supported by the Exascale Computing Project (ECP) that is combining advanced algorithms with adaptive mesh refinement to allow challenging simulations of a multi-stage plasma-based TeV acceleration relevant for future high-energy physics discoveries. WarpX relies on the ECP co-design center for mesh refinement AMReX, and runs on CPU and GPU-accelerated computers. Production simulation have run on Cori KNL at NERSC and
Summit at OLCF. In this poster, recent results and strategies on GPU will be presented, along with recent performance results.

Best Poster Finalist: no

**Poster 50: Implementing an Adaptive Sparse Grid Discretization (ASGarD) for High Dimensional Advection-Diffusion Problems on Exascale Architectures**

M. Graham Lopez (Oak Ridge National Laboratory), David L. Green (Oak Ridge National Laboratory), Lin Mu (University of Georgia), Ed D’Azevedo (Oak Ridge National Laboratory), Wael Elwasif (Oak Ridge National Laboratory), Tyler McDaniel (University of Tennessee), Timothy Younkin (University of Tennessee), Adam McDaniel (Oak Ridge National Laboratory), Diego Del-Castillo-Negrete (Oak Ridge National Laboratory)

Many scientific domains require the solution of high dimensional PDEs. Traditional grid- or mesh-based methods for solving such systems in a noise-free manner quickly become intractable due to the scaling of the degrees of freedom going as $O(N^d)$ sometimes called "the curse of dimensionality." We are developing an arbitrarily high-order discontinuous-Galerkin finite-element solver that leverages an adaptive sparse-grid discretization whose degrees of freedom scale as $O(N\times\log_2 N^{D-1})$. This method and its subsequent reduction in the required resources is being applied to several PDEs including time-domain Maxwell's equations (3D), the Vlasov equation (in up to 6D) and a Fokker-Planck-like problem in ongoing related efforts. Here we present our implementation which is designed to run on multiple accelerated architectures, including distributed systems. Our implementation takes advantage of a system matrix decomposed as the Kronecker product of many smaller matrices which is implemented as batched operations.

Best Poster Finalist: no

**Poster 51: SmartK: Efficient, Scalable, and Winning Parallel MCTS**

Michael S. Davinroy (Swarthmore College), Shawn Pan (Swarthmore College), Bryce Wiedenbeck (Swarthmore College, Davidson College), Tia Newhall (Swarthmore College)

SmartK is our efficient and scalable parallel algorithm for Monte Carlo Tree Search (MCTS), an approximation technique for game searches. MCTS is also used to solve problems as diverse as planning under uncertainty, combinatorial optimization, and high-energy physics. In these problems, the solution search space is significantly large, necessitating parallel solutions. Shared memory parallel approaches do not scale well beyond the size of a single node's RAM. SmartK is a distributed memory parallelization that takes advantage of both inter-node and intra-node parallelism and a large
cumulative RAM found in clusters. SmartK’s novel selection algorithm combined with its ability to efficiently search the solution space, results in better solutions than other MCTS parallel approaches. Results of an MPI implementation of SmartK for the game of Hex, show SmartK yields a better win percentage than other parallel algorithms, and that its performance scales to larger search spaces and high degrees of parallelism.

Best Poster Finalist: no

Poster 70: Numerical Method and Parallelization for the Computation of Coherent Synchrotron Radiation
Boqian Shen (Rice University, Los Alamos National Laboratory)

The purpose of this work is to develop and parallelize an accurate and efficient numerical method for the computation of synchrotron radiation from relativistic electrons in the near field. The high-brilliance electron beam and coherent short-wavelength light source provide a powerful method to understand the microscopic structure and dynamics of materials. Such a method supports a wide range of applications including matter physics, structural biology, and medicine development. To understand the interaction between the beam and synchrotron radiation, an accurate and efficient numerical simulation is needed. With millions of electrons, the computational cost of the field would be large. Thus, multilevel parallelism and performance portability are desired since modern supercomputers are getting more complex and heterogeneous. The performance model and performance analysis are presented.

Best Poster Finalist: no

Poster 146: AI Matrix: A Deep Learning Benchmark for Alibaba Data Centers
Wei Zhang (Alibaba Inc), Wei Wei (Alibaba Inc), Lingjie Xu (Alibaba Inc), Lingling Jin (Alibaba Inc), Cheng Li (University of Illinois)

This work introduces AI Matrix, an in-house Deep Learning (DL) benchmark suite developed specifically for Alibaba’s e-commerce environment. AI Matrix results from a full investigation of the DL applications used inside Alibaba and aims to cover the typical DL applications that account for more than 90% of the GPU usage in Alibaba data centers. This benchmark suite collects DL models that are either directly used or closely resemble the models used in the company’s real e-commerce applications. It also collects the real e-commerce applications if no similar DL models are not available. Through the high coverage and close resemblance to real applications, AI Matrix fully represents the DL workloads on Alibaba data centers. The collected benchmarks mainly fall into three categories:
computer vision, recommendation, and language processing, which consist of the most majority of DL applications in Alibaba. AI Matrix is made open source, hoping it can benefit the public.

Best Poster Finalist: no

**Poster 100: Comparison of Array Management Library Performance - A Neuroscience Use Case**
Donghe Kang (Ohio State University), Oliver Rübel (Lawrence Berkeley National Laboratory), Suren Byna (Lawrence Berkeley National Laboratory), Spyros Blanas (Ohio State University)

Array management libraries, such as HDF5, Zarr, etc., depend on a complex software stack that consists of parallel I/O middleware (MPI-IO), POSIX-IO, and file systems. Components in the stack are interdependent, such that effort in tuning the parameters in these software libraries for optimal performance is non-trivial. On the other hand, it is challenging to choose an array management library based on the array configuration and access patterns. In this poster, we investigate the performance aspect of two array management libraries, i.e., HDF5 and Zarr, in the context of a neuroscience use case. We highlight the performance variability of HDF5 and Zarr in our preliminary results and discuss potential optimization strategies.

Best Poster Finalist: no

**Poster 118: Self-Driving Reconfigurable Silicon Photonic Interconnects (Flex-LIONS) with Deep Reinforcement Learning**
Roberto Proietti (University of California, Davis), Yu Shang (University of California, Davis), Xian Xiao (University of California, Davis), Xiaoliang Chen (University of California, Davis), Yu Zhang (University of California, Davis), SJ Ben Yoo (University of California, Davis)

We propose a self-driving reconfigurable optical interconnect architecture for HPC systems exploiting a deep reinforcement learning (DRL) algorithm and a reconfigurable silicon photonic (SiPh) switching fabric to adapt the interconnect topology to different traffic demands. Preliminary simulation results show that after training, the DRL-based SiPh fabric provides the lowest average end-to-end latency for time-varying traffic patterns.

Best Poster Finalist: no

**Poster 147: Extremely Accelerated Deep Learning: ResNet-50 Training in 70.4 Seconds**
Akihiro Tabuchi (Fujitsu Laboratories Ltd), Akihiko Kasagi (Fujitsu Laboratories Ltd), Masafumi
Distributed deep learning using a large mini-batch is a key technology to accelerate training in deep learning. However, it is difficult to achieve a high scalability and maintain validation accuracy in distributed learning on large clusters. We introduce two optimizations, reducing the computation time and overlapping the communication with the computation. By applying the techniques and using 2,048 GPUs, we achieved the world's fastest ResNet-50 training in MLPerf, which is a de facto standard DNN benchmark (as of July 2019).

Best Poster Finalist: no

Poster 111: Multiple HPC Environments-Aware Container Image Configuration for Bioinformatics Application
Kento Aoyama (Tokyo Institute of Technology, National Institute of Advanced Industrial Science and Technology (AIST)), Hiroki Watanabe (Tokyo Institute of Technology, National Institute of Advanced Industrial Science and Technology (AIST)), Masahito Ohue (Tokyo Institute of Technology), Yutaka Akiyama (Tokyo Institute of Technology)

Containers have a considerable advantage for application portability in different environments by isolating process with a small performance overhead; thus it has been rapidly getting popular in a wide range of science fields. However, there are problems in container image configuration when run in multiple HPC environments, and it requires users to have knowledge of systems, container runtimes, container image format, and library compatibilities in HPC environments.

In this study, we introduce our HPC container workflow in multiple supercomputing environments that have different system/library specifications (ABCI, TSUBAME3.0). Our workflow provides custom container image configurations for HPC environments by taking into account differences in container runtime, container image, and library compatibility between the host and inside of the container. We also show the parallel performance of our application in each HPC environment.

Best Poster Finalist: no

Poster 134: Minimal-Precision Computing for High-Performance, Energy-Efficient, and Reliable Computations
Daichi Mukunoki (RIKEN Center for Computational Science (R-CCS)), Toshiyuki Imamura (RIKEN Center for Computational Science (R-CCS)), Yiyu Tan (RIKEN Center for Computational Science (R-CCS)), Atsushi Koshiba (RIKEN Center for Computational Science (R-CCS)), Jens Huthmann (RIKEN Center for Computational Science (R-CCS)), Kentaro Sano (RIKEN Center for Computational Science (R-CCS)), Fabienne Jézéquel (Sorbonne University), Stef Graillat (Sorbonne University), Roman Iakymchuk (Sorbonne University), Norihisa Fujita (University of Tsukuba), Taisuke Boku (University of Tsukuba)

In numerical computations, the precision of floating-point computations is a key factor to determine the performance (speed and energy-efficiency) as well as the reliability (accuracy and reproducibility). However, the precision generally plays a contrary role for both. Therefore, the ultimate concept for maximizing both at the same time is the minimal-precision computation through precision-tuning, which adjusts the optimal precision for each operation and data. Several studies have been already conducted for it so far, but the scope of those studies is limited to the precision-tuning alone. In this study, we propose a more broad concept of the minimal-precision computing with precision-tuning, involving both hardware and software stack.

Best Poster Finalist: no

Poster 112: Building Complex Software Applications Inside Containers
Calvin D. Seamons (Los Alamos National Laboratory)

High performance computing (HPC) scientific applications require complex dependencies to operate. As user demand for HPC systems increases, it becomes unrealistic to support every unique dependency request. Containers can offer the ability to satisfy the users’ dependency request while simultaneously offering HPC portability across systems. By “containerizing” Model for Prediction Across Scales (MPAS, a large atmospheric simulation suite), we show that it is possible to containerize and run complex software. Furthermore, the container can be run across different HPC systems with nearly identical results (21 bytes difference over 2.1 gigabytes). Containers have the possibility to bring flexibility to code teams in HPC by helping to meet the demand for user defined software stacks (UDSS), and giving teams the ability to choose their software, independently of what is offered by the HPC system.

Best Poster Finalist: no

Poster 101: Job Performance Overview of Apache Flink and Apache Spark Applications
Jan Frenzel (Technical University Dresden), René Jäkel (Technical University Dresden)
Apache Spark and Apache Flink are two Big Data frameworks used for fast data exploration and analysis. Both frameworks provide the runtime of program sections and performance metrics, such as the number of bytes read or written, via an integrated dashboard. Performance metrics available in the dashboard lack timely information and are only shown aggregated in a separate part of the dashboard. However, performance investigations and optimizations would benefit from an integrated view with detailed performance metric events. Thus, we propose a system that samples metrics at runtime and collects information about the program sections after the execution finishes. The performance data is stored in an established format independent from Spark and Flink versions and can be viewed with state-of-the-art performance tools, i.e. Vampir. The overhead depends on the sampling interval and was below 10% in our experiments.

Best Poster Finalist: no

**Poster 113: Improvements Toward the Release of the Pavilion 2.0 Test Harness**  
*Kody J. Everson (Los Alamos National Laboratory, Dakota State University), Maria Francine Lapid (Los Alamos National Laboratory)*

High-performance computing production support entails thorough testing in order to evaluate the efficacy of a system for production-grade workloads. There are various phases of a system’s life-cycle to assess, requiring different methods to accomplish effective evaluation of performance and correctness. Due to the unique and distributed nature of an HPC-system, the necessity for sophisticated tools to automatically harness and assess test results, all while interacting with schedulers and programming environment software, requires a customizable, extensible, and lightweight system to manage concurrent testing. Beginning with the recently refactored codebase of Pavilion 1.0, we assisted with the finishing touches on readying this software for open-source release and production usage. Pavilion 2.0 is a Python 3-based testing framework for HPC clusters that facilitates the building, running, and analysis of tests through an easy-to-use, flexible, YAML-based configuration system. This enables users to write their own tests by simply wrapping everything in Pavilion’s well-defined format.

Best Poster Finalist: no

**Poster 119: Toward Lattice QCD on Fugaku: SVE Compiler Studies and Micro-Benchmarks in the RIKEN Fugaku Processor Simulator**  
*Nils Meyer (University of Regensburg, Bavaria), Tilo Wettig (University of Regensburg, Bavaria), Yuetsu Kodama (RIKEN Center for Computational Science (R-CCS)), Mitsuhisa Sato (RIKEN Center*
The Fugaku supercomputer, successor to the Japanese flagship K-Computer, will start operation in 2021. Fugaku incorporates the Fujitsu A64FX processor, which is the first hardware implementation supporting the Arm SVE instruction set, in this case a 512-bit version. Real hardware is not accessible today, but RIKEN has designed a simulator of the A64FX. We present micro-benchmarks relevant for Lattice QCD obtained in the RIKEN Fugaku processor simulator and compare three different SVE compilers.

Poster 62: Emulating Multi-Pattern Quantum Grover’s Search on a High-Performance Reconfigurable Computer
Naveed Mahmud (University of Kansas), Bennett Haase-Divine (University of Kansas), Bailey K. Srimoungchanh (University of Kansas), Nolan Blankenau (University of Kansas), Annika Kuhnke (University of Kansas), Esam El-Araby (University of Kansas)

Grover’s search (GS) is a widely studied quantum algorithm that can be employed for both single and multi-pattern search problems and potentially provides quadratic speedup over existing classical search algorithms. In this paper, we propose a multi-pattern quantum search methodology based on a modified GS quantum circuit. The proposed method combines classical post-processing permutations with a modified Grover’s circuit to efficiently search for given single/multiple input patterns. Our proposed methodology reduces quantum circuit complexity, realizes space-efficient emulation hardware and improves overall system configurability for dynamic, multi-pattern search. We use a high-performance reconfigurable computer to emulate multi-pattern GS(MGS) and present scalable emulation architectures of a complete multi-pattern search system. We validate the system and provide analysis of experimental results in terms of FPGA resource utilization and emulation time. Our results include a successful hardware architecture that is capable of emulating MGS algorithm up to 32 fully-entangled quantum bits on a single FPGA.

Best Poster Finalist: no

Poster 107: Exploring Interprocess Work Stealing for Balanced MPI Communication
Kaiming Ouyang (University of California, Riverside), Min Si (Argonne National Laboratory), Zizhong Chen (University of California, Riverside)

Workload balance among MPI processes is a critical consideration during the development of HPC
applications. However, because of many factors such as complex network interconnections and irregularity of HPC applications, fully achieving workload balance in practice is nearly impossible. Although interprocess job stealing is a promising solution, existing shared-memory techniques that lack necessary flexibility or cause inefficiency during data access cannot provide an applicable job-stealing implementation. To solve this problem, we propose a new process-in-process (PiP) interprocess job-stealing method to balance communication workload among processes on MPI layers. Our initial experimental results show PiP-based job stealing can efficiently help amortize workload, reduce imbalance, and greatly improve intra- and intersocket ping-pong performance compared with original MPI.

Best Poster Finalist: no

**Poster 127: sFlow Monitoring for Security and Reliability**

Xava A. Grooms (Los Alamos National Laboratory, University of Kentucky), Robert V. Rollins (Los Alamos National Laboratory, Michigan Technological University), Collin T. Rumpca (Los Alamos National Laboratory, Dakota State University)

In the past ten years, High Performance Computing (HPC) has moved far beyond the terascale performance, making petascale systems the new standard. The drastic improvement in performance has been largely unmatched with insignificant improvements in system monitoring. Thus, there is an immediate need for practical and scalable monitoring solutions to ensure the effectiveness of costly compute clusters. This project aims to explore the viability and impact of sFlow enabled switches in cluster network monitoring for security and reliability. A series of tests and exploits were performed to target specific network abnormalities on a nine-node HPC cluster. The results present web-based dashboards that can aid network administrators in improving a cluster’s security and reliability.

Best Poster Finalist: no

**Poster 77: Extreme Scale Phase-Field Simulations of Sintering Processes**

Johannes Hötzer (Karlsruhe University of Applied Sciences), Henrik Hierl (Karlsruhe University of Applied Sciences), Marco Seiz (Karlsruhe Institute of Technology), Andreas Reiter (Karlsruhe Institute of Technology), Britta Nestler (Karlsruhe Institute of Technology)

The sintering process, which turns loose powders into dense materials, is naturally found in the formation of glaciers, but is also the indispensable process to manufacture ceramic materials. This process is described by a dynamically evolving microstructure, which largely influences the resulting material properties.
To investigate this complex three-dimensional, scale-bridging evolution in realistic domain sizes, a highly optimized and parallelized multiphysics phase-field solver is developed. The solver is optimized in a holistic way, from the application level over the time integration and parallelization, down to the hardware. Optimizations include communication hiding, explicit vectorization, implicit schemes, and local reduction of degrees of freedom.

With this, we are able to investigate large-scale, three-dimensional domains, and long integration times. We have achieved a single-core peak performance of 32.5%, scaled up to 98304 cores on Hazel Hen and SuperMUC-NG, and simulated a multimillion particle system.

Best Poster Finalist: no

**Poster 140: Toward Automatic Function Call Generation for Deep Learning**  
Shizhi Tang (Tsinghua University, China), Jidong Zhai (Tsinghua University, China)

Mainstream deep learning frameworks are commonly implemented by invoking underlying high performance tensor libraries on various architectures. However, as these libraries provide increasingly complex semantics including operator fusions, in-place operations, and various memory layouts, the gap between mathematical deep learning models and the underlying libraries becomes larger. In this paper, inspired by the classic problem of Instruction Selection, we design a theorem solver guided exhausted search algorithm to select functions for complex tensor computations. Preliminary results with some micro-benchmarks and a real model show that our approach can outperform both Tensorflow and Tensor Comprehensions at run time.

Best Poster Finalist: no

**Poster 83: ETL: Elastic Training Layer for Deep Learning**  
Lei Xie (Tsinghua University, China), Jidong Zhai (Tsinghua University, China)

Due to the rising of deep learning, clusters for deep learning training are widely deployed in production. However, static task configuration and resource fragmentation problems in existing clusters result in low efficiency and poor quality of service. We propose ETL, an elastic training layer for deep learning, to help address them once for all. ETL adopts many novel mechanisms, such as lightweight and configurable report primitive and asynchronous, parallel and IO-free state replication, to achieve both high elasticity and efficiency. The evaluation demonstrates the low overhead and high efficiency of these mechanisms and reveals the advantages of elastic deep learning supported
Poster 130: Deep Learning-Based Feature-Aware Data Modeling for Complex Physics Simulations
Qun Liu (Louisiana State University), Subhashis Hazarika (Ohio State University), John M. Patchett (Los Alamos National Laboratory), James P. Ahrens (Los Alamos National Laboratory), Ayan Biswas (Los Alamos National Laboratory)

Data modeling and reduction for in situ is important. Feature-driven methods for in situ data analysis and reduction are a priority for future exascale machines as there are currently very few such methods. We investigate a deep-learning-based workflow that targets in situ data processing using autoencoders. We employ integrated skip connections to obtain higher performance compared to the existing autoencoders. Our experiments demonstrate the initial success of the proposed framework and create optimism for the in situ use case.

Best Poster Finalist: no

Poster 125: Physics Informed Generative Adversarial Networks for Virtual Mechanical Testing
Julian Cuevas (NASA, University of Puerto Rico at Mayaguez), Patrick Leser (NASA), James Warner (NASA), Geoffrey Bomarito (NASA), William Leser (NASA)

Physics-informed generative adversarial networks (PI-GANs) are used to learn the underlying probability distributions of spatially-varying material properties (e.g., microstructure variability in a polycrystalline material). While standard GANs rely solely on data for training, PI-GANs encode physics in the form of stochastic differential equations using automatic differentiation. The goal here is to show that experimental data from a limited number of material tests can be used with PI-GANs to enable unlimited virtual testing for aerospace applications. Preliminary results using synthetically generated data are provided to demonstrate the proposed framework. Deep learning and automatic differentiation capabilities in Tensorflow were implemented on Nvidia Tesla V100 GPUs.

Best Poster Finalist: no

Poster 141: ExaGeoStatR: Harnessing HPC Capabilities for Large Scale Geospatial Modeling Using R
Large-scale simulations and parallel computing techniques are becoming essential in Gaussian process calculations to lessen the complexity of geostatistics applications. The log-likelihood function is used in such applications to evaluate the model associated with a given set of measurements in existing n geographic locations. The evaluation of such a function requires O(n^2) memory and O(n^3) computation, which is infeasible for large datasets with existing software tools.

We present ExaGeoStatR, a package for large-scale geostatistics in R that computes the log-likelihood function on shared and distributed-memory, possibly equipped with GPU, using advanced linear algebra techniques. The package provides a high-level abstraction of the underlying architecture while enhancing the R developers’ productivity. We demonstrate ExaGeoStatR package by illustrating its implementation details, analyzing its performance on various parallel architectures, and assessing its accuracy using synthetic datasets and a sea surface temperature dataset. The performance evaluation involves spatial datasets with up to 250K observations.

Poster 48: Runtime System for GPU-Based Hierarchical LU Factorization
Qianxiang Ma (Tokyo Institute of Technology), Rio Yokota (Tokyo Institute of Technology, Global Scientific Information and Computing Center; Tokyo Institute of Technology)

Hierarchical low-rank approximation can reduce both the storage and computation costs of dense matrices, but its implementation is challenging. In this research, we tackle one of the most difficult problems of GPU parallelization of the factorization of these hierarchical matrices. To this end, we are developing a new runtime system for GPUs that can schedule all tasks into one GPU kernel. Other existing runtime systems, like cuGraph and Standford Legion, can only manage streams and kernel-level parallelism. Even without too much tuning, we achieved 4x better performance in H-LU factorization with a single GPU when comparing with a well-tuned CPU-based hierarchical matrix library, HLIBpro, on moderately sized matrices. Additionally, we have significantly less runtime overheads exposed when processing smaller matrices.

Best Poster Finalist: no
**Poster 145: Improving Data Compression with Deep Predictive Neural Network for Time Evolutional Data**

Rupak Roy (Florida State University), Kento Sato (RIKEN Center for Computational Science (R-CCS)), Jian Guo (RIKEN Center for Computational Science (R-CCS)), Jens Domke (RIKEN Center for Computational Science (R-CCS)), Weikuan Yu (Florida State University), Takaki Hatsui (RIKEN SPring-8 Center), Yasumasa Joti (Japan Synchrotron Radiation Research Institute)

Scientific applications/simulations periodically generate huge intermediate data. Storing or transferring such a large scale of data is critical. Fast I/O is important for making this process faster. One of the approaches to achieve fast I/O is data compression. Our goal is to achieve a delta technique that can improve the performance of existing data compression algorithms for time evolutional intermediate data.

In our approach, we compute the delta values from original data and data predicted by the deep predictive neural network. We pass these delta values through three phases which are preprocessing phase, partitioned entropy coding phase, and density-based spatial delta encoding phase.

In our poster, we present how our predictive delta technique can leverage the time evolutional data to produce highly concentrated small values. We show the improvement in compression ratio when our technique, combined with existing compression algorithms, are applied on the intermediate data for different datasets.

Best Poster Finalist: no

**Poster 144: Optimizing Asynchronous Multi-Level Checkpoint/Restart Configurations with Machine Learning**

Tonmoy Dey (Florida State University), Kento Sato (RIKEN Center for Computational Science (R-CCS)), Jian Guo (RIKEN Center for Computational Science (R-CCS)), Bogdan Nicolae (Argonne National Laboratory), Jens Domke (RIKEN Center for Computational Science (R-CCS)), Weikuan Yu (Florida State University), Franck Cappello (Argonne National Laboratory), Kathryn Mohror (Lawrence Livermore National Laboratory)

With the emergence of fast local storage, multi-level checkpointing (MLC) has become a common approach for efficient checkpointing. To utilize MLC efficiently, it is important to determine the optimal configuration for the checkpoint/restart (CR). There are mainly two approaches for determining the optimal configuration for CR, namely modeling and simulation approach. However, with MLC, CR
becomes more complicated making the modeling approach inaccurate and the simulation approach though accurate, very slow. In this poster, we focus on optimizing the performance of CR by predicting the optimized checkpoint count and interval. This was achieved by combining the simulation approach with machine learning and neural network to leverage its accuracy without spending time on simulating different CR parameters. We demonstrate that our models can predict the optimized parameter values with minimal error when compared to the simulation approach.

Best Poster Finalist: no

**Poster 132: Optimizing Performance at Runtime Using Binary Rewriting**  
Alexis Engelke (Technical University Munich), David Hildenbrand (Technical University Munich), Martin Schulz (Technical University Munich)

In addition to scalability, performance of sequential code in applications is an important factor in HPC. Typically, programs are compiled once, at which time optimizations are applied, and are then run several times. However, not all information relevant for performance optimizations are available at compile-time, restricting optimization possibilities. The generation of specialized code at runtime allows for further optimizations. Performing such specialization on binary code allows for initial code to be generated at compile-time with only the relevant parts being rewritten at runtime, reducing the optimization overhead. For targeted optimizations and effective use of known runtime information, the rewriting process needs to be guided by the application itself, exploiting information only known to the developer.

We describe three approaches for self-guided binary rewriting explicitly guided by the running application and evaluate the performance of the optimized code as well as the performance of the rewriting process itself.

Best Poster Finalist: no

**Poster 53: Unstructured Mesh Technologies for Fusion Simulations**  
Cameron Smith (Rensselaer Polytechnic Institute (RPI)), Gerrett Diamond (Rensselaer Polytechnic Institute (RPI)), Gopan Perumpilly (Rensselaer Polytechnic Institute (RPI)), Chonglin Zhang (Rensselaer Polytechnic Institute (RPI)), Agnieszka Truszkowska (Rensselaer Polytechnic Institute (RPI)), Onkar Sahni (Rensselaer Polytechnic Institute (RPI)), Morteza Hakimi (Rensselaer Polytechnic Institute (RPI)), Mark Shephard (Rensselaer Polytechnic Institute (RPI)), Eisung Yoon (Ulsan National Institute of Science and Technology, South Korea), Daniel Ibanez (Sandia National Laboratories)
Multiple unstructured mesh technologies are needed to define and execute plasma physics simulations. The domains of interest combine model features defined from physical fields within 3D CAD of the tokamak vessel with an antenna assembly, and 2D cross sections of the tokamak vessel. Mesh generation technologies must satisfy these geometric constraints and additional constraints imposed by the numerical models. Likewise, fusion simulations over these domains study a range of timescales and physical phenomena within a tokamak.

XGCm studies the development of plasma turbulence in the reactor vessel, GITRm studies impurity transport, and PetraM simulations model RF wave propagation in scrape off layer plasmas. GITRm and XGCm developments are using the PUMIpic infrastructure to manage the storage and access of non-uniform particle distributions in unstructured meshes on GPUs. PetraM combines PUMI adaptive unstructured mesh control with MFEM using CAD models and meshes defined with Simmetrix tools.

Best Poster Finalist: no

**Poster 99: Eithne: A Framework for Benchmarking Micro-Core Accelerators**

Maurice C. Jamieson (Edinburgh Parallel Computing Centre, University of Edinburgh), Nick Brown (Edinburgh Parallel Computing Centre, University of Edinburgh)

Running existing HPC benchmarks as-is on micro-core architectures is at best difficult and most often impossible as they have a number of architectural features that makes them significantly different from traditional CPUs: tiny amounts on-chip RAM (c. 32KB), low-level knowledge specific to each device (including the host/device communications interface), limited communications bandwidth and complex or no device debugging environment. In order to compare and contrast different the micro-core architectures, a benchmark framework is required to abstract much of this complexity.

The modular Eithne framework supports the comparison of a number of micro-core architectures. The framework separates the actual benchmark from the details of how this is executed on the different technologies. The framework was evaluated by running the LINPACK benchmark on the Adapteva Epiphany, PicoRV32 and VectorBlox Orca RISC-V soft-cores, NXP RV32M1, ARM Cortex-A9, and Xilinx MicroBlaze soft-core, and comparing resulting performance and power consumption.

Best Poster Finalist: no

**Poster 133: Portable Resilience with Kokkos**

Jeffery Miles (Sandia National Laboratories), Nicolas Morales (Sandia National Laboratories), Carson
Mould (Sandia National Laboratories), Keita Teranishi (Sandia National Laboratories)

The Kokkos ecosystem is a programming environment that provides performance and portability to many scientific applications that run on DOE supercomputers as well as other smaller scale systems. Leveraging software abstraction concepts within Kokkos, software resilience for end user code is made portable with abstractions and concepts while implementing the most efficient resilience algorithms internally. This addition enables an application to manage hardware failures reducing the cost of interruption without drastically increasing the software maintenance cost. Two main resilience methodologies have been added to the Kokkos ecosystem to validate the resilience abstractions: 1. Checkpointing includes an automatic mode supporting other checkpointing libraries and a manual mode which leverages the data abstraction and memory space concepts. 2. The redundant execution model anticipates failures by replicating data and execution paths. The design and implementation of these additions are illustrated, and appropriate examples are included to demonstrate the simplicity of use.

Best Poster Finalist: no

Poster 79: The HPC PowerStack: A Community-Wide Collaboration Toward an Energy Efficient Software Stack
Siddhartha Jana (Energy Efficient HPC Working Group, Intel Corporation), Stephanie Brink (Lawrence Livermore National Laboratory), Christopher Cantalupo (Intel Corporation), Jonathan Eastep (Intel Corporation), Masaaki Kondo (RIKEN Advanced Institute for Computational Science (AICS), University of Tokyo), Matthias Maiterth (Intel Corporation), Aniruddha Marathe (Lawrence Livermore National Laboratory), Tapasya Patki (Lawrence Livermore National Laboratory), Barry Rountree (Lawrence Livermore National Laboratory), Ryuichi Sakamoto (University of Tokyo), Martin Schulz (Technical University Munich, Leibniz Supercomputing Centre), Carsten Trinitis (Technical University Munich), Josef Weidendorfer (Technical University Munich, Leibniz Supercomputing Centre)

This poster highlights an ongoing community-wide effort among vendors, labs, and academia, to incorporate power-awareness within system-stacks in upcoming exascale machines. HPC PowerStack is the first-and-only community-driven vendor-neutral effort to identify what power optimization software actors are key within the modern-day stack; discuss their interoperability, and work toward gluing together existing open source projects to engineer cost-effective, but cohesive, portable implementations.

This poster disseminates key insights acquired in the project, provides prototyping status updates, highlights open questions, and solicits participation addressing the imminent exascale power challenge.
**Poster 87: Parallelizing Simulations of Large Quantum Circuits**

Michael A. Perlin (University of Colorado, National Institute of Standards and Technology (NIST)), Teague Tomesh (Princeton University), Bradley Pearlman (University of Colorado, National Institute of Standards and Technology (NIST)), Wei Tang (Princeton University), Yuri Alexeev (Argonne National Laboratory), Martin Suchara (Argonne National Laboratory)

We present a parallelization scheme for classical simulations of quantum circuits. Our scheme is based on a recent method to "cut" large quantum circuits into smaller sub-circuits that can be simulated independently, and whose simulation results can in turn be re-combined to infer the output of the original circuit. The exponentially smaller classical computing resources needed to simulate smaller circuits are counterbalanced by exponential overhead in terms of classical post-processing costs. We discuss how this overhead can be massively parallelized to reduce classical computing costs.

**Poster 120: ILP-Based Scheduling for Linear-Tape Model Trapped-Ion Quantum Computers**

Xin-Chuan Wu (University of Chicago), Yongshan Ding (University of Chicago), Yunong Shi (University of Chicago), Yuri Alexeev (Argonne National Laboratory), Hal Finkel (Argonne National Laboratory), Kibaek Kim (Argonne National Laboratory), Frederic T. Chong (University of Chicago)

Quantum computing (QC) is emerging as a potential post-Moore high-performance computing (HPC) technology. Trapped-ion quantum bits (qubits) are among the most leading technologies to reach scalable quantum computers that would solve certain problems beyond the capabilities of even the largest classical supercomputers. In trapped-ion QC, qubits can physically move on the ion trap. The state-of-the-art architecture, linear-tape model, only requires a few laser beams to interact with the entire qubits by physically moving the interacting ions to the execution zone. Since the laser beams are limited resources, the ion chain movement and quantum gate scheduling are critical for the circuit latency. To harness the emerging architecture, we present our mathematical model for scheduling the qubit movements and quantum gates in order to minimize the circuit latency. In our experiment, our scheduling reduces 29.47% circuit latency on average. The results suggest classical HPC would further improve the quantum circuit optimization.
**Poster 55: MPI+OpenMP Parallelization of DFT Method in GAMESS**

Vladimir Mironov (Lomonosov Moscow State University), Yuri Alexeev (Argonne National Laboratory), Dmitry Fedorov (National Institute of Advanced Industrial Science and Technology (AIST))

In this work, the Density Functional Theory (DFT) method is parallelized with MPI-OpenMP in the quantum chemistry package GAMESS. It has been implemented in both regular and Fragment Molecular Orbital (FMO) based DFT codes. The scalability of the FMO-DFT code was demonstrated on Cray XC40 Theta supercomputer. We demonstrated excellent scalability of the code up 2,048 Intel Xeon Phi nodes (131,072 cores). Moreover, the developed DFT code is about twice as fast as the original code because of our new grid integration algorithm.

Best Poster Finalist: no

---

**Poster 71: AI-Solver: Uncertainty in Prediction and Error Estimation for AI in Engineering**

Ahmed Al-Jarro (Fujitsu Laboratories of Europe Ltd), Loic Beheshti (Fujitsu Laboratories of Europe Ltd), Serban Georgescu (Fujitsu Laboratories of Europe Ltd), Koichi Shirahata (Fujitsu Laboratories Ltd), Yasumoto Tomita (Fujitsu Laboratories Ltd), Nakashima Kouta (Fujitsu Laboratories Ltd)

The AI-Solver is a deep learning platform that learns from simulation data to extract general behavior based on physical parameters. The AI-Solver can handle a wide variety of classes of problems including those commonly identified in FEA, CFD and CEM, to name a few, with speedups of up to 250,000X and extremely low error rate of 2-3%. In this work, we build on this recent effort. We first integrate uncertainty quantification, via exploiting the approximation of Bayesian Deep Learning. Second, we develop bespoke error estimation mechanisms capable of processing this uncertainty to provide instant feedback on the confidence in predictions without relying on the availability of ground truth data. To our knowledge, the ability to estimate the discrepancy in predictions without labels is a first in the field of AI for Engineering.

Best Poster Finalist: no

---

**Poster 72: Kokkos and Fortran in the Exascale Computing Project Plasma Physics Code XGC**

Aaron Scheinberg (Princeton Plasma Physics Laboratory), Guangye Chen (Los Alamos National Laboratory), Stephane Ethier (Princeton Plasma Physics Laboratory), Stuart Slattery (Oak Ridge National Laboratory), Robert Bird (Los Alamos National Laboratory), Pat Worley (PHWorley Consulting), Choong-Seock Chang (Princeton Plasma Physics Laboratory)
Numerical plasma physics models such as the particle-in-cell XGC code are important tools to understand phenomena encountered in experimental fusion devices. Adequately resolved simulations are computationally expensive, so optimization is essential. To address the need for consistent high performance by cutting-edge scientific software applications, frameworks such as Kokkos have been developed to enable portability as new architectures require hardware-specific coding implementation for best performance. Cabana, a recent extension to Kokkos developed with the ECP-CoPA project, is a library of common kernels and operations typically necessary for particle-based codes. The Kokkos/Cabana framework enables intuitive construction of particle-based codes, while maintaining portability between architectures. Here, we summarize the adoption by XGC of the execution and data layout patterns offered by this framework. We demonstrate a method for Fortran codes to adopt Kokkos and show that it can provide a single, portable code base that performs well on both GPUs and multicore machines.

Best Poster Finalist: no

**Poster 106: Optimizing Hybrid Access Virtual Memory System Using SCM/DRAM Unified Memory Management Unit**

Yusuke Shirota (Toshiba Corporation), Shiyo Yoshimura (Toshiba Corporation), Satoshi Shirai (Toshiba Corporation), Tatsunori Kanai (Toshiba Corporation)

In HPC systems, expectations for storage-class memory (SCM) are increasing in large-scale in-memory processing. While SCM can deliver higher capacity and lower standby power than DRAM, it is slower and the dynamic power is higher. Therefore, in order to realize high-speed, low-power and scalable main memory, it is necessary to build an SCM/DRAM unified memory, and dynamically optimize data placement between the two memories according to the memory access pattern.

In this poster, we describe a new hybrid access type virtual memory method using TLB-extended unified memory management unit which enables collecting and extracting fine-grained memory access locality characteristics. We show that with the proposed method, Hybrid Access control, which is a memory hierarchy control that selectively uses Direct Access to bus attached byte-addressable SCM and low power Aggressive Paging using small DRAM as cache, can be made more accurate, and the efficiency of memory access can be significantly improved.

Best Poster Finalist: no

**Holistic Measurement Driven System Assessment**

Saurabh Jha (University of Illinois), Mike Showerman (National Center for Supercomputing)
HPC users deploy a suite of monitors to observe patterns of failures and performance anomalies to improve operational efficiency, achieve higher application performance and inform the design of future systems. However, the promises and the potential of monitoring data have largely been not realized due to various challenges such as inadequacy in monitoring, limited availability of data, lack of methods for fusing monitoring data at time-scales necessary for enabling human-in-the-loop or machine-in-the-loop feedback. To address above challenges, in this work we developed a monitoring fabric Holistic Measurement Driven System Assessment (HMDSA) for large-scale HPC facilities, independent of major component vendor, and within budget constraints of money, space, and power. We accomplish this through development and deployment of scalable, platform-independent, open-source tools and techniques for monitoring, coupled with statistical and machine-learning based runtime analysis and feedback, which enables highly efficient HPC system operation and usage and also informs future system improvements.

Best Poster Finalist: no

Poster 139: Model Identification of Pressure Drop in Membrane Channels with Multilayer Artificial Neural Networks

Jiang-hang Gu (Sun Yat-sen University, Zhuhai, School of Chemical Engineering and Technology), Jiu Luo (Sun Yat-sen University, Guangzhou, School of Materials Science and Engineering), Ming-heng Li (California State Polytechnic University, Pomona), Yi Heng (Sun Yat-sen University, Guangzhou, School of Data and Computer Science; Sun Yat-sen University, Guangzhou, China)

This poster presents the work of identifying a data-driven model of pressure drop in spacer-filled reverse osmosis membrane channels and conducting CFD simulations. The established model correlates the pressure drop with a wide range of design objectives, which enables a quantitative description of the geometric structures and operation conditions for improvement. This way, it aims at optimizing the spacer geometry with minimal effort. Furthermore, a high-performance computing strategy is employed to tackle the resulted intractable computational task in the identification procedure and CFD simulations.
Poster 61: Fast 3D Diffeomorphic Image Registration on GPUs
Malte Brunn (University of Stuttgart), Naveen Himthani (University of Texas), George Biros (University of Texas), Miriam Mehl (University of Stuttgart), Andreas Mang (University of Houston)

3D image registration is one of the most fundamental and computationally expensive operations in medical image analysis. Here, we present a mixed-precision, Gauss-Newton-Krylov solver for diffeomorphic registration. Our work extends the publicly available CLAIRE library to GPU architectures. Despite the importance of image registration, only a few implementations of large deformation diffeomorphic registration packages support GPUs. Our contributions are new algorithms and dedicated computational kernels to significantly reduce the runtime of the main computational kernels in CLAIRE: derivatives and interpolation. We deploy (i) highly-optimized, mixed-precision GPU-kernels for the evaluation of scattered-data interpolation, (ii) replace FFT-based first-order derivatives with optimized 8th-order finite differences, and (iii) compare with state-of-the-art CPU and GPU implementations. As a highlight, we demonstrate that we can register 256^3 clinical images in less than 6 seconds on a single NVIDIA Tesla V100. This amounts to over 20x speed-up over CLAIRE and over 30x speed-up over existing GPU implementations.

Poster 138: Across-Stack Profiling and Characterization of State-of-the-Art Machine Learning Models on GPUs
Cheng Li (University of Illinois), Abdul Dakkak (University of Illinois), Wei Wei (Alibaba Inc), Jinjun Xiong (IBM Research), Lingjie Xu (Alibaba Inc), Wei Zhang (Alibaba Inc), Wen-mei Hwu (University of Illinois)

The past few years have seen a surge of using Machine Learning (ML) and Deep Learning (DL) algorithms for traditional HPC tasks such as feature detection, numerical analysis, and graph analytics. While ML and DL enable solving HPC tasks, their adoption has been hampered due to the lack of understanding of how they utilize systems. Optimizing these algorithms requires characterizing their performance across the hardware/software (HW/SW) stack, but the lack of simple tools to automate the process and the reliance on researchers to perform manual characterization is a bottleneck. To alleviate this, we propose an across-stack profiling scheme and integrate it within MLModelScope — a hardware and software agnostic tool for evaluating and benchmarking ML/DL at scale. We demonstrate MLModelScope’s ability to characterize state-of-art ML/DL models and give insights that are only possible obtained by performing across-stack profiling.
Poster 60: Massively Parallel Large-Scale Multi-Model Simulation of Tumor Development  
Marco Berghoff (Karlsruhe Institute of Technology), Jakob Rosenbauer (Forschungszentrum Juelich), Alexander Schug (Forschungszentrum Juelich)

The temporal and spatial resolution in the microscopy of tissues has increased significantly within the last years, yielding new insights into the dynamics of tissue development and the role of the single-cell within it. A thorough theoretical description of the connection of single-cell processes to macroscopic tissue reorganizations is still lacking. Especially in tumor development, single cells play a crucial role in advance of tumor properties.

We developed a simulation framework that can model tissue development up to the centimeter scale with micrometer resolution of single cells. Through a full parallelization, it enables the efficient use of HPC systems, therefore enabling detailed simulations on a large scale. We developed a generalized tumor model that respects adhesion driven cell migration, cell-to-cell signaling, and mutation-driven tumor heterogeneity. We scan the response of the tumor development depending on division inhibiting substances such as cytostatic agents.

Best Poster Finalist: no

Poster 114: Optimizing Recommendation System Inference Performance Based on GPU  
Xiaowei Shen (Alibaba Inc), Junrui Zhou (Alibaba Inc), Kan Liu (Alibaba Inc), Lingling Jin (Alibaba Inc), Pengfei Fan (Alibaba Inc), Wei Zhang (Alibaba Inc), Jun Yang (University of Pittsburgh)

Neural network-based recommendation models have been widely applied on tracking personalization and recommendation tasks at large Internet companies such as e-commerce companies and social media companies. Alibaba recommendation system deploys WDL (wide and deep learning) models for product recommendation tasks. The WDL model consists of two main parts: embedding lookup and neural network-based feature ranking model that ranks different products for different users. As more and more products and users the model need to rank, the feature length and batch size of the models are increased. The computation of models is also increased so that traditional model inference implementation on CPU cannot meet the requirement of QPS (query per second) and latency of recommendation tasks. In this poster, we develop a GPU based system to speedup recommendation system inference performance. By model quantization and graph transformation, we can achieve 3.9x performance speedup when compared with a baseline GPU implementation.
**Poster 59: Accelerating BFS and SSSP on a NUMA Machine for the Graph500 Challenge**  
Tanuj K. Aasawat (RIKEN), Kazuki Yoshizoe (RIKEN), Tahsin Reza (University of British Columbia), Matei Ripeanu (University of British Columbia)

The NUMA architecture is the design choice for modern multi-CPU shared memory systems. In many ways, a NUMA system resembles a shared-nothing distributed system: memory accesses to remote NUMA domains are more expensive than local accesses.

In this work, we explore how improved data locality and reduced expensive remote communication can be achieved by exploiting "distributed" shared-memory of NUMA machines to develop shared-memory graph processing solutions optimized for NUMA systems. We introduce a novel hybrid design for memory accesses that handles the burst mode in traversal based algorithms, like BFS and SSSP, and reduces the number of remote accesses and updates. We demonstrate that our designs offer up to 84% speedup over our NUMA-oblivious framework Totem and 2.86x over shared-nothing distributed design, for BFS and SSSP algorithms.

Best Poster Finalist: no

---

**Poster 47: Decomposition Algorithms for Scalable Quantum Annealing**  
Elijah Pelofske (Los Alamos National Laboratory), Georg Hahn (Harvard University), Hristo Djidjev (Los Alamos National Laboratory)

Commercial adiabatic quantum annealers such as D-Wave 2000Q have the potential to solve NP-complete optimization problems efficiently. One of the primary constraints of such devices is the limited number and connectivity of their qubits. This research presents two exact decomposition methods (for the Maximum Clique and the Minimum Vertex Cover problem) that allow us to solve problems of arbitrarily large sizes by splitting them up recursively into a series of arbitrarily small subproblems. Those subproblems are then solved exactly or approximately using a quantum annealer. Whereas some previous approaches are based on heuristics that do not guarantee optimality of their solutions, our decomposition algorithms have the property that the optimal solution of the input problem can be reconstructed given all generated subproblems are solved optimally as well. We investigate various heuristic and exact bounds as well as reduction methods that help to increase the scalability of our approaches.

Best Poster Finalist: no
**Poster 90: You Have to Break It to Make It: How On-Demand, Ephemeral Public Cloud Projects with Alces Flight Compute Resulted in the Open-Source OpenFlightHPC Project**

Cristin Merritt (Alces Flight Limited; Alces Software Ltd, UK), Wil Mayers (Alces Flight Limited), Stu Franks (Alces Flight Limited)

Over three years ago the Alces Flight team made a decision to explore on-demand public cloud consumption for High Performance Computing (HPC). Our premise was simple, create a fully-featured, scalable HPC environment for research and scientific computing and provide it as a free subscription service to researchers in order to study how they would approach and consume this then new technology. This tool, Alces Flight Compute, would set out to chart how far away from the traditional bare-metal platforms our subscribers were willing to go. What we didn’t expect was that to get to their destination, our users would proceed to take our tool apart. This deconstruction has resulted in a new approach to HPC environment creation (the open-source OpenFlightHPC project), helped us better understand cloud adoption strategies, and handed over a set of guidelines to help those looking to bring public cloud into their HPC solution.

Best Poster Finalist: no

**Poster 126: Enforcing Crash Consistency of Scientific Applications in Non-Volatile Main Memory Systems**

Tyler Coy (Washington State University, Vancouver), Xuechen Zhang (Washington State University, Vancouver)

This poster presents a compiler-assistant technique, NVPath, to automatically generates NVMM-aware persistent data structures which provide the same level of guarantee of crash consistency compared to the baseline code. Compiler-assistant code annotation and transformation is general and can be applied to applications using various data structures. Our experimental results with real-world scientific applications show that the performance of the annotated programs is commensurate with the version using the manual code transformation on the Titan supercomputer.

Best Poster Finalist: no

**Poster 137: Warwick Data Store: A HPC Library for Flexible Data Storage in Multi-Physics Applications**

Richard O. Kirk (University of Warwick), Timothy R. Law (Atomic Weapons Establishment (AWE),
With the increasing complexity of memory architectures and multi-physics applications, developing data structures that are performant, portable, scalable, and support developer productivity, is difficult. In order to manage these complexities and allow rapid prototyping of different approaches we are building a lightweight and extensible C++ template library called the Warwick Data Store (WDS). WDS is designed to abstract details of the data structure away from the user, thus easing application development and optimisation. We show that WDS generates minimal performance overhead, via a variety of different scientific benchmarks and proxy-applications.

**Poster 76: HPChain: An MPI-Based Blockchain Framework for High Performance Computing Systems**

Abdullah Al-Mamun (University of Nevada, Reno; Lawrence Berkeley National Laboratory), Tonglin Li (Lawrence Berkeley National Laboratory), Mohammad Sadoghi (University of California, Davis), Linhua Jiang (Fudan University, Shanghai), Haoting Shen (University of Nevada, Reno), Dongfang Zhao (University of Nevada, Reno; University of California, Davis)

Data fidelity is of prominent importance for scientific experiments and simulations. The state-of-the-art mechanism to ensure data fidelity is through data provenance. However, the provenance data itself may as well exhibit unintentional human errors and malicious data manipulation. To enable a trustworthy and reliable data fidelity service, we advocate achieving the immutability and decentralization of scientific data provenance through blockchains. Specifically, we propose HPChain, a new blockchain framework specially designed for HPC systems. HPChain employs a new consensus protocol compatible with and optimized for HPC systems. Furthermore, HPChain was implemented with MPI and integrated with an off-chain distributed provenance service to tolerate the failures caused by faulty MPI ranks. The HPChain prototype system has been deployed to 500 cores at the University of Nevada’s HPC center and demonstrated strong resilience and scalability while outperforming state-of-the-art blockchains by orders of magnitude.

**Poster 104: An Adaptive Checkpoint Model For Large-Scale HPC Systems**

Subhendu S. Behera (North Carolina State University), Lipeng Wan (Oak Ridge National Laboratory), Frank Mueller (North Carolina State University), Matthew Wolf (Oak Ridge National Laboratory), Scott
Checkpoin Restart is a widely used Fault Tolerance technique for application resilience. However, failures and the overhead of saving application state for future recovery upon failure reduces the application efficiency significantly. This work contributes a failure analysis and prediction model making decisions for checkpoint data placement, recovery, and techniques for reducing checkpoint frequency. We also demonstrate a reduction in application overhead by taking proactive measures guided by failure prediction.

Best Poster Finalist: no

**Poster 123: Cloud-Native SmartX Intelligence Cluster for AI-Inspired HPC/HPDA Workloads**

Jungsu Han (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science), Jun-Sik Shin (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science), JinCheol Kwon (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science), JongWon Kim (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science)

In this poster, we introduce Cloud-native SmartX Intelligence Cluster for flexibly supporting AI-inspired HPC (high performance computing) / HPDA (high performance data analytics) workloads. This work has been continuously refined from 2013 with a futuristic vision for operating 100 petascale data center. Then, we discuss issues and approaches that come with building a Cloud-native SmartX Intelligence Cluster.

Best Poster Finalist: no

**Poster 86: High-Performance Custom Computing with FPGA Cluster as an Off-Loading Engine**

Takaaki Miyajima (RIKEN Center for Computational Science (R-CCS)), Tomohiro Ueno (RIKEN Center for Computational Science (R-CCS)), Jens Huthmann (RIKEN Center for Computational Science (R-CCS)), Atsushi Koshiba (RIKEN Center for Computational Science (R-CCS)), Kentaro Sano (RIKEN Center for Computational Science (R-CCS)), Mitsuhsisa Sato (RIKEN Center for Computational Science (R-CCS))

A heterogeneous system with Field Programmable Gate Arrays (FPGAs) is gathering attention in High-Performance Computing (HPC) area. We have been researching and developing a high-performance custom computing with FPGA cluster as an off-loading engine for an existing HPC system. In this research poster, we describe the motivation of our research and present research
topics on a software bridge between the FPGA cluster and existing HPC servers, and dedicated inter-FPGA networks.

Best Poster Finalist: no

**Poster 69: Optimization for Quantum Computer Simulation**

Naoki Yoshioka (RIKEN Center for Computational Science (R-CCS)), Hajime Inaoka (RIKEN Center for Computational Science (R-CCS)), Nobuyasu Ito (RIKEN Center for Computational Science (R-CCS)), Fengping Jin (Forschungszentrum Juelich), Kristel Michielsen (Forschungszentrum Juelich), Hans De Raedt (University of Groningen)

Simulator of quantum circuits is developed for massively parallel classical computers, and it is tested on the K computer in RIKEN R-CCS up to 45 qubits. Two optimization techniques are proposed in order to improve performance of the simulator. The "page method" reduces unnecessary copies in each node. It is found that this method makes approximately 17% speed-up maximum. Initial permutation of qubits is also studied how it affects performance of the simulator. It is found that a simple permutation in ascending order of the number of operations for each qubit is sufficient in the case of simulations of quantum adder circuits.

Best Poster Finalist: no

**Poster 52: Design and Specification of Large-Scale Simulations for GPUs Using FFTX**

Anuva Kulkarni (Carnegie Mellon University), Daniele Spampinato (Carnegie Mellon University), Franz Franchetti (Carnegie Mellon University)

Large-scale scientific simulations can be ported to heterogeneous environments with GPUs using domain decomposition. However, Fast Fourier Transform (FFT) based simulations require all-to-all communication and large memory, which is beyond the capacity of on-chip GPU memory. To overcome this, domain decomposition solutions are combined with adaptive sampling or pruning around the domain to reduce storage. Expression of such operations is a challenge in existing FFT libraries like FFTW, and thus it is difficult to get a high performance implementation of such methods. We demonstrate algorithm specification for one such simulation (Hooke’s law) using FFTX, an emerging API with a SPIRAL-based code generation back-end, and suggest future extensions useful for GPU-based scientific computing.

Best Poster Finalist: no
Poster 136: CHAMELEON: Reactive Load Balancing and Migratable Tasks for Hybrid MPI+OpenMP Applications

Jannis Klinkenberg (RWTH Aachen University), Philipp Samfaß (Technical University Munich), Michael Bader (Technical University Munich), Karl Fürlinger (Ludwig Maximilian University of Munich), Christian Terboven (RWTH Aachen University), Matthias S. Müller (RWTH Aachen University)

Many HPC applications are designed based on underlying performance and execution models. These models could successfully be employed in the past for balancing load within and between compute nodes. However, the increasing complexity of modern software and hardware makes performance predictability and load balancing much more difficult. Tackling these challenges in search for a generic solution, we present a novel library for fine-granular task-based reactive load balancing in distributed memory based on MPI and OpenMP. Our concept allows creating individual migratable tasks that can be executed on any MPI rank. Migration decisions are performed at run time based on online performance or load data. Two fundamental approaches to balance load and at the same time overlap computation and communication are compared. We evaluate our concept under enforced power caps and clock frequency changes using a synthetic benchmark and demonstrate robustness against work-induced imbalances for an AMR application.

Best Poster Finalist: no

Poster 124: Porting Finite State Automata Traversal from GPU to FPGA: Exploring the Implementation Space

Marziyeh Nourian (North Carolina State University), Mostafa Eghbali Zarch (North Carolina State University), Michela Becchi (North Carolina State University)

While FPGAs are traditionally considered hard to program, recently there are efforts to allow using high-level programming models intended for multi-core CPUs and GPUs to program FPGAs. For example, both Intel and Xilinx are now providing OpenCL-to-FPGA toolchains. However, since GPU and FPGA devices offer different parallelism models, OpenCL code optimized for GPU can prove inefficient on FPGA, in terms of both performance and hardware resource utilization.

In this poster, we explore this problem on an emerging workload: finite state automata traversal. Specifically, we explore a set of structural code changes, custom, and best-practice optimizations to retarget an OpenCL NFA engine designed for GPU to FPGA. Our evaluation, which covers traversal throughput and resource utilization, shows that our optimizations lead, on a single execution pipeline, to speedups up to 4x over an already optimized baseline that uses one of the proposed code changes to fit the original code on FPGA.
Poster 68: Linking a Next-Gen Remap Library into a Long-Lived Production Code
Charles R. Ferenbaugh (Los Alamos National Laboratory), Brendan K. Krueger (Los Alamos National Laboratory)

LANL's long-lived production application xRage contains a remapper capability that maps mesh fields from its native AMR mesh to the GEM mesh format used by some third-party libraries. The current remapper was implemented in a short timeframe and is challenging to maintain. Meanwhile, our next-generation code project has developed a modern remapping library Portage, and the xRage team wanted to link in Portage as an alternate mapper option. But the two codes are very different from each other, and connecting the two required us to deal with a number of challenges. This poster describes the codes, the challenges we worked through, current status, and some initial performance statistics.

Best Poster Finalist: no

Poster 131: Efficiency of Algorithmic Structures
Julian Miller (RWTH Aachen University), Lukas Trümper (RWTH Aachen University), Christian Terboven (RWTH Aachen University), Matthias S. Müller (RWTH Aachen University)

The implementation of high-performance parallel software is challenging and raises issues not seen in serial programs before. It requires a strategy of parallel execution which preserves correctness but maximizes scalability. Efficiently deriving well-scaling solutions remains an unsolved problem especially with the quickly-evolving hardware landscape of high-performance computing (HPC).

This work proposes a framework for classifying the efficiency of parallel programs. It bases on a strict separation between the algorithmic structure of a program and its executed functions. By decomposing parallel programs into a hierarchical structure of parallel patterns, a high-level abstraction is provided which leads to equivalence classes over parallel programs. Each equivalence class possesses efficiency properties, mainly communication and synchronization, dataflow and architecture efficiency. This classification allows for wide application areas and a workflow for structural optimization of parallel algorithms is proposed.

Best Poster Finalist: no
Poster 98: INSPECT Intranode Stencil Performance Evaluation Collection
Julian Hammer (University of Erlangen-Nuremberg), Julian Hornich (University of Erlangen-Nuremberg), Georg Hager (University of Erlangen-Nuremberg), Thomas Gruber (University of Erlangen-Nuremberg), Gerhard Wellein (University of Erlangen-Nuremberg)

Modeling and presenting performance data---even for simple kernels such as stencils---is not trivial. We therefore present an overview on how to interpret and what to learn from an INSPECT report, as well as highlighting best practices for performance data reporting.

INSPECT is the "Intranode Stencil Performance Evaluation Collection", which compiles performance benchmarks and reports of various stencil and streaming kernels on a variety of architectures. The goal is to aid performance-aware developers with reference material and a methodology to analyze their own codes.

INSPECT set out to cover these topics and compile a summary of all necessary information to allow reproduction of the performance results, their interpretation and discussion.

Best Poster Finalist: no

Poster 97: Optimizing Multigrid Poisson Solver of Cartesian CFD Code CUBE
Kazuto Ando (RIKEN Center for Computational Science (R-CCS)), Rahul Bale (RIKEN), Keiji Onishi (RIKEN Center for Computational Science (R-CCS)), Kiyoshi Kumahata (RIKEN Center for Computational Science (R-CCS)), Kazuo Minami (RIKEN Center for Computational Science (R-CCS)), Makoto Tsubokura (Kobe University, RIKEN Center for Computational Science (R-CCS))

We demonstrate an optimization of multigrid Poisson solver of Cartesian CFD code “CUBE (Complex Unified Building cube method)”. CUBE is a simulation framework for complex industrial flow problem, such as aerodynamics of vehicles, based on hierarchical Cartesian mesh. In incompressible CFD simulation, solving pressure Poisson equation is the most time-consuming part. In this study, we use a cavity flow simulation as a benchmark problem. With this problem, multigrid Poisson solver dominates 91% of execution time of the time-step loop. Specifically, we evaluate the performance of Gauss-Seidel loop as a computational kernel based on “Byte per Flop” approach. With optimization of the kernel, we achieved 9.8x speedup and peak floating point performance ratio increased from 0.4% to 4.0%. We also measured parallel performance up to 8,192 nodes (65,536 cores) on the K computer. With optimization of the parallel performance, we achieved 2.9x–3.9x sustainable speedup in the time-step loop.

Best Poster Finalist: no
Poster 85: Hybrid Computing Platform for Combinatorial Optimization with the Coherent Ising Machine
Junya Arai (Nippon Telegraph and Telephone Corporation), Yagi Satoshi (Nippon Telegraph and Telephone Corporation), Hiroyuki Uchiyama (Nippon Telegraph and Telephone Corporation), Toshimori Honjo (Nippon Telegraph and Telephone Corporation), Takahiro Inagaki (Nippon Telegraph and Telephone Corporation), Kensuke Inaba (Nippon Telegraph and Telephone Corporation), Takuya Ikuta (Nippon Telegraph and Telephone Corporation), Hiroki Takesue (Nippon Telegraph and Telephone Corporation), Keitaro Horikawa (Nippon Telegraph and Telephone Corporation)

Several institutes are operating cloud platforms that offer Web API access to Ising computers such as quantum annealing machines. Platform users can solve complex combinatorial optimization problems by using hybrid algorithms that utilize both users' conventional digital computers and remote Ising computers. However, communication via the Internet takes an order of magnitude longer time than optimization on Ising computers. This overheads seriously degrade the performance of hybrid algorithms since they involve frequent communication. In this poster, we first state issues in the design of Ising computing platforms, including communication overheads. Then, we answer the issues by introducing the computing platform for the coherent Ising machine (CIM), an Ising computer based on photonics technologies. Our platform offers efficient CIM-digital communication by allowing users to execute their program on digital computers co-located with the CIM. We have released the platform to our research collaborators in this autumn and started the evaluation.

Best Poster Finalist: no

Poster 117: A New Polymorphic Computing Architecture Based on Fine-Grained Instruction Mobility
David Hentrich (Illinois Institute of Technology), Erdal Oruklu (Illinois Institute of Technology), Jafar Saniei (Illinois Institute of Technology)

This is a summary of the base concepts behind David Hentrich’s May 2018 Ph.D. dissertation in Polymorphic Computing. Polymorphic Computing is the emerging field of changing the computer architecture around the software, rather than vice versa. The main contribution is a new polymorphic computing architecture. The key idea behind the architecture is to create an array of processors where a program’s instructions can be individually and arbitrarily assigned/mobilized to any processor, even during runtime. The key enablers of this architecture are a dataflow instruction set that is conducive to instruction migration, a microarchitectural block called an “operation cell” (“op-cell”), a processor built around the instruction set and the “op-cells”, and arrays of these processors.
Best Poster Finalist: no

**Poster 67: Genie: an MPEG-G Conformant Software to Compress Genomic Data.**

Brian E. Bliss (University of Illinois), Joshua M. Allen (University of Illinois), Saurabh Baheti (Mayo Clinic), Matthew A. Bockol (Mayo Clinic), Shubham Chandak (Stanford University), Jaime Delgado (Polytechnic University of Catalonia), Jan Fostier (Ghent University), Josep L. Gelpi (University of Barcelona), Steven N. Hart (Mayo Clinic), Mikel Hernaez Arrazola (University of Illinois), Matthew E. Hudson (University of Illinois), Michael T. Kalmbach (Mayo Clinic), Eric W. Klee (Mayo Clinic), Liudmila S. Mainzer (University of Illinois), Fabian Müntefering (Leibniz University), Daniel Naro (Barcelona Supercomputing Center), Idoia Ochoa-Alvarez (University of Illinois), Jörn Ostermann (Leibniz University), Tom Paridaens (Ghent University), Christian A. Ross (Mayo Clinic), Jan Voges (Leibniz University), Eric D. Wieben (Mayo Clinic), Mingyu Yang (University of Illinois), Tsachy Weissman (Stanford University), Mathieu Wiepert (Mayo Clinic)

Precision medicine has unprecedented potential for accurate diagnosis and effective treatment. It is supported by an explosion of genomic data, which continues to accumulate at accelerated pace. Yet storage and analysis of petascale genomic data is expensive, and that cost will ultimately be borne by the patients and citizens. The Moving Picture Experts Group (MPEG) has developed MPEG-G, a new open standard to compress, store, transmit and process genomic sequencing data that provides an evolved and superior alternative to currently used genomic file formats. Our poster will showcase software package GENIE, the first open source implementation of an encoder-decoder pair that is compliant with the MPEG-G specifications and delivers all its benefits: efficient compression, selective access, transport and analysis, guarantee of long-term support, and embedded mechanisms for annotation and encryption of compressed information. GENIE will create a step-change in medical genomics by reducing the cost of data storage and analysis.

Best Poster Finalist: no

**Poster 82: A View from the Facility Operations Side on the Water/Air Cooling System of the K Computer**

Jorji Nonaka (RIKEN Center for Computational Science (R-CCS)), Keiji Yamamoto (RIKEN Center for Computational Science (R-CCS)), Akiyoshi Kuroda (RIKEN Center for Computational Science (R-CCS)), Toshiyuki Tsukamoto (RIKEN Center for Computational Science (R-CCS)), Kazuki Koiso (Kobe University, RIKEN Center for Computational Science (R-CCS)), Naohisa Sakamoto (Kobe University, RIKEN Center for Computational Science (R-CCS))

The Operations and Computer Technologies Division at the RIKEN R-CCS is responsible for the
operations of the entire K computer facility, which includes the auxiliary subsystems such as the power supply and water/air cooling systems. It is worth noting that part of these subsystems will be reused in the next supercomputer (Fugaku), thus a better understanding of the operational behavior as well as the potential impacts especially on the hardware failure and energy consumption would be greatly beneficial. In this poster, we will present some preliminary impressions of the impact of the water/air cooling system on the K computer system, focusing on the potential benefits of the use of low water/air temperature respectively for the CPU and DRAM memory modules produced by the cooling system. We expect that the obtained knowledge will be helpful for the decision support and/or operation planning of the next supercomputer.

Best Poster Finalist: no

**Poster 135: High-Performance Deep Learning via a Single Building Block**
Evangelos Georganas (Intel Corporation), Kunal Banerjee (Intel Corporation), Dhiraj Kalamkar (Intel Corporation), Sasikanth Avancha (Intel Corporation), Anand Venkat (Intel Corporation), Michael Anderson (Intel Corporation), Greg Henry (Intel Corporation), Hans Pabst (Intel Corporation), Alexander Heinecke (Intel Corporation)

Deep learning (DL) is one of the most prominent branches of machine learning. Due to the immense computational cost of DL workloads, industry and academia have developed DL libraries with highly-specialized kernels for each workload/architecture, leading to numerous, complex code-bases that strive for performance, yet they are hard to maintain and do not generalize. In this work, we introduce the batch-reduce-GEMM kernel and show how the most popular DL algorithms can be formulated with this kernel as basic building-block. Consequently, the DL library-development degenerates to mere (potentially automatic) tuning of loops around this sole optimized kernel. By exploiting our kernel we implement Recurrent Neural Networks, Convolution Neural Networks and Multilayer Perceptron training and inference primitives in just 3K lines of high-level-code. Our primitives outperform vendor-optimized libraries on multi-node CPU-Clusters. We also provide CNN kernels targeting GPUs. Finally, we demonstrate that batch-reduce-GEMM kernel within a tensor compiler yields high-performance CNN primitives.

Best Poster Finalist: no

**Poster 56: Reinforcement Learning for Quantum Approximate Optimization**
Sami Khairy (Illinois Institute of Technology), Ruslan Shaydulin (Clemson University), Lukasz Cincio (Los Alamos National Laboratory), Yuri Alexeev (Argonne National Laboratory), Prasanna Balaprakash (Argonne National Laboratory)
The Quantum Approximate Optimization Algorithm (QAOA) is one of the leading candidates for demonstrating quantum advantage. The quality of the solution obtained by QAOA depends on the performance of the classical optimization routine used to optimize the variational parameters. In this work, we propose a Reinforcement Learning (RL) based approach to drastically reduce the number of evaluations needed to find high-quality variational parameters. We train an RL agent on small 8-qubit Max-Cut problem instances on an Intel Xeon Phi supercomputer Bebop, and use (transfer) the learned optimization policy to quickly find high-quality solutions for other larger problem instances coming from different distributions and graph classes. The preliminary results show that our RL based approach is able to improve the quality of the obtained solution by up to 10% within a fixed budget of function evaluations and demonstrate learned optimization policy transferability between different graph classes and sizes.

Best Poster Finalist: no


John Shalf (Lawrence Berkeley National Laboratory), Dilip Vasudevan (Lawrence Berkeley National Laboratory), David Donofrio (Lawrence Berkeley National Laboratory), Anastasia Butko (Lawrence Berkeley National Laboratory), Andrew Chien (University of Chicago), Yuanwei Fang (University of Chicago), Arjun Rawal (University of Chicago), Chen Zou (University of Chicago), Raymond Bair (Argonne National Laboratory), Kristopher Keipert (Argonne National Laboratory), Arun Rodriguez (Sandia National Laboratories), Maya Gokhale (Lawrence Livermore National Laboratory), Scott Lloyd (Lawrence Livermore National Laboratory), Xiaochen Guo (Lehigh University), Yuan Zeng (Lehigh University)

Accelerating technology disruptions and architectural change create growing opportunities and urgency to reduce the latency in for new architectural innovations to be deployed in extreme scale systems. We are exploring new architectural features that improve memory system performance including word-wise scratchpad memory, a flexible Recode engine, hardware message queues, and the data rearrangement engine (DRE). Performance results are promising yielding as much as 20x benefit. Project 38 is a cross-agency effort undertaken by the US Department of Energy (DOE) and Department of Defense (DoD).

Best Poster Finalist: no

**Poster 128: Identifying Time Series Similarity in Large-Scale Earth System Datasets**
Scientific data volumes are growing every day and instrument configurations, quality control and software updates result in changes to the data. This study focuses on developing algorithms that detect changes in time series datasets in the context of the Deduce project. We propose a combination of methods that include dimensionality reduction and clustering to evaluate similarity measuring algorithms. This methodology can be used to discover existing patterns and correlations within a dataset. The current results indicate that the Euclidean Distance metric provides the best results in terms of internal cluster validity measures for multi-variable analyses of large-scale earth system datasets. The poster will include details on our methodology, results, and future work.

Best Poster Finalist: no

**Poster 151: Three-Dimensional Characterization on Edge AI Processors with Object Detection Workloads**

Yujie Hui (Ohio State University), Jeffrey Lien (NovuMind Inc), Xiaoyi Lu (Ohio State University)

The Deep Learning inference applications are moving to the edge side, as edge-side AI platforms are cheap and energy-efficient. Different edge AI processors are diversified, since these processors are designed with different approaches. However, it is hard for customers to select an edge AI processor without an overall evaluation of these processors. We propose a three-dimensional characterization (i.e., accuracy, latency, and energy efficiency) approach on three different kinds of edge AI processors (i.e., Edge TPU, NVIDIA Xavier, and NovuTensor). We deploy YOLOv2 and Tiny-YOLO, which are two YOLO-based object detection systems, on these edge AI platforms with Microsoft COCO dataset. I will present our work starting from the problem statement. And then I'll introduce our experiments setup and hardware configuration. Lastly, I'll conclude our experimental results and current work status, as well as the future work.

Best Poster Finalist: no

**Poster 148: Unsupervised Clustering of Golden Eagle Telemetry Data**

Natalya I. Rapstine (US Geological Survey), Jeff A. Tracey (US Geological Survey), Janice M. Gordon
We use a recurrent autoencoder neural network to encode sequential California golden eagle telemetry data. The encoding is followed by an unsupervised clustering technique, Deep Embedded Clustering (DEC), to iteratively cluster the data into a chosen number of behavior classes. We apply the method to simulated movement data sets and telemetry data for a Golden Eagle. The DEC achieves better unsupervised clustering accuracy scores for the simulated data sets as compared to the baseline K-means clustering result.

Best Poster Finalist: no

Poster 66: Hybrid CPU/GPU FE2 Multi-Scale Implementation Coupling Alya and Microp
Guido Giuntoli (Barcelona Supercomputing Center), Judicaël Grasset (Science and Technology Facilities Council (STFC)), Alejandro Figueroa (George Mason University), Charles Moulinec (Science and Technology Facilities Council (STFC)), Mariano Vázquez (Barcelona Supercomputing Center), Guillaume Houzeaux (Barcelona Supercomputing Center), Stephen Longshaw (Science and Technology Facilities Council (STFC)), Sergio Oller (Polytechnic University of Catalonia)

This poster exposes the results of a new implementation of the FE2 multi-scale algorithm that is achieved by coupling the multi-physics and massively parallel code Alya with the GPU-based code microp. The coupled code is mainly designed to solve large scale and realistic composite material problems for the aircraft industry. Alya is responsible of solving the macro-scale equations and microp for solving the representation of fibres at the microscopic level. The poster shows computational performance results that demonstrate that the technique is scalable for real size industrial problems and also how the execution time is dramatically reduced using GPU-based clusters.

Best Poster Finalist: no

Poster 129: Understanding I/O Behavior in Scientific Workflows on High Performance Computing Systems
Fahim Tahmid Chowdhury (Florida State University, Lawrence Livermore National Laboratory), Francesco Di Natale (Lawrence Livermore National Laboratory), Adam Moody (Lawrence Livermore National Laboratory), Elsa Goniorowski (Lawrence Livermore National Laboratory), Kathryn Mohror (Lawrence Livermore National Laboratory), Weikuan Yu (Florida State University)

Leadership high performance computing (HPC) systems have the capability to execute workflows of
In this poster, we describe our efforts to extract the I/O characteristics of various HPC workflows and develop strategies to improve I/O performance by leveraging heterogeneous storage systems. We have implemented an emulator to mimic different types of I/O requirements posed by HPC application workflows. We have analyzed the workflow of Cancer Moonshot Pilot 2 (CMP2) project to determine possible I/O inefficiencies. To date, we have performed a systematic characterization and evaluation on the workloads generated by the workflow emulator and a small scale adaptation of the CMP2 workflow.

Best Poster Finalist: no

Poster 116: Advancements in Ultrasound Simulations Enabled by High-Bandwidth GPU Interconnects
Filip Vaverka (Brno University of Technology), Bradley E. Treeby (University College London), Jiri Jaros (Brno University of Technology)

Realistic ultrasound simulations are becoming integral part of many novel medical procedures such as photoacoustic screening and non-invasive treatment planning. The common denominator of all these applications is the need for cheap and relatively large-scale ultrasound simulations with sufficient accuracy. Typical medical applications require full-wave simulations which take frequency-dependent absorption and non-linearity into account.

This poster investigates the benefits of high-bandwidth low-latency interconnects to k-Wave acoustic toolbox in dense multi-GPU environment. The k-Wave multi-GPU code is based on a variant of the local Fourier basis domain decomposition. The poster compares the behavior of the code on a typical PCI-E 3.0 machine with 8 Nvidia Tesla P40 GPUs and a Nvidia DGX-2 server. The performance constraints of PCI-E platforms built around multiple socket servers on multi-GPU applications are deeply explored. Finally, it is shown the k-Wave toolbox can efficiently utilize NVlink 2.0 and achieve over 4x speedup compared to PCI-E systems.

Best Poster Finalist: no

Poster 65: Comparing Granular Dynamics vs. Fluid Dynamics via Large DOF-Count Parallel
Simulation on the GPU
Milad Rakhsha (University of Wisconsin), Conlain Kelly (Georgia Institute of Technology), Nicholas Olsen (University of Wisconsin), Lijing Yang (University of Wisconsin), Radu Serban (University of Wisconsin), Dan Negrut (University of Wisconsin)

In understanding granular dynamics, the commonly-used discrete modeling approach that tracks the motion of all particles is computationally demanding, especially with large system size. In such cases, one can contemplate switching to continuum models that are computationally less expensive. In order to assess when such a discrete to continuum switch is justified, we compare granular and fluid dynamics that scales to handle more than 1 billion degrees of freedom (DOFs); i.e., two orders of magnitude higher than the state-of-the-art. On the granular side, we solve the Newton-Euler equations of motion; on the fluid side, we solve the Navier-Stokes equations. Both solvers leverage parallel computing on the GPU, and are publicly available on GitHub as part of an open-source code called Chrono. We report similarities and differences between the dynamics of the discrete, fully-resolved system and the continuum model via numerical experiments including both static and highly transient scenarios.

Best Poster Finalist: no

Poster 78: Understanding HPC Application I/O Behavior Using System Level Statistics
Arnab K. Paul (Virginia Tech), Olaf Faaland (Lawrence Livermore National Laboratory), Adam Moody (Lawrence Livermore National Laboratory), Elsa Gonsiorowski (Lawrence Livermore National Laboratory), Kathryn Mohror (Lawrence Livermore National Laboratory), Ali R. Butt (Virginia Tech)

The processor performance of high performance computing (HPC) systems is increasing at a much higher rate than storage performance. Storage and file system designers therefore require a deep understanding of how HPC application I/O behavior affects current storage system installations in order to improve storage performance. In this work, we contribute to this understanding using application-agnostic file system statistics gathered on compute nodes as well as metadata and object storage file system servers. We analyze file system statistics of more than 4 million jobs over a period of three years on two systems at Lawrence Livermore National Laboratory that include a 15 PiB Lustre file system for storage. Some key observations in our study show that more than 65% HPC users perform significant I/O which are mostly writes; and less than 22% of HPC users who submit write-intensive jobs perform efficient writes to the file system.

Best Poster Finalist: no
Load imbalance is the major source of performance degradation in computationally-intensive applications that frequently consist of parallel loops. Efficient scheduling can improve the performance of such programs. OpenMP is the de-facto standard for parallel programming on shared-memory systems. The current OpenMP specification provides only three choices for loop scheduling which are insufficient in scenarios with irregular loops, system-induced interference, or both. Therefore, this work augments the LLVM OpenMP runtime library implementation with eleven ready to use scheduling techniques. We tested existing and added scheduling strategies on several applications from NAS, SPEC OMP 2012, and CORAL2 benchmark suites. Experiments show that implemented scheduling techniques outperform others in certain application and system configurations. We measured performance gains of up to 6% compared to the fastest standard scheduling technique. This work aims to be a convincing step toward beyond-standard scheduling options in OpenMP for the benefit of evolving applications executing on multicore architectures.

Best Poster Finalist: no

Poster 143: Quantum Natural Language Processing
Lee James O'Riordan (Irish Centre for High-End Computing (ICHEC), National University of Ireland Galway (NUIG)), Myles Doyle (Irish Centre for High-End Computing (ICHEC), National University of Ireland Galway (NUIG)), Venkatesh Kannan (Irish Centre for High-End Computing (ICHEC), National University of Ireland Galway (NUIG)), Fabio Baruffa (Intel Corporation)

Natural language processing (NLP) algorithms that operate over strings of words are limited since they analyse meanings of the component words in a corpus without information about grammatical rules of the language. Consequently, they often produce unsatisfactory results with increase in problem complexity.

The "distributed compositional semantics" (DisCo) model incorporates grammatical structure of sentences into the algorithms, and offers significant improvements to the quality of results. However, their main challenge is the need for large classical computational resources. The DisCo model presents two quantum algorithms which lower storage and compute requirements compared to a classic HPC implementation.
In this project, we implement the two DisCo model quantum algorithms on the Intel Quantum Simulator deployed on the Irish national supercomputer. We target corpuses with ~1000 most-common words using up to 36 qubits simulation. The solution will be able to compute the meanings of two sentences and decide if their meanings match.

Best Poster Finalist: no

**Poster 152: Deep Domain Adaptation for Runtime Prediction in Dynamic Workload Scheduler**  
Hoang H. Nguyen (National Center for Atmospheric Research (NCAR); University of Illinois, Chicago), Ben Matthews (National Center for Atmospheric Research (NCAR)), Irfan Elahi (National Center for Atmospheric Research (NCAR))

In HPC systems, users' requested runtime for submitted jobs plays a crucial role in efficiency. While underestimation of job runtime could terminate jobs before completion, overestimation could result in long queuing of other jobs in HPC systems. In reality, runtime prediction in HPC is challenging due to the complexity and dynamics of running workloads. Most of the current predictive runtime models are trained on static workloads. This poses a risk of over-fitting the predictions with bias from the learned workload distribution. In this work, we propose an adaptation of Correlation Alignment method in our deep neural network architecture (DCORAL) to alleviate the domain shift between workloads for better runtime predictions. Experiments on both standard benchmark workloads and NCAR real-time production workloads reveal that our proposed method results in a more stable training model across different workloads with low accuracy variance as compared to the other state-of-the-art methods.

Best Poster Finalist: no

**Poster 75: libCEED - Lightweight High-Order Finite Elements Library with Performance Portability and Extensibility**  
Jeremy Thompson (University of Colorado), Valeria Barra (University of Colorado), Yohann Dudouit (Lawrence Livermore National Laboratory), Oana Marin (Argonne National Laboratory), Jed Brown (University of Colorado)

High-order numerical methods are widely used in PDE solvers, but software packages that have provided high-performance implementations have often been special-purpose and intrusive. libCEED is a new library that offers a purely algebraic interface for matrix-free operator representation and supports run-time selection of implementations tuned for a variety of
computational device types, including CPUs and GPUs. We introduce the libCEED API and demonstrate how it can be used in standalone code or integrated with other packages (e.g., PETSc, MFEM, Nek5000) to solve examples of problems that often arise in the scientific computing community, ranging from fast solvers via geometric multigrid methods to Computational Fluid Dynamics (CFD) applications.

Best Poster Finalist: no

**Progress on the Exascale Transition of the VSim Multiphysics PIC code**
Benjamin M. Cowan (Tech-X Corporation), Sergey N. Averkin (Tech-X Corporation), John R. Cary (Tech-X Corporation), Jarrod Leddy (Tech-X Corporation), Scott W. Sides (Tech-X Corporation), Ilya A. Zilberter (Tech-X Corporation)

The highly performant, flexible plasma simulation code VSim was designed nearly 20 years ago (originally as Vorpal), with its first applications roughly four years later. Using object oriented methods, VSim was designed to allow runtime selection from multiple field solvers, particle dynamics, and reactions. It has been successful in modeling for many areas of physics, including fusion plasmas, particle accelerators, microwave devices, and RF and dielectric structures. Now it is critical to move to exascale systems, with their compute accelerator architectures, massive threading, and advanced instruction sets. Here we discuss how we are moving this complex, multiphysics computational application to the new computing paradigm, and how it is done in a way that kept the application producing physics during the move. We present performance results showing significant speedups in all parts of the PIC loop, including field updates, particle pushes, and reactions.

Best Poster Finalist: no

**Poster 58: Lock-Free van Emde Boas Array**
Ziyuan Guo (University of Tokyo)

Lock-based data structures have some potential issues such as deadlock, livelock, and priority inversion, and the progress can be delayed indefinitely if the thread that is holding locks cannot acquire a timeslice from the scheduler. Lock-free data structures, which guarantees the progress of some method call, can be used to avoid these problems. This poster introduces the first lock-free concurrent van Emde Boas Array which is a variant of van Emde Boas Tree. It is linearizable, and the benchmark shows significant performance improvement comparing to other lock-free search trees when the date set is large and dense enough.
Poster 63: Adaptive Execution Planning in Biomedical Workflow Management Systems
Marta Jaros (Brno University of Technology), Bradley E. Treeby (University College London), Jiri Jaros (Brno University of Technology)

Biomedical simulations require very powerful computers. Their execution is described by a workflow consisting of a number of different cooperating tasks. The manual execution of individual tasks may be tedious for expert users, but prohibiting for most inexperienced clinicians. k-Dispatch offers a ‘run and forget’ approach where the users are completely screened out from the complexity of HPC systems. k-Dispatch provides task scheduling, execution, monitoring, and fault tolerance. Since the task execution configuration strongly affects the final tasks mapping on the computational resources, the execution planning is of the highest priority. Unlike other tools, k-Dispatch considers a variable amount of computational resources per individual tasks. Since the scaling of the individual HPC codes is never perfect, k-Dispatch may find such a good mapping even an experienced user would miss. The proposed adaptive execution planning is based on collected performance data and the current cluster utilization monitoring.

Poster 142: Training Deep Neural Networks Directly on Hundred-Million-Pixel Histopathology Images on a Large-Scale GPU Cluster
Chi-Chung Chen (AetherAI, Taiwan), Wen-Yu Chuang (Chang-Gung Memorial Hospital, Taiwan), Wei-Hsiang Yu (AetherAI, Taiwan), Hsi-Ching Lin (National Center for High-Performance Computing (NCHC), Taiwan), Shuen-Tai Wang (National Center for High-Performance Computing (NCHC), Taiwan), Fang-An Kuo (National Center for High-Performance Computing (NCHC), Taiwan), Chao-Chun Chuang (National Center for High-Performance Computing (NCHC), Taiwan), Chao-Yuan Yeh (AetherAI, Taiwan)

Deep learning for digital pathology is challenging because the resolution of whole-slide-images (WSI) is extremely high, often in billions. The most common approach is patch-based method, where WSIs are divided into small patches to train convolutional neural networks (CNN). This approach has significant drawbacks. To have ground truth for individual patches, detailed annotations by pathologists are required. This laborious process has become the major impediment to the development of digital pathology AI. End-to-end WSI training, however, faces the difficulties of fitting the task into limited GPU memory. In this work, we improved the efficiency of using system
memory for GPU compute by 411% through memory optimization and deployed the training pipeline on 8 nodes, totally 32 GPUs distributed system, achieving 147.28x speedup. We demonstrated that CNN is capable of learning features without detailed annotations. The trained CNN can correctly classify cancerous specimen, with performance level closely matching the patch-based methods.

Best Poster Finalist: no

**Poster 95: A Heterogeneous HEVC Video Encoder Based on OpenPOWER Acceleration Platform**
Chenhao Gu (Fudan University, Shanghai, State Key Laboratory of ASIC and System; IBM Corporation), Yang Chen (Fudan University, Shanghai, State Key Laboratory of ASIC and System; IBM Corporation), Yanheng Lu (IBM Corporation), Pengfei Gou (IBM Corporation), Yong Lu (IBM Corporation), Yang Dai (IBM Corporation), Yue Xu (IBM Corporation), Yang Liu (IBM Corporation), Yibo Fan (Fudan University, Shanghai, State Key Laboratory of ASIC and System)

This poster describes a heterogeneous HEVC video encoder system based on the OpenPOWER platform. Our design leverages the Coherent Accelerator Processor Interface (CAPI) on the OpenPOWER, which provides cache-coherent access for FPGA. This technology highly improves CPU-FPGA data communication bandwidth and programming efficiency. X265 is optimized on the OpenPOWER platform to improve its performance with both architecture specific methods and hardware-acceleration methods. For hardware acceleration, frame-level acceleration and functional-unit-level acceleration are introduced and evaluated in this work.

Best Poster Finalist: no

**Poster 102: Fast Training of an AI Radiologist: Leveraging Data Pipelining to Efficiently Utilize GPUs**
Rakshith Vasudev (Dell EMC), John A. Lockman III (Dell EMC), Lucas A. Wilson (Dell EMC), Srinivas Varadharajan (Dell EMC), Frank Han (Dell EMC), Rengan Xu (Dell EMC), Quy Ta (Dell EMC)

In a distributed deep learning training setting, using accelerators such as GPUs can be challenging to develop a high throughput model. If the accelerators are not utilized effectively, this could mean more time to solution, and thus the model's throughput is low. To use accelerators effectively across multiple nodes, we need to utilize an effective data pipelining mechanism that handles scaling gracefully so GPUs can be exploited of their parallelism. We study the effect of using the correct pipelining mechanism that is followed by tensorflow official models vs a naive pipelining
mechanism that doesn't scale well, on two image classification models. Both the models using the optimized data pipeline demonstrate effective linear scaling when GPUs are added. We also show that converting to TF Records is not always necessary.

Best Poster Finalist: no

Poster 94: Multi-GPU Optimization of a Non-Hydrostatic Numerical Ocean Model with Multigrid Preconditioned Conjugate Gradient Method
Takateru Yamagishi (Research Organization for Information Science and Technology, Japan), Yoshimasa Matsumura (University of Tokyo, Atmosphere and Ocean Research Institute), Hiroyasu Hasumi (University of Tokyo, Atmosphere and Ocean Research Institute)

The conjugate gradient method with multigrid preconditioners (MGCG) is used in scientific applications because of its high performance and scalability with many computational nodes. GPUs are thought to be good candidates for accelerating such applications with many meshes where an MGCG solver could show high performance. No previous studies have evaluated and discussed the numerical character of an MGCG solver on GPUs. Consequently, we have implemented and optimized our “kinaco” numerical ocean model with an MGCG solver on GPUs. We evaluated its performance and discussed inter-GPU communications on a coarse grid on which GPUs could be intrinsically problematic. We achieved 3.9 times speedup compared to CPUs and learned how inter-GPU communications depended on the number of GPUs and the aggregation level of information in a multigrid method.

Best Poster Finalist: no

Poster 108: Power Prediction for High-Performance Computing
Shigeto Suzuki (Fujitsu Laboratories Ltd), Michiko Hiraoka (Fujitsu Ltd), Takashi Shiraishi (Fujitsu Laboratories Ltd), Enxhi Kreshpa (Fujitsu Laboratories Ltd), Takuji Yamamoto (Fujitsu Laboratories Ltd), Hiroyuki Fukuda (Fujitsu Laboratories Ltd), Shuji Matsui (Fujitsu Ltd), Masahide Fujisaki (Fujitsu Ltd), Atsuya Uno (RIKEN Center for Computational Science (R-CCS))

Exascale computers consume large amounts of power both for computing and cooling-units. As power of the computer varies dynamically corresponding to the load change, cooling-units are desirable to follow it for effective energy management. Because of time lags in cooling-unit operations, advance control is inevitable and an accurate prediction is a key for it. Conventional prediction methods make use of the similarity between job information while in queue. The prediction fails if there is no previously similar job. We developed two models to correct the
prediction after queued jobs start running. By taking power histories into account, power-correlated
topic model reselects more suitable candidate and recurrent-neural-network model considering
variable network sizes predicts power variation from shape features of it. We integrated these into
a single algorithm and demonstrated high-precision prediction with an average relative error of
5.7% in K computer as compared to the 18.0% obtained using the conventional method.

Best Poster Finalist: no

Poster 80: Sharing and Replicability of Notebook-Based Research on Open Testbeds
Maxine V. King (University of Chicago), Jason Anderson (University of Chicago), Kate Keahey
(Argonne National Laboratory, University of Chicago)

We seek to facilitate replicability by creating a way to share experiments easily in and out of
notebook-based, open testbed environments and a sharing platform for such experiments in order
to allow researchers to combine shareability, consistency of code environment, and well-
documented process.

Best Poster Finalist: no

Poster 121: HFlush: Realtime Flushing for Modern Storage Environments
Jaime Cernuda (Illinois Institute of Technology), Hugo Trivino (Illinois Institute of Technology),
Hariharan Devarajan (Illinois Institute of Technology), Anthony Kougkas (Illinois Institute of
Technology), Xian-He Sun (Illinois Institute of Technology)

Due to the unparalleled magnitude of data movement in extreme scale computing, I/O has become
a central challenge. Modern storage environments have proposed the use of multiple layers
between applications and the PFS. Nonetheless, the difference in capacities and speeds between
storage layers makes it extremely challenging to evict data from upper layers to lower layers
efficiently. However, current solutions are executed in batches, compromising latency; are also
push-based implementations, compromising resource utilization. Hence, we propose HFlush, a
continuous data eviction mechanism built on a streaming architecture that is pull-based and in
which each component is decoupled and executed in parallel. Initial results have shown RFlush to
obtain a 7X latency reduction and a 2X bandwidth improvement over a baseline batch-based
system. Therefore, RFlush is a promising solution to the growing challenges of extreme scale data
generation and eviction shortcomings when archiving data across multiple tiers of storage.

Best Poster Finalist: no
Poster 88: HPC Container Runtime Performance Overhead: At First Order, There Is None
Alfred Torrez (Los Alamos National Laboratory), Reid Priedhorsky (Los Alamos National Laboratory), Timothy Randles (Los Alamos National Laboratory)

Linux containers are an increasingly popular method used by HPC centers to meet increasing demand for greater software flexibility. A common concern is that containers may introduce application performance overhead. Prior work has not tested a broad set of HPC container technologies on a broad set of benchmarks. This poster addresses the gap by comparing performance of the three HPC container implementations (Charliecloud, Shifter, and Singularity) and bare metal on multiple dimensions using industry-standard benchmarks.

We found no meaningful performance differences between the four environments with the possible exception of modest variation in memory usage, which is broadly consistent with prior results. This result suggests that HPC users should feel free to containerize their applications without concern about performance degradation, regardless of the container technology used. It is an encouraging development on the path towards greater adoption of user-defined software stacks to increase the flexibility of HPC.

Best Poster Finalist: no

Tsuyoshi Ichimura (University of Tokyo, RIKEN), Kohei Fujita (University of Tokyo, RIKEN), Takuma Yamaguchi (University of Tokyo), Akira Naruse (Nvidia Corporation), Jack C. Wells (Oak Ridge National Laboratory), Christopher J. Zimmer (Oak Ridge National Laboratory), Tjerk P. Straatsma (Oak Ridge National Laboratory), Takane Hori (Japan Agency for Marine-Earth Science and Technology), Simone Puel (University of Texas), Thorsten W. Becker (University of Texas), Muneo Hori (Japan Agency for Marine-Earth Science and Technology), Naonori Ueda (RIKEN)

We propose herein an approach for reformulating an equation-based modeling algorithm to an algorithm similar to that of training artificial intelligence (AI) and accelerate this algorithm using high-performance accelerators to reduce the huge computational costs encountered for physics equation-based modeling in earthquake disaster mitigation. A fast scalable equation-based implicit solver on unstructured finite elements is accelerated with a Tensor Core-enabled matrix-vector product kernel. The developed kernel attains 1.10 ExaFLOPS, leading to 416 PFLOPS for the whole
solver on full Summit. This corresponds to a 75-fold speedup from a previous state-of-the-art
solver running on full Piz Daint. This result could lead to breakthroughs in earthquake disaster
mitigation. Our new idea in the HPC algorithm design of combining equation-based modeling with
AI is expected to have broad impacts in other earth science and industrial problems.

Best Poster Finalist: no

**Poster 91: FreeCompilerCamp: Online Training for Extending Compilers**
Alok Mishra (Stony Brook University, Lawrence Livermore National Laboratory), Anjia Wang
(University of North Carolina, Charlotte; Lawrence Livermore National Laboratory), Chunhua Liao
(Lawrence Livermore National Laboratory), Yonghong Yan (University of North Carolina, Charlotte),
Barbara Chapman (Stony Brook University)

In this presentation, we introduce an ongoing effort of an online training platform aimed to
automate the training of developers to quickly extend compilers - FreeCompilerCamp.org. Our free
and open platform allows anyone who is interested in developing compilers to learn the necessary
skills. A live training website, built on top of Play-With-Docker, is set up so that anyone with
internet access and a web browser will be able to take this training. The entire training system is
open-source and developers with relevant skills can contribute new tutorials and deploy it on a
private server, workstation or even laptop. We have created some initial tutorials on how to extend
the Clang/LLVM or ROSE compilers to support new OpenMP features. Using a web interface
consisting of two side-by-side panels, users can follow the tutorials on one side and immediately
practice what they learned in a terminal sandbox embedded on the other.

Best Poster Finalist: yes

**Poster 92: Nanoporous Flow Simulations on the Summit Supercomputer**
Yidong Xia (Idaho National Laboratory), Lixiang Luo (IBM - TJ Watson Research Center), Ansel
Blumers (Brown University), Joshua Kane (Idaho National Laboratory), Jan Goral (University of Utah),
Yu-Hang Tang (Lawrence Berkeley National Laboratory), Zhen Li (Clemson University, Brown
University), Hai Huang (Idaho National Laboratory), Milind Deo (University of Utah)

Fluid flow behaviors in nanoporous materials are distinct from those following the continuum
physics. Numerical simulations can be a complement to laboratory experiments. This work presents
a dissipative particle dynamics (DPD) package for GPU-accelerated mesoscale flow simulations in
nanoporous materials. In an ideal benchmark that minimizes load imbalance, the package delivered
nearly perfect strong- and weak-scaling (with up to 4 billion DPD particles) on up to 1,536 V100
GPUs on Oak Ridge National Laboratory’s Summit supercomputer. More remarkably, in a benchmark to measure its usefulness with realistic nanopores in SBA-15 silica, the package exhibited more than 20x speedup over its LAMMPS-based CPU counterpart with the same number nodes (e.g., 384 V100 GPUs vs. 2,688 POWER9 cores). It is worth highlighting that the NVLink2 Host-to-Device interconnects kept the cost of CPU-GPU memory copy as low as only 10% of GPU activity time per rank: 4 times less than their PCIe counterparts.

Best Poster Finalist: yes

**Poster 96: TSQR on TensorCores**
Hiroyuki Ootomo (Tokyo Institute of Technology), Rio Yokota (Tokyo Institute of Technology)

Tall-Skinny QR (TSQR) is an efficient algorithm for calculating the QR decomposition of $m \times n$ matrices where $m \ll n$, which is done by recursively performing QR decomposition on subdivided blocks of the tall and skinny matrix. Such operations are useful for low-rank approximation methods, which are replacing more and more dense linear algebra in both scientific computing and machine learning fields. The present work focuses on the implementation of this important algorithm on Tensor Cores, which are available on the latest NVIDIA GPUs. We evaluate the speed, accuracy, and stability of TSQR on TensorCores.

Best Poster Finalist: yes

**Poster 110: Hierarchical Data Prefetching in Multi-Tiered Storage Environments**
Hariharan Devarajan (Illinois Institute of Technology), Anthony Kougkas (Illinois Institute of Technology), Xian-He Sun (Illinois Institute of Technology)

In the era of data-intensive computing, accessing data with a high-throughput and low-latency is very imperative. Data prefetching is used for hiding read latency by requesting data before it is needed to move it from a high-latency medium to a low-latency one. However, existing solutions do not consider the multi-tiered storage and also suffer from under-utilization of prefetching resources and unnecessary evictions. Additionally, existing approaches implement a client-pull model where understanding the application’s I/O behavior drives prefetching decisions. Moving toward exascale, where machines run multiple applications concurrently by accessing files in a workflow, a more data-centric approach resolves challenges such as cache pollution and redundancy. In this study, we present HFetch, a truly hierarchical data prefetcher that adopts a server-push approach to data prefetching. We demonstrate the benefits of such an approach. Results show 10-35% performance gains over existing prefetchers and over 50% when compared to systems with no prefetching.
**Poster 54: Massively Parallel Eigensolvers Based on Unconstrained Energy Functionals Methods**

*Osni Marques (Lawrence Berkeley National Laboratory), Mauro Del Ben (Lawrence Berkeley National Laboratory), Andrew Canning (Lawrence Berkeley National Laboratory)*

This poster focuses on a preconditioned conjugate gradient based iterative eigensolver using an unconstrained energy functional minimization scheme. This scheme avoids an explicit reorthogonalization of the trial eigenvectors and becomes an attractive alternative for the solution of very large problems. The unconstrained formulation is implemented in the first-principles materials and chemistry CP2K code, which performs electronic structure calculations based on a density functional theory approximation to the solution of the many-body Schrödinger equation. The systems we use in our studies have a number of atoms ranging from 2,247 to 12,288. We study the convergence of the unconstrained formulation and its scaling on a Cray XC40 (a partition with 9,688 Intel KNL nodes). We show that there is a trade-off between the preconditioner that leads to fast convergence and lower cost preconditioners that lead to best time to solution.

**8:30 am - 5:00 pm**

**Scientific Visualization & Data Analytics Showcase Posters Display**

**Visualization of Entrainment and Mixing Phenomena at Cloud Edges**

*Bipin Kumar (Indian Institute of Tropical Meteorology), Matt Rehme (National Center for Atmospheric Research (NCAR)), Neethi Suresh (Indian Institute of Tropical Meteorology), Nihanth Cherukuru (National Center for Atmospheric Research (NCAR)), Stanislaw Jaroszynski (National Center for Atmospheric Research (NCAR)), Shaomeng Li (National Center for Atmospheric Research (NCAR)), Scott Pearse (National Center for Atmospheric Research (NCAR)), Tim Scheitlin (National Center for Atmospheric Research (NCAR))*

Entrainment and mixing processes occur during the entire life of a cloud. These processes change the droplet size distribution, which determines rain formation and radiative properties. Since it is a microphysical process, it cannot be resolved in large scale weather forecasting models. Small scale simulations such as Direct Numerical Simulations (DNS) are required to resolve the smallest scale of
these processes. However, it has been a challenge to visualize these processes in a 3D domain as it generates petabytes of data. Visualization plays a vital role in analyzing and understanding these huge data outputs. Here, we present different techniques for 3D visualization of data obtained from DNS carried out at Indian Institute of Tropical Meteorology (IITM) to understand cloud microphysical properties more closely.

Multiple tools were used to conduct a visual analysis of this data. Two of these tools are well established and tested technologies: ParaView and VAPOR. The others are emergent technologies created at the National Center for Atmospheric Research (NCAR) and are in the development phase. This data simulation, in addition to exploring DNS as mentioned above, provided an opportunity to test and improve development of several tools and methods.

**Visualizing Deep Learning at Scale for the Construction of Galaxy Catalogs in the Dark Energy Survey**
Janet Y. K. Knowles (Argonne National Laboratory), Joseph A. Insley (Argonne National Laboratory, Northern Illinois University), Silvio Rizzi (Argonne National Laboratory), Elise Jennings (Argonne National Laboratory), Asad Khan (University of Illinois), Eliu Huerta (University of Illinois), Sibo Wang (University of Illinois), Robert Gruendl (University of Illinois), Huihuo Zheng (Argonne National Laboratory)

The advent of machine and deep learning algorithms on petascale supercomputers is accelerating the pace of discovery in astrophysics and poses significant challenges to the interpretability of these deep neural networks. We present a novel visualization of a deep neural network output during training as it is learning to classify galaxy images as either spiral or elliptical. The network is trained using labeled datasets from the citizen science campaign, Galaxy Zoo, adopted by the Sloan Digital Sky Survey. These trained neural network models can then be used to classify galaxies in the Dark Energy Survey that overlap the footprint of both surveys. Visualizing a reduced representation of the network output, projected into 3-D parameter space, reveals how the network has discovered two distinct clusters of features which allows it to classify galaxies into two groups. These visualizations of the neural network during training aid in the interpretability of the black box of deep learning and reveal how the network responds to the input images at various stages of training. Finally, it allows a wider net to be cast to a general audience, thereby generating interest in and visibility to an otherwise highly specialized field.

**NVIDIA IndeX Accelerated Computing for Visualizing Cholla’s Galactic Winds**
Christopher Lux (Nvidia Corporation), Alexander Kuhn (Nvidia Corporation), Evan Schneider (Princeton University), Brant Robertson (University of California, Santa Cruz), Marc Nienhaus (Nvidia
Galactic winds – outflows of gas driven out of galaxies by the combined effects of thousands of supernovae – are a crucial feature of galaxy evolution. Despite their importance, a complete theoretical picture of these winds has been elusive. Simulating the complicated interaction between the hot, high pressure gas created by supernovae and the cooler, high density gas in the galaxy disk requires massive computational resources and sophisticated software. For this purpose, Computational Hydrodynamics On Parallel Architectures (Cholla) has been demonstrated to be a scalable and efficient tool that operates in large, distributed multi-GPU environments at high levels of performance. This additional performance results in unprecedented resolution for this type of simulation and massive amounts of generated data. This raises the requirements for analysis tools that can cope with scale and complexity of the simulated physical processes. To address those challenges, we utilize NVIDIA IndeX as a scalable framework to visualize the simulation output: NVIDIA IndeX features a streaming-based architecture to interactively explore simulation results in large-scale, multi-GPU environments. We utilize customized sampling programs for multi-volume and surface rendering to address analysis questions of galactic wind simulations. This combination of massively parallel simulation and analysis allows us to utilize recent supercomputer capabilities and to speed up the exploration of galactic wind simulations.

Visualizing the World's Largest Turbulence Simulation
Salvatore Cielo (Leibniz Supercomputing Centre), Luigi Iapichino (Leibniz Supercomputing Centre), Johannes Günther (Intel Corporation), Christoph Federrath (Australian National University, Research School of Astronomy and Astrophysics), Elisabeth Mayer (Leibniz Supercomputing Centre), Markus Wiedemann (Leibniz Supercomputing Centre)

We present the visualization of the largest interstellar turbulence simulations ever performed, unravelling key astrophysical processes concerning the formation of stars and the relative role of magnetic fields. The simulations, including pure hydrodynamical (HD) and magneto-hydrodynamical (MHD) runs, up to a size of $10048^3$ grid elements, were produced on the supercomputers of the Leibniz Supercomputing Centre and visualized using the hybrid parallel (MPI + TBB) ray-tracing engine OSPRay associated with VisIt. Besides revealing features of turbulence with an unprecedented resolution, the visualizations brilliantly showcase the stretching-and-folding mechanisms through which astrophysical processes such as supernova explosions drive turbulence and amplify the magnetic field in the interstellar gas, and how the first structures, the seeds of newborn stars are shaped by this process.

An Accessible Visual Narrative for the Primary Energy Source of Life from the Fulldome Show
Conversion of sunlight into chemical energy, namely photosynthesis, is the primary energy source of life on Earth. An explanatory visualization depicting this process is presented in the form of an excerpt from the fulldome show Birth of Planet Earth. This accessible visual narrative shows a lay audience, especially children, how the energy of sunlight is captured, converted, and stored through a chain of proteins to power living cells. The visualization is the result of a multi-year collaboration among biophysicists, visualization scientists, and artists, which, in turn, is based on a decade-long experimental-computational collaboration on structural and functional modeling that produced an atomic detail description of a bacterial bioenergetic organelle, the chromatophore. The energy conversion steps depicted feature an integration of function from electronic to cell levels, spanning nearly 12 orders of magnitude in time scales modeled with multi-scale computational approaches. This atomic detail description uniquely enables a modern retelling of one of humanity’s earliest stories---the interplay between light and life.

As part of the Oak Ridge Leadership Computing Facility (OLCF) Early Science program, NASA ran a series of simulations on the new, GPU-accelerated system Summit, using the flow solver FUN3D to produce solutions for a set of representative descent conditions. The raw output from these simulations was over a petabyte of data. We produced visualizations at the NASA Advanced Supercomputing Division at Ames Research Center using parallel techniques where feasible to accelerate the process. Included with this summary is an animation with some of the highlights of our visualizations so far. We discuss some of our visualization design choices, and plans for future work.
**Poster 91: FreeCompilerCamp: Online Training for Extending Compilers**
Alok Mishra (Stony Brook University, Lawrence Livermore National Laboratory), Anjia Wang (University of North Carolina, Charlotte; Lawrence Livermore National Laboratory), Chunhua Liao (Lawrence Livermore National Laboratory), Yonghong Yan (University of North Carolina, Charlotte), Barbara Chapman (Stony Brook University)

In this presentation, we introduce an ongoing effort of an online training platform aimed to automate the training of developers to quickly extend compilers - FreeCompilerCamp.org. Our free and open platform allows anyone who is interested in developing compilers to learn the necessary skills. A live training website, built on top of Play-With-Docker, is set up so that anyone with internet access and a web browser will be able to take this training. The entire training system is open-source and developers with relevant skills can contribute new tutorials and deploy it on a private server, workstation or even laptop. We have created some initial tutorials on how to extend the Clang/LLVM or ROSE compilers to support new OpenMP features. Using a web interface consisting of two side-by-side panels, users can follow the tutorials on one side and immediately practice what they learned in a terminal sandbox embedded on the other.

Best Poster Finalist: yes

**Poster 92: Nanoporous Flow Simulations on the Summit Supercomputer**
Yidong Xia (Idaho National Laboratory), Lixiang Luo (IBM - TJ Watson Research Center), Ansel Blumers (Brown University), Joshua Kane (Idaho National Laboratory), Jan Goral (University of Utah), Yu-Hang Tang (Lawrence Berkeley National Laboratory), Zhen Li (Clemson University, Brown University), Hai Huang (Idaho National Laboratory), Milind Deo (University of Utah)

Fluid flow behaviors in nanoporous materials are distinct from those following the continuum physics. Numerical simulations can be a complement to laboratory experiments. This work presents a dissipative particle dynamics (DPD) package for GPU-accelerated mesoscale flow simulations in nanoporous materials. In an ideal benchmark that minimizes load imbalance, the package delivered nearly perfect strong- and weak-scaling (with up to 4 billion DPD particles) on up to 1,536 V100 GPUs on Oak Ridge National Laboratory’s Summit supercomputer. More remarkably, in a benchmark to measure its usefulness with realistic nanopores in SBA-15 silica, the package exhibited more than 20x speedup over its LAMMPS-based CPU counterpart with the same number nodes (e.g., 384 V100 GPUs vs. 2,688 POWER9 cores). It is worth highlighting that the NVLink2 Host-to-Device interconnects kept the cost of CPU-GPU memory copy as low as only 10% of GPU activity time per rank: 4 times less than their PCIe counterparts.

Best Poster Finalist: yes
Poster 96: TSQR on TensorCores
Hiroyuki Ootomo (Tokyo Institute of Technology), Rio Yokota (Tokyo Institute of Technology)

Tall-Skinny QR (TSQR) is an efficient algorithm for calculating the QR decomposition of m x n matrices where m << n, which is done by recursively performing QR decomposition on subdivided blocks of the tall and skinny matrix. Such operations are useful for low-rank approximation methods, which are replacing more and more dense linear algebra in both scientific computing and machine learning fields. The present work focuses on the implementation of this important algorithm on Tensor Cores, which are available on the latest NVIDIA GPUs. We evaluate the speed, accuracy, and stability of TSQR on TensorCores.

Best Poster Finalist: yes

Poster 110: Hierarchical Data Prefetching in Multi-Tiered Storage Environments
Hariharan Devarajan (Illinois Institute of Technology), Anthony Kougkas (Illinois Institute of Technology), Xian-He Sun (Illinois Institute of Technology)

In the era of data-intensive computing, accessing data with a high-throughput and low-latency is very imperative. Data prefetching is used for hiding read latency by requesting data before it is needed to move it from a high-latency medium to a low-latency one. However, existing solutions do not consider the multi-tiered storage and also suffer from under-utilization of prefetching resources and unnecessary evictions. Additionally, existing approaches implement a client-pull model where understanding the application’s I/O behavior drives prefetching decisions. Moving toward exascale, where machines run multiple applications concurrently by accessing files in a workflow, a more data-centric approach resolves challenges such as cache pollution and redundancy. In this study, we present HFetch, a truly hierarchical data prefetcher that adopts a server-push approach to data prefetching. We demonstrate the benefits of such an approach. Results show 10-35% performance gains over existing prefetchers and over 50% when compared to systems with no prefetching.

Best Poster Finalist: yes

Poster 54: Massively Parallel Eigensolvers Based on Unconstrained Energy Functionals Methods
Osni Marques (Lawrence Berkeley National Laboratory), Mauro Del Ben (Lawrence Berkeley National Laboratory), Andrew Canning (Lawrence Berkeley National Laboratory)
This poster focuses on a preconditioned conjugate gradient based iterative eigensolver using an unconstrained energy functional minimization scheme. This scheme avoids an explicit reorthogonalization of the trial eigenvectors and becomes an attractive alternative for the solution of very large problems. The unconstrained formulation is implemented in the first-principles materials and chemistry CP2K code, which performs electronic structure calculations based on a density functional theory approximation to the solution of the many-body Schrödinger equation. The systems we use in our studies have a number of atoms ranging from 2,247 to 12,288. We study the convergence of the unconstrained formulation and its scaling on a Cray XC40 (a partition with 9,688 Intel KNL nodes). We show that there is a trade-off between the preconditioner that leads to fast convergence and lower cost preconditioners that lead to best time to solution.

Best Poster Finalist: yes

1:30 pm - 3:00 pm

Scientific Visualization & Data Analytics Showcase Posters Presentations

Visualization of Entrainment and Mixing Phenomena at Cloud Edges
Bipin Kumar (Indian Institute of Tropical Meteorology), Matt Rehme (National Center for Atmospheric Research (NCAR)), Neethi Suresh (Indian Institute of Tropical Meteorology), Nihanth Cherukuru (National Center for Atmospheric Research (NCAR)), Stanislaw Jaroszynski (National Center for Atmospheric Research (NCAR)), Shaomeng Li (National Center for Atmospheric Research (NCAR)), Scott Pearse (National Center for Atmospheric Research (NCAR)), Tim Scheitlin (National Center for Atmospheric Research (NCAR))

Entrainment and mixing processes occur during the entire life of a cloud. These processes change the droplet size distribution, which determines rain formation and radiative properties. Since it is a microphysical process, it cannot be resolved in large scale weather forecasting models. Small scale simulations such as Direct Numerical Simulations (DNS) are required to resolve the smallest scale of these processes. However, it has been a challenge to visualize these processes in a 3D domain as it generates petabytes of data. Visualization plays a vital role in analyzing and understanding these huge data outputs. Here, we present different techniques for 3D visualization of data obtained from DNS carried out at Indian Institute of Tropical Meteorology (IITM) to understand cloud microphysical properties more closely.

Multiple tools were used to conduct a visual analysis of this data. Two of these tools are well established and tested technologies: ParaView and VAPOR. The others are emergent technologies
created at the National Center for Atmospheric Research (NCAR) and are in the development phase. This data simulation, in addition to exploring DNS as mentioned above, provided an opportunity to test and improve development of several tools and methods.

**Visualizing Deep Learning at Scale for the Construction of Galaxy Catalogs in the Dark Energy Survey**
Janet Y. K. Knowles (Argonne National Laboratory), Joseph A. Insley (Argonne National Laboratory, Northern Illinois University), Silvio Rizzi (Argonne National Laboratory), Elise Jennings (Argonne National Laboratory), Asad Khan (University of Illinois), Eliu Huerta (University of Illinois), Sibo Wang (University of Illinois), Robert Gruendl (University of Illinois), Huihuo Zheng (Argonne National Laboratory)

The advent of machine and deep learning algorithms on petascale supercomputers is accelerating the pace of discovery in astrophysics and poses significant challenges to the interpretability of these deep neural networks. We present a novel visualization of a deep neural network output during training as it is learning to classify galaxy images as either spiral or elliptical. The network is trained using labeled datasets from the citizen science campaign, Galaxy Zoo, adopted by the Sloan Digital Sky Survey. These trained neural network models can then be used to classify galaxies in the Dark Energy Survey that overlap the footprint of both surveys. Visualizing a reduced representation of the network output, projected into 3-D parameter space, reveals how the network has discovered two distinct clusters of features which allows it to classify galaxies into two groups. These visualizations of the neural network during training aid in the interpretability of the black box of deep learning and reveal how the network responds to the input images at various stages of training. Finally, it allows a wider net to be cast to a general audience, thereby generating interest in and visibility to an otherwise highly specialized field.

**NVIDIA IndeX Accelerated Computing for Visualizing Cholla’s Galactic Winds**
Christopher Lux (Nvidia Corporation), Alexander Kuhn (Nvidia Corporation), Evan Schneider (Princeton University), Brant Robertson (University of California, Santa Cruz), Marc Nienhaus (Nvidia Corporation)

Galactic winds – outflows of gas driven out of galaxies by the combined effects of thousands of supernovae – are a crucial feature of galaxy evolution. Despite their importance, a complete theoretical picture of these winds has been elusive. Simulating the complicated interaction between the hot, high pressure gas created by supernovae and the cooler, high density gas in the galaxy disk requires massive computational resources and sophisticated software. For this purpose, Computational Hydrodynamics On Parallel Architectures (Cholla) has been demonstrated to be a
scalable and efficient tool that operates in large, distributed multi-GPU environments at high levels of performance. This additional performance results in unprecedented resolution for this type of simulation and massive amounts of generated data. This raises the requirements for analysis tools that can cope with scale and complexity of the simulated physical processes. To address those challenges, we utilize NVIDIA IndeX as a scalable framework to visualize the simulation output: NVIDIA IndeX features a streaming-based architecture to interactively explore simulation results in large-scale, multi-GPU environments. We utilize customized sampling programs for multi-volume and surface rendering to address analysis questions of galactic wind simulations. This combination of massively parallel simulation and analysis allows us to utilize recent supercomputer capabilities and to speed up the exploration of galactic wind simulations.

Visualizing the World’s Largest Turbulence Simulation
Salvatore Cielo (Leibniz Supercomputing Centre), Luigi Iapichino (Leibniz Supercomputing Centre), Johannes Günther (Intel Corporation), Christoph Federrath (Australian National University, Research School of Astronomy and Astrophysics), Elisabeth Mayer (Leibniz Supercomputing Centre), Markus Wiedemann (Leibniz Supercomputing Centre)

We present the visualization of the largest interstellar turbulence simulations ever performed, unravelling key astrophysical processes concerning the formation of stars and the relative role of magnetic fields. The simulations, including pure hydrodynamical (HD) and magneto-hydrodynamical (MHD) runs, up to a size of 10048^3 grid elements, were produced on the supercomputers of the Leibniz Supercomputing Centre and visualized using the hybrid parallel (MPI + TBB) ray-tracing engine OSPRay associated with VisIt. Besides revealing features of turbulence with an unprecedented resolution, the visualizations brilliantly showcase the stretching-and-folding mechanisms through which astrophysical processes such as supernova explosions drive turbulence and amplify the magnetic field in the interstellar gas, and how the first structures, the seeds of newborn stars are shaped by this process.

An Accessible Visual Narrative for the Primary Energy Source of Life from the Fulldome Show Birth of Planet Earth
Melih Sener (University of Illinois), Stuart Levy (University of Illinois), AJ Christensen (University of Illinois), Robert Patterson (University of Illinois), Kalina Borkiewicz (University of Illinois), John E. Stone (University of Illinois), Barry Isralewitz (University of Illinois), Jeffrey Carpenter (University of Illinois), Donna Cox (University of Illinois)

Conversion of sunlight into chemical energy, namely photosynthesis, is the primary energy source of life on Earth. An explanatory visualization depicting this process is presented in the form of an
excerpt from the fulldome show Birth of Planet Earth. This accessible visual narrative shows a lay audience, especially children, how the energy of sunlight is captured, converted, and stored through a chain of proteins to power living cells. The visualization is the result of a multi-year collaboration among biophysicists, visualization scientists, and artists, which, in turn, is based on a decade-long experimental-computational collaboration on structural and functional modeling that produced an atomic detail description of a bacterial bioenergetic organelle, the chromatophore. The energy conversion steps depicted feature an integration of function from electronic to cell levels, spanning nearly 12 orders of magnitude in time scales modeled with multi-scale computational approaches. This atomic detail description uniquely enables a modern retelling of one of humanity’s earliest stories---the interplay between light and life.

Visualizing Supersonic Retropropulsion for Mars: The Summit Early Science Results
Patrick Moran (NASA Ames Research Center), Timothy Sandstrom (NASA Ames Research Center), Ashley Korzun (NASA Langley Research Center), Eric Nielsen (NASA Langley Research Center), Aaron Walden (NASA Langley Research Center)

As part of the Oak Ridge Leadership Computing Facility (OLCF) Early Science program, NASA ran a series of simulations on the new, GPU-accelerated system Summit, using the flow solver FUN3D to produce solutions for a set of representative descent conditions. The raw output from these simulations was over a petabyte of data. We produced visualizations at the NASA Advanced Supercomputing Division at Ames Research Center using parallel techniques where feasible to accelerate the process. Included with this summary is an animation with some of the highlights of our visualizations so far. We discuss some of our visualization design choices, and plans for future work.

3:30 pm - 5:00 pm

ACM Student Research Semi-Finalists Presentations

Thursday, November 21

8:30 am - 5:00 pm

ACM Student Research Competition Posters Display
Poster 12: Investigating Input Sensitivity in GPU Graph Analytics with Machine Learning
Jacob M. Hope (Texas State University), Trisha Nag (Texas Tech University)

Graph algorithms are at the core of data-intensive applications. As such, efficient graph processing is of great importance. Irregularity in real-world graphs can make performance unpredictable and non-portable across different inputs and architectures. Given a type of graph, the same optimized implementation of an algorithm can produce performance numbers that differ by orders-of-magnitude. We conduct extensive analysis on a set of 1238 graphs to identify input-sensitive performance inefficiencies, including two that have not been studied: (i) register pressure and (ii) CPU-GPU data movement via demand paging. We then build a multiclass decision tree classifier that characterizes the irregular properties of graphs from our data and maps them to an optimal control parameterization at the compiler, system and algorithmic layers, that yield the highest overall algorithmic performance. We then integrate the classifier into a system where it will process a new graph and generate a kernel on the predicted optimal configuration.

Poster 13: Scalable Load-Balancing Concurrent Queues on Many-Core Architectures
Caleb Lehman (Ohio State University)

As the core counts of computing platforms continue to rise, parallel runtime systems with support for very fine-grained tasks become increasingly necessary to fully utilize the available resources. A critical feature of such task-based parallel runtime systems is the ability to balance work evenly and quickly across available cores. We highlight this by studying XTask, a custom parallel runtime system based on XQueue, which is a novel lock-less concurrent queuing system with relaxed ordering semantics that is geared to realizing scalability to hundreds of concurrent threads. We demonstrate the lack of adequate load balancing in the original XQueue design and present several solutions for improving load balancing. We also evaluate the corresponding improvements in performance on two sample workloads, computation of Fibonacci numbers and computation of Cholesky factorization. Finally, we compare the performance of several versions of XTask along with several implementations of the popular OpenMP runtime system.

Poster 6: Analyzing the Performance of ZFP Compressed Arrays on HPC Kernels
Pavlo D. Triantafyllides (Clemson University)

Per-node memory capacity limits the maximal problem size of HPC applications. Naïve data compression alleviates the memory impact, but requires full decompression before the data is accessed. ZFP compressed arrays reduce the memory footprint, by independently compressing data in fixed sized blocks. Thus, decompressing individual blocks and caching them enables random
access and a reduction in decompressions on the critical path. The performance of ZFP compressed arrays is dependent on several key variables: software cache size, cache policy, and compression rate. In this poster, we explore the sensitivity of these ZFP parameters on runtime performance for the matrix-matrix multiplication algorithm. Results show that selection of cache size, policy, and rate yields 8% performance improvement over the default ZFP configuration.

Poster 1: Early Experiences on OpenPOWER Architecture: Analysis of Billion-Scale Atomistic Datasets
Yuya Kawakami (Grinnell College, Oak Ridge National Laboratory)

We present performance analysis on OpenPOWER architecture of an algorithm to generate transversal views of atomistic models. The algorithm was implemented with data parallel primitives in NVIDIA Thrust for architecture portability. We report performance results on IBM Power9 CPUs (OpenMP, Intel Threading Blocks) and NVIDIA Volta GPUs (single and multi GPU). We also evaluate CUDA unified memory performance, exposed by NVIDIA RAPIDS Memory Manager library (RMM).

Poster 2: An Efficient Parallel Algorithm for Dominator Detection
Daniel Giger (University of Massachusetts, Lowell), Hang Liu (Stevens Institute of Technology)

In graph theory, a vertex \(v\) dominates a vertex \(u\) if every path from the entry vertex to \(u\) must go through vertex \(v\). This algorithm is called dominator detection and holds a wide range of applications, such as compiler design, circuit testing, and social network analysis. While the performance of many other graph algorithms soars with respect to the increase of the hardware parallelism, dominator detection algorithm experiences very little advancement due to the hardship of parallelism. This work thus introduces an efficient parallel dominator detection algorithm that is inspired by Breadth-First Search (BFS), which bests SEMI-NCA on large graphs.

Poster 9: Machine Specific Symbolic Code Generation
Robert King (University of Utah)

New discoveries in science and engineering are primarily driven by numerical simulations of underlying governing equations. Manually written codes to evaluate the main computational kernels lack portability, prone to human errors, ability to perform code optimizations due to the complexity of the underlying equations. In this work we present a symbolic code generation framework, which generates architecture optimized code for different platforms. As the driving application, we primarily use computational relativity where computations of Einstein equations become
complicated due to the presence of curvature in spacetime.

The presented framework is based on SymPy with additional modules written to handle complicated partial differential equations (PDEs). The symbolically written equations are converted to a computational graph, which enables to perform architecture (cache, register optimizations) and language specific (SIMD vectorization, CUDA) optimizations. By computing common subexpressions, we reduce the number of compute operations needed, by storing them in temporary variables.

**Poster 5: Evaluating Lossy Compressors for Inline Compression**
Donald T. Elmore (Clemson University)

HPC applications require massive amounts of data. The data required is growing faster than memory capabilities. An example of this is pySDC, a framework for solving collocation problems iteratively using parallel-in-time methods. pySDC requires storing and exchange of 3D volume data for each parallel point in time. We evaluate several state-of-the-art lossy compressors such as SZ and ZFP for their applicability to inline compression for pySDC. We evaluate the compressors based on compression ratio, compression bandwidth, decompression bandwidth, and overall error introduced.

This poster evaluates state-of-the-art lossy compressors for use in pySDC; shows lossy data compression is an effective tool for reducing memory requirements for pySDC; and highlights current compression/decompression bandwidth is not fast enough for inline lossy compression yet. Results show using SZ with an error bound of 1e-5, we reduce the memory footprint by a factor of 311.99 while maintaining an acceptable level of loss.

**Poster 3: A Deep Learning Approach to Noise Prediction and Circuit Optimization for Near-Term Quantum Devices**
Alexander Zlokapa (California Institute of Technology), Alexandru Gheorghiu (California Institute of Technology)

Noisy intermediate-scale quantum (NISQ) devices face challenges in achieving high-fidelity computations due to hardware-specific noise. As a basis for noise mitigation, we develop a convolutional neural network noise model to estimate the difference in noise between a given pair of equivalent quantum circuits. On a classically simulated dataset of 1.6 million pairs of quantum circuits with a simplified noise model calibrated to IBM Q hardware, the deep learning approach shows a significant improvement in noise prediction over linear gate count models. A greedy
peephole optimization procedure is proposed to minimize noise using the deep learning model as an objective function, showing further improvement in noise mitigation compared to commonly used gate count minimization heuristics.

**Poster 10: Learning Page Access Patterns for Algorithms Programmed with GPU UVM**
Bennett Cooper (Clemson University), Derek Rodriguez (Clemson University)

We aim to mitigate the performance bottleneck of migrating data between host and device memory in GPU applications by accurately predicting application access patterns using deep neural networks. We model the memory access pattern of any given application by collecting page faults that trigger data migration to the GPU and feed this time series as input to a neural network that outputs the next several page faults. We evaluate these predictions on the basis of what makes a useful prefetch in our context for the GPU. Current work has looked at trivial GPU applications, such as matrix operations, and moving toward real, complex applications. Our work will be presented by bringing attention to the predictive capability of our neural network on the current applications tested.

**Poster 8: Mitigating Communication Bottlenecks in MPI-Based Distributed Learning**
Abdullah B. Nauman (Ward Melville High School, Brookhaven National Laboratory)

Current commercial and scientific facilities generate and maintain vast amounts of complex data. While machine learning (ML) techniques can provide crucial insight, developing these models is often impractical on a single process. Distributed learning techniques mitigate this problem; however, current models contain significant performance bottlenecks. Here, we conduct a detailed performance analysis of MPI_Learn, a widespread distributed ML framework for high-energy physics (HEP) applications, on the Summit supercomputer, by training a network to classify simulated collision events from high-energy particle detectors at the CERN Large Hadron Collider (LHC).

We conclude that these bottlenecks occur as a result of increasing communication time between the different processes, and to mitigate the bottlenecks we propose the implementation of a new distributed algorithm for stochastic gradient descent (SGD). We provide a proof of concept by demonstrating better scalability with results on 250 GPUs, and with hyperparameter optimization, show a ten-fold decrease in training time.

**Hearing Single- and Multi-Threaded Program Behavior**
There are many examples of visual program representation, but there is an unexplored area of sonic representation—audialization—of programs. The Thread Safe Audio Library (TSAL) provides the necessary tools to create such program audializations. TSAL is a platform-independent, object-oriented C++ library that provides thread safe classes for synthesizing sound from a program. By adding TSAL calls to a sequential or parallel program, its behavior can be represented with sound. Audializations require a unique interpretation that may greatly benefit auditory learners and other programmers who are looking for a different way to profile a sequential or parallel program.

**Poster 11: Walking the Cost-Accuracy Tightrope: Balancing Trade-Offs in Data-Intensive Genomics**

Kathryn Leung (Princeton University), Meghan Kimball (DePaul University)

Scientific applications often exhibit a trade-off between cost and accuracy. However, measuring and predicting cost and accuracy in a way that users can understand these trade-offs is challenging. To address these needs, we present predictive cost and accuracy models for data-intensive genomics applications. We use these models to create a trade-off graph, which researchers can use to selectively trade-off cost and accuracy.

**Poster 31: Designing High-Performance Erasure Coding Schemes for Next-Generation Storage Systems**

Haiyang Shi (Ohio State University), Xiaoyi Lu (Ohio State University)

Erasure Coding (EC) NIC offload is a promising technology for designing next-generation distributed storage systems. To alleviate the EC overhead, various hardware-based Erasure Coding schemes have been proposed to leverage the advanced compute capabilities on modern HPC clusters. However, this poster has identified some critical limitations of current-generation EC NIC offload schemes on modern SmartNICs and the missing of exploiting heterogeneous hardware available on modern HPC clusters in parallel. To address these limitations, this poster first proposes a unified benchmark suite to benchmark, measure and characterize hardware-optimized erasure coders. Then, it proposes a Multi-Rail EC concept which enables upper-layer applications to leverage heterogeneous hardware to perform EC operations simultaneously. Finally, it proposes a new EC NIC offload paradigm based on the tripartite graph model, namely TriEC. TriEC supports both encode-and-send and receive-and-decode offload primitives efficiently.
Poster 27: A Framework for Measuring Hardware Gather-Scatter Support
Patrick Lavin (Georgia Institute of Technology), Jeffrey Young (Georgia Institute of Technology), Richard Vuduc (Georgia Institute of Technology)

In recent years, we have seen the re-addition of vector units to CPUs. While these units easily give speedups for easily vectorized applications with dense memory access, it can be hard to characterize how different access patterns will affect the performance of vectorized code.

We have developed Spatter, a benchmark which allows us to test and investigate the gather-scatter units available on current and upcoming hardware. The information that Spatter reveals to users is of use to everyone from hardware vendors who wish to compare gather-scatter units across platforms, to application developers who wish to test memory access pattern performance in their vectorized code.

Poster 30: A Fast and Efficient Incremental Approach toward Dynamic Community Detection
Neda Zarayeneh (Washington State University)

Community detection is a discovery tool to identify groups of vertices which are tightly connected, called communities. Most of the literature on this subject caters to the static use-case where the underlying network does not change. However, many emerging real-world use-cases give rise to a need to incorporate dynamic graphs as inputs.

We present a fast and efficient incremental approach toward dynamic community detection. The key contribution is a generic technique called $\Delta$-screening, which examines the most recent batch of changes made to an input graph and selects a subset of vertices to reevaluate for potential community (re)assignment. This technique can be incorporated into any of the community detection methods that use modularity as its objective function for clustering. For demonstration purposes, we incorporated the technique into two well-known community detection tools. Our experiments demonstrate that our approach is able to generate performance speedups without compromising on the output quality.

Poster 29: Optimal Routing for a Family of Scalable Interconnection Networks
Zhipeng Xu (Stony Brook University; Sun Yat-sen University, Guangzhou, China)

We propose a scheme to construct a family of large and high-performance interconnection networks that are scalable, low-radix, minimum diameters. These networks, whose diameters grow linearly as their sizes grow exponentially, are generated by using the Cartesian products of smaller optimal
networks of minimum diameters. For the smaller base networks, we design the vertex-balanced routing algorithm by considering the forwarding pressure at each vertex. Comparative benchmarks on a Beowulf cluster show significant improvement in performance after using the new routing algorithm. Each node of the new network generated from base graphs with low-diameter can also sustain balanced forwarding loadings if we apply optimal routing algorithms to the base network. Simulation results for larger networks show that the optimal routing algorithms achieve the gain of communication performance.

Poster 16: perf-taint: Taint Analysis for Automatic Many-Parameter Performance Modeling
Marcin Copik (ETH Zurich)

Performance modeling is a well-known technique for understanding the scaling behavior of an application. Although the modeling process is today often automatic, it still relies on a domain expert selecting program parameters and deciding relevant sampling intervals. Since existing empirical methods attempt blackbox modeling, the decision on which parameters influence a selected part of the program is based on measured data, making empirical modeling sensitive to human errors and instrumentation noise. We introduce a hybrid analysis to mitigate the current limitations of empirical modeling, combining the confidence of static analysis with the ability of dynamic taint analysis to capture the effects of control-flow and memory operations. We construct models of computation and communication volumes that help the modeler to remove effects of noise and improve the correctness of estimated models. Our automatic analysis prunes irrelevant program parameters and brings an understanding of parameter dependencies which helps in designing the experiment.

Poster 24: Fingerprinting Anomalous Computation with RNN for GPU-Accelerated HPC Machines
Pengfei Zou (Clemson University)

This paper presents a workload classification framework that discriminates illicit computation from authorized workloads on GPU-accelerated HPC systems. As such systems become more and more powerful, they are exploited by attackers to run malicious and for-profit programs that typically require extremely high computing ability to be successful. Our classification framework leverages the distinctive signatures between illicit and authorized workloads, and explore machine learning methods to learn the workloads and classify them. The framework uses lightweight, non-intrusive workload profiling to collect model input data, and explores multiple machine learning methods, particularly recurrent neural network (RNN) that is suitable for online anomalous workload detection. Evaluation results on four generations of GPU machines demonstrate that the workload classification framework can tell apart the illicit authorized workloads with high accuracy of over 95%. The
collected dataset, detection framework, and neural network models will be made available on GitHub.

**Poster 33: Enhancing Neural Architecture Search with Speciation and Inter-Epoch Crossover**
Matthew E. Baughman (University of Chicago)

As deep learning continues to expand into new areas of application, the demand for efficient use of our HPC resources increase. For new problem domains, new model architectures are developed through a neural architecture search (NAS), which consist of iteratively training many neural networks. To combat the computational waste and maximize compute efficiency for NAS, we demonstrate that the use of genetic algorithms with speciation can be used to both shorten training time and increase accuracy at each iteration.

**Poster 14: Data Reuse Analysis for GPU Offloading Using OpenMP**
Alok Mishra (Stony Brook University, Lawrence Livermore National Laboratory)

More researchers and developers desire to port their applications to GPU-based clusters, due to their abundant parallelism and energy efficiency. Unfortunately porting or writing an application for accelerators, such as GPUs, requires extensive knowledge of the underlying architectures, the application/algorithm and the interfacing programming model (e.g. OpenMP). Often applications spend a significant portion of their execution time on data transfer. Exploiting data reuse opportunities in an application can reduce its overall execution time. In this research, we present an approach to automatically recognize data reuse opportunities in an application which uses OpenMP for exploiting GPU parallelism, and consequently insert pertinent code to take advantage of data reuse on GPU. Using our approach we were able to retain reused data on the GPU and reduce the overall execution time of multiple benchmark application.

**Poster 17: Exploiting Multi-Resource Scheduling for HPC**
Yuping Fan (Illinois Institute of Technology)

High-performance computing (HPC) is undergoing significant changes. Next generation HPC systems are equipped with diverse global/local resources. HPC job scheduler plays a crucial role in efficient use of resources. However, traditional job schedulers are single-objective and fail to efficient use of other resources. In our previous work, we present a job scheduling framework named BBSched to schedule CPUs and burst buffers. As we are heading toward exascale computing, a variety of heterogeneous resources are deployed in HPC systems. In this poster, we extend BBSched for managing multiple resources beyond CPUs and burst buffers. We formulate multi-resource
scheduling as a general multi-objective optimization (MOO) problem, present a heuristic method to solve the NP-hard MOO problem, and provide a preliminary evaluation for scheduling up to ten resources. The proposed multi-resource scheduling design is intended to enhance Cobalt, a production job scheduler deployed on HPC systems at Argonne Leadership Computing Facility (ALCF).

**Poster 15: Cost-Aware Cloudlet Placement in Edge Computing Systems**  
Dixit Bhatta (University of Delaware)

A well-known challenge in Edge Computing is strategic placement of cloudlets. The fundamental goals of this challenge are to minimize the cloudlet deployment cost and to guarantee minimum latency to the users of edge services. We address this challenge by designing a cost-aware cloudlet placement approach that ensures user latency requirements while covering all devices in the service region. We first mathematically formulate the problem as a multi-objective integer programming model in a general deployment scenario, which is computationally NP-hard. We then propose a genetic algorithm-based approach, GACP, to find heuristic solutions in significantly reduced time. We investigate the effectiveness of GACP by performing extensive experiments on multiple deployment scenarios based on New York City OpenData. The results presented in the poster show that our approach obtains close to optimal cost solutions with significant time reduction.

**Poster 19: Accelerating Real-World Stencil Computations Using Temporal Blocking: Handling Sparse Sources and Receivers**  
George Bisbas (Imperial College, London)

This paper concerns performance optimization in finite-difference solvers found in seismic imaging. We tackle a complicating factor that is arising in seismic inversion problems: the addition of waves injected from sources distributed sparsely over the 2D/3D domain, and the need for receivers that interpolate data measurements at a set of points, again distributed sparsely across the domain. In this work we show how to overcome this limitation. We introduce an algorithm for a loop nest transformation policy applied to wave propagation in order to improve data locality and optimize our cache memory use. Our algorithm uses an inspector/executor scheme capable of inspecting measurement operations at sparse locations and then compute the field updates. The ultimate goal of this ongoing work is to automate this scheme for stencil codes generated by Devito.

Tirthak Patel (Northeastern University), Devesh Tiwari (Northeastern University)

Large-scale computing systems are becoming increasingly more power-constrained, but these systems employ hardware over-provisioning to achieve higher system throughput because HPC applications often do not consume the peak power capacity of nodes. Unfortunately, focusing on system throughput alone can lead to severe unfairness among multiple concurrently-running applications. This paper introduces PERQ, a new feedback-based principled approach to improve system throughput while achieving fairness among concurrent applications.

**Poster 18: Using Non Volatile Memories to Build Energy- and Cost-Efficient Clusters**
Onkar Patil (North Carolina State University)

Non-Volatile Memory (NVM) is a byte-addressable, high capacity, high latency, and persistent form of memory that can extend the primary memory hierarchy by another level. It allows clusters to have significantly greater memory capacity per node. Intel’s Optane DC Persistent Memory Module (PMM) is a NVM device that can be used to increase the memory density of high performance computing (HPC) system. With higher memory density, we can run scientific computing applications with larger problem sizes on fewer compute nodes than on current HPC systems. Smaller HPC clusters will reduce the cost of running scientific simulations. We test our hypothesis by running a HPC application with large problem sizes on a single node with NVM and on multiple nodes connected by a high speed interconnect with equivalent amount of volatile memory. We compare the performance characteristics and power consumption of both systems.

**Poster 22: Fast Profiling-Based Performance Modeling of Distributed GPU Applications**
Jaemin Choi (University of Illinois, Lawrence Livermore National Laboratory)

An increasing number of applications utilize GPUs to accelerate computation, with MPI responsible for communication in distributed environments. Existing performance models only focus on either modeling GPU kernels or MPI communication; few that do model the entire application are often too specialized for a single application and require extensive input from the programmer.

To be able to quickly model different types of distributed GPU applications, we propose a profiling-based methodology for creating performance models. We build upon the roofline performance model for GPU kernels and analytical models for MPI communication, with a significant reduction in profiling time. We also develop a benchmark to model 3D halo exchange that occurs in many scientific applications. Our proposed model for the main iteration loops of MiniFE achieves 6-7% prediction error on LLNL Lassen and 1-2% error on PSC Bridges, with minimal code inspection required to
model MPI communication.

Poster 26: Neural Networks for the Benchmarking of Detection Algorithms
Silvia Miramontes (Lawrence Berkeley National Laboratory; University of California, Berkeley)

There are several automated methods to detect objects from grayscale images. However, materials scientists still lack basic tools to compare different detection results, particularly when working with microtomography. This poster introduces FibCAM, a convolutional neural network (CNN)-based method using TensorFlow that allows benchmarking fiber detection algorithms. Our contribution is three-fold: (a) the design of a computational framework to compare automated fiber detection models with curated datasets through classification; (b) lossless data reduction by embedding prior knowledge into data-driven models; (c) a scheme to decompose computation into embarrassingly parallel processes for future analysis at scale. Our results show how FibCAM classifies different structures, and how it illustrates the material’s composition and frequency distribution of microstructures for improved interpretability of machine learning models. The proposed algorithms support probing the specimen content from gigabyte-sized volumes and enable pinpointing inconsistencies between real structures known a priori and results derived from automated detections.

Poster 21: Optimization of GPU Kernels for Sparse Matrix Computations in Hypre
Chaoyu Zhang (Arkansas State University)

The acceleration of sparse matrix computations on GPUs can significantly enhance the performance of iterative methods for solving linear systems. In this work, we consider the kernels of Sparse Matrix Vector Multiplications (SpMV), Sparse Triangular Matrix Solves (SpTrSv) and Sparse Matrix Matrix Multiplications (SpMM), which are often demanded by Algebraic Multigrid (AMG) solvers. With the CUDA and the hardware support of the Volta GPUs on Sierra, the existing kernels should be further optimized to fully take the advantage of the new hardware, and the optimizations have shown significant performance improvement. The presented kernels have been put in HYPRE for solving large scale linear systems on HPC equipped with GPUs. These shared-memory kernels for single GPU are the building blocks of distributed matrix operations required by the solver across multiple GPUs and compute nodes. The implementations of these kernels in Hypre and the code optimizations will be discussed.

Poster 20: From IoT to Cloud: Research Platform for IoT/Cloud Experiments
Jinfeng Lin (University of Notre Dame)
IoT studies leverage a wide range of lightweight hardware for collecting and processing data in the field. Limited by the resources on the devices, IoT systems have to interoperate with cloud platforms for addressing computation intensive tasks such as image processing, application backend supporting and centralized data storage. Therefore, a testbed for IoT/Cloud experiments should provide infrastructure for IoT to cloud communication, computation deployment, and hardware resource management. With these functionalities, users can focus on research problems without distraction from manually constructing experiment environments. Though cloud providers such as Google, Amazon, and Microsoft all provide IoT to Cloud solutions in general, this commercial model is not entirely compatible with research purpose. We propose a framework named Chameleon IoT testbed (CHIoT) that extends the Chameleon bare-metal cloud resources to support general IoT to Cloud experiments.

Poster 32: OSU INAM: A Profiling and Visualization Tool for Scalable and In-Depth Analysis of High-Performance GPU-Enabled HPC Clusters
Pouya Kousha (Ohio State University)

The lack of low-overhead and scalable monitoring tools have prevented a comprehensive study of efficiency and utilization of emerging NVLink-enabled GPU clusters. We address this by proposing and designing an in-depth, real-time analysis, profiling, and visualization tool for high-performance GPU-enabled clusters with NVLinks on the top of the OSU INAM. The proposed tool is capable of presenting a unified and holistic view of MPI-level and fabric level information for emerging NVLink-enabled high-performance GPU clusters. It also provides insights into the efficiency and utilization of underlying interconnects for different communication patterns. We also designed a low overhead and scalable modules to discover the fabric topology and gather fabric metrics by using different levels of threading, bulk insertions and deletions for storage, and using parallel components for fabric discovery and port metric inquiry.

Zheng Miao (Clemson University)

High-Performance Computing systems must simultaneously address both resilience and power. In heterogeneous systems, the trade-offs between resilience and energy-efficiency are more complex for applications using both CPUs and GPUs. A deep understanding of the interplay among energy efficiency, resilience, and performance is required for heterogeneous systems to address them simultaneously.
In this work, we present a new framework for resilient and energy-efficient computing in GPU-accelerated systems. This framework supports partial or full redundancy and checkpointing for resilience, and provides users with flexible hardware resource selection, adjustable precision and power management to improve performance and energy-efficiency. We further perform CUDA-aware MPI to reduce resilience overhead, mainly in message communication between GPUs. Using CG as an example, we show that our framework provides about 40% time and 45% energy savings, comparing to simple extension of RedMPI, a redundancy based resilience framework for homogeneous CPU systems.

**Poster 25: Leveraging Smart Data Transfer and Metadirective in Adaptive Computing**

*Anjia Wang (University of North Carolina, Charlotte; Lawrence Livermore National Laboratory)*

In this work, we propose smart data transfer (SDT) along with extensions to metadirective and map constructs in OpenMP 5.0 to improve adaptive computing. The Smith-Waterman algorithm is used as an example, whose naïve implementation does not conduct data transfer efficiently. SDT is used to solve this issue with the following advantages: (1) SDT only transfers necessary data to GPU instead of the whole data set, resulting in 4.5x of speedup in our initial experiments. (2) Depending on computing vs. data transfer requirements of a program, SDT will transfer the output of each iteration from GPU to host either immediately or all together after the last GPU kernel call. (3) It supports large data exceeding GPU device memory’s size via data tiling. We propose to extend metadirective's context selector to obtain similar improvement by enabling target enter/exit data and on-demand data access.

**Poster 36: Modeling Non-Determinism in HPC Applications**

*Dylan Chapp (University of Delaware, University of Tennessee)*

As HPC applications migrate from the petascale systems of today to the exascale systems of tomorrow, the increasing need to embrace asynchronous, irregular, and dynamic communication patterns will lead to a corresponding decrease in application-level determinism. Two critical challenges emerge from this trend. First, unchecked non-determinism coupled with the non-associativity of floating-point arithmetic undermines numerical reproducibility of scientific applications. Second, the prevalence of non-determinism amplifies the cost of debugging, both in
terms of computing resources and human effort. In this thesis, we present a modeling methodology to quantify and characterize communication non-determinism in parallel applications. Our methodology consists of three core components. First, we build graph-structured models of relevant communication events from execution traces. Second, we apply similarity metrics based on graph kernels to quantify run-to-run variability and thus identify the regions of executions where non-determinism manifests most prominently. Third, we leverage our notion of execution similarity to characterize applications via clustering, anomaly detection, and extraction of representative patterns of non-deterministic communication which we dub "non-determinism motifs". Our work will amplify the effectiveness of software tools that target mitigation or control of application-level non-determinism (e.g., record-and-replay tools) by providing them with a common metric for quantifying communication non-determinism in parallel applications and a common language for describing it.

Poster 35: Scaling Up Pipeline Optimization with High Performance Computing
Robert Lim (University of Oregon)

My research focuses on developing a pipeline optimization infrastructure that automates the design and code generation of neural networks through the use of high-performance computing. The problem has the following objectives: unify automated machine learning (AutoML) and compilation, archive profiles for creation of a knowledge base for a data-driven approach toward search, explore various search optimizations for model design and code generation. The field of automated deep learning includes hyperparameter optimization and neural architecture search (NAS), which requires domain expertise in designing a model, in addition to the tuning parameters related to learning and the model itself. The search space is complex and deciding which parameters factor into the overall accuracy of a model is a non-trivial task. Once a model is trained, the next step compiles the model, which maps to the backend of a targeted architecture, whether GPU, embedded mobile phones, or FPGA. The compilation phase also involves choices, in terms of optimizations applied, which can include a set of flags, or direct code transformations. Various efficient search techniques are explored, archiving results along the way for facilitating in code generation options, which could possibly inform the model design process. Our previous work reduced the search space for code generation of various domain kernels by 92%, and this work investigates whether the same approach can be applied in all aspects of AutoML design and code generation.

Poster 38: High-Performance Backpropagation in Scientific Computing
Navjot Kukreja (Imperial College, London)

Devito is a domain-specific language for the automatic generation of high-performance solvers for
finite difference equations provided in a high-level symbolic representation. The primary use of Devito is to enable Full-waveform inversion, which is an adjoint-based optimization problem that uses the wave equation as part of its objective function. This doctoral project consists of three elements of Devito that are key to solving adjoint-based optimization problems (i.e. the back propagation). The first is automatic differentiation of stencil loops. This allows the automatic generation of high-performance code implementing the derivative of any provided function. This is essential to enable the use of a wider range of physical equations - in order to use better physical approximations. An essential feature of the generated derivatives is that they show scaling and performance behaviour that is very similar to the forward function. A common issue in solving adjoint-based optimization problems is the prohibitively high memory requirement. The second contribution is a runtime for automatic execution of checkpoint-recompute schedules (called pyRevolve) to alleviate this memory requirement by trading it off for re-computations. These schedules may involve storing some intermediate states on disk, some in memory, and recomputing others. The third contribution is the use of lossy compression (ZFP/SZ), which is a second means of trading off memory and compute, automatically as part of checkpointing in pyRevolve. This is the first time it has been shown that lossy compression combined with checkpoint-recomputation can provide a much better error-memory tradeoff than using checkpoint-recomputation or lossy compression alone.

Poster 37: Large Scale Ultrasound Simulations on Accelerated Clusters
Filip Vaverka (Brno University of Technology)

Many emerging non-invasive medical procedures such as high intensity focused ultrasound treatment (HIFU), photoacoustic imaging, or transcranial ultrasound therapy require ultrasound wave propagation simulations. Typically soft tissue can be modeled as fluid, weakly heterogeneous medium with frequency dependent absorption and non-linear effects at high ultrasound intensities. The k-Wave acoustic toolbox, widely used in medical community, uses highly efficient k-space pseudo-spectral time-domain (KSTD) discretization of acoustic equations to solve these problems. However, it is difficult to efficiently map the KSTD method onto modern cluster architectures with accelerators such as GPUs. The present thesis analyzes shortcomings of the KSTD method in respect to modern clusters and proposes local Fourier basis approach to improve scaling of the method. It is shown that the proposed method is able to achieve 5x speedup, while having sufficient accuracy for these medical applications. Behavior of the method is analyzed across variety of GPU and MIC (Intel Xeon Phi) accelerated clusters and results are presented.

Poster 34: Analysis of Automata Processing Acceleration on Disparate Hardware Technologies
Marziyeh Nourian (North Carolina State University)
Pattern matching is a computation that maps naturally onto finite automata (FA) abstractions. There has been a substantial amount of work on accelerating FA processing on various parallel platforms. However, the advantages and disadvantages of different automata processing accelerators and the innovation space in this area are still unclear. We target this problem and propose a compiler toolchain that automates the deployment of non-deterministic finite automata (NFAs) onto different target platforms. Using this toolchain, we perform an apples-to-apples comparison between AP, GPU- and FPGA-based NFA accelerator designs on large-scale datasets. Specifically, we observe that memory-based designs are limited by memory size and bandwidth. To address this issue, we target fixed-topology NFAs and propose a memory-efficient design that embeds the automata topology in code and stores only the transition symbols in memory. Our solution is suitable for SIMD architectures and is called SIMD_NFA. We design a compiler that automates the deployment of this design on SIMD platforms. We showcase our compiler framework on GPU and Intel platforms. Additionally, we observe that for NFAs with a grid-like fixed-topology (e.g., NFAs for Levenshtein and Hamming distance-based matching), transitions do not need to be encoded within the traversal code but can be inferred from the reference string to be matched and the knowledge of the NFA topology. Lastly, SIMD_NFA is a good fit for FPGA deployment using OpenCL-to-FPGA toolchains. We investigate the deployment of the OpenCL version of SIMD_NFA, on FPGA and explore a set of optimizations techniques to retarget SIMD_NFA to FPGA.

**Poster 42: Power and Thermal Management Runtimes for HPC Applications in the Era of Exascale Computing**

Daniele Cesarini (University of Bologna, CINECA)

In the scope of technical and scientific computing, the rush toward larger simulations has been so far assisted by a steady downsizing of micro-processing units, which has allowed to increase the compute capacity of general-purpose architectures at constant power. As side effects of the end of Dennard's scaling, this process is now hitting its ultimate power limits and is just about to come to an end. This implies an increase in the energy cost for computation, performance loss due to a design based on worst-case power consumption, and performance loss due to overheating, and thermal gradients. As result, thermal and power bound supercomputing machines show performance degradation and heterogeneity which limit the peak performance of the system. This doctoral showcase presents software strategies to tackle the main bottlenecks induced by power and thermal issues that affect next-generation supercomputers. To respond to the above challenges, my work shows that propagating workload requirements from application to the runtime and operating system levels is the key to provide efficiency. This is possible only if the proposed software methodologies cause little or no overhead in term of application performance. With this in mind in my work, I have designed application-aware node-level optimal thermal
management algorithms and runtimes, lazy node-level power capping, and energy reduction runtime. The experimental results show a significant step forward with respect to the current state-of-the-art solutions in power and thermal control of HPC systems.

Poster 41: Co-Designing Communication Middleware and Deep Learning Frameworks for High-Performance DNN Training on HPC Systems
Ammar Ahmad Awan (Ohio State University)

Recent advances in Machine/Deep Learning techniques have triggered key success stories in many application domains like Computer Vision, Speech Comprehension and Recognition, and Natural Language Processing. Large-scale Deep Neural Networks (DNNs), that are at the core of state-of-the-art AI technologies, have been the primary drivers of this success. Training very complicated and large DNN architectures using a large number of training examples (data) is compute-intensive and can take from weeks to months to achieve state-of-the-art prediction capabilities. To achieve higher accuracy, making the DNN deeper is also a common strategy. These requirements have led to a simple but powerful approach called Data Parallelism to achieve shorter training times. This has resulted in various research studies and ML/DL software like TensorFlow and PyTorch as well as distributed-training middleware like Horovod. In addition, for DNNs that do not fit the GPU memory, a new DNN workloads are emerging that we call Out-of-Core DNNs and different strategies (out-of-core training and model-parallelism) are needed to train them. Clearly, large-scale DNN training brings forward new requirements for computation runtimes like CUDA and communication middleware like the MVAPICH2 MPI library. In this thesis, we broadly explore three different strategies to train DNNs on modern CPU and GPU architectures: 1) Data Parallelism, 2) Model Parallelism, and 3) Out-of-Core Training. We address the key challenge: How to co-design computation and communication in modern ML/DL frameworks with execution runtimes like CUDA and communication middleware like MVAPICH2 to enable scalable, high-performance, and efficient training of DNNs on large-scale HPC systems?

Poster 40: Performance, Portability, and Productivity for Data-Parallel Computations on Multi- and Many-Core Architectures
Ari Rasch (University of Münster)

This thesis presents an approach to performance, portability, and productivity for data-parallel computations on multi- and many-core architectures, e.g., Intel CPU and NVIDIA GPU. We introduce the algebraic formalism of Multi-Dimensional Homomorphisms (MDHs) – a class of functions that cover important data-parallel computations, e.g., linear algebra routines (BLAS) and stencil computations. For our MDHs, we propose a Domain-Specific Language (DSL), based on
patterns of parallelism (a.k.a. algorithmic skeletons), to enable conveniently expressing MDH functions. We introduce a code generation approach for our DSL to automatically generate for MDHs optimized program code targeting multi- and many-core architectures. Our code generation approach relies on OpenCL – an emerging de-facto standard for uniformly programming parallel architectures, such as CPU and GPU. A major feature of our generated code is that it is targeted to OpenCL’s abstract device models (rather than a particular architecture) by being parameterized in performance-critical parameters of these abstract models (e.g., the number of threads and size of tiles). With our code generation approach, we enable both high performance and performance portability: we fully automatically optimize our generated code -- for any given combination of an MDH function, architecture, and input size -- by automatically choosing (auto-tuning) optimized values of our code’s performance-critical parameters using our own Auto-Tuning Framework (ATF). Our experimental results on CPU and GPU demonstrate competitive and often significantly better performance of our MDH+ATF approach as compared to the currently best-performing competitors, e.g., Intel MKL/MKL-DNN, NVIDIA cuBLAS/cuDNN, and Facebook’s Tensor Comprehensions framework.

Poster 45: Characterization and Modeling of Error Resilience in HPC Applications
Luanzheng Guo (University of California, Merced)

As high-performance computing systems scale in size and computational power, the occurrence of transient faults grows. Without protection by efficient and effective fault tolerance mechanisms, transient errors can cause incorrect execution outcome and even lead to a catastrophe in safety-critical HPC applications. Previous work attributes error resilience in HPC applications at a high-level to either the probabilistic or iterative nature of the application, whereas the community still lacks the fundamental understanding of the program constructs that result in natural error resilience. We design FlipTracker, a framework to analytically track error propagation and to provide a fine-grained understanding of the propagation and tolerance of errors. After running FlipTracker on representative HPC applications, we summarize six resilience computation patterns that lead to nature error resilience in HPC applications. With a better understanding of natural resilience in HPC applications, we aim to model application resilience on data objects to transient faults. Many common application-level fault tolerance mechanisms focus on data objects. Understanding application resilience on data objects can be helpful to direct those mechanisms. The common practice to understand application resilience (random fault injection) gives us little knowledge of how and where errors are tolerated. Understanding "how" and "where" is necessary to understand how to apply application-level fault tolerance mechanisms effectively and efficiently. We design a practical model (MOARD) to measure application resilience on data objects by analytically quantifying error masking events happening to the data object. Using our model, users can compare application resilience on different data objects with different data types.
Poster 43: Efficient and Scalable Communication Middleware for Emerging Dense-GPU Clusters  
Ching-Hsiang Chu (Ohio State University)

In the era of post Moore’s law, the traditional CPU is not able to keep the pace up and provide the computing power demanded by the modern compute-intensive and highly parallelizable applications. Under this context, various accelerator architectures such as general-purpose graphics processing unit (GPU), which equipped with the high-bandwidth memory (HBM) and massive parallelizable streaming multiprocessors, has been widely adopted in high-performance computing (HPC) and cloud systems to significantly accelerate numerous scientific and emerging machine/deep learning applications. Message Passing Interface (MPI), the standard programming model for parallel applications, has been widely used for GPU communication. However, the state-of-the-art MPI libraries are only optimizing GPU communication by leveraging advanced technology like Remote Direct Memory Access (RDMA) and not fully utilizing the computational power of GPUs. In this work, we propose GPU-enabled communication schemes to harness GPU computational resources, and cutting-edge interconnects such as NVIDIA NVLink for communication operations on the emerging heterogeneous systems. In this work, three primary MPI operations are addressed. First, intelligent communication scheduling, efficient packing/unpacking, and packing-free schemes are proposed to accelerate non-contiguous data transfer in scientific HPC applications. Second, scalable broadcast operations are presented to leverage the low-level hardware multicast feature to speed up GPU communication at scale. Finally, we also design topology-aware, link-efficient, and cooperative GPU kernels to significantly accelerate All-reduce operation, which is the primary performance bottleneck in deep learning applications. The proposed designs demonstrate significant performance improvements over the state-of-the-art communication schemes for various HPC and deep learning applications.

Poster 39: Designing Next-Generation Communication Middlewares for Many-Core Architectures  
Jahanzeb Maqbool Hashmi (Ohio State University)

Modern multi-petaflop HPC systems are powered by dense multi-/many-core architectures and this trend is expected to grow for the next-generation supercomputing systems. This rapid adoption of high core-density architectures by the current- and next-generation HPC systems is further fueled by the emerging application trends such as Deep Learning. This is putting more emphasis on middleware designers to optimize various communication protocols to meet the diverse needs of the applications. While the novelties in the processor architectures have led to the increased on-
chip parallelism, they come at the cost of rendering traditional designs, employed by the communication runtimes such as MPI, to suffer from higher degree of intra-node communication latencies. Tackling the computation and communication challenges that accompany these dense multi-/many-cores garner special design considerations. The proposed work in this thesis tries to address the performance challenges posed by a diverse range of applications and the lacking support in state-of-the-art communication libraries such as MPI to exploit high-concurrency architectures. The author first proposes a "shared-address-spaces"-based communication substrate to derive intra-node communication in MPI. Atop this framework, the author has re-designed various MPI primitives such as point-to-point communication protocols (e.g., user-space zero-copy rendezvous transfer), collective communication (e.g., load/store based collectives, truly zero-copy and partitioning-based reduction algorithms), and efficient MPI derived datatypes processing (e.g., memoization-based "packing-free" communication) to exploit the potential of emerging multi-/many-core architectures and high throughput networks. The proposed designs have demonstrated significant improvement over state-of-the-art for various scientific and deep learning applications.

8:30 am - 5:00 pm

Research Posters Display

Poster 74: Enabling Code Portability of a Parallel and Distributed Smooth-Particle Hydrodynamics Application, FleCSPH
Suyash Tandon (University of Michigan), Nicholas Stegmeier (University of Illinois), Vasu Jaganath (University of Wyoming), Jennifer Ranta (Michigan State University), Rathish Ratnasingam (Newcastle University), Elizabeth Carlson (University of Nebraska), Julien Loiseau (Los Alamos National Laboratory), Vinay Ramakrishnaiah (Los Alamos National Laboratory), Robert Pavel (Los Alamos National Laboratory)

Core-collapse supernovae (CCSNe) are integral to the formation and distribution of heavy elements across the universe. However, CCSNe are highly complex and inherently non-linear phenomena. Large-scale simulations of these cosmic events can provide us a glimpse of their hydrodynamic and nucleosynthetic processes which are difficult to observe. To enable these massive numerical simulations on high-performance computing (HPC) centers, this study uses FleCSPH, a parallel and distributed code, based on the smooth-particle hydrodynamics (SPH) formulation. In the recent years, the HPC architecture has evolved and the next generation of exascale computers are expected to feature heterogenous architecture. Therefore, it is important to maintain code portability across platforms. This work demonstrates code portability of FleCSPH through the incorporation of
Poster 73: Accelerating Large-Scale GW Calculations on Hybrid CPU-GPU Architectures

Mauro Del Ben (Lawrence Berkeley National Laboratory), Charlene Yang (National Energy Research Scientific Computing Center (NERSC)), Felipe Jornada (University of California, Berkeley; Lawrence Berkeley National Laboratory), Steven G. Louie (University of California, Berkeley; Lawrence Berkeley National Laboratory), Jack Deslippe (National Energy Research Scientific Computing Center (NERSC))

In this poster, we present the strategy, progress, and performance while GPU porting one of the major modules, epsilon, of the electronic structure code BerkeleyGW. Epsilon represents the most time-consuming routines in the BerkeleyGW workflow for large-scale material science simulations. Some of the porting/optimization strategies include, changing our original data layout to efficiently use libraries such as cuBLAS and cuFFT, implementation of specific CUDA kernels to minimize data copies between host/device and keeping data on device, efficient use of data streams to leverage high concurrency on the device, asynchronous memory copies and overlapping (MPI) communication on the host and computation on the device. Preliminary results are presented in terms of the speedup compare to the CPU-only implementation, strong/weak scaling, and power efficiency. Excellent acceleration is demonstrated: up to 30x for specific kernels. Our port also exhibits good scalability and about 16x higher FLOPs/watt efficiency compared to the CPU-only implementation.

Poster 89: BeeCWL: A CWL Compliant Workflow Management System

Betis Baheri (Kent State University), Steven Anaya (New Mexico Institute of Mining and Technology), Patricia Grubel (Los Alamos National Laboratory), Qiang Guan (Kent State University), Timothy Randles (Los Alamos National Laboratory)

Scientific workflows are used widely to carry out complex and hierarchical experiments. Although there are many trends to extend the functionality of workflow management systems to cover all possible requirements that may arise from a user community, one unified standard over cloud and HPC systems is still missing. In this paper, we propose a Common Workflow Language (CWL) compliant workflow management system. BeeCWL is a parser to derive meaningful information such as requirements, steps, relationships, etc. from CWL files and to create a graph database from
those components. Generated graphs can be passed to an arbitrary scheduler and management system to decide whether there are enough resources to optimize and execute the workflow. Lastly, the user can have control over workflow execution, collecting logs, and restart or rerun some part of a complex workflow.

Best Poster Finalist: no

**Poster 150: A Machine Learning Approach to Understanding HPC Application Performance Variation**
Burak Aksar (Boston University, Sandia National Laboratories), Benjamin Schwaller (Sandia National Laboratories), Omar Aaziz (Sandia National Laboratories), Emre Ates (Boston University), Jim Brandt (Sandia National Laboratories), Ayse K. Coskun (Boston University), Manuel Egele (Boston University), Vitus Leung (Sandia National Laboratories)

Performance anomalies are difficult to detect because often a “healthy system” is vaguely defined, and the ground truth for how a system should be operating is evasive. As we move to exascale, however, detection of performance anomalies will become increasingly important with the increase in size and complexity of systems. There are very few accepted ways of detecting anomalies in the literature, and there are no published and labeled sets of anomalous HPC behavior. In this research, we develop a suite of applications that represent HPC workloads and use data from a lightweight metric collection service to train machine learning models to predict the future behavior of metrics. In the future, this work will be used to predict anomalous runs in compute nodes and determine some root causes of performance issues to help improve the efficiency of HPC system administrators and users.

Best Poster Finalist: no

**Poster 81: Performance of Devito on HPC-Optimised ARM Processors**
Hermes Senger (Federal University of São Carlos, Brazil; University of São Paulo), Jaime Freire de Souza (Federal University of São Carlos, Brazil), Edson Satoshi Gomi (University of São Paulo), Fabio Luporini (Imperial College, London), Gerard Gorman (Imperial College, London)

We evaluate the performance of Devito, a domain specific language (DSL) for finite differences on Arm ThunderX2 processors. Experiments with two common seismic computational kernels demonstrate that Devito can apply automatic code generation and optimization across Arm and Intel platforms. The code transformations include: parallelism, and SIMD vectorization (OpenMP >=4); loop tiling (with best block shape obtained via auto-tuning); domain-specific symbolic
optimisation such as common sub-expression elimination and factorisation for Flop reduction, polynomial approximations for trigonometry terms, and heuristic hoisting of time-invariant expressions. Results show that Devito can achieve performance on Arm processors which is competitive to other Intel Xeon processors.

Best Poster Finalist: no

Poster 103: LIKWID 5: Lightweight Performance Tools
Thomas Gruber (Erlangen Regional Computing Center), Jan Eitzinger (Erlangen Regional Computing Center), Georg Hager (Erlangen Regional Computing Center), Gerhard Wellein (Erlangen Regional Computing Center)

LIKWID is a tool suite for performance oriented programmers with a worldwide user group. It is developed by the HPC group of the University Erlangen-Nuremberg since 2009 to support them in their daily research and performance engineering of user codes. The HPC landscape has become more and more diverse over the last years with clusters using non-x86 architectures and being equipped with accelerators. With the new major version, the architectural support of LIKWID is extended to ARM and POWER CPUs with the same functionality and features as for x86 architectures. Besides the CPU monitoring, the new version provides access the hardware counting facilities of Nvidia GPUs. This poster introduces the new features and shows the successes of applying LIKWID to identify performance bottlenecks and to test optimizations. Furthermore, the poster gives an overview of how users can integrate the LIKWID tools in their application using a lightweight add-once-and-reuse instrumentation API.

Best Poster Finalist: no

Poster 149: Solving Phase-Field Equations in Space-Time: Adaptive Space-Time Meshes and Stabilized Variational Formulations
Kumar Saurabh (Iowa State University), Biswajit Khara (Iowa State University), Milinda Fernando (University of Utah), Masado Ishii (University of Utah), Hari Sundar (University of Utah), Baskar Ganapathysubramanian (Iowa State University)

We seek to efficiently solve a generalized class of partial differential equations called the phase-field equations. These non-linear PDE’s model phase transition (solidification, melting, phase-separation) phenomena which exhibit spatially and temporally localized regions of steep gradients. We consider time as an additional dimension and simultaneously solve for the unknown in large blocks of time (i.e. in space-time), instead of the standard approach of sequential time-stepping. We
use variational multiscale (VMS) based finite element approach to solve the ensuing space-time equations. This allows us to (a) exploit parallelism not only in space but also in time, (b) gain high order accuracy in time, and (c) exploit adaptive refinement approaches to locally refine region of interest in both space and time. We illustrate this approach with several canonical problems including melting and solidification of complex snow flake structures.

Best Poster Finalist: no

**Poster 84: ESTEE: A Simulation Toolkit for Distributed Workflow Execution**
Vojtěch Cima (IT4Innovations, Czech Republic), Jakub Beránek (IT4Innovations, Czech Republic), Stanislav Böhm (IT4Innovations, Czech Republic)

Task graphs provide a simple way to describe scientific workflows (sets of tasks with dependencies) that can be executed on both HPC clusters and in the cloud. An important aspect of executing such graphs is the used scheduling algorithm. Many scheduling heuristics have been proposed in existing works; nevertheless, they are often tested in oversimplified environments. We introduce a simulation environment designed for prototyping and benchmarking task schedulers. Our simulation environment, scheduler source codes, and graph datasets are open in order to be fully reproducible. To demonstrate usage of Estee, as an example, we compare the performance of various workflow schedulers in an environment using two different network models.

Best Poster Finalist: no

**Poster 93: Robust Data-Driven Power Simulator for Fast Cooling Control Optimization of a Large-Scale Computing System**
Takashi Shiraishi (Fujitsu Laboratories Ltd), Hiroshi Endo (Fujitsu Laboratories Ltd), Takaaki Hineno (Fujitsu Laboratories Ltd), Hiroyuki Fukuda (Fujitsu Laboratories Ltd)

Power of large-scale systems such as an HPC or a datacenter is a significant issue. Cooling units consume 30% of the total power. General control policies for cooling units are local and static (manual overall optimization nearly once a week). However, free cooling and IT-load fluctuation may change hourly optimum control variables of the cooling units. In this work, we present a deep neural network (DNN) power simulator that can learn from actual operating logs and can quickly identify the optimum control variables. We demonstrated the power simulator of an actual large-scale system with 4.7-MW-power IT load. Our robust simulator predicted the total power with error of 4.8% without retraining during one year. We achieved optimization by the simulator within 80 seconds that was drastically faster than previous works. The dynamic control optimization each hour showed a 15%
power reduction compared to that of conventional policy in the actual system.

Best Poster Finalist: no

Poster 115: sDNA: Software-Defined Network Accelerator Based on Optical Interconnection Architecture

En Shao (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences), Guangming Tan (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences), Zhan Wang (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences), Guojun Yuan (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences), Ninghui Sun (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences)

Software-Defined Network Accelerator (sDNA) is a new accelerated system for the exascale computer. Inspired by the edge forwarding index (EFI), the main contribution of our work is that it presents an extended EFI-based optical interconnection method with slow switching optical device. In our work, we found that sDNA based on extended EFI evaluation is not only able to offload the traffic from an electrical link to an optical link but is also able to avoid congestion inherent to electrical link.

Best Poster Finalist: no

Poster 49: WarpX: Toward Exascale Modeling of Plasma Particle Accelerators on GPU

Maxence Thevenet (Lawrence Berkeley National Laboratory), Jean-Luc Vay (Lawrence Berkeley National Laboratory), Ann Almgren (Lawrence Berkeley National Laboratory), Diana Amorim (Lawrence Berkeley National Laboratory), John Bell (Lawrence Berkeley National Laboratory), Axel Huebl (Lawrence Berkeley National Laboratory), Revathi Jambunathan (Lawrence Berkeley National Laboratory), Rémi Lehe (Lawrence Berkeley National Laboratory), Andrew Myers (Lawrence Berkeley National Laboratory), Jaehong Park (Lawrence Berkeley National Laboratory), Olga Shapoval (Lawrence Berkeley National Laboratory), Weiqun Zhang (Lawrence Berkeley National Laboratory), Lixin Ge (SLAC National Accelerator Laboratory), Mark Hogan (SLAC National Accelerator Laboratory), Cho Ng (SLAC National Accelerator Laboratory), David Grote (Lawrence Livermore National Laboratory)

Particle accelerators are a vital part of the DOE-supported infrastructure of discovery science and applications, but we need game-changing improvements in the size and cost for future accelerators. Plasma-based particle accelerators stand apart in their potential for these improvements. Turning this
from a promising technology into mainstream scientific tools depends critically on high-performance, high-fidelity modeling of complex processes.

WarpX is an open-source particle-in-cell (PIC) code supported by the Exascale Computing Project (ECP) that is combining advanced algorithms with adaptive mesh refinement to allow challenging simulations of a multi-stage plasma-based TeV acceleration relevant for future high-energy physics discoveries. WarpX relies on the ECP co-design center for mesh refinement AMReX, and runs on CPU and GPU-accelerated computers. Production simulation have run on Cori KNL at NERSC and Summit at OLCF. In this poster, recent results and strategies on GPU will be presented, along with recent performance results.

Best Poster Finalist: no

Poster 50: Implementing an Adaptive Sparse Grid Discretization (ASGarD) for High Dimensional Advection-Diffusion Problems on Exascale Architectures
M. Graham Lopez (Oak Ridge National Laboratory), David L. Green (Oak Ridge National Laboratory), Lin Mu (University of Georgia), Ed D’Azevedo (Oak Ridge National Laboratory), Wael Elwasif (Oak Ridge National Laboratory), Tyler McDaniel (University of Tennessee), Timothy Younkin (University of Tennessee), Adam McDaniel (Oak Ridge National Laboratory), Diego Del-Castillo-Negrete (Oak Ridge National Laboratory)

Many scientific domains require the solution of high dimensional PDEs. Traditional grid- or mesh-based methods for solving such systems in a noise-free manner quickly become intractable due to the scaling of the degrees of freedom going as $O(N^d)$ sometimes called "the curse of dimensionality." We are developing an arbitrarily high-order discontinuous-Galerkin finite-element solver that leverages an adaptive sparse-grid discretization whose degrees of freedom scale as $O(N \cdot \log_2 N^{D-1})$. This method and its subsequent reduction in the required resources is being applied to several PDEs including time-domain Maxwell’s equations (3D), the Vlasov equation (in up to 6D) and a Fokker-Planck-like problem in ongoing related efforts. Here we present our implementation which is designed to run on multiple accelerated architectures, including distributed systems. Our implementation takes advantage of a system matrix decomposed as the Kronecker product of many smaller matrices which is implemented as batched operations.

Best Poster Finalist: no

Poster 91: FreeCompilerCamp: Online Training for Extending Compilers
Alok Mishra (Stony Brook University, Lawrence Livermore National Laboratory), Anjia Wang
In this presentation, we introduce an ongoing effort of an online training platform aimed to automate the training of developers to quickly extend compilers - FreeCompilerCamp.org. Our free and open platform allows anyone who is interested in developing compilers to learn the necessary skills. A live training website, built on top of Play-With-Docker, is set up so that anyone with internet access and a web browser will be able to take this training. The entire training system is open-source and developers with relevant skills can contribute new tutorials and deploy it on a private server, workstation or even laptop. We have created some initial tutorials on how to extend the Clang/LLVM or ROSE compilers to support new OpenMP features. Using a web interface consisting of two side-by-side panels, users can follow the tutorials on one side and immediately practice what they learned in a terminal sandbox embedded on the other.

Best Poster Finalist: yes

**Poster 51: SmartK: Efficient, Scalable, and Winning Parallel MCTS**
Michael S. Davinroy (Swarthmore College), Shawn Pan (Swarthmore College), Bryce Wiedenbeck (Swarthmore College, Davidson College), Tia Newhall (Swarthmore College)

SmartK is our efficient and scalable parallel algorithm for Monte Carlo Tree Search (MCTS), an approximation technique for game searches. MCTS is also used to solve problems as diverse as planning under uncertainty, combinatorial optimization, and high-energy physics. In these problems, the solution search space is significantly large, necessitating parallel solutions. Shared memory parallel approaches do not scale well beyond the size of a single node's RAM. SmartK is a distributed memory parallelization that takes advantage of both inter-node and intra-node parallelism and a large cumulative RAM found in clusters. SmartK's novel selection algorithm combined with its ability to efficiently search the solution space, results in better solutions than other MCTS parallel approaches. Results of an MPI implementation of SmartK for the game of Hex, show SmartK yields a better win percentage than other parallel algorithms, and that its performance scales to larger search spaces and high degrees of parallelism.

Best Poster Finalist: no

**Poster 70: Numerical Method and Parallelization for the Computation of Coherent Synchrotron Radiation**
Boqian Shen (Rice University, Los Alamos National Laboratory)

The purpose of this work is to develop and parallelize an accurate and efficient numerical method for the computation of synchrotron radiation from relativistic electrons in the near field. The high-brilliance electron beam and coherent short-wavelength light source provide a powerful method to understand the microscopic structure and dynamics of materials. Such a method supports a wide range of applications including matter physics, structural biology, and medicine development. To understand the interaction between the beam and synchrotron radiation, an accurate and efficient numerical simulation is needed. With millions of electrons, the computational cost of the field would be large. Thus, multilevel parallelism and performance portability are desired since modern supercomputers are getting more complex and heterogeneous. The performance model and performance analysis are presented.

Best Poster Finalist: no

Poster 146: AI Matrix: A Deep Learning Benchmark for Alibaba Data Centers
Wei Zhang (Alibaba Inc), Wei Wei (Alibaba Inc), Lingjie Xu (Alibaba Inc), Lingling Jin (Alibaba Inc), Cheng Li (University of Illinois)

This work introduces AI Matrix, an in-house Deep Learning (DL) benchmark suite developed specifically for Alibaba’s e-commerce environment. AI Matrix results from a full investigation of the DL applications used inside Alibaba and aims to cover the typical DL applications that account for more than 90% of the GPU usage in Alibaba data centers. This benchmark suite collects DL models that are either directly used or closely resemble the models used in the company’s real e-commerce applications. It also collects the real e-commerce applications if no similar DL models are not available. Through the high coverage and close resemblance to real applications, AI Matrix fully represents the DL workloads on Alibaba data centers. The collected benchmarks mainly fall into three categories: computer vision, recommendation, and language processing, which consist of the most majority of DL applications in Alibaba. AI Matrix is made open source, hoping it can benefit the public.

Best Poster Finalist: no

Poster 54: Massively Parallel Eigensolvers Based on Unconstrained Energy Functionals Methods
Osni Marques (Lawrence Berkeley National Laboratory), Mauro Del Ben (Lawrence Berkeley National Laboratory), Andrew Canning (Lawrence Berkeley National Laboratory)

This poster focuses on a preconditioned conjugate gradient based iterative eigensolver using an
unconstrained energy functional minimization scheme. This scheme avoids an explicit reorthogonalization of the trial eigenvectors and becomes an attractive alternative for the solution of very large problems. The unconstrained formulation is implemented in the first-principles materials and chemistry CP2K code, which performs electronic structure calculations based on a density functional theory approximation to the solution of the many-body Schrödinger equation. The systems we use in our studies have a number of atoms ranging from 2,247 to 12,288. We study the convergence of the unconstrained formulation and its scaling on a Cray XC40 (a partition with 9,688 Intel KNL nodes). We show that there is a trade-off between the preconditioner that leads to fast convergence and lower cost preconditioners that lead to best time to solution.

Best Poster Finalist: yes

**Poster 100: Comparison of Array Management Library Performance - A Neuroscience Use Case**

Donghe Kang (Ohio State University), Oliver Rübel (Lawrence Berkeley National Laboratory), Suren Byna (Lawrence Berkeley National Laboratory), Spyros Blanas (Ohio State University)

Array management libraries, such as HDF5, Zarr, etc., depend on a complex software stack that consists of parallel I/O middleware (MPI-IO), POSIX-IO, and file systems. Components in the stack are interdependent, such that effort in tuning the parameters in these software libraries for optimal performance is non-trivial. On the other hand, it is challenging to choose an array management library based on the array configuration and access patterns. In this poster, we investigate the performance aspect of two array management libraries, i.e., HDF5 and Zarr, in the context of a neuroscience use case. We highlight the performance variability of HDF5 and Zarr in our preliminary results and discuss potential optimization strategies.

Best Poster Finalist: no

**Poster 118: Self-Driving Reconfigurable Silicon Photonic Interconnects (Flex-LIONS) with Deep Reinforcement Learning**

Roberto Proietti (University of California, Davis), Yu Shang (University of California, Davis), Xian Xiao (University of California, Davis), Xiaoliang Chen (University of California, Davis), Yu Zhang (University of California, Davis), SJ Ben Yoo (University of California, Davis)

We propose a self-driving reconfigurable optical interconnect architecture for HPC systems exploiting a deep reinforcement learning (DRL) algorithm and a reconfigurable silicon photonic (SiPh) switching fabric to adapt the interconnect topology to different traffic demands. Preliminary simulation results show that after training, the DRL-based SiPh fabric provides the lowest average end-to-end latency.
for time-varying traffic patterns.

Best Poster Finalist: no

**Poster 147: Extremely Accelerated Deep Learning: ResNet-50 Training in 70.4 Seconds**

Akihiro Tabuchi (Fujitsu Laboratories Ltd), Akihiko Kasagi (Fujitsu Laboratories Ltd), Masafumi Yamazaki (Fujitsu Laboratories Ltd), Takumi Honda (Fujitsu Laboratories Ltd), Masahiro Miwa (Fujitsu Laboratories Ltd), Takashi Shiraishi (Fujitsu Laboratories Ltd), Motohiro Kosaki (Fujitsu Laboratories Ltd), Naoto Fukumoto (Fujitsu Laboratories Ltd), Tsuguchika Tabaru (Fujitsu Laboratories Ltd), Atsushi Ike (Fujitsu Laboratories Ltd), Kohta Nakashima (Fujitsu Laboratories Ltd)

Distributed deep learning using a large mini-batch is a key technology to accelerate training in deep learning. However, it is difficult to achieve a high scalability and maintain validation accuracy in distributed learning on large clusters. We introduce two optimizations, reducing the computation time and overlapping the communication with the computation. By applying the techniques and using 2,048 GPUs, we achieved the world's fastest ResNet-50 training in MLPerf, which is a de facto standard DNN benchmark (as of July 2019).

Best Poster Finalist: no

**Poster 111: Multiple HPC Environments-Aware Container Image Configuration for Bioinformatics Application**

Kento Aoyama (Tokyo Institute of Technology, National Institute of Advanced Industrial Science and Technology (AIST)), Hiroki Watanabe (Tokyo Institute of Technology, National Institute of Advanced Industrial Science and Technology (AIST)), Masahito Ohue (Tokyo Institute of Technology), Yutaka Akiyama (Tokyo Institute of Technology)

Containers have a considerable advantage for application portability in different environments by isolating process with a small performance overhead; thus it has been rapidly getting popular in a wide range of science fields. However, there are problems in container image configuration when run in multiple HPC environments, and it requires users to have knowledge of systems, container runtimes, container image format, and library compatibilities in HPC environments.

In this study, we introduce our HPC container workflow in multiple supercomputing environments that have different system/library specifications (ABCI, TSUBAME3.0). Our workflow provides custom container image configurations for HPC environments by taking into account differences in container runtime, container image, and library compatibility between the host and inside of the container. We
also show the parallel performance of our application in each HPC environment.

Best Poster Finalist: no

**Poster 134: Minimal-Precision Computing for High-Performance, Energy-Efficient, and Reliable Computations**

Daichi Mukunoki (RIKEN Center for Computational Science (R-CCS)), Toshiyuki Imamura (RIKEN Center for Computational Science (R-CCS)), Yiyu Tan (RIKEN Center for Computational Science (R-CCS)), Atsushi Koshiba (RIKEN Center for Computational Science (R-CCS)), Jens Huthmann (RIKEN Center for Computational Science (R-CCS)), Kentaro Sano (RIKEN Center for Computational Science (R-CCS)), Fabienne Jézéquel (Sorbonne University), Stef Graillat (Sorbonne University), Roman Iakymchuk (Sorbonne University), Norihisa Fujita (University of Tsukuba), Taisuke Boku (University of Tsukuba)

In numerical computations, the precision of floating-point computations is a key factor to determine the performance (speed and energy-efficiency) as well as the reliability (accuracy and reproducibility). However, the precision generally plays a contrary role for both. Therefore, the ultimate concept for maximizing both at the same time is the minimal-precision computation through precision-tuning, which adjusts the optimal precision for each operation and data. Several studies have been already conducted for it so far, but the scope of those studies is limited to the precision-tuning alone. In this study, we propose a more broad concept of the minimal-precision computing with precision-tuning, involving both hardware and software stack.

Best Poster Finalist: no

**Poster 112: Building Complex Software Applications Inside Containers**

Calvin D. Seamons (Los Alamos National Laboratory)

High performance computing (HPC) scientific applications require complex dependencies to operate. As user demand for HPC systems increases, it becomes unrealistic to support every unique dependency request. Containers can offer the ability to satisfy the users’ dependency request while simultaneously offering HPC portability across systems. By “containerizing” Model for Prediction Across Scales (MPAS, a large atmospheric simulation suite), we show that it is possible to containerize and run complex software. Furthermore, the container can be run across different HPC systems with nearly identical results (21 bytes difference over 2.1 gigabytes). Containers have the possibility to bring flexibility to code teams in HPC by helping to meet the demand for user defined software stacks (UDSS), and giving teams the ability to choose their software, independently of what
Poster 101: Job Performance Overview of Apache Flink and Apache Spark Applications
Jan Frenzel (Technical University Dresden), René Jäkel (Technical University Dresden)

Apache Spark and Apache Flink are two Big Data frameworks used for fast data exploration and analysis. Both frameworks provide the runtime of program sections and performance metrics, such as the number of bytes read or written, via an integrated dashboard. Performance metrics available in the dashboard lack timely information and are only shown aggregated in a separate part of the dashboard. However, performance investigations and optimizations would benefit from an integrated view with detailed performance metric events. Thus, we propose a system that samples metrics at runtime and collects information about the program sections after the execution finishes. The performance data is stored in an established format independent from Spark and Flink versions and can be viewed with state-of-the-art performance tools, i.e. Vampir. The overhead depends on the sampling interval and was below 10% in our experiments.

Poster 113: Improvements Toward the Release of the Pavilion 2.0 Test Harness
Kody J. Everson (Los Alamos National Laboratory, Dakota State University), Maria Francine Lapid (Los Alamos National Laboratory)

High-performance computing production support entails thorough testing in order to evaluate the efficacy of a system for production-grade workloads. There are various phases of a system’s life-cycle to assess, requiring different methods to accomplish effective evaluation of performance and correctness. Due to the unique and distributed nature of an HPC-system, the necessity for sophisticated tools to automatically harness and assess test results, all while interacting with schedulers and programming environment software, requires a customizable, extensible, and lightweight system to manage concurrent testing. Beginning with the recently refactored codebase of Pavilion 1.0, we assisted with the finishing touches on readying this software for open-source release and production usage. Pavilion 2.0 is a Python 3-based testing framework for HPC clusters that facilitates the building, running, and analysis of tests through an easy-to-use, flexible, YAML-based configuration system. This enables users to write their own tests by simply wrapping everything in Pavilion’s well-defined format.
Poster 119: Toward Lattice QCD on Fugaku: SVE Compiler Studies and Micro-Benchmarks in the RIKEN Fugaku Processor Simulator

Nils Meyer (University of Regensburg, Bavaria), Tilo Wettig (University of Regensburg, Bavaria), Yuetsu Kodama (RIKEN Center for Computational Science (R-CCS)), Mitsuhisa Sato (RIKEN Center for Computational Science (R-CCS))

The Fugaku supercomputer, successor to the Japanese flagship K-Computer, will start operation in 2021. Fugaku incorporates the Fujitsu A64FX processor, which is the first hardware implementation supporting the Arm SVE instruction set, in this case a 512-bit version. Real hardware is not accessible today, but RIKEN has designed a simulator of the A64FX. We present micro-benchmarks relevant for Lattice QCD obtained in the RIKEN Fugaku processor simulator and compare three different SVE compilers.

Poster 62: Emulating Multi-Pattern Quantum Grover’s Search on a High-Performance Reconfigurable Computer

Naveed Mahmud (University of Kansas), Bennett Haase-Divine (University of Kansas), Bailey K. Srimoungchanh (University of Kansas), Nolan Blankenau (University of Kansas), Annika Kuhnke (University of Kansas), Esam El-Araby (University of Kansas)

Grover’s search(GS) is a widely studied quantum algorithm that can be employed for both single and multi-pattern search problems and potentially provides quadratic speedup over existing classical search algorithms. In this paper, we propose a multi-pattern quantum search methodology based on a modified GS quantum circuit. The proposed method combines classical post-processing permutations with a modified Grover’s circuit to efficiently search for given single/multiple input patterns. Our proposed methodology reduces quantum circuit complexity, realizes space-efficient emulation hardware and improves overall system configurability for dynamic, multi-pattern search. We use a high-performance reconfigurable computer to emulate multi-pattern GS(MGS) and present scalable emulation architectures of a complete multi-pattern search system. We validate the system and provide analysis of experimental results in terms of FPGA resource utilization and emulation time. Our results include a successful hardware architecture that is capable of emulating MGS algorithm up to 32 fully-entangled quantum bits on a single FPGA.
**Poster 107: Exploring Interprocess Work Stealing for Balanced MPI Communication**  
Kaiming Ouyang (University of California, Riverside), Min Si (Argonne National Laboratory), Zizhong Chen (University of California, Riverside)

Workload balance among MPI processes is a critical consideration during the development of HPC applications. However, because of many factors such as complex network interconnections and irregularity of HPC applications, fully achieving workload balance in practice is nearly impossible. Although interprocess job stealing is a promising solution, existing shared-memory techniques that lack necessary flexibility or cause inefficiency during data access cannot provide an applicable job-stealing implementation. To solve this problem, we propose a new process-in-process (PiP) interprocess job-stealing method to balance communication workload among processes on MPI layers. Our initial experimental results show PiP-based job stealing can efficiently help amortize workload, reduce imbalance, and greatly improve intra- and intersocket ping-pong performance compared with original MPI.

Best Poster Finalist: no

**Poster 127: sFlow Monitoring for Security and Reliability**  
Xava A. Grooms (Los Alamos National Laboratory, University of Kentucky), Robert V. Rollins (Los Alamos National Laboratory, Michigan Technological University), Collin T. Rumpca (Los Alamos National Laboratory, Dakota State University)

In the past ten years, High Performance Computing (HPC) has moved far beyond the terascale performance, making petascale systems the new standard. The drastic improvement in performance has been largely unmatched with insignificant improvements in system monitoring. Thus, there is an immediate need for practical and scalable monitoring solutions to ensure the effectiveness of costly compute clusters. This project aims to explore the viability and impact of sFlow enabled switches in cluster network monitoring for security and reliability. A series of tests and exploits were performed to target specific network abnormalities on a nine-node HPC cluster. The results present web-based dashboards that can aid network administrators in improving a cluster’s security and reliability.

Best Poster Finalist: no

**Poster 77: Extreme Scale Phase-Field Simulations of Sintering Processes**  
Johannes Hötzer (Karlsruhe University of Applied Sciences), Henrik Hierl (Karlsruhe University of...
The sintering process, which turns loose powders into dense materials, is naturally found in the formation of glaciers, but is also the indispensable process to manufacture ceramic materials. This process is described by a dynamically evolving microstructure, which largely influences the resulting material properties.

To investigate this complex three-dimensional, scale-bridging evolution in realistic domain sizes, a highly optimized and parallelized multiphysics phase-field solver is developed. The solver is optimized in a holistic way, from the application level over the time integration and parallelization, down to the hardware. Optimizations include communication hiding, explicit vectorization, implicit schemes, and local reduction of degrees of freedom.

With this, we are able to investigate large-scale, three-dimensional domains, and long integration times. We have achieved a single-core peak performance of 32.5%, scaled up to 98304 cores on Hazel Hen and SuperMUC-NG, and simulated a multimillion particle system.

Best Poster Finalist: no

**Poster 140: Toward Automatic Function Call Generation for Deep Learning**

*Shizhi Tang (Tsinghua University, China), Jidong Zhai (Tsinghua University, China)*

Mainstream deep learning frameworks are commonly implemented by invoking underlying high performance tensor libraries on various architectures. However, as these libraries provide increasingly complex semantics including operator fusions, in-place operations, and various memory layouts, the gap between mathematical deep learning models and the underlying libraries becomes larger. In this paper, inspired by the classic problem of Instruction Selection, we design a theorem solver guided exhausted search algorithm to select functions for complex tensor computations. Preliminary results with some micro-benchmarks and a real model show that our approach can outperform both Tensorflow and Tensor Comprehensions at run time.

Best Poster Finalist: no

**Poster 83: ETL: Elastic Training Layer for Deep Learning**

*Lei Xie (Tsinghua University, China), Jidong Zhai (Tsinghua University, China)*

Mainstream deep learning frameworks are commonly implemented by invoking underlying high performance tensor libraries on various architectures. However, as these libraries provide increasingly complex semantics including operator fusions, in-place operations, and various memory layouts, the gap between mathematical deep learning models and the underlying libraries becomes larger. In this paper, inspired by the classic problem of Instruction Selection, we design a theorem solver guided exhausted search algorithm to select functions for complex tensor computations. Preliminary results with some micro-benchmarks and a real model show that our approach can outperform both Tensorflow and Tensor Comprehensions at run time.
Due to the rising of deep learning, clusters for deep learning training are widely deployed in production. However, static task configuration and resource fragmentation problems in existing clusters result in low efficiency and poor quality of service. We propose ETL, an elastic training layer for deep learning, to help address them once for all. ETL adopts many novel mechanisms, such as lightweight and configurable report primitive and asynchronous, parallel and IO-free state replication, to achieve both high elasticity and efficiency. The evaluation demonstrates the low overhead and high efficiency of these mechanisms and reveals the advantages of elastic deep learning supported by ETL.

Best Poster Finalist: no

**Poster 130: Deep Learning-Based Feature-Aware Data Modeling for Complex Physics Simulations**

Qun Liu (Louisiana State University), Subhashis Hazarika (Ohio State University), John M. Patchett (Los Alamos National Laboratory), James P. Ahrens (Los Alamos National Laboratory), Ayan Biswas (Los Alamos National Laboratory)

Data modeling and reduction for in situ is important. Feature-driven methods for in situ data analysis and reduction are a priority for future exascale machines as there are currently very few such methods. We investigate a deep-learning-based workflow that targets in situ data processing using autoencoders. We employ integrated skip connections to obtain higher performance compared to the existing autoencoders. Our experiments demonstrate the initial success of the proposed framework and create optimism for the in situ use case.

Best Poster Finalist: no

**Poster 125: Physics Informed Generative Adversarial Networks for Virtual Mechanical Testing**

Julian Cuevas (NASA, University of Puerto Rico at Mayaguez), Patrick Leser (NASA), James Warner (NASA), Geoffrey Bomarito (NASA), William Leser (NASA)

Physics-informed generative adversarial networks (PI-GANs) are used to learn the underlying probability distributions of spatially-varying material properties (e.g., microstructure variability in a polycrystalline material). While standard GANs rely solely on data for training, PI-GANs encode physics in the form of stochastic differential equations using automatic differentiation. The goal here is to show that experimental data from a limited number of material tests can be used with PI-GANs to enable unlimited virtual testing for aerospace applications. Preliminary results using synthetically generated data are provided to demonstrate the proposed framework. Deep learning and automatic
differentiation capabilities in Tensorflow were implemented on Nvidia Tesla V100 GPUs.

Best Poster Finalist: no

**Poster 141: ExaGeoStatR: Harnessing HPC Capabilities for Large Scale Geospatial Modeling Using R**

Sameh Abdulah (King Abdullah University of Science and Technology (KAUST)), Yuxiao Li (King Abdullah University of Science and Technology (KAUST)), Jian Cao (King Abdullah University of Science and Technology (KAUST)), Hatem Ltaief (King Abdullah University of Science and Technology (KAUST)), David Keyes (King Abdullah University of Science and Technology (KAUST)), Marc Genton (King Abdullah University of Science and Technology (KAUST)), Ying Sun (King Abdullah University of Science and Technology (KAUST))

Large-scale simulations and parallel computing techniques are becoming essential in Gaussian process calculations to lessen the complexity of geostatistics applications. The log-likelihood function is used in such applications to evaluate the model associated with a given set of measurements in existing n geographic locations. The evaluation of such a function requires $O(n^2)$ memory and $O(n^3)$ computation, which is infeasible for large datasets with existing software tools.

We present ExaGeoStatR, a package for large-scale geostatistics in R that computes the log-likelihood function on shared and distributed-memory, possibly equipped with GPU, using advanced linear algebra techniques. The package provides a high-level abstraction of the underlying architecture while enhancing the R developers' productivity. We demonstrate ExaGeoStatR package by illustrating its implementation details, analyzing its performance on various parallel architectures, and assessing its accuracy using synthetic datasets and a sea surface temperature dataset. The performance evaluation involves spatial datasets with up to 250K observations.

Best Poster Finalist: no

**Poster 48: Runtime System for GPU-Based Hierarchical LU Factorization**

Qianxiang Ma (Tokyo Institute of Technology), Rio Yokota (Tokyo Institute of Technology, Global Scientific Information and Computing Center; Tokyo Institute of Technology)

Hierarchical low-rank approximation can reduce both the storage and computation costs of dense matrices, but its implementation is challenging. In this research, we tackle one of the most difficult problems of GPU parallelization of the factorization of these hierarchical matrices. To this end, we are developing a new runtime system for GPUs that can schedule all tasks into one GPU kernel. Other
existing runtime systems, like cuGraph and Standford Legion, can only manage streams and kernel-level parallelism. Even without too much tuning, we achieved 4x better performance in H-LU factorization with a single GPU when comparing with a well-tuned CPU-based hierarchical matrix library, HLIBpro, on moderately sized matrices. Additionally, we have significantly less runtime overheads exposed when processing smaller matrices.

Best Poster Finalist: no

**Poster 145: Improving Data Compression with Deep Predictive Neural Network for Time Evolitional Data**
Rupak Roy (Florida State University), Kento Sato (RIKEN Center for Computational Science (R-CCS)), Jian Guo (RIKEN Center for Computational Science (R-CCS)), Jens Domke (RIKEN Center for Computational Science (R-CCS)), Weikuan Yu (Florida State University), Takaki Hatsui (RIKEN SPring-8 Center), Yasumasa Joti (Japan Synchrotron Radiation Research Institute)

Scientific applications/simulations periodically generate huge intermediate data. Storing or transferring such a large scale of data is critical. Fast I/O is important for making this process faster. One of the approaches to achieve fast I/O is data compression. Our goal is to achieve a delta technique that can improve the performance of existing data compression algorithms for time evolional intermediate data.

In our approach, we compute the delta values from original data and data predicted by the deep predictive neural network. We pass these delta values through three phases which are preprocessing phase, partitioned entropy coding phase, and density-based spatial delta encoding phase.

In our poster, we present how our predictive delta technique can leverage the time evolional data to produce highly concentrated small values. We show the improvement in compression ratio when our technique, combined with existing compression algorithms, are applied on the intermediate data for different datasets.

Best Poster Finalist: no

**Poster 144: Optimizing Asynchronous Multi-Level Checkpoint/Restart Configurations with Machine Learning**
Tonmoy Dey (Florida State University), Kento Sato (RIKEN Center for Computational Science (R-CCS)), Jian Guo (RIKEN Center for Computational Science (R-CCS)), Bogdan Nicolae (Argonne National Laboratory), Jens Domke (RIKEN Center for Computational Science (R-CCS)), Weikuan Yu
With the emergence of fast local storage, multi-level checkpointing (MLC) has become a common approach for efficient checkpointing. To utilize MLC efficiently, it is important to determine the optimal configuration for the checkpoint/restart (CR). There are mainly two approaches for determining the optimal configuration for CR, namely modeling and simulation approach. However, with MLC, CR becomes more complicated making the modeling approach inaccurate and the simulation approach though accurate, very slow. In this poster, we focus on optimizing the performance of CR by predicting the optimized checkpoint count and interval. This was achieved by combining the simulation approach with machine learning and neural network to leverage its accuracy without spending time on simulating different CR parameters. We demonstrate that our models can predict the optimized parameter values with minimal error when compared to the simulation approach.

Best Poster Finalist: no

Poster 132: Optimizing Performance at Runtime Using Binary Rewriting
Alexis Engelke (Technical University Munich), David Hildenbrand (Technical University Munich), Martin Schulz (Technical University Munich)

In addition to scalability, performance of sequential code in applications is an important factor in HPC. Typically, programs are compiled once, at which time optimizations are applied, and are then run several times. However, not all information relevant for performance optimizations are available at compile-time, restricting optimization possibilities. The generation of specialized code at runtime allows for further optimizations. Performing such specialization on binary code allows for initial code to be generated at compile-time with only the relevant parts being rewritten at runtime, reducing the optimization overhead. For targeted optimizations and effective use of known runtime information, the rewriting process needs to be guided by the application itself, exploiting information only known to the developer.

We describe three approaches for self-guided binary rewriting explicitly guided by the running application and evaluate the performance of the optimized code as well as the performance of the rewriting process itself.

Best Poster Finalist: no

Poster 53: Unstructured Mesh Technologies for Fusion Simulations
Multiple unstructured mesh technologies are needed to define and execute plasma physics simulations. The domains of interest combine model features defined from physical fields within 3D CAD of the tokamak vessel with an antenna assembly, and 2D cross sections of the tokamak vessel. Mesh generation technologies must satisfy these geometric constraints and additional constraints imposed by the numerical models. Likewise, fusion simulations over these domains study a range of timescales and physical phenomena within a tokamak.

XGCm studies the development of plasma turbulence in the reactor vessel, GITRm studies impurity transport, and PetraM simulations model RF wave propagation in scrape off layer plasmas. GITRm and XGCm developments are using the PUMIpic infrastructure to manage the storage and access of non-uniform particle distributions in unstructured meshes on GPUs. PetraM combines PUMI adaptive unstructured mesh control with MFEM using CAD models and meshes defined with Simmetrix tools.

Best Poster Finalist: no

**Poster 99: Eithne: A Framework for Benchmarking Micro-Core Accelerators**

*Maurice C. Jamieson (Edinburgh Parallel Computing Centre, University of Edinburgh), Nick Brown (Edinburgh Parallel Computing Centre, University of Edinburgh)*

Running existing HPC benchmarks as-is on micro-core architectures is at best difficult and most often impossible as they have a number of architectural features that makes them significantly different from traditional CPUs: tiny amounts on-chip RAM (c. 32KB), low-level knowledge specific to each device (including the host/device communications interface), limited communications bandwidth and complex or no device debugging environment. In order to compare and contrast different the micro-core architectures, a benchmark framework is required to abstract much of this complexity.

The modular Eithne framework supports the comparison of a number of micro-core architectures. The framework separates the actual benchmark from the details of how this is executed on the different technologies. The framework was evaluated by running the LINPACK benchmark on the Adapteva Epiphany, PicoRV32 and VectorBlox Orca RISC-V soft-cores, NXP RV32M1, ARM Cortex-A9, and
Xilinx MicroBlaze soft-core, and comparing resulting performance and power consumption.

Best Poster Finalist: no

**Poster 133: Portable Resilience with Kokkos**
Jeffery Miles (Sandia National Laboratories), Nicolas Morales (Sandia National Laboratories), Carson Mould (Sandia National Laboratories), Keita Teranishi (Sandia National Laboratories)

The Kokkos ecosystem is a programming environment that provides performance and portability to many scientific applications that run on DOE supercomputers as well as other smaller scale systems. Leveraging software abstraction concepts within Kokkos, software resilience for end user code is made portable with abstractions and concepts while implementing the most efficient resilience algorithms internally. This addition enables an application to manage hardware failures reducing the cost of interruption without drastically increasing the software maintenance cost. Two main resilience methodologies have been added to the Kokkos ecosystem to validate the resilience abstractions: 1. Checkpointing includes an automatic mode supporting other checkpointing libraries and a manual mode which leverages the data abstraction and memory space concepts. 2. The redundant execution model anticipates failures by replicating data and execution paths. The design and implementation of these additions are illustrated, and appropriate examples are included to demonstrate the simplicity of use.

Best Poster Finalist: no

**Poster 79: The HPC PowerStack: A Community-Wide Collaboration Toward an Energy Efficient Software Stack**
Siddhartha Jana (Energy Efficient HPC Working Group, Intel Corporation), Stephanie Brink (Lawrence Livermore National Laboratory), Christopher Cantalupo (Intel Corporation), Jonathan Eastep (Intel Corporation), Masaaki Kondo (RIKEN Advanced Institute for Computational Science (AICS), University of Tokyo), Matthias Maiterth (Intel Corporation), Aniruddha Marathe (Lawrence Livermore National Laboratory), Tapasya Patki (Lawrence Livermore National Laboratory), Barry Rountree (Lawrence Livermore National Laboratory), Ryuichi Sakamoto (University of Tokyo), Martin Schulz (Technical University Munich, Leibniz Supercomputing Centre), Carsten Trinitis (Technical University Munich), Josef Weidendorfer (Technical University Munich, Leibniz Supercomputing Centre)

This poster highlights an ongoing community-wide effort among vendors, labs, and academia, to incorporate power-awareness within system-stacks in upcoming exascale machines. HPC PowerStack is the first-and-only community-driven vendor-neutral effort to identify what power
optimization software actors are key within the modern-day stack; discuss their interoperability, and work toward gluing together existing open source projects to engineer cost-effective, but cohesive, portable implementations.

This poster disseminates key insights acquired in the project, provides prototyping status updates, highlights open questions, and solicits participation addressing the imminent exascale power challenge.

Best Poster Finalist: no

**Poster 87: Parallelizing Simulations of Large Quantum Circuits**  
*Michael A. Perlin (University of Colorado, National Institute of Standards and Technology (NIST)), Teague Tomesh (Princeton University), Bradley Pearlman (University of Colorado, National Institute of Standards and Technology (NIST)), Wei Tang (Princeton University), Yuri Alexeev (Argonne National Laboratory), Martin Suchara (Argonne National Laboratory)*

We present a parallelization scheme for classical simulations of quantum circuits. Our scheme is based on a recent method to "cut" large quantum circuits into smaller sub-circuits that can be simulated independently, and whose simulation results can in turn be re-combined to infer the output of the original circuit. The exponentially smaller classical computing resources needed to simulate smaller circuits are counterbalanced by exponential overhead in terms of classical post-processing costs. We discuss how this overhead can be massively parallelized to reduce classical computing costs.

Best Poster Finalist: no

**Poster 120: ILP-Based Scheduling for Linear-Tape Model Trapped-Ion Quantum Computers**  
*Xin-Chuan Wu (University of Chicago), Yongshan Ding (University of Chicago), Yunong Shi (University of Chicago), Yuri Alexeev (Argonne National Laboratory), Hal Finkel (Argonne National Laboratory), Kibaek Kim (Argonne National Laboratory), Frederic T. Chong (University of Chicago)*

Quantum computing (QC) is emerging as a potential post-Moore high-performance computing (HPC) technology. Trapped-ion quantum bits (qubits) are among the most leading technologies to reach scalable quantum computers that would solve certain problems beyond the capabilities of even the largest classical supercomputers. In trapped-ion QC, qubits can physically move on the ion trap. The state-of-the-art architecture, linear-tape model, only requires a few laser beams to interact with the entire qubits by physically moving the interacting ions to the execution zone. Since the laser beams
are limited resources, the ion chain movement and quantum gate scheduling are critical for the circuit latency. To harness the emerging architecture, we present our mathematical model for scheduling the qubit movements and quantum gates in order to minimize the circuit latency. In our experiment, our scheduling reduces 29.47% circuit latency on average. The results suggest classical HPC would further improve the quantum circuit optimization.

Best Poster Finalist: no

**Poster 55: MPI+OpenMP Parallelization of DFT Method in GAMESS**

Vladimir Mironov (Lomonosov Moscow State University), Yuri Alexeev (Argonne National Laboratory), Dmitry Fedorov (National Institute of Advanced Industrial Science and Technology (AIST))

In this work, the Density Functional Theory (DFT) method is parallelized with MPI-OpenMP in the quantum chemistry package GAMESS. It has been implemented in both regular and Fragment Molecular Orbital (FMO) based DFT codes. The scalability of the FMO-DFT code was demonstrated on Cray XC40 Theta supercomputer. We demonstrated excellent scalability of the code up 2,048 Intel Xeon Phi nodes (131,072 cores). Moreover, the developed DFT code is about twice as fast as the original code because of our new grid integration algorithm.

Best Poster Finalist: no

**Poster 71: AI-Solver: Uncertainty in Prediction and Error Estimation for AI in Engineering**

Ahmed Al-Jarro (Fujitsu Laboratories of Europe Ltd), Loic Beheshti (Fujitsu Laboratories of Europe Ltd), Serban Georgescu (Fujitsu Laboratories of Europe Ltd), Koichi Shirahata (Fujitsu Laboratories Ltd), Yasumoto Tomita (Fujitsu Laboratories Ltd), Nakashima Kouta (Fujitsu Laboratories Ltd)

The AI-Solver is a deep learning platform that learns from simulation data to extract general behavior based on physical parameters. The AI-Solver can handle a wide variety of classes of problems including those commonly identified in FEA, CFD and CEM, to name a few, with speedups of up to 250,000X and extremely low error rate of 2-3%. In this work, we build on this recent effort. We first integrate uncertainty quantification, via exploiting the approximation of Bayesian Deep Learning. Second, we develop bespoke error estimation mechanisms capable of processing this uncertainty to provide instant feedback on the confidence in predictions without relying on the availability of ground truth data. To our knowledge, the ability to estimate the discrepancy in predictions without labels is a first in the field of AI for Engineering.

Best Poster Finalist: no
Numerical plasma physics models such as the particle-in-cell XGC code are important tools to understand phenomena encountered in experimental fusion devices. Adequately resolved simulations are computationally expensive, so optimization is essential. To address the need for consistent high performance by cutting-edge scientific software applications, frameworks such as Kokkos have been developed to enable portability as new architectures require hardware-specific coding implementation for best performance. Cabana, a recent extension to Kokkos developed with the ECP-CoPA project, is a library of common kernels and operations typically necessary for particle-based codes. The Kokkos/Cabana framework enables intuitive construction of particle-based codes, while maintaining portability between architectures. Here, we summarize the adoption by XGC of the execution and data layout patterns offered by this framework. We demonstrate a method for Fortran codes to adopt Kokkos and show that it can provide a single, portable code base that performs well on both GPUs and multicore machines.

Best Poster Finalist: no

In HPC systems, expectations for storage-class memory (SCM) are increasing in large-scale in-memory processing. While SCM can deliver higher capacity and lower standby power than DRAM, it is slower and the dynamic power is higher. Therefore, in order to realize high-speed, low-power and scalable main memory, it is necessary to build an SCM/DRAM unified memory, and dynamically optimize data placement between the two memories according to the memory access pattern.

In this poster, we describe a new hybrid access type virtual memory method using TLB-extended unified memory management unit which enables collecting and extracting fine-grained memory access locality characteristics. We show that with the proposed method, Hybrid Access control, which is a memory hierarchy control that selectively uses Direct Access to bus attached byte-
addressable SCM and low power Aggressive Paging using small DRAM as cache, can be made more accurate, and the efficiency of memory access can be significantly improved.

Best Poster Finalist: no

**Holistic Measurement Driven System Assessment**
Saurabh Jha (University of Illinois), Mike Showerman (National Center for Supercomputing Applications (NCSA), University of Illinois), Aaron Saxton (National Center for Supercomputing Applications (NCSA), University of Illinois), Jeremy Enos (National Center for Supercomputing Applications (NCSA), University of Illinois), Greg Bauer (National Center for Supercomputing Applications (NCSA), University of Illinois), Zbigniew Kalbarczyk (University of Illinois), Ann Gentile (Sandia National Laboratories), Jim Brandt (Sandia National Laboratories), Ravi Iyer (University of Illinois), William T. Kramer (University of Illinois, National Center for Supercomputing Applications (NCSA))

HPC users deploy a suite of monitors to observe patterns of failures and performance anomalies to improve operational efficiency, achieve higher application performance and inform the design of future systems. However, the promises and the potential of monitoring data have largely been not realized due to various challenges such as inadequacy in monitoring, limited availability of data, lack of methods for fusing monitoring data at time-scales necessary for enabling human-in-the-loop or machine-in-the-loop feedback. To address above challenges, in this work we developed a monitoring fabric Holistic Measurement Driven System Assessment (HMDSA) for large-scale HPC facilities, independent of major component vendor, and within budget constraints of money, space, and power. We accomplish this through development and deployment of scalable, platform-independent, open-source tools and techniques for monitoring, coupled with statistical and machine-learning based runtime analysis and feedback, which enables highly efficient HPC system operation and usage and also informs future system improvements.

Best Poster Finalist: no

**Poster 139: Model Identification of Pressure Drop in Membrane Channels with Multilayer Artificial Neural Networks**
Jiang-hang Gu (Sun Yat-sen University, Zhuhai, School of Chemical Engineering and Technology), Jiu Luo (Sun Yat-sen University, Guangzhou, School of Materials Science and Engineering), Ming-heng Li (California State Polytechnic University, Pomona), Yi Heng (Sun Yat-sen University, Guangzhou, School of Data and Computer Science; Sun Yat-sen University, Guangzhou, China)
This poster presents the work of identifying a data-driven model of pressure drop in spacer-filled reverse osmosis membrane channels and conducting CFD simulations. The established model correlates the pressure drop with a wide range of design objectives, which enables a quantitative description of the geometric structures and operation conditions for improvement. This way, it aims at optimizing the spacer geometry with minimal effort. Furthermore, a high-performance computing strategy is employed to tackle the resulted intractable computational task in the identification procedure and CFD simulations.

Best Poster Finalist: no

**Poster 61: Fast 3D Diffeomorphic Image Registration on GPUs**
Malte Brunn (University of Stuttgart), Naveen Himthani (University of Texas), George Biros (University of Texas), Miriam Mehl (University of Stuttgart), Andreas Mang (University of Houston)

3D image registration is one of the most fundamental and computationally expensive operations in medical image analysis. Here, we present a mixed-precision, Gauss-Newton-Krylov solver for diffeomorphic registration. Our work extends the publicly available CLAIRE library to GPU architectures. Despite the importance of image registration, only a few implementations of large deformation diffeomorphic registration packages support GPUs. Our contributions are new algorithms and dedicated computational kernels to significantly reduce the runtime of the main computational kernels in CLAIRE: derivatives and interpolation. We deploy (i) highly-optimized, mixed-precision GPU-kernels for the evaluation of scattered-data interpolation, (ii) replace FFT-based first-order derivatives with optimized 8th-order finite differences, and (iii) compare with state-of-the-art CPU and GPU implementations. As a highlight, we demonstrate that we can register 256^3 clinical images in less than 6 seconds on a single NVIDIA Tesla V100. This amounts to over 20x speed-up over CLAIRE and over 30x speed-up over existing GPU implementations.

Best Poster Finalist: no

**Poster 138: Across-Stack Profiling and Characterization of State-of-the-Art Machine Learning Models on GPUs**
Cheng Li (University of Illinois), Abdul Dakkak (University of Illinois), Wei Wei (Alibaba Inc), Jinjun Xiong (IBM Research), Lingjie Xu (Alibaba Inc), Wei Zhang (Alibaba Inc), Wen-mei Hwu (University of Illinois)

The past few years have seen a surge of using Machine Learning (ML) and Deep Learning (DL) algorithms for traditional HPC tasks such as feature detection, numerical analysis, and graph
analytics. While ML and DL enable solving HPC tasks, their adoption has been hampered due to the lack of understanding of how they utilize systems. Optimizing these algorithms requires characterizing their performance across the hardware/software (HW/SW) stack, but the lack of simple tools to automate the process and the reliance on researchers to perform manual characterization is a bottleneck. To alleviate this, we propose an across-stack profiling scheme and integrate it within MLModelScope—a hardware and software agnostic tool for evaluating and benchmarking ML/DL at scale. We demonstrate MLModelScope’s ability to characterize state-of-art ML/DL models and give insights that are only possible obtained by performing across-stack profiling.

Poster 60: Massively Parallel Large-Scale Multi-Model Simulation of Tumor Development
Marco Berghoff (Karlsruhe Institute of Technology), Jakob Rosenbauer (Forschungszentrum Juelich), Alexander Schug (Forschungszentrum Juelich)

The temporal and spatial resolution in the microscopy of tissues has increased significantly within the last years, yielding new insights into the dynamics of tissue development and the role of the single-cell within it. A thorough theoretical description of the connection of single-cell processes to macroscopic tissue reorganizations is still lacking. Especially in tumor development, single cells play a crucial role in advance of tumor properties.

We developed a simulation framework that can model tissue development up to the centimeter scale with micrometer resolution of single cells. Through a full parallelization, it enables the efficient use of HPC systems, therefore enabling detailed simulations on a large scale. We developed a generalized tumor model that respects adhesion driven cell migration, cell-to-cell signaling, and mutation-driven tumor heterogeneity. We scan the response of the tumor development depending on division inhibiting substances such as cytostatic agents.

Poster 114: Optimizing Recommendation System Inference Performance Based on GPU
Xiaowei Shen (Alibaba Inc), Junrui Zhou (Alibaba Inc), Kan Liu (Alibaba Inc), Lingling Jin (Alibaba Inc), Pengfei Fan (Alibaba Inc), Wei Zhang (Alibaba Inc), Jun Yang (University of Pittsburgh)

Neural network-based recommendation models have been widely applied on tracking personalization and recommendation tasks at large Internet companies such as e-commerce companies and social media companies. Alibaba recommendation system deploys WDL (wide and deep learning) models
for product recommendation tasks. The WDL model consists of two main parts: embedding lookup and neural network-based feature ranking model that ranks different products for different users. As more and more products and users the model need to rank, the feature length and batch size of the models are increased. The computation of models is also increased so that traditional model inference implementation on CPU cannot meet the requirement of QPS (query per second) and latency of recommendation tasks. In this poster, we develop a GPU based system to speedup recommendation system inference performance. By model quantization and graph transformation, we can achieve 3.9x performance speedup when compared with a baseline GPU implementation.

Best Poster Finalist: no

Poster 59: Accelerating BFS and SSSP on a NUMA Machine for the Graph500 Challenge
Tanuj K. Aasawat (RIKEN), Kazuki Yoshizoe (RIKEN), Tahsin Reza (University of British Columbia), Matei Ripeanu (University of British Columbia)

The NUMA architecture is the design choice for modern multi-CPU shared memory systems. In many ways, a NUMA system resembles a shared-nothing distributed system: memory accesses to remote NUMA domains are more expensive than local accesses.

In this work, we explore how improved data locality and reduced expensive remote communication can be achieved by exploiting "distributed" shared-memory of NUMA machines to develop shared-memory graph processing solutions optimized for NUMA systems. We introduce a novel hybrid design for memory accesses that handles the burst mode in traversal based algorithms, like BFS and SSSP, and reduces the number of remote accesses and updates. We demonstrate that our designs offer up to 84% speedup over our NUMA-oblivious framework Totem and 2.86x over shared-nothing distributed design, for BFS and SSSP algorithms.

Best Poster Finalist: no

Poster 47: Decomposition Algorithms for Scalable Quantum Annealing
Elijah Pelofske (Los Alamos National Laboratory), Georg Hahn (Harvard University), Hristo Djidjev (Los Alamos National Laboratory)

Commercial adiabatic quantum annealers such as D-Wave 2000Q have the potential to solve NP-complete optimization problems efficiently. One of the primary constraints of such devices is the limited number and connectivity of their qubits. This research presents two exact decomposition methods (for the Maximum Clique and the Minimum Vertex Cover problem) that allow us to solve
problems of arbitrarily large sizes by splitting them up recursively into a series of arbitrarily small subproblems. Those subproblems are then solved exactly or approximately using a quantum annealer. Whereas some previous approaches are based on heuristics that do not guarantee optimality of their solutions, our decomposition algorithms have the property that the optimal solution of the input problem can be reconstructed given all generated subproblems are solved optimally as well. We investigate various heuristic and exact bounds as well as reduction methods that help to increase the scalability of our approaches.

Best Poster Finalist: no

Poster 92: Nanoporous Flow Simulations on the Summit Supercomputer
Yidong Xia (Idaho National Laboratory), Lixiang Luo (IBM - TJ Watson Research Center), Ansel Blumers (Brown University), Joshua Kane (Idaho National Laboratory), Jan Goral (University of Utah), Yu-Hang Tang (Lawrence Berkeley National Laboratory), Zhen Li (Clemson University, Brown University), Hai Huang (Idaho National Laboratory), Milind Deo (University of Utah)

Fluid flow behaviors in nanoporous materials are distinct from those following the continuum physics. Numerical simulations can be a complement to laboratory experiments. This work presents a dissipative particle dynamics (DPD) package for GPU-accelerated mesoscale flow simulations in nanoporous materials. In an ideal benchmark that minimizes load imbalance, the package delivered nearly perfect strong- and weak-scaling (with up to 4 billion DPD particles) on up to 1,536 V100 GPUs on Oak Ridge National Laboratory’s Summit supercomputer. More remarkably, in a benchmark to measure its usefulness with realistic nanopores in SBA-15 silica, the package exhibited more than 20x speedup over its LAMMPS-based CPU counterpart with the same number nodes (e.g., 384 V100 GPUs vs. 2,688 POWER9 cores). It is worth highlighting that the NVLink2 Host-to-Device interconnects kept the cost of CPU-GPU memory copy as low as only 10% of GPU activity time per rank: 4 times less than their PCIe counterparts.

Best Poster Finalist: yes

Poster 90: You Have to Break It to Make It: How On-Demand, Ephemeral Public Cloud Projects with Alces Flight Compute Resulted in the Open-Source OpenFlightHPC Project
Cristin Merritt (Alces Flight Limited; Alces Software Ltd, UK), Wil Mayers (Alces Flight Limited), Stu Franks (Alces Flight Limited)

Over three years ago the Alces Flight team made a decision to explore on-demand public cloud consumption for High Performance Computing (HPC). Our premise was simple, create a fully-
featured, scalable HPC environment for research and scientific computing and provide it as a free subscription service to researchers in order to study how they would approach and consume this then new technology. This tool, Alces Flight Compute, would set out to chart how far away from the traditional bare-metal platforms our subscribers were willing to go. What we didn’t expect was that to get to their destination, our users would proceed to take our tool apart. This deconstruction has resulted in a new approach to HPC environment creation (the open-source OpenFlightHPC project), helped us better understand cloud adoption strategies, and handed over a set of guidelines to help those looking to bring public cloud into their HPC solution.

Best Poster Finalist: no

**Poster 126: Enforcing Crash Consistency of Scientific Applications in Non-Volatile Main Memory Systems**

Tyler Coy (Washington State University, Vancouver), Xuechen Zhang (Washington State University, Vancouver)

This poster presents a compiler-assistant technique, NVPath, to automatically generates NVMM-aware persistent data structures which provide the same level of guarantee of crash consistency compared to the baseline code. Compiler-assistant code annotation and transformation is general and can be applied to applications using various data structures. Our experimental results with real-world scientific applications show that the performance of the annotated programs is commensurate with the version using the manual code transformation on the Titan supercomputer.

Best Poster Finalist: no

**Poster 137: Warwick Data Store: A HPC Library for Flexible Data Storage in Multi-Physics Applications**

Richard O. Kirk (University of Warwick), Timothy R. Law (Atomic Weapons Establishment (AWE), UK), Satheesh Maheswaran (Atomic Weapons Establishment (AWE), UK), Stephen A. Jarvis (University of Warwick)

With the increasing complexity of memory architectures and multi-physics applications, developing data structures that are performant, portable, scalable, and support developer productivity, is difficult. In order to manage these complexities and allow rapid prototyping of different approaches we are building a lightweight and extensible C++ template library called the Warwick Data Store (WDS). WDS is designed to abstract details of the data structure away from the user, thus easing application development and optimisation. We show that WDS generates minimal performance overhead, via a
variety of different scientific benchmarks and proxy-applications.

Best Poster Finalist: no

**Poster 76: HPChain: An MPI-Based Blockchain Framework for High Performance Computing Systems**

Abdullah Al-Mamun (University of Nevada, Reno; Lawrence Berkeley National Laboratory), Tonglin Li (Lawrence Berkeley National Laboratory), Mohammad Sadoghi (University of California, Davis), Linhua Jiang (Fudan University, Shanghai), Haoting Shen (University of Nevada, Reno), Dongfang Zhao (University of Nevada, Reno; University of California, Davis)

Data fidelity is of prominent importance for scientific experiments and simulations. The state-of-the-art mechanism to ensure data fidelity is through data provenance. However, the provenance data itself may as well exhibit unintentional human errors and malicious data manipulation. To enable a trustworthy and reliable data fidelity service, we advocate achieving the immutability and decentralization of scientific data provenance through blockchains. Specifically, we propose HPChain, a new blockchain framework specially designed for HPC systems. HPChain employs a new consensus protocol compatible with and optimized for HPC systems. Furthermore, HPChain was implemented with MPI and integrated with an off-chain distributed provenance service to tolerate the failures caused by faulty MPI ranks. The HPChain prototype system has been deployed to 500 cores at the University of Nevada’s HPC center and demonstrated strong resilience and scalability while outperforming state-of-the-art blockchains by orders of magnitude.

Best Poster Finalist: no

**Poster 104: An Adaptive Checkpoint Model For Large-Scale HPC Systems**

Subhendu S. Behera (North Carolina State University), Lipeng Wan (Oak Ridge National Laboratory), Frank Mueller (North Carolina State University), Matthew Wolf (Oak Ridge National Laboratory), Scott Klasky (Oak Ridge National Laboratory)

Checkpoint/Restart is a widely used Fault Tolerance technique for application resilience. However, failures and the overhead of saving application state for future recovery upon failure reduces the application efficiency significantly. This work contributes a failure analysis and prediction model making decisions for checkpoint data placement, recovery, and techniques for reducing checkpoint frequency. We also demonstrate a reduction in application overhead by taking proactive measures guided by failure prediction.
Poster 123: Cloud-Native SmartX Intelligence Cluster for AI-Inspired HPC/HPDA Workloads
Jungsu Han (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science), Jun-Sik Shin (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science), JinCheol Kwon (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science), JongWon Kim (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science)

In this poster, we introduce Cloud-native SmartX Intelligence Cluster for flexibly supporting AI-inspired HPC (high performance computing) / HPDA (high performance data analytics) workloads. This work has been continuously refined from 2013 with a futuristic vision for operating 100 petascale data center. Then, we discuss issues and approaches that come with building a Cloud-native SmartX Intelligence Cluster.

Poster 86: High-Performance Custom Computing with FPGA Cluster as an Off-Loading Engine
Takaaki Miyajima (RIKEN Center for Computational Science (R-CCS)), Tomohiro Ueno (RIKEN Center for Computational Science (R-CCS)), Jens Huthmann (RIKEN Center for Computational Science (R-CCS)), Atsushi Koshiba (RIKEN Center for Computational Science (R-CCS)), Kentaro Sano (RIKEN Center for Computational Science (R-CCS)), Mitsuhisa Sato (RIKEN Center for Computational Science (R-CCS))

A heterogeneous system with Field Programmable Gate Arrays (FPGAs) is gathering attention in High-Performance Computing (HPC) area. We have been researching and developing a high-performance custom computing with FPGA cluster as an off-loading engine for an existing HPC system. In this research poster, we describe the motivation of our research and present research topics on a software bridge between the FPGA cluster and existing HPC servers, and dedicated inter-FPGA networks.

Poster 69: Optimization for Quantum Computer Simulation
Naoki Yoshioka (RIKEN Center for Computational Science (R-CCS)), Hajime Inaoka (RIKEN Center for Computational Science (R-CCS)), Nobuyasu Ito (RIKEN Center for Computational Science (R-CCS)), 
Simulator of quantum circuits is developed for massively parallel classical computers, and it is tested on the K computer in RIKEN R-CCS up to 45 qubits. Two optimization techniques are proposed in order to improve performance of the simulator. The "page method" reduces unnecessary copies in each node. It is found that this method makes approximately 17% speed-up maximum. Initial permutation of qubits is also studied how it affects performance of the simulator. It is found that a simple permutation in ascending order of the number of operations for each qubit is sufficient in the case of simulations of quantum adder circuits.

**Poster 110: Hierarchical Data Prefetching in Multi-Tiered Storage Environments**

Hariharan Devarajan (Illinois Institute of Technology), Anthony Kougkas (Illinois Institute of Technology), Xian-He Sun (Illinois Institute of Technology)

In the era of data-intensive computing, accessing data with a high-throughput and low-latency is very imperative. Data prefetching is used for hiding read latency by requesting data before it is needed to move it from a high-latency medium to a low-latency one. However, existing solutions do not consider the multi-tiered storage and also suffer from under-utilization of prefetching resources and unnecessary evictions. Additionally, existing approaches implement a client-pull model where understanding the application's I/O behavior drives prefetching decisions. Moving toward exascale, where machines run multiple applications concurrently by accessing files in a workflow, a more data-centric approach resolves challenges such as cache pollution and redundancy. In this study, we present HFetch, a truly hierarchical data prefetcher that adopts a server-push approach to data prefetching. We demonstrate the benefits of such an approach. Results show 10-35% performance gains over existing prefetchers and over 50% when compared to systems with no prefetching.

Best Poster Finalist: yes

**Poster 52: Design and Specification of Large-Scale Simulations for GPUs Using FFTX**

Anuva Kulkarni (Carnegie Mellon University), Daniele Spampinato (Carnegie Mellon University), Franz Franchetti (Carnegie Mellon University)

Large-scale scientific simulations can be ported to heterogeneous environments with GPUs using domain decomposition. However, Fast Fourier Transform (FFT) based simulations require all-to-all
communication and large memory, which is beyond the capacity of on-chip GPU memory. To overcome this, domain decomposition solutions are combined with adaptive sampling or pruning around the domain to reduce storage. Expression of such operations is a challenge in existing FFT libraries like FFTW, and thus it is difficult to get a high performance implementation of such methods. We demonstrate algorithm specification for one such simulation (Hooke’s law) using FFTX, an emerging API with a SPIRAL-based code generation back-end, and suggest future extensions useful for GPU-based scientific computing.

Best Poster Finalist: no

Poster 136: CHAMELEON: Reactive Load Balancing and Migratable Tasks for Hybrid MPI+OpenMP Applications
Jannis Klinkenberg (RWTH Aachen University), Philipp Samfä (Technical University Munich), Michael Bader (Technical University Munich), Karl Fürlinger (Ludwig Maximilian University of Munich), Christian Terboven (RWTH Aachen University), Matthias S. Müller (RWTH Aachen University)

Many HPC applications are designed based on underlying performance and execution models. These models could successfully be employed in the past for balancing load within and between compute nodes. However, the increasing complexity of modern software and hardware makes performance predictability and load balancing much more difficult. Tackling these challenges in search for a generic solution, we present a novel library for fine-granular task-based reactive load balancing in distributed memory based on MPI and OpenMP. Our concept allows creating individual migratable tasks that can be executed on any MPI rank. Migration decisions are performed at run time based on online performance or load data. Two fundamental approaches to balance load and at the same time overlap computation and communication are compared. We evaluate our concept under enforced power caps and clock frequency changes using a synthetic benchmark and demonstrate robustness against work-induced imbalances for an AMR application.

Best Poster Finalist: no

Poster 124: Porting Finite State Automata Traversal from GPU to FPGA: Exploring the Implementation Space
Marziyeh Nourian (North Carolina State University), Mostafa Eghbali Zarch (North Carolina State University), Michela Becchi (North Carolina State University)

While FPGAs are traditionally considered hard to program, recently there are efforts to allow using high-level programming models intended for multi-core CPUs and GPUs to program FPGAs. For
example, both Intel and Xilinx are now providing OpenCL-to-FPGA toolchains. However, since GPU and FPGA devices offer different parallelism models, OpenCL code optimized for GPU can prove inefficient on FPGA, in terms of both performance and hardware resource utilization.

In this poster, we explore this problem on an emerging workload: finite state automata traversal. Specifically, we explore a set of structural code changes, custom, and best-practice optimizations to retarget an OpenCL NFA engine designed for GPU to FPGA. Our evaluation, which covers traversal throughput and resource utilization, shows that our optimizations lead, on a single execution pipeline, to speedups up to 4x over an already optimized baseline that uses one of the proposed code changes to fit the original code on FPGA.

Best Poster Finalist: no

**Poster 68: Linking a Next-Gen Remap Library into a Long-Lived Production Code**

Charles R. Ferenbaugh (Los Alamos National Laboratory), Brendan K. Krueger (Los Alamos National Laboratory)

LANL's long-lived production application xRage contains a remapper capability that maps mesh fields from its native AMR mesh to the GEM mesh format used by some third-party libraries. The current remapper was implemented in a short timeframe and is challenging to maintain. Meanwhile, our next-generation code project has developed a modern remapping library Portage, and the xRage team wanted to link in Portage as an alternate mapper option. But the two codes are very different from each other, and connecting the two required us to deal with a number of challenges. This poster describes the codes, the challenges we worked through, current status, and some initial performance statistics.

Best Poster Finalist: no

**Poster 131: Efficiency of Algorithmic Structures**

Julian Miller (RWTH Aachen University), Lukas Trümper (RWTH Aachen University), Christian Terboven (RWTH Aachen University), Matthias S. Müller (RWTH Aachen University)

The implementation of high-performance parallel software is challenging and raises issues not seen in serial programs before. It requires a strategy of parallel execution which preserves correctness but maximizes scalability. Efficiently deriving well-scaling solutions remains an unsolved problem especially with the quickly-evolving hardware landscape of high-performance computing (HPC).
This work proposes a framework for classifying the efficiency of parallel programs. It bases on a strict separation between the algorithmic structure of a program and its executed functions. By decomposing parallel programs into a hierarchical structure of parallel patterns, a high-level abstraction is provided which leads to equivalence classes over parallel programs. Each equivalence class possesses efficiency properties, mainly communication and synchronization, dataflow and architecture efficiency. This classification allows for wide application areas and a workflow for structural optimization of parallel algorithms is proposed.

Best Poster Finalist: no

**Poster 98: INSPECT Intranode Stencil Performance Evaluation Collection**

Julian Hammer (University of Erlangen-Nuremberg), Julian Hornich (University of Erlangen-Nuremberg), Georg Hager (University of Erlangen-Nuremberg), Thomas Gruber (University of Erlangen-Nuremberg), Gerhard Wellein (University of Erlangen-Nuremberg)

Modeling and presenting performance data---even for simple kernels such as stencils---is not trivial. We therefore present an overview on how to interpret and what to learn from an INSPECT report, as well as highlighting best practices for performance data reporting.

INSPECT is the "Intranode Stencil Performance Evaluation Collection", which compiles performance benchmarks and reports of various stencil and streaming kernels on a variety of architectures. The goal is to aid performance-aware developers with reference material and a methodology to analyze their own codes.

INSPECT set out to cover these topics and compile a summary of all necessary information to allow reproduction of the performance results, their interpretation and discussion.

Best Poster Finalist: no

**Poster 97: Optimizing Multigrid Poisson Solver of Cartesian CFD Code CUBE**

Kazuto Ando (RIKEN Center for Computational Science (R-CCS)), Rahul Bale (RIKEN), Keiji Onishi (RIKEN Center for Computational Science (R-CCS)), Kiyoshi Kumahata (RIKEN Center for Computational Science (R-CCS)), Kazuo Minami (RIKEN Center for Computational Science (R-CCS)), Makoto Tsubokura (Kobe University, RIKEN Center for Computational Science (R-CCS))

We demonstrate an optimization of multigrid Poisson solver of Cartesian CFD code "CUBE (Complex Unified Building cubE method)". CUBE is a simulation framework for complex industrial flow problem,
such as aerodynamics of vehicles, based on hierarchical Cartesian mesh. In incompressible CFD simulation, solving pressure Poisson equation is the most time-consuming part. In this study, we use a cavity flow simulation as a benchmark problem. With this problem, multigrid Poisson solver dominates 91% of execution time of the time-step loop. Specifically, we evaluate the performance of Gauss-Seidel loop as a computational kernel based on “Byte per Flop” approach. With optimization of the kernel, we achieved 9.8x speedup and peak floating point performance ratio increased from 0.4% to 4.0%. We also measured parallel performance up to 8,192 nodes (65,536 cores) on the K computer. With optimization of the parallel performance, we achieved 2.9x–3.9x sustainable speedup in the time-step loop.

Best Poster Finalist: no

Poster 85: Hybrid Computing Platform for Combinatorial Optimization with the Coherent Ising Machine
Junya Arai (Nippon Telegraph and Telephone Corporation), Yagi Satoshi (Nippon Telegraph and Telephone Corporation), Hiroyuki Uchiyama (Nippon Telegraph and Telephone Corporation), Toshimori Honjo (Nippon Telegraph and Telephone Corporation), Takahiro Inagaki (Nippon Telegraph and Telephone Corporation), Kensuke Inaba (Nippon Telegraph and Telephone Corporation), Takuya Ikuta (Nippon Telegraph and Telephone Corporation), Hiroki Takesue (Nippon Telegraph and Telephone Corporation), Keitaro Horikawa (Nippon Telegraph and Telephone Corporation)

Several institutes are operating cloud platforms that offer Web API access to Ising computers such as quantum annealing machines. Platform users can solve complex combinatorial optimization problems by using hybrid algorithms that utilize both users' conventional digital computers and remote Ising computers. However, communication via the Internet takes an order of magnitude longer time than optimization on Ising computers. This overheads seriously degrade the performance of hybrid algorithms since they involve frequent communication. In this poster, we first state issues in the design of Ising computing platforms, including communication overheads. Then, we answer the issues by introducing the computing platform for the coherent Ising machine (CIM), an Ising computer based on photonics technologies. Our platform offers efficient CIM-digital communication by allowing users to execute their program on digital computers co-located with the CIM. We have released the platform to our research collaborators in this autumn and started the evaluation.

Best Poster Finalist: no

Poster 117: A New Polymorphic Computing Architecture Based on Fine-Grained Instruction Mobility
David Hentrich (Illinois Institute of Technology), Erdal Oruklu (Illinois Institute of Technology), Jafar Saniee (Illinois Institute of Technology)

This is a summary of the base concepts behind David Hentrich’s May 2018 Ph.D. dissertation in Polymorphic Computing. Polymorphic Computing is the emerging field of changing the computer architecture around the software, rather than vice versa. The main contribution is a new polymorphic computing architecture. The key idea behind the architecture is to create an array of processors where a program’s instructions can be individually and arbitrarily assigned/mobilized to any processor, even during runtime. The key enablers of this architecture are a dataflow instruction set that is conducive to instruction migration, a microarchitectural block called an “operation cell” (“op-cell”), a processor built around the instruction set and the “op-cells”, and arrays of these processors.

Best Poster Finalist: no

Poster 67: Genie: an MPEG-G Conformant Software to Compress Genomic Data.
Brian E. Bliss (University of Illinois), Joshua M. Allen (University of Illinois), Saurabh Baheti (Mayo Clinic), Matthew A. Bockol (Mayo Clinic), Shubham Chandak (Stanford University), Jaime Delgado (Polytechnic University of Catalonia), Jan Fostier (Ghent University), Josep L. Gelpi (University of Barcelona), Steven N. Hart (Mayo Clinic), Mikel Hernaez Arrazola (University of Illinois), Matthew E. Hudson (University of Illinois), Michael T. Kalmbach (Mayo Clinic), Eric W. Klee (Mayo Clinic), Liudmila S. Mainzer (University of Illinois), Fabian Müntefering (Leibniz University), Daniel Naro (Barcelona Supercomputing Center), Idoia Ochoa-Alvarez (University of Illinois), Jörn Ostermann (Leibniz University), Tom Paridaens (Ghent University), Christian A. Ross (Mayo Clinic), Jan Voges (Leibniz University), Eric D. Wieben (Mayo Clinic), Mingyu Yang (University of Illinois), Tsachy Weissman (Stanford University), Mathieu Wiepert (Mayo Clinic)

Precision medicine has unprecedented potential for accurate diagnosis and effective treatment. It is supported by an explosion of genomic data, which continues to accumulate at accelerated pace. Yet storage and analysis of petascale genomic data is expensive, and that cost will ultimately be borne by the patients and citizens. The Moving Picture Experts Group (MPEG) has developed MPEG-G, a new open standard to compress, store, transmit and process genomic sequencing data that provides an evolved and superior alternative to currently used genomic file formats. Our poster will showcase software package GENIE, the first open source implementation of an encoder-decoder pair that is compliant with the MPEG-G specifications and delivers all its benefits: efficient compression, selective access, transport and analysis, guarantee of long-term support, and embedded mechanisms for annotation and encryption of compressed information. GENIE will create a step-change in medical genomics by reducing the cost of data storage and analysis.
**Poster 82: A View from the Facility Operations Side on the Water/Air Cooling System of the K Computer**

Jorji Nonaka (RIKEN Center for Computational Science (R-CCS)), Keiji Yamamoto (RIKEN Center for Computational Science (R-CCS)), Akiyoshi Kuroda (RIKEN Center for Computational Science (R-CCS)), Toshiyuki Tsukamoto (RIKEN Center for Computational Science (R-CCS)), Kazuki Koiso (Kobe University, RIKEN Center for Computational Science (R-CCS)), Naohisa Sakamoto (Kobe University, RIKEN Center for Computational Science (R-CCS))

The Operations and Computer Technologies Division at the RIKEN R-CCS is responsible for the operations of the entire K computer facility, which includes the auxiliary subsystems such as the power supply and water/air cooling systems. It is worth noting that part of these subsystems will be reused in the next supercomputer (Fugaku), thus a better understanding of the operational behavior as well as the potential impacts especially on the hardware failure and energy consumption would be greatly beneficial. In this poster, we will present some preliminary impressions of the impact of the water/air cooling system on the K computer system, focusing on the potential benefits of the use of low water/air temperature respectively for the CPU and DRAM memory modules produced by the cooling system. We expect that the obtained knowledge will be helpful for the decision support and/or operation planning of the next supercomputer.

**Poster 135: High-Performance Deep Learning via a Single Building Block**

Evangelos Georganas (Intel Corporation), Kunal Banerjee (Intel Corporation), Dhiraj Kalamkar (Intel Corporation), Sasikanth Avancha (Intel Corporation), Anand Venkat (Intel Corporation), Michael Anderson (Intel Corporation), Greg Henry (Intel Corporation), Hans Pabst (Intel Corporation), Alexander Heinecke (Intel Corporation)

Deep learning (DL) is one of the most prominent branches of machine learning. Due to the immense computational cost of DL workloads, industry and academia have developed DL libraries with highly-specialized kernels for each workload/architecture, leading to numerous, complex code-bases that strive for performance, yet they are hard to maintain and do not generalize. In this work, we introduce the batch-reduce-GEMM kernel and show how the most popular DL algorithms can be formulated with this kernel as basic building-block. Consequently, the DL library-development degenerates to mere (potentially automatic) tuning of loops around this sole optimized kernel. By exploiting our kernel we implement Recurrent Neural Networks, Convolution Neural Networks and Multilayer Perceptron.
training and inference primitives in just 3K lines of high-level-code. Our primitives outperform vendor-optimized libraries on multi-node CPU-Clusters. We also provide CNN kernels targeting GPUs. Finally, we demonstrate that batch-reduce-GEMM kernel within a tensor compiler yields high-performance CNN primitives.

Best Poster Finalist: no

**Poster 56: Reinforcement Learning for Quantum Approximate Optimization**

Sami Khairy (Illinois Institute of Technology), Ruslan Shaydulin (Clemson University), Lukasz Cincio (Los Alamos National Laboratory), Yuri Alexeev (Argonne National Laboratory), Prasanna Balaprakash (Argonne National Laboratory)

The Quantum Approximate Optimization Algorithm (QAOA) is one of the leading candidates for demonstrating quantum advantage. The quality of the solution obtained by QAOA depends on the performance of the classical optimization routine used to optimize the variational parameters. In this work, we propose a Reinforcement Learning (RL) based approach to drastically reduce the number of evaluations needed to find high-quality variational parameters. We train an RL agent on small 8-qubit Max-Cut problem instances on an Intel Xeon Phi supercomputer Bebop, and use (transfer) the learned optimization policy to quickly find high-quality solutions for other larger problem instances coming from different distributions and graph classes. The preliminary results show that our RL based approach is able to improve the quality of the obtained solution by up to 10% within a fixed budget of function evaluations and demonstrate learned optimization policy transferability between different graph classes and sizes.

Best Poster Finalist: no


John Shalf (Lawrence Berkeley National Laboratory), Dilip Vasudevan (Lawrence Berkeley National Laboratory), David Donofrio (Lawrence Berkeley National Laboratory), Anastasia Butko (Lawrence Berkeley National Laboratory), Andrew Chien (University of Chicago), Yuanwei Fang (University of Chicago), Arjun Rawal (University of Chicago), Chen Zou (University of Chicago), Raymond Bair (Argonne National Laboratory), Kristopher Keipert (Argonne National Laboratory), Arun Rodriguez (Sandia National Laboratories), Maya Gokhale (Lawrence Livermore National Laboratory), Scott Lloyd (Lawrence Livermore National Laboratory), Xiaochen Guo (Lehigh University), Yuan Zeng (Lehigh University)
Accelerating technology disruptions and architectural change create growing opportunities and urgency to reduce the latency in for new architectural innovations to be deployed in extreme scale systems. We are exploring new architectural features that improve memory system performance including word-wise scratchpad memory, a flexible Recode engine, hardware message queues, and the data rearrangement engine (DRE). Performance results are promising yielding as much as 20x benefit. Project 38 is a cross-agency effort undertaken by the US Department of Energy (DOE) and Department of Defense (DoD).

Best Poster Finalist: no

**Poster 128: Identifying Time Series Similarity in Large-Scale Earth System Datasets**

Payton Linton (Youngstown State University), William Melodia (Youngstown State University), Alina Lazar (Youngstown State University), Deborah Agarwal (Lawrence Berkeley National Laboratory), Ludovico Bianchi (Lawrence Berkeley National Laboratory), Devarshi Ghoshal (Lawrence Berkeley National Laboratory), Kesheng Wu (Lawrence Berkeley National Laboratory), Gilberto Pastorello (Lawrence Berkeley National Laboratory), Lavanya Ramakrishnan (Lawrence Berkeley National Laboratory)

Scientific data volumes are growing every day and instrument configurations, quality control and software updates result in changes to the data. This study focuses on developing algorithms that detect changes in time series datasets in the context of the Deduce project. We propose a combination of methods that include dimensionality reduction and clustering to evaluate similarity measuring algorithms. This methodology can be used to discover existing patterns and correlations within a dataset. The current results indicate that the Euclidean Distance metric provides the best results in terms of internal cluster validity measures for multi-variable analyses of large-scale earth system datasets. The poster will include details on our methodology, results, and future work.

Best Poster Finalist: no

**Poster 151: Three-Dimensional Characterization on Edge AI Processors with Object Detection Workloads**

Yujie Hui (Ohio State University), Jeffrey Lien (NovuMind Inc), Xiaoyi Lu (Ohio State University)

The Deep Learning inference applications are moving to the edge side, as edge-side AI platforms are cheap and energy-efficient. Different edge AI processors are diversified, since these processors are designed with different approaches. However, it is hard for customers to select an edge AI processor without an overall evaluation of these processors. We propose a three-dimensional characterization
(i.e., accuracy, latency, and energy efficiency) approach on three different kinds of edge AI processors (i.e., Edge TPU, NVIDIA Xavier, and NovuTensor). We deploy YOLOv2 and Tiny-YOLO, which are two YOLO-based object detection systems, on these edge AI platforms with Microsoft COCO dataset. I will present our work starting from the problem statement. And then I’ll introduce our experiments setup and hardware configuration. Lastly, I’ll conclude our experimental results and current work status, as well as the future work.

Best Poster Finalist: no

Poster 148: Unsupervised Clustering of Golden Eagle Telemetry Data

We use a recurrent autoencoder neural network to encode sequential California golden eagle telemetry data. The encoding is followed by an unsupervised clustering technique, Deep Embedded Clustering (DEC), to iteratively cluster the data into a chosen number of behavior classes. We apply the method to simulated movement data sets and telemetry data for a Golden Eagle. The DEC achieves better unsupervised clustering accuracy scores for the simulated data sets as compared to the baseline K-means clustering result.

Best Poster Finalist: no

Poster 66: Hybrid CPU/GPU FE2 Multi-Scale Implementation Coupling Alya and Micropp
Guido Giuntoli (Barcelona Supercomputing Center), Judicaël Grasset (Science and Technology Facilities Council (STFC)), Alejandro Figueroa (George Mason University), Charles Moulinec (Science and Technology Facilities Council (STFC)), Mariano Vázquez (Barcelona Supercomputing Center), Guillaume Houzeaux (Barcelona Supercomputing Center), Stephen Longshaw (Science and Technology Facilities Council (STFC)), Sergio Oller (Polytechnic University of Catalonia)

This poster exposes the results of a new implementation of the FE2 multi-scale algorithm that is achieved by coupling the multi-physics and massively parallel code Alya with the GPU-based code micropp. The coupled code is mainly designed to solve large scale and realistic composite material problems for the aircraft industry. Alya is responsible of solving the macro-scale equations and micropp for solving the representation of fibres at the microscopic level. The poster shows computational performance results that demonstrate that the technique is scalable for real size industrial problems and also how the execution time is dramatically reduced using GPU-based clusters.
Best Poster Finalist: no

Poster 129: Understanding I/O Behavior in Scientific Workflows on High Performance Computing Systems
Fahim Tahmid Chowdhury (Florida State University, Lawrence Livermore National Laboratory), Francesco Di Natale (Lawrence Livermore National Laboratory), Adam Moody (Lawrence Livermore National Laboratory), Elsa Gonsiorowski (Lawrence Livermore National Laboratory), Kathryn Mohror (Lawrence Livermore National Laboratory), Weikuan Yu (Florida State University)

Leadership high performance computing (HPC) systems have the capability to execute workflows of scientific, research or industry applications. Complex HPC workflows can have significant data transfer and I/O requirements. Heterogeneous storage systems in supercomputers equipped with bleeding-edge non-volatile persistent storage devices can be leveraged to handle these data transfer and I/O requirements efficiently.

In this poster, we describe our efforts to extract the I/O characteristics of various HPC workflows and develop strategies to improve I/O performance by leveraging heterogeneous storage systems. We have implemented an emulator to mimic different types of I/O requirements posed by HPC application workflows. We have analyzed the workflow of Cancer Moonshot Pilot 2 (CMP2) project to determine possible I/O inefficiencies. To date, we have performed a systematic characterization and evaluation on the workloads generated by the workflow emulator and a small scale adaptation of the CMP2 workflow.

Best Poster Finalist: no

Poster 116: Advancements in Ultrasound Simulations Enabled by High-Bandwidth GPU Interconnects
Filip Vaverka (Brno University of Technology), Bradley E. Treeby (University College London), Jiri Jaros (Brno University of Technology)

Realistic ultrasound simulations are becoming integral part of many novel medical procedures such as photoacoustic screening and non-invasive treatment planning. The common denominator of all these applications is the need for cheap and relatively large-scale ultrasound simulations with sufficient accuracy. Typical medical applications require full-wave simulations which take frequency-dependent absorption and non-linearity into account.
This poster investigates the benefits of high-bandwidth low-latency interconnects to k-Wave acoustic toolbox in dense multi-GPU environment. The k-Wave multi-GPU code is based on a variant of the local Fourier basis domain decomposition. The poster compares the behavior of the code on a typical PCI-E 3.0 machine with 8 Nvidia Tesla P40 GPUs and a Nvidia DGX-2 server. The performance constraints of PCI-E platforms built around multiple socket servers on multi-GPU applications are deeply explored. Finally, it is shown the k-Wave toolbox can efficiently utilize NVlink 2.0 and achieve over 4x speedup compared to PCI-E systems.

Best Poster Finalist: no

**Poster 65: Comparing Granular Dynamics vs. Fluid Dynamics via Large DOF-Count Parallel Simulation on the GPU**

Milad Rakhsha (University of Wisconsin), Conlain Kelly (Georgia Institute of Technology), Nicholas Olsen (University of Wisconsin), Lijing Yang (University of Wisconsin), Radu Serban (University of Wisconsin), Dan Negrut (University of Wisconsin)

In understanding granular dynamics, the commonly-used discrete modeling approach that tracks the motion of all particles is computationally demanding, especially with large system size. In such cases, one can contemplate switching to continuum models that are computationally less expensive. In order to assess when such a discrete to continuum switch is justified, we compare granular and fluid dynamics that scales to handle more than 1 billion degrees of freedom (DOFs); i.e., two orders of magnitude higher than the state-of-the-art. On the granular side, we solve the Newton-Euler equations of motion; on the fluid side, we solve the Navier-Stokes equations. Both solvers leverage parallel computing on the GPU, and are publicly available on GitHub as part of an open-source code called Chrono. We report similarities and differences between the dynamics of the discrete, fully-resolved system and the continuum model via numerical experiments including both static and highly transient scenarios.

Best Poster Finalist: no

**Poster 78: Understanding HPC Application I/O Behavior Using System Level Statistics**

Arnab K. Paul (Virginia Tech), Olaf Faaland (Lawrence Livermore National Laboratory), Adam Moody (Lawrence Livermore National Laboratory), Elsa Gonsiorowski (Lawrence Livermore National Laboratory), Kathryn Mohror (Lawrence Livermore National Laboratory), Ali R. Butt (Virginia Tech)

The processor performance of high performance computing (HPC) systems is increasing at a much higher rate than storage performance. Storage and file system designers therefore require a deep
understanding of how HPC application I/O behavior affects current storage system installations in order to improve storage performance. In this work, we contribute to this understanding using application-agnostic file system statistics gathered on compute nodes as well as metadata and object storage file system servers. We analyze file system statistics of more than 4 million jobs over a period of three years on two systems at Lawrence Livermore National Laboratory that include a 15 PiB Lustre file system for storage. Some key observations in our study show that more than 65% HPC users perform significant I/O which are mostly writes; and less than 22% of HPC users who submit write-intensive jobs perform efficient writes to the file system.

Best Poster Finalist: no

**Poster 109: A Runtime Approach for Dynamic Load Balancing of OpenMP Parallel Loops in LLVM**

Jonas H. Müller Korndörfer (University of Basel, Switzerland), Florina M. Ciorba (University of Basel, Switzerland), Akan Yilmaz (University of Basel, Switzerland), Christian Iwainsky (Technical University Darmstadt), Johannes Doerfert (Argonne National Laboratory), Hal Finkel (Argonne National Laboratory), Vivek Kale (Brookhaven National Laboratory), Michael Klemm (Intel Corporation)

Load imbalance is the major source of performance degradation in computationally-intensive applications that frequently consist of parallel loops. Efficient scheduling can improve the performance of such programs. OpenMP is the de-facto standard for parallel programming on shared-memory systems. The current OpenMP specification provides only three choices for loop scheduling which are insufficient in scenarios with irregular loops, system-induced interference, or both. Therefore, this work augments the LLVM OpenMP runtime library implementation with eleven ready to use scheduling techniques. We tested existing and added scheduling strategies on several applications from NAS, SPEC OMP 2012, and CORAL2 benchmark suites. Experiments show that implemented scheduling techniques outperform others in certain application and system configurations. We measured performance gains of up to 6% compared to the fastest standard scheduling technique. This work aims to be a convincing step toward beyond-standard scheduling options in OpenMP for the benefit of evolving applications executing on multicore architectures.

Best Poster Finalist: no

**Poster 143: Quantum Natural Language Processing**

Lee James O'Riordan (Irish Centre for High-End Computing (ICHEC), National University of Ireland Galway (NUIG)), Myles Doyle (Irish Centre for High-End Computing (ICHEC), National University of Ireland Galway (NUIG)), Venkatesh Kannan (Irish Centre for High-End Computing (ICHEC), National
Natural language processing (NLP) algorithms that operate over strings of words are limited since they analyse meanings of the component words in a corpus without information about grammatical rules of the language. Consequently, they often produce unsatisfactory results with increase in problem complexity.

The "distributed compositional semantics" (DisCo) model incorporates grammatical structure of sentences into the algorithms, and offers significant improvements to the quality of results. However, their main challenge is the need for large classical computational resources. The DisCo model presents two quantum algorithms which lower storage and compute requirements compared to a classic HPC implementation.

In this project, we implement the two DisCo model quantum algorithms on the Intel Quantum Simulator deployed on the Irish national supercomputer. We target corpuses with ~1000 most-common words using up to 36 qubits simulation. The solution will be able to compute the meanings of two sentences and decide if their meanings match.

Best Poster Finalist: no

---

**Poster 152: Deep Domain Adaptation for Runtime Prediction in Dynamic Workload Scheduler**

Hoang H. Nguyen (National Center for Atmospheric Research (NCAR); University of Illinois, Chicago), Ben Matthews (National Center for Atmospheric Research (NCAR)), Irfan Elahi (National Center for Atmospheric Research (NCAR))

In HPC systems, users' requested runtime for submitted jobs plays a crucial role in efficiency. While underestimation of job runtime could terminate jobs before completion, overestimation could result in long queuing of other jobs in HPC systems. In reality, runtime prediction in HPC is challenging due to the complexity and dynamics of running workloads. Most of the current predictive runtime models are trained on static workloads. This poses a risk of over-fitting the predictions with bias from the learned workload distribution. In this work, we propose an adaptation of Correlation Alignment method in our deep neural network architecture (DCORAL) to alleviate the domain shift between workloads for better runtime predictions. Experiments on both standard benchmark workloads and NCAR real-time production workloads reveal that our proposed method results in a more stable training model across different workloads with low accuracy variance as compared to the other state-of-the-art methods.

Best Poster Finalist: no
Poster 75: libCEED - Lightweight High-Order Finite Elements Library with Performance Portability and Extensibility
Jeremy Thompson (University of Colorado), Valeria Barra (University of Colorado), Yohann Dudouit (Lawrence Livermore National Laboratory), Oana Marin (Argonne National Laboratory), Jed Brown (University of Colorado)

High-order numerical methods are widely used in PDE solvers, but software packages that have provided high-performance implementations have often been special-purpose and intrusive. libCEED is a new library that offers a purely algebraic interface for matrix-free operator representation and supports run-time selection of implementations tuned for a variety of computational device types, including CPUs and GPUs. We introduce the libCEED API and demonstrate how it can be used in standalone code or integrated with other packages (e.g., PETSc, MFEM, Nek5000) to solve examples of problems that often arise in the scientific computing community, ranging from fast solvers via geometric multigrid methods to Computational Fluid Dynamics (CFD) applications.

Best Poster Finalist: no

Progress on the Exascale Transition of the VSim Multiphysics PIC code
Benjamin M. Cowan (Tech-X Corporation), Sergey N. Averkin (Tech-X Corporation), John R. Cary (Tech-X Corporation), Jarrod Leddy (Tech-X Corporation), Scott W. Sides (Tech-X Corporation), Ilya A. Zilberter (Tech-X Corporation)

The highly performant, flexible plasma simulation code VSim was designed nearly 20 years ago (originally as Vorpal), with its first applications roughly four years later. Using object oriented methods, VSim was designed to allow runtime selection from multiple field solvers, particle dynamics, and reactions. It has been successful in modeling for many areas of physics, including fusion plasmas, particle accelerators, microwave devices, and RF and dielectric structures. Now it is critical to move to exascale systems, with their compute accelerator architectures, massive threading, and advanced instruction sets. Here we discuss how we are moving this complex, multiphysics computational application to the new computing paradigm, and how it is done in a way that kept the application producing physics during the move. We present performance results showing significant speedups in all parts of the PIC loop, including field updates, particle pushes, and reactions.

Best Poster Finalist: no
Poster 58: Lock-Free van Emde Boas Array
Ziyuan Guo (University of Tokyo)

Lock-based data structures have some potential issues such as deadlock, livelock, and priority inversion, and the progress can be delayed indefinitely if the thread that is holding locks cannot acquire a timeslice from the scheduler. Lock-free data structures, which guarantees the progress of some method call, can be used to avoid these problems. This poster introduces the first lock-free concurrent van Emde Boas Array which is a variant of van Emde Boas Tree. It is linearizable, and the benchmark shows significant performance improvement comparing to other lock-free search trees when the data set is large and dense enough.

Best Poster Finalist: no

Poster 63: Adaptive Execution Planning in Biomedical Workflow Management Systems
Marta Jaros (Brno University of Technology), Bradley E. Treeby (University College London), Jiri Jaros (Brno University of Technology)

Biomedical simulations require very powerful computers. Their execution is described by a workflow consisting of a number of different cooperating tasks. The manual execution of individual tasks may be tedious for expert users, but prohibiting for most inexperienced clinicians. k-Dispatch offers a 'run and forget' approach where the users are completely screened out from the complexity of HPC systems. k-Dispatch provides task scheduling, execution, monitoring, and fault tolerance. Since the task execution configuration strongly affects the final tasks mapping on the computational resources, the execution planning is of the highest priority. Unlike other tools, k-Dispatch considers a variable amount of computational resources per individual tasks. Since the scaling of the individual HPC codes is never perfect, k-Dispatch may find such a good mapping even an experienced user would miss. The proposed adaptive execution planning is based on collected performance data and the current cluster utilization monitoring.

Best Poster Finalist: no

Poster 142: Training Deep Neural Networks Directly on Hundred-Million-Pixel Histopathology Images on a Large-Scale GPU Cluster
Chi-Chung Chen (AetherAI, Taiwan), Wen-Yu Chuang (Chang-Gung Memorial Hospital, Taiwan), Wei-Hsiang Yu (AetherAI, Taiwan), Hsi-Ching Lin (National Center for High-Performance Computing (NCHC), Taiwan), Shuen-Tai Wang (National Center for High-Performance Computing
Deep learning for digital pathology is challenging because the resolution of whole-slide-images (WSI) is extremely high, often in billions. The most common approach is patch-based method, where WSIs are divided into small patches to train convolutional neural networks (CNN). This approach has significant drawbacks. To have ground truth for individual patches, detailed annotations by pathologists are required. This laborious process has become the major impediment to the development of digital pathology AI. End-to-end WSI training, however, faces the difficulties of fitting the task into limited GPU memory. In this work, we improved the efficiency of using system memory for GPU compute by 411% through memory optimization and deployed the training pipeline on 8 nodes, totally 32 GPUs distributed system, achieving 147.28x speedup. We demonstrated that CNN is capable of learning features without detailed annotations. The trained CNN can correctly classify cancerous specimen, with performance level closely matching the patch-based methods.

Best Poster Finalist: no

**Poster 96: TSQR on TensorCores**
Hiroyuki Ootomo (Tokyo Institute of Technology), Rio Yokota (Tokyo Institute of Technology)

Tall-Skinny QR (TSQR) is an efficient algorithm for calculating the QR decomposition of m x n matrices where m << n, which is done by recursively performing QR decomposition on subdivided blocks of the tall and skinny matrix. Such operations are useful for low-rank approximation methods, which are replacing more and more dense linear algebra in both scientific computing and machine learning fields. The present work focuses on the implementation of this important algorithm on Tensor Cores, which are available on the latest NVIDIA GPUs. We evaluate the speed, accuracy, and stability of TSQR on TensorCores.

Best Poster Finalist: yes

**Poster 95: A Heterogeneous HEVC Video Encoder Based on OpenPOWER Acceleration Platform**
Chenhao Gu (Fudan University, Shanghai, State Key Laboratory of ASIC and System; IBM Corporation), Yang Chen (Fudan University, Shanghai, State Key Laboratory of ASIC and System; IBM Corporation), Yanheng Lu (IBM Corporation), Pengfei Gou (IBM Corporation), Yong Lu (IBM Corporation), Fang-An Kuo (National Center for High-Performance Computing (NCHC), Taiwan), Chao-Chun Chuang (National Center for High-Performance Computing (NCHC), Taiwan), Chao-Yuan Yeh (AetherAI, Taiwan)
This poster describes a heterogeneous HEVC video encoder system based on the OpenPOWER platform. Our design leverages the Coherent Accelerator Processor Interface (CAPI) on the OpenPOWER, which provides cache-coherent access for FPGA. This technology highly improves CPU-FPGA data communication bandwidth and programming efficiency. X265 is optimized on the OpenPOWER platform to improve its performance with both architecture specific methods and hardware-acceleration methods. For hardware acceleration, frame-level acceleration and functional-unit-level acceleration are introduced and evaluated in this work.

Best Poster Finalist: no
applications because of its high performance and scalability with many computational nodes. GPUs are thought to be good candidates for accelerating such applications with many meshes where an MGCG solver could show high performance. No previous studies have evaluated and discussed the numerical character of an MGCG solver on GPUs. Consequently, we have implemented and optimized our "kinaco" numerical ocean model with an MGCG solver on GPUs. We evaluated its performance and discussed inter-GPU communications on a coarse grid on which GPUs could be intrinsically problematic. We achieved 3.9 times speedup compared to CPUs and learned how inter-GPU communications depended on the number of GPUs and the aggregation level of information in a multigrid method.

Best Poster Finalist: no

Poster 108: Power Prediction for High-Performance Computing
Shigeto Suzuki (Fujitsu Laboratories Ltd), Michiko Hiraoka (Fujitsu Ltd), Takashi Shiraiishi (Fujitsu Laboratories Ltd), Enxhi Kreshpa (Fujitsu Laboratories Ltd), Takuji Yamamoto (Fujitsu Laboratories Ltd), Hiroyuki Fukuda (Fujitsu Laboratories Ltd), Shuji Matsui (Fujitsu Ltd), Masahide Fujisaki (Fujitsu Ltd), Atsuya Uno (RIKEN Center for Computational Science (R-CCS))

Exascale computers consume large amounts of power both for computing and cooling-units. As power of the computer varies dynamically corresponding to the load change, cooling-units are desirable to follow it for effective energy management. Because of time lags in cooling-unit operations, advance control is inevitable and an accurate prediction is a key for it. Conventional prediction methods make use of the similarity between job information while in queue. The prediction fails if there is no previously similar job. We developed two models to correct the prediction after queued jobs start running. By taking power histories into account, power-correlated topic model reselects more suitable candidate and recurrent-neural-network model considering variable network sizes predicts power variation from shape features of it. We integrated these into a single algorithm and demonstrated high-precision prediction with an average relative error of 5.7% in K computer as compared to the 18.0% obtained using the conventional method.

Best Poster Finalist: no

Poster 80: Sharing and Replicability of Notebook-Based Research on Open Testbeds
Maxine V. King (University of Chicago), Jason Anderson (University of Chicago), Kate Keahey (Argonne National Laboratory, University of Chicago)

We seek to facilitate replicability by creating a way to share experiments easily in and out of
notebook-based, open testbed environments and a sharing platform for such experiments in order to allow researchers to combine shareability, consistency of code environment, and well-documented process.

Best Poster Finalist: no

**Poster 121: HFlush: Realtime Flushing for Modern Storage Environments**  
Jaime Cernuda (Illinois Institute of Technology), Hugo Trivino (Illinois Institute of Technology), Hariharan Devarajan (Illinois Institute of Technology), Anthony Kougkas (Illinois Institute of Technology), Xian-He Sun (Illinois Institute of Technology)

Due to the unparalleled magnitude of data movement in extreme scale computing, I/O has become a central challenge. Modern storage environments have proposed the use of multiple layers between applications and the PFS. Nonetheless, the difference in capacities and speeds between storage layers makes it extremely challenging to evict data from upper layers to lower layers efficiently. However, current solutions are executed in batches, compromising latency; are also push-based implementations, compromising resource utilization. Hence, we propose HFlush, a continuous data eviction mechanism built on a streaming architecture that is pull-based and in which each component is decoupled and executed in parallel. Initial results have shown HFlush to obtain a 7X latency reduction and a 2X bandwidth improvement over a baseline batch-based system. Therefore, HFlush is a promising solution to the growing challenges of extreme scale data generation and eviction shortcomings when archiving data across multiple tiers of storage.

Best Poster Finalist: no

**Poster 88: HPC Container Runtime Performance Overhead: At First Order, There Is None**  
Alfred Torrez (Los Alamos National Laboratory), Reid Priedhorsky (Los Alamos National Laboratory), Timothy Randles (Los Alamos National Laboratory)

Linux containers are an increasingly popular method used by HPC centers to meet increasing demand for greater software flexibility. A common concern is that containers may introduce application performance overhead. Prior work has not tested a broad set of HPC container technologies on a broad set of benchmarks. This poster addresses the gap by comparing performance of the three HPC container implementations (Charliecloud, Shifter, and Singularity) and bare metal on multiple dimensions using industry-standard benchmarks.

We found no meaningful performance differences between the four environments with the possible
exception of modest variation in memory usage, which is broadly consistent with prior results. This result suggests that HPC users should feel free to containerize their applications without concern about performance degradation, regardless of the container technology used. It is an encouraging development on the path towards greater adoption of user-defined software stacks to increase the flexibility of HPC.

Best Poster Finalist: no


Tsuyoshi Ichimura (University of Tokyo, RIKEN), Kohei Fujita (University of Tokyo, RIKEN), Takuma Yamaguchi (University of Tokyo), Akira Naruse (Nvidia Corporation), Jack C. Wells (Oak Ridge National Laboratory), Christopher J. Zimmer (Oak Ridge National Laboratory), Tjerk P. Straatsma (Oak Ridge National Laboratory), Takane Hori (Japan Agency for Marine-Earth Science and Technology), Simone Puel (University of Texas), Thorsten W. Becker (University of Texas), Muneo Hori (Japan Agency for Marine-Earth Science and Technology), Naonori Ueda (RIKEN)

We propose herein an approach for reformulating an equation-based modeling algorithm to an algorithm similar to that of training artificial intelligence (AI) and accelerate this algorithm using high-performance accelerators to reduce the huge computational costs encountered for physics equation-based modeling in earthquake disaster mitigation. A fast scalable equation-based implicit solver on unstructured finite elements is accelerated with a Tensor Core-enabled matrix-vector product kernel. The developed kernel attains 1.10 ExaFLOPS, leading to 416 PFLOPS for the whole solver on full Summit. This corresponds to a 75-fold speedup from a previous state-of-the-art solver running on full Piz Daint. This result could lead to breakthroughs in earthquake disaster mitigation. Our new idea in the HPC algorithm design of combining equation-based modeling with AI is expected to have broad impacts in other earth science and industrial problems.

Best Poster Finalist: no

8:30 am - 5:00 pm

Scientific Visualization & Data Analytics Showcase Posters Display

Visualization of Entrainment and Mixing Phenomena at Cloud Edges
Entrainment and mixing processes occur during the entire life of a cloud. These processes change the droplet size distribution, which determines rain formation and radiative properties. Since it is a microphysical process, it cannot be resolved in large scale weather forecasting models. Small scale simulations such as Direct Numerical Simulations (DNS) are required to resolve the smallest scale of these processes. However, it has been a challenge to visualize these processes in a 3D domain as it generates petabytes of data. Visualization plays a vital role in analyzing and understanding these huge data outputs. Here, we present different techniques for 3D visualization of data obtained from DNS carried out at Indian Institute of Tropical Meteorology (IITM) to understand cloud microphysical properties more closely.

Multiple tools were used to conduct a visual analysis of this data. Two of these tools are well established and tested technologies: ParaView and VAPOR. The others are emergent technologies created at the National Center for Atmospheric Research (NCAR) and are in the development phase. This data simulation, in addition to exploring DNS as mentioned above, provided an opportunity to test and improve development of several tools and methods.
the network output, projected into 3-D parameter space, reveals how the network has discovered two distinct clusters of features which allows it to classify galaxies into two groups. These visualizations of the neural network during training aid in the interpretability of the black box of deep learning and reveal how the network responds to the input images at various stages of training. Finally, it allows a wider net to be cast to a general audience, thereby generating interest in and visibility to an otherwise highly specialized field.

**NVIDIA IndeX Accelerated Computing for Visualizing Cholla's Galactic Winds**
Christopher Lux (Nvidia Corporation), Alexander Kuhn (Nvidia Corporation), Evan Schneider (Princeton University), Brant Robertson (University of California, Santa Cruz), Marc Nienhaus (Nvidia Corporation)

Galactic winds – outflows of gas driven out of galaxies by the combined effects of thousands of supernovae – are a crucial feature of galaxy evolution. Despite their importance, a complete theoretical picture of these winds has been elusive. Simulating the complicated interaction between the hot, high pressure gas created by supernovae and the cooler, high density gas in the galaxy disk requires massive computational resources and sophisticated software. For this purpose, Computational Hydrodynamics On Parallel Architectures (Cholla) has been demonstrated to be a scalable and efficient tool that operates in large, distributed multi-GPU environments at high levels of performance. This additional performance results in unprecedented resolution for this type of simulation and massive amounts of generated data. This raises the requirements for analysis tools that can cope with scale and complexity of the simulated physical processes. To address those challenges, we utilize NVIDIA IndeX as a scalable framework to visualize the simulation output: NVIDIA IndeX features a streaming-based architecture to interactively explore simulation results in large-scale, multi-GPU environments. We utilize customized sampling programs for multi-volume and surface rendering to address analysis questions of galactic wind simulations. This combination of massively parallel simulation and analysis allows us to utilize recent supercomputer capabilities and to speed up the exploration of galactic wind simulations.

**Visualizing the World's Largest Turbulence Simulation**
Salvatore Cielo (Leibniz Supercomputing Centre), Luigi Iapichino (Leibniz Supercomputing Centre), Johannes Günther (Intel Corporation), Christoph Federrath (Australian National University, Research School of Astronomy and Astrophysics), Elisabeth Mayer (Leibniz Supercomputing Centre), Markus Wiedemann (Leibniz Supercomputing Centre)

We present the visualization of the largest interstellar turbulence simulations ever performed, unravelling key astrophysical processes concerning the formation of stars and the relative role of
magnetic fields. The simulations, including pure hydrodynamical (HD) and magneto-hydrodynamical (MHD) runs, up to a size of $10048^3$ grid elements, were produced on the supercomputers of the Leibniz Supercomputing Centre and visualized using the hybrid parallel (MPI + TBB) ray-tracing engine OSPRay associated with VisIt. Besides revealing features of turbulence with an unprecedented resolution, the visualizations brilliantly showcase the stretching-and-folding mechanisms through which astrophysical processes such as supernova explosions drive turbulence and amplify the magnetic field in the interstellar gas, and how the first structures, the seeds of newborn stars are shaped by this process.

An Accessible Visual Narrative for the Primary Energy Source of Life from the Fulldome Show Birth of Planet Earth

Melih Sener (University of Illinois), Stuart Levy (University of Illinois), AJ Christensen (University of Illinois), Robert Patterson (University of Illinois), Kalina Borkiewicz (University of Illinois), John E. Stone (University of Illinois), Barry Isralewitz (University of Illinois), Jeffrey Carpenter (University of Illinois), Donna Cox (University of Illinois)

Conversion of sunlight into chemical energy, namely photosynthesis, is the primary energy source of life on Earth. An explanatory visualization depicting this process is presented in the form of an excerpt from the fulldome show Birth of Planet Earth. This accessible visual narrative shows a lay audience, especially children, how the energy of sunlight is captured, converted, and stored through a chain of proteins to power living cells. The visualization is the result of a multi-year collaboration among biophysicists, visualization scientists, and artists, which, in turn, is based on a decade-long experimental-computational collaboration on structural and functional modeling that produced an atomic detail description of a bacterial bioenergetic organelle, the chromatophore. The energy conversion steps depicted feature an integration of function from electronic to cell levels, spanning nearly 12 orders of magnitude in time scales modeled with multi-scale computational approaches. This atomic detail description uniquely enables a modern retelling of one of humanity’s earliest stories---the interplay between light and life.

Visualizing Supersonic Retropropulsion for Mars: The Summit Early Science Results

Patrick Moran (NASA Ames Research Center), Timothy Sandstrom (NASA Ames Research Center), Ashley Korzun (NASA Langley Research Center), Eric Nielsen (NASA Langley Research Center), Aaron Walden (NASA Langley Research Center)

As part of the Oak Ridge Leadership Computing Facility (OLCF) Early Science program, NASA ran a series of simulations on the new, GPU-accelerated system Summit, using the flow solver FUN3D to produce solutions for a set of representative descent conditions. The raw output from these
simulations was over a petabyte of data. We produced visualizations at the NASA Advanced Supercomputing Division at Ames Research Center using parallel techniques where feasible to accelerate the process. Included with this summary is an animation with some of the highlights of our visualizations so far. We discuss some of our visualization design choices, and plans for future work.

10:30 am - 12:00 pm

Doctoral Showcase I Presentations

Poster 42: Power and Thermal Management Runtimes for HPC Applications in the Era of Exascale Computing
Daniele Cesarini (University of Bologna, CINECA)

In the scope of technical and scientific computing, the rush toward larger simulations has been so far assisted by a steady downsizing of micro-processing units, which has allowed to increase the compute capacity of general-purpose architectures at constant power. As side effects of the end of Dennard's scaling, this process is now hitting its ultimate power limits and is just about to come to an end. This implies an increase in the energy cost for computation, performance loss due to a design based on worst-case power consumption, and performance loss due to overheating, and thermal gradients. As result, thermal and power bound supercomputing machines show performance degradation and heterogeneity which limit the peak performance of the system. This doctoral showcase presents software strategies to tackle the main bottlenecks induced by power and thermal issues that affect next-generation supercomputers. To respond to the above challenges, my work shows that propagating workload requirements from application to the runtime and operating system levels is the key to provide efficiency. This is possible only if the proposed software methodologies cause little or no overhead in term of application performance. With this in mind in my work, I have designed application-aware node-level optimal thermal management algorithms and runtimes, lazy node-level power capping, and energy reduction runtime. The experimental results show a significant step forward with respect to the current state-of-the-art solutions in power and thermal control of HPC systems.

Poster 45: Characterization and Modeling of Error Resilience in HPC Applications
Luanzheng Guo (University of California, Merced)

As high-performance computing systems scale in size and computational power, the occurrence of
Transient faults grow. Without protection by efficient and effective fault tolerance mechanisms, transient errors can cause incorrect execution outcome and even lead to a catastrophe in safety-critical HPC applications. Previous work attributes error resilience in HPC applications at a high-level to either the probabilistic or iterative nature of the application, whereas the community still lacks the fundamental understanding of the program constructs that result in natural error resilience. We design FlipTracker, a framework to analytically track error propagation and to provide a fine-grained understanding of the propagation and tolerance of errors. After running FlipTracker on representative HPC applications, we summarize six resilience computation patterns that lead to nature error resilience in HPC applications. With a better understanding of natural resilience in HPC applications, we aim to model application resilience on data objects to transient faults. Many common application-level fault tolerance mechanisms focus on data objects. Understanding application resilience on data objects can be helpful to direct those mechanisms. The common practice to understand application resilience (random fault injection) gives us little knowledge of how and where errors are tolerated. Understanding "how" and "where" is necessary to understand how to apply application-level fault tolerance mechanisms effectively and efficiently. We design a practical model (MOARD) to measure application resilience on data objects by analytically quantifying error masking events happening to the data object. Using our model, users can compare application resilience on different data objects with different data types.

Poster 39: Designing Next-Generation Communication Middlewares for Many-Core Architectures
Jahanzeb Maqbool Hashmi (Ohio State University)

Modern multi-petaflop HPC systems are powered by dense multi-/many-core architectures and this trend is expected to grow for the next-generation supercomputing systems. This rapid adoption of high core-density architectures by the current- and next-generation HPC systems is further fueled by the emerging application trends such as Deep Learning. This is putting more emphasis on middleware designers to optimize various communication protocols to meet the diverse needs of the applications. While the novelties in the processor architectures have led to the increased on-chip parallelism, they come at the cost of rendering traditional designs, employed by the communication runtimes such as MPI, to suffer from higher degree of intra-node communication latencies. Tackling the computation and communication challenges that accompany these dense multi-/many-cores garner special design considerations. The proposed work in this thesis tries to address the performance challenges posed by a diverse range of applications and the lacking support in state-of-the-art communication libraries such as MPI to exploit high-concurrency architectures. The author first proposes a "shared-address-spaces"-based communication substrate to derive intra-node communication in MPI. Atop this framework, the author has re-designed various MPI primitives such as point-to-point communication protocols (e.g., user-space
zero-copy rendezvous transfer), collective communication (e.g., load/store based collectives, truly zero-copy and partitioning-based reduction algorithms), and efficient MPI derived datatypes processing (e.g., memoization-based "packing-free" communication) to exploit the potential of emerging multi-/many-core architectures and high throughput networks. The proposed designs have demonstrated significant improvement over state-of-the-art for various scientific and deep learning applications.

1:30 pm - 3:00 pm

**Doctoral Showcase II Presentations**

**Poster 41: Co-Designing Communication Middleware and Deep Learning Frameworks for High-Performance DNN Training on HPC Systems**

Ammar Ahmad Awan (Ohio State University)

Recent advances in Machine/Deep Learning techniques have triggered key success stories in many application domains like Computer Vision, Speech Comprehension and Recognition, and Natural Language Processing. Large-scale Deep Neural Networks (DNNs), that are at the core of state-of-the-art AI technologies, have been the primary drivers of this success. Training very complicated and large DNN architectures using a large number of training examples (data) is compute-intensive and can take from weeks to months to achieve state-of-the-art prediction capabilities. To achieve higher accuracy, making the DNN deeper is also a common strategy. These requirements have led to a simple but powerful approach called Data Parallelism to achieve shorter training times. This has resulted in various research studies and ML/DL software like TensorFlow and PyTorch as well as distributed-training middleware like Horovod. In addition, for DNNs that do not fit the GPU memory, a new DNN workloads are emerging that we call Out-of-Core DNNs and different strategies (out-of-core training and model-parallelism) are needed to train them. Clearly, large-scale DNN training brings forward new requirements for computation runtimes like CUDA and communication middleware like the MVAPICH2 MPI library. In this thesis, we broadly explore three different strategies to train DNNs on modern CPU and GPU architectures: 1) Data Parallelism, 2) Model Parallelism, and 3) Out-of-Core Training. We address the key challenge: How to co-design computation and communication in modern ML/DL frameworks with execution runtimes like CUDA and communication middleware like MVAPICH2 to enable scalable, high-performance, and efficient training of DNNs on large-scale HPC systems?

**Poster 40: Performance, Portability, and Productivity for Data-Parallel Computations on Multi-
and Many-Core Architectures
Ari Rasch (University of Münster)

This thesis presents an approach to performance, portability, and productivity for data-parallel computations on multi- and many-core architectures, e.g., Intel CPU and NVIDIA GPU. We introduce the algebraic formalism of Multi-Dimensional Homomorphisms (MDHs) – a class of functions that cover important data-parallel computations, e.g., linear algebra routines (BLAS) and stencil computations. For our MDHs, we propose a Domain-Specific Language (DSL), based on patterns of parallelism (a.k.a. algorithmic skeletons), to enable conveniently expressing MDH functions. We introduce a code generation approach for our DSL to automatically generate for MDHs optimized program code targeting multi- and many-core architectures. Our code generation approach relies on OpenCL – an emerging de-facto standard for uniformly programming parallel architectures, such as CPU and GPU. A major feature of our generated code is that it is targeted to OpenCL’s abstract device models (rather than a particular architecture) by being parameterized in performance-critical parameters of these abstract models (e.g., the number of threads and size of tiles). With our code generation approach, we enable both high performance and performance portability: we fully automatically optimize our generated code -- for any given combination of an MDH function, architecture, and input size -- by automatically choosing (auto-tuning) optimized values of our code’s performance-critical parameters using our own Auto-Tuning Framework (ATF). Our experimental results on CPU and GPU demonstrate competitive and often significantly better performance of our MDH+ATF approach as compared to the currently best-performing competitors, e.g., Intel MKL/MKL-DNN, NVIDIA cuBLAS/cuDNN, and Facebook's Tensor Comprehensions framework.

Poster 43: Efficient and Scalable Communication Middleware for Emerging Dense-GPU Clusters
Ching-Hsiang Chu (Ohio State University)

In the era of post Moore’s law, the traditional CPU is not able to keep the pace up and provide the computing power demanded by the modern compute-intensive and highly parallelizable applications. Under this context, various accelerator architectures such as general-purpose graphics processing unit (GPU), which equipped with the high-bandwidth memory (HBM) and massive parallelizable streaming multiprocessors, has been widely adopted in high-performance computing (HPC) and cloud systems to significantly accelerate numerous scientific and emerging machine/deep learning applications. Message Passing Interface (MPI), the standard programming model for parallel applications, has been widely used for GPU communication. However, the state-of-the-art MPI libraries are only optimizing GPU communication by leveraging advanced technology like Remote Direct Memory Access (RDMA) and not fully utilizing the computational power of
GPUs. In this work, we propose GPU-enabled communication schemes to harness GPU computational resources, and cutting-edge interconnects such as NVIDIA NVLink for communication operations on the emerging heterogeneous systems. In this work, three primary MPI operations are addressed. First, intelligent communication scheduling, efficient packing/unpacking, and packing-free schemes are proposed to accelerate non-contiguous data transfer in scientific HPC applications. Second, scalable broadcast operations are presented to leverage the low-level hardware multicast feature to speed up GPU communication at scale. Finally, we also design topology-aware, link-efficient, and cooperative GPU kernels to significantly accelerate All-reduce operation, which is the primary performance bottleneck in deep learning applications. The proposed designs demonstrate significant performance improvements over the state-of-the-art communication schemes for various HPC and deep learning applications.

Poster 38: High-Performance Backpropagation in Scientific Computing
Navjot Kukreja (Imperial College, London)

Devito is a domain-specific language for the automatic generation of high-performance solvers for finite difference equations provided in a high-level symbolic representation. The primary use of Devito is to enable Full-waveform inversion, which is an adjoint-based optimization problem that uses the wave equation as part of its objective function. This doctoral project consists of three elements of Devito that are key to solving adjoint-based optimization problems (i.e. the back propagation). The first is automatic differentiation of stencil loops. This allows the automatic generation of high-performance code implementing the derivative of any provided function. This is essential to enable the use of a wider range of physical equations - in order to use better physical approximations. An essential feature of the generated derivatives is that they show scaling and performance behaviour that is very similar to the forward function. A common issue in solving adjoint-based optimization problems is the prohibitively high memory requirement. The second contribution is a runtime for automatic execution of checkpoint-recompute schedules (called pyRevolve) to alleviate this memory requirement by trading it off for re-computations. These schedules may involve storing some intermediate states on disk, some in memory, and recomputing others. The third contribution is the use of lossy compression (ZFP/SZ), which is a second means of trading off memory and compute, automatically as part of checkpointing in pyRevolve. This is the first time it has been shown that lossy compression combined with checkpoint-recomputation can provide a much better error-memory tradeoff than using checkpoint-recomputation or lossy compression alone.

3:30 pm - 5:00 pm
Doctoral Showcase III Presentations

**Poster 35: Scaling Up Pipeline Optimization with High Performance Computing**  
Robert Lim (University of Oregon)

My research focuses on developing a pipeline optimization infrastructure that automates the design and code generation of neural networks through the use of high-performance computing. The problem has the following objectives: unify automated machine learning (AutoML) and compilation, archive profiles for creation of a knowledge base for a data-driven approach toward search, explore various search optimizations for model design and code generation. The field of automated deep learning includes hyperparameter optimization and neural architecture search (NAS), which requires domain expertise in designing a model, in addition to the tuning parameters related to learning and the model itself. The search space is complex and deciding which parameters factor into the overall accuracy of a model is a non-trivial task. Once a model is trained, the next step compiles the model, which maps to the backend of a targeted architecture, whether GPU, embedded mobile phones, or FPGA. The compilation phase also involves choices, in terms of optimizations applied, which can include a set of flags, or direct code transformations. Various efficient search techniques are explored, archiving results along the way for facilitating in code generation options, which could possibly inform the model design process. Our previous work reduced the search space for code generation of various domain kernels by 92%, and this work investigates whether the same approach can be applied in all aspects of AutoML design and code generation.

**Poster 34: Analysis of Automata Processing Acceleration on Disparate Hardware Technologies**  
Marziyeh Nourian (North Carolina State University)

Pattern matching is a computation that maps naturally onto finite automata (FA) abstractions. There has been a substantial amount of work on accelerating FA processing on various parallel platforms. However, the advantages and disadvantages of different automata processing accelerators and the innovation space in this area are still unclear. We target this problem and propose a compiler toolchain that automates the deployment of non-deterministic finite automata (NFAs) onto different target platforms. Using this toolchain, we perform an apples-to-apples comparison between AP, GPU- and FPGA-based NFA accelerator designs on large-scale datasets. Specifically, we observe that memory-based designs are limited by memory size and bandwidth. To address this issue, we target fixed-topology NFAs and propose a memory-efficient design that embeds the automata topology in code and stores only the transition symbols in memory. Our solution is suitable for SIMD architectures and is called SIMD_NFA. We design a compiler that automates the deployment of this design on SIMD platforms. We showcase our compiler framework on GPU and Intel platforms.
Additionally, we observe that for NFAs with a grid-like fixed-topology (e.g., NFAs for Levenshtein and Hamming distance-based matching), transitions do not need to be encoded within the traversal code but can be inferred from the reference string to be matched and the knowledge of the NFA topology. Lastly, SIMD_NFA is a good fit for FPGA deployment using OpenCL-to-FPGA toolchains. We investigate the deployment of the OpenCL version of SIMD_NFA, on FPGA and explore a set of optimizations techniques to retarget SIMD_NFA to FPGA.

Poster 36: Modeling Non-Determinism in HPC Applications
Dylan Chapp (University of Delaware, University of Tennessee)

As HPC applications migrate from the petascale systems of today to the exascale systems of tomorrow, the increasing need to embrace asynchronous, irregular, and dynamic communication patterns will lead to a corresponding decrease in application-level determinism. Two critical challenges emerge from this trend. First, unchecked non-determinism coupled with the non-associativity of floating-point arithmetic undermines numerical reproducibility of scientific applications. Second, the prevalence of non-determinism amplifies the cost of debugging, both in terms of computing resources and human effort. In this thesis, we present a modeling methodology to quantify and characterize communication non-determinism in parallel applications. Our methodology consists of three core components. First, we build graph-structured models of relevant communication events from execution traces. Second, we apply similarity metrics based on graph kernels to quantify run-to-run variability and thus identify the regions of executions where non-determinism manifests most prominently. Third, we leverage our notion of execution similarity to characterize applications via clustering, anomaly detection, and extraction of representative patterns of non-deterministic communication which we dub "non-determinism motifs". Our work will amplify the effectiveness of software tools that target mitigation or control of application-level non-determinism (e.g., record-and-replay tools) by providing them with a common metric for quantifying communication non-determinism in parallel applications and a common language for describing it.

Poster 37: Large Scale Ultrasound Simulations on Accelerated Clusters
Filip Vaverka (Brno University of Technology)

Many emerging non-invasive medical procedures such as high intensity focused ultrasound treatment (HIFU), photoacoustic imaging, or transcranial ultrasound therapy require ultrasound wave propagation simulations. Typically soft tissue can be modeled as fluid, weakly heterogeneous medium with frequency dependent absorption and non-linear effects at high ultrasound intensities. The k-Wave acoustic toolbox, widely used in medical community, uses highly efficient k-space
pseudo-spectral time-domain (KSTD) discretization of acoustic equations to solve these problems. However, it is difficult to efficiently map the KSTD method onto modern cluster architectures with accelerators such as GPUs. The present thesis analyzes shortcomings of the KSTD method in respect to modern clusters and proposes local Fourier basis approach to improve scaling of the method. It is shown that the proposed method is able to achieve 5x speedup, while having sufficient accuracy for these medical applications. Behavior of the method is analyzed across variety of GPU and MIC (Intel Xeon Phi) accelerated clusters and results are presented.

Friday, November 22
8:30 am - 12:00 pm
Scientific Visualization & Data Analytics Showcase Posters Display

Visualization of Entrainment and Mixing Phenomena at Cloud Edges
Bipin Kumar (Indian Institute of Tropical Meteorology), Matt Rehme (National Center for Atmospheric Research (NCAR)), Neethi Suresh (Indian Institute of Tropical Meteorology), Nihanth Cherukuru (National Center for Atmospheric Research (NCAR)), Stanislaw Jaroszynski (National Center for Atmospheric Research (NCAR)), Shaomeng Li (National Center for Atmospheric Research (NCAR)), Scott Pearse (National Center for Atmospheric Research (NCAR)), Tim Scheitlin (National Center for Atmospheric Research (NCAR))

Entrainment and mixing processes occur during the entire life of a cloud. These processes change the droplet size distribution, which determines rain formation and radiative properties. Since it is a microphysical process, it cannot be resolved in large scale weather forecasting models. Small scale simulations such as Direct Numerical Simulations (DNS) are required to resolve the smallest scale of these processes. However, it has been a challenge to visualize these processes in a 3D domain as it generates petabytes of data. Visualization plays a vital role in analyzing and understanding these huge data outputs. Here, we present different techniques for 3D visualization of data obtained from DNS carried out at Indian Institute of Tropical Meteorology (IITM) to understand cloud microphysical properties more closely.

Multiple tools were used to conduct a visual analysis of this data. Two of these tools are well established and tested technologies: ParaView and VAPOR. The others are emergent technologies created at the National Center for Atmospheric Research (NCAR) and are in the development phase. This data simulation, in addition to exploring DNS as mentioned above, provided an opportunity to test and improve development of several tools and methods.
Visualizing Deep Learning at Scale for the Construction of Galaxy Catalogs in the Dark Energy Survey
Janet Y. K. Knowles (Argonne National Laboratory), Joseph A. Insley (Argonne National Laboratory, Northern Illinois University), Silvio Rizzi (Argonne National Laboratory), Elise Jennings (Argonne National Laboratory), Asad Khan (University of Illinois), Eliu Huerta (University of Illinois), Sibo Wang (University of Illinois), Robert Gruendl (University of Illinois), Huihuo Zheng (Argonne National Laboratory)

The advent of machine and deep learning algorithms on petascale supercomputers is accelerating the pace of discovery in astrophysics and poses significant challenges to the interpretability of these deep neural networks. We present a novel visualization of a deep neural network output during training as it is learning to classify galaxy images as either spiral or elliptical. The network is trained using labeled datasets from the citizen science campaign, Galaxy Zoo, adopted by the Sloan Digital Sky Survey. These trained neural network models can then be used to classify galaxies in the Dark Energy Survey that overlap the footprint of both surveys. Visualizing a reduced representation of the network output, projected into 3-D parameter space, reveals how the network has discovered two distinct clusters of features which allows it to classify galaxies into two groups. These visualizations of the neural network during training aid in the interpretability of the black box of deep learning and reveal how the network responds to the input images at various stages of training. Finally, it allows a wider net to be cast to a general audience, thereby generating interest in and visibility to an otherwise highly specialized field.

NVIDIA IndeX Accelerated Computing for Visualizing Cholla's Galactic Winds
Christopher Lux (Nvidia Corporation), Alexander Kuhn (Nvidia Corporation), Evan Schneider (Princeton University), Brant Robertson (University of California, Santa Cruz), Marc Nienhaus (Nvidia Corporation)

Galactic winds – outflows of gas driven out of galaxies by the combined effects of thousands of supernovae – are a crucial feature of galaxy evolution. Despite their importance, a complete theoretical picture of these winds has been elusive. Simulating the complicated interaction between the hot, high pressure gas created by supernovae and the cooler, high density gas in the galaxy disk requires massive computational resources and sophisticated software. For this purpose, Computational Hydrodynamics On Parallel Architectures (Cholla) has been demonstrated to be a scalable and efficient tool that operates in large, distributed multi-GPU environments at high levels of performance. This additional performance results in unprecedented resolution for this type of simulation and massive amounts of generated data. This raises the requirements for analysis tools
that can cope with scale and complexity of the simulated physical processes. To address those challenges, we utilize NVIDIA IndeX as a scalable framework to visualize the simulation output. NVIDIA IndeX features a streaming-based architecture to interactively explore simulation results in large-scale, multi-GPU environments. We utilize customized sampling programs for multi-volume and surface rendering to address analysis questions of galactic wind simulations. This combination of massively parallel simulation and analysis allows us to utilize recent supercomputer capabilities and to speed up the exploration of galactic wind simulations.

Visualizing the World's Largest Turbulence Simulation
Salvatore Cielo (Leibniz Supercomputing Centre), Luigi Iapichino (Leibniz Supercomputing Centre), Johannes Günther (Intel Corporation), Christoph Federrath (Australian National University, Research School of Astronomy and Astrophysics), Elisabeth Mayer (Leibniz Supercomputing Centre), Markus Wiedemann (Leibniz Supercomputing Centre)

We present the visualization of the largest interstellar turbulence simulations ever performed, unravelling key astrophysical processes concerning the formation of stars and the relative role of magnetic fields. The simulations, including pure hydrodynamical (HD) and magneto-hydrodynamical (MHD) runs, up to a size of 10048^3 grid elements, were produced on the supercomputers of the Leibniz Supercomputing Centre and visualized using the hybrid parallel (MPI + TBB) ray-tracing engine OSPRay associated with VisIt. Besides revealing features of turbulence with an unprecedented resolution, the visualizations brilliantly showcase the stretching-and-folding mechanisms through which astrophysical processes such as supernova explosions drive turbulence and amplify the magnetic field in the interstellar gas, and how the first structures, the seeds of newborn stars are shaped by this process.

An Accessible Visual Narrative for the Primary Energy Source of Life from the Fulldome Show Birth of Planet Earth
Melih Sener (University of Illinois), Stuart Levy (University of Illinois), AJ Christensen (University of Illinois), Robert Patterson (University of Illinois), Kalina Borkiewicz (University of Illinois), John E. Stone (University of Illinois), Barry Israelewitz (University of Illinois), Jeffrey Carpenter (University of Illinois), Donna Cox (University of Illinois)

Conversion of sunlight into chemical energy, namely photosynthesis, is the primary energy source of life on Earth. An explanatory visualization depicting this process is presented in the form of an excerpt from the fulldome show Birth of Planet Earth. This accessible visual narrative shows a lay audience, especially children, how the energy of sunlight is captured, converted, and stored through a chain of proteins to power living cells. The visualization is the result of a multi-year collaboration
among biophysicists, visualization scientists, and artists, which, in turn, is based on a decade-long experimental-computational collaboration on structural and functional modeling that produced an atomic detail description of a bacterial bioenergetic organelle, the chromatophore. The energy conversion steps depicted feature an integration of function from electronic to cell levels, spanning nearly 12 orders of magnitude in time scales modeled with multi-scale computational approaches. This atomic detail description uniquely enables a modern retelling of one of humanity's earliest stories---the interplay between light and life.

**Visualizing Supersonic Retropropulsion for Mars: The Summit Early Science Results**
Patrick Moran (NASA Ames Research Center), Timothy Sandstrom (NASA Ames Research Center), Ashley Korzun (NASA Langley Research Center), Eric Nielsen (NASA Langley Research Center), Aaron Walden (NASA Langley Research Center)

As part of the Oak Ridge Leadership Computing Facility (OLCF) Early Science program, NASA ran a series of simulations on the new, GPU-accelerated system Summit, using the flow solver FUN3D to produce solutions for a set of representative descent conditions. The raw output from these simulations was over a petabyte of data. We produced visualizations at the NASA Advanced Supercomputing Division at Ames Research Center using parallel techniques where feasible to accelerate the process. Included with this summary is an animation with some of the highlights of our visualizations so far. We discuss some of our visualization design choices, and plans for future work.
Reception

**Monday, November 18**

7:00 pm - 9:00 pm

**Gala Opening Reception**

**Session Description:** SC19 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 19**

5:15 pm - 7:00 pm

**Poster Reception**
Poster 74: Enabling Code Portability of a Parallel and Distributed Smooth-Particle Hydrodynamics Application, FleCSPH

Suyash Tandon (University of Michigan), Nicholas Stegmeier (University of Illinois), Vasu Jaganath (University of Wyoming), Jennifer Ranta (Michigan State University), Rathish Ratnasingam (Newcastle University), Elizabeth Carlson (University of Nebraska), Julien Loiseau (Los Alamos National Laboratory), Vinay Ramakrishnaiah (Los Alamos National Laboratory), Robert Pavel (Los Alamos National Laboratory)

Core-collapse supernovae (CCSNe) are integral to the formation and distribution of heavy elements across the universe. However, CCSNe are highly complex and inherently non-linear phenomena. Large-scale simulations of these cosmic events can provide us a glimpse of their hydrodynamic and nucleosynthetic processes which are difficult to observe. To enable these massive numerical simulations on high-performance computing (HPC) centers, this study uses FleCSPH, a parallel and distributed code, based on the smooth-particle hydrodynamics (SPH) formulation. In the recent years, the HPC architecture has evolved and the next generation of exascale computers are expected to feature heterogenous architecture. Therefore, it is important to maintain code portability across platforms. This work demonstrates code portability of FleCSPH through the incorporation of Kokkos C++ library and containers using Charliecloud.

Best Poster Finalist: no

Poster 73: Accelerating Large-Scale GW Calculations on Hybrid CPU-GPU Architectures

Mauro Del Ben (Lawrence Berkeley National Laboratory), Charlene Yang (National Energy Research Scientific Computing Center (NERSC)), Felipe Jornada (University of California, Berkeley; Lawrence Berkeley National Laboratory), Steven G. Louie (University of California, Berkeley; Lawrence
In this poster, we present the strategy, progress, and performance while GPU porting one of the major modules, epsilon, of the electronic structure code BerkeleyGW. Epsilon represents the most time-consuming routines in the BerkeleyGW workflow for large-scale material science simulations. Some of the porting/optimization strategies include, changing our original data layout to efficiently use libraries such as cuBLAS and cuFFT, implementation of specific CUDA kernels to minimize data copies between host/device and keeping data on device, efficient use of data streams to leverage high concurrency on the device, asynchronous memory copies and overlapping (MPI) communication on the host and computation on the device. Preliminary results are presented in terms of the speedup compare to the CPU-only implementation, strong/weak scaling, and power efficiency. Excellent acceleration is demonstrated: up to 30x for specific kernels. Our port also exhibits good scalability and about 16x higher FLOPs/watt efficiency compared to the CPU-only implementation.

Best Poster Finalist: no

**Poster 89: BeeCWL: A CWL Compliant Workflow Management System**
Betis Baheri (Kent State University), Steven Anaya (New Mexico Institute of Mining and Technology), Patricia Grubel (Los Alamos National Laboratory), Qiang Guan (Kent State University), Timothy Randles (Los Alamos National Laboratory)

Scientific workflows are used widely to carry out complex and hierarchical experiments. Although there are many trends to extend the functionality of workflow management systems to cover all possible requirements that may arise from a user community, one unified standard over cloud and HPC systems is still missing. In this paper, we propose a Common Workflow Language (CWL) compliant workflow management system. BeeCWL is a parser to derive meaningful information such as requirements, steps, relationships, etc. from CWL files and to create a graph database from those components. Generated graphs can be passed to an arbitrary scheduler and management system to decide whether there are enough resources to optimize and execute the workflow. Lastly, the user can have control over workflow execution, collecting logs, and restart or rerun some part of a complex workflow.

Best Poster Finalist: no

**Poster 150: A Machine Learning Approach to Understanding HPC Application Performance**
Variation

Burak Aksar (Boston University, Sandia National Laboratories), Benjamin Schwaller (Sandia National Laboratories), Omar Aaziz (Sandia National Laboratories), Emre Ates (Boston University), Jim Brandt (Sandia National Laboratories), Ayse K. Coskun (Boston University), Manuel Egele (Boston University), Vitus Leung (Sandia National Laboratories)

Performance anomalies are difficult to detect because often a “healthy system” is vaguely defined, and the ground truth for how a system should be operating is evasive. As we move to exascale, however, detection of performance anomalies will become increasingly important with the increase in size and complexity of systems. There are very few accepted ways of detecting anomalies in the literature, and there are no published and labeled sets of anomalous HPC behavior. In this research, we develop a suite of applications that represent HPC workloads and use data from a lightweight metric collection service to train machine learning models to predict the future behavior of metrics. In the future, this work will be used to predict anomalous runs in compute nodes and determine some root causes of performance issues to help improve the efficiency of HPC system administrators and users.

Best Poster Finalist: no

Poster 81: Performance of Devito on HPC-Optimised ARM Processors

Hermes Senger (Federal University of São Carlos, Brazil; University of São Paulo), Jaime Freire de Souza (Federal University of São Carlos, Brazil), Edson Satoshi Gomi (University of São Paulo), Fabio Luporini (Imperial College, London), Gerard Gorman (Imperial College, London)

We evaluate the performance of Devito, a domain specific language (DSL) for finite differences on Arm ThunderX2 processors. Experiments with two common seismic computational kernels demonstrate that Devito can apply automatic code generation and optimization across Arm and Intel platforms. The code transformations include: parallelism, and SIMD vectorization (OpenMP >=4); loop tiling (with best block shape obtained via auto-tuning); domain-specific symbolic optimisation such as common sub-expression elimination and factorisation for Flop reduction, polynomial approximations for trigonometry terms, and heuristic hoisting of time-invariant expressions. Results show that Devito can achieve performance on Arm processors which is competitive to other Intel Xeon processors.

Best Poster Finalist: no

Poster 103: LIKWID 5: Lightweight Performance Tools
LIKWID is a tool suite for performance oriented programmers with a worldwide user group. It is developed by the HPC group of the University Erlangen-Nuremberg since 2009 to support them in their daily research and performance engineering of user codes. The HPC landscape has become more and more diverse over the last years with clusters using non-x86 architectures and being equipped with accelerators. With the new major version, the architectural support of LIKWID is extended to ARM and POWER CPUs with the same functionality and features as for x86 architectures. Besides the CPU monitoring, the new version provides access the hardware counting facilities of Nvidia GPUs. This poster introduces the new features and shows the successes of applying LIKWID to identify performance bottlenecks and to test optimizations. Furthermore, the poster gives an overview of how users can integrate the LIKWID tools in their application using a lightweight add-once-and-reuse instrumentation API.

Best Poster Finalist: no

**Poster 149: Solving Phase-Field Equations in Space-Time: Adaptive Space-Time Meshes and Stabilized Variational Formulations**

*Kumar Saurabh (Iowa State University), Biswajit Khara (Iowa State University), Milinda Fernando (University of Utah), Masado Ishii (University of Utah), Hari Sundar (University of Utah), Baskar Ganapathysubramanian (Iowa State University)*

We seek to efficiently solve a generalized class of partial differential equations called the phase-field equations. These non-linear PDE’s model phase transition (solidification, melting, phase-separation) phenomena which exhibit spatially and temporally localized regions of steep gradients. We consider time as an additional dimension and simultaneously solve for the unknown in large blocks of time (i.e. in space-time), instead of the standard approach of sequential time-stepping. We use variational multiscale (VMS) based finite element approach to solve the ensuing space-time equations. This allows us to (a) exploit parallelism not only in space but also in time, (b) gain high order accuracy in time, and (c) exploit adaptive refinement approaches to locally refine region of interest in both space and time. We illustrate this approach with several canonical problems including melting and solidification of complex snow flake structures.

Best Poster Finalist: no
Poster 84: ESTEE: A Simulation Toolkit for Distributed Workflow Execution
Vojtěch Cima (IT4Innovations, Czech Republic), Jakub Beránek (IT4Innovations, Czech Republic), Stanislav Böhm (IT4Innovations, Czech Republic)

Task graphs provide a simple way to describe scientific workflows (sets of tasks with dependencies) that can be executed on both HPC clusters and in the cloud. An important aspect of executing such graphs is the used scheduling algorithm. Many scheduling heuristics have been proposed in existing works; nevertheless, they are often tested in oversimplified environments. We introduce a simulation environment designed for prototyping and benchmarking task schedulers. Our simulation environment, scheduler source codes, and graph datasets are open in order to be fully reproducible. To demonstrate usage of Estee, as an example, we compare the performance of various workflow schedulers in an environment using two different network models.

Best Poster Finalist: no

Poster 93: Robust Data-Driven Power Simulator for Fast Cooling Control Optimization of a Large-Scale Computing System
Takashi Shiraishi (Fujitsu Laboratories Ltd), Hiroshi Endo (Fujitsu Laboratories Ltd), Takaaki Hineno (Fujitsu Laboratories Ltd), Hiroyuki Fukuda (Fujitsu Laboratories Ltd)

Power of large-scale systems such as an HPC or a datacenter is a significant issue. Cooling units consume 30% of the total power. General control policies for cooling units are local and static (manual overall optimization nearly once a week). However, free cooling and IT-load fluctuation may change hourly optimum control variables of the cooling units. In this work, we present a deep neural network (DNN) power simulator that can learn from actual operating logs and can quickly identify the optimum control variables. We demonstrated the power simulator of an actual large-scale system with 4.7-MW-power IT load. Our robust simulator predicted the total power with error of 4.8% without retraining during one year. We achieved optimization by the simulator within 80 seconds that was drastically faster than previous works. The dynamic control optimization each hour showed a 15% power reduction compared to that of conventional policy in the actual system.

Best Poster Finalist: no

Poster 115: sDNA: Software-Defined Network Accelerator Based on Optical Interconnection Architecture
En Shao (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences), Guangming Tan (Institute of Computing Technology, Chinese Academy of Sciences;
Software-Defined Network Accelerator (sDNA) is a new accelerated system for the exascale computer. Inspired by the edge forwarding index (EFI), the main contribution of our work is that it presents an extended EFI-based optical interconnection method with slow switching optical device. In our work, we found that sDNA based on extended EFI evaluation is not only able to offload the traffic from an electrical link to an optical link but is also able to avoid congestion inherent to electrical link.

Poster 49: WarpX: Toward Exascale Modeling of Plasma Particle Accelerators on GPU
Maxence Thevenet (Lawrence Berkeley National Laboratory), Jean-Luc Vay (Lawrence Berkeley National Laboratory), Ann Almgren (Lawrence Berkeley National Laboratory), Diana Amorim (Lawrence Berkeley National Laboratory), John Bell (Lawrence Berkeley National Laboratory), Axel Huebl (Lawrence Berkeley National Laboratory), Revathi Jambunathan (Lawrence Berkeley National Laboratory), Rémi Lehe (Lawrence Berkeley National Laboratory), Andrew Myers (Lawrence Berkeley National Laboratory), Jaehong Park (Lawrence Berkeley National Laboratory), Olga Shapoval (Lawrence Berkeley National Laboratory), Weiqun Zhang (Lawrence Berkeley National Laboratory), Lixin Ge (SLAC National Accelerator Laboratory), Mark Hogan (SLAC National Accelerator Laboratory), Cho Ng (SLAC National Accelerator Laboratory), David Grote (Lawrence Livermore National Laboratory)

Particle accelerators are a vital part of the DOE-supported infrastructure of discovery science and applications, but we need game-changing improvements in the size and cost for future accelerators. Plasma-based particle accelerators stand apart in their potential for these improvements. Turning this from a promising technology into mainstream scientific tools depends critically on high-performance, high-fidelity modeling of complex processes.

WarpX is an open-source particle-in-cell (PIC) code supported by the Exascale Computing Project (ECP) that is combining advanced algorithms with adaptive mesh refinement to allow challenging simulations of a multi-stage plasma-based TeV acceleration relevant for future high-energy physics discoveries. WarpX relies on the ECP co-design center for mesh refinement AMReX, and runs on CPU and GPU-accelerated computers. Production simulation have run on Cori KNL at NERSC and Summit at OLCF. In this poster, recent results and strategies on GPU will be presented, along with
Poster 50: Implementing an Adaptive Sparse Grid Discretization (ASGarD) for High Dimensional Advection-Diffusion Problems on Exascale Architectures

M. Graham Lopez (Oak Ridge National Laboratory), David L. Green (Oak Ridge National Laboratory), Lin Mu (University of Georgia), Ed D'Azevedo (Oak Ridge National Laboratory), Wael Elwasif (Oak Ridge National Laboratory), Tyler McDaniel (University of Tennessee), Timothy Younkin (University of Tennessee), Adam McDaniel (Oak Ridge National Laboratory), Diego Del-Castillo-Negrete (Oak Ridge National Laboratory)

Many scientific domains require the solution of high dimensional PDEs. Traditional grid- or mesh-based methods for solving such systems in a noise-free manner quickly become intractable due to the scaling of the degrees of freedom going as $O(N^d)$ sometimes called "the curse of dimensionality." We are developing an arbitrarily high-order discontinuous-Galerkin finite-element solver that leverages an adaptive sparse-grid discretization whose degrees of freedom scale as $O(N\log_2 N^{D-1})$. This method and its subsequent reduction in the required resources is being applied to several PDEs including time-domain Maxwell's equations (3D), the Vlasov equation (in up to 6D) and a Fokker-Planck-like problem in ongoing related efforts. Here we present our implementation which is designed to run on multiple accelerated architectures, including distributed systems. Our implementation takes advantage of a system matrix decomposed as the Kronecker product of many smaller matrices which is implemented as batched operations.

Best Poster Finalist: no

Poster 91: FreeCompilerCamp: Online Training for Extending Compilers

Alok Mishra (Stony Brook University, Lawrence Livermore National Laboratory), Anjia Wang (University of North Carolina, Charlotte; Lawrence Livermore National Laboratory), Chunhua Liao (Lawrence Livermore National Laboratory), Yonghong Yan (University of North Carolina, Charlotte), Barbara Chapman (Stony Brook University)

In this presentation, we introduce an ongoing effort of an online training platform aimed to automate the training of developers to quickly extend compilers - FreeCompilerCamp.org. Our free and open platform allows anyone who is interested in developing compilers to learn the necessary skills. A live training website, built on top of Play-With-Docker, is set up so that anyone with internet access and a web browser will be able to take this training. The entire training system is open-source and
developers with relevant skills can contribute new tutorials and deploy it on a private server, workstation or even laptop. We have created some initial tutorials on how to extend the Clang/LLVM or ROSE compilers to support new OpenMP features. Using a web interface consisting of two side-by-side panels, users can follow the tutorials on one side and immediately practice what they learned in a terminal sandbox embedded on the other.

Best Poster Finalist: yes

Poster 51: SmartK: Efficient, Scalable, and Winning Parallel MCTS
Michael S. Davinroy (Swarthmore College), Shawn Pan (Swarthmore College), Bryce Wiedenbeck (Swarthmore College, Davidson College), Tia Newhall (Swarthmore College)

SmartK is our efficient and scalable parallel algorithm for Monte Carlo Tree Search (MCTS), an approximation technique for game searches. MCTS is also used to solve problems as diverse as planning under uncertainty, combinatorial optimization, and high-energy physics. In these problems, the solution search space is significantly large, necessitating parallel solutions. Shared memory parallel approaches do not scale well beyond the size of a single node's RAM. SmartK is a distributed memory parallelization that takes advantage of both inter-node and intra-node parallelism and a large cumulative RAM found in clusters. SmartK’s novel selection algorithm combined with its ability to efficiently search the solution space, results in better solutions than other MCTS parallel approaches. Results of an MPI implementation of SmartK for the game of Hex, show SmartK yields a better win percentage than other parallel algorithms, and that its performance scales to larger search spaces and high degrees of parallelism.

Best Poster Finalist: no

Poster 70: Numerical Method and Parallelization for the Computation of Coherent Synchrotron Radiation
Boqian Shen (Rice University, Los Alamos National Laboratory)

The purpose of this work is to develop and parallelize an accurate and efficient numerical method for the computation of synchrotron radiation from relativistic electrons in the near field. The high-brilliance electron beam and coherent short-wavelength light source provide a powerful method to understand the microscopic structure and dynamics of materials. Such a method supports a wide range of applications including matter physics, structural biology, and medicine development. To understand the interaction between the beam and synchrotron radiation, an accurate and efficient numerical simulation is needed. With millions of electrons, the computational cost of the field would
be large. Thus, multilevel parallelism and performance portability are desired since modern supercomputers are getting more complex and heterogeneous. The performance model and performance analysis are presented.

Best Poster Finalist: no

**Poster 146: AI Matrix: A Deep Learning Benchmark for Alibaba Data Centers**

Wei Zhang (Alibaba Inc), Wei Wei (Alibaba Inc), Lingjie Xu (Alibaba Inc), Lingling Jin (Alibaba Inc), Cheng Li (University of Illinois)

This work introduces AI Matrix, an in-house Deep Learning (DL) benchmark suite developed specifically for Alibaba's e-commerce environment. AI Matrix results from a full investigation of the DL applications used inside Alibaba and aims to cover the typical DL applications that account for more than 90% of the GPU usage in Alibaba data centers. This benchmark suite collects DL models that are either directly used or closely resemble the models used in the company's real e-commerce applications. It also collects the real e-commerce applications if no similar DL models are not available. Through the high coverage and close resemblance to real applications, AI Matrix fully represents the DL workloads on Alibaba data centers. The collected benchmarks mainly fall into three categories: computer vision, recommendation, and language processing, which consist of the most majority of DL applications in Alibaba. AI Matrix is made open source, hoping it can benefit the public.

Best Poster Finalist: no

**Poster 54: Massively Parallel Eigensolvers Based on Unconstrained Energy Functionals Methods**

Osni Marques (Lawrence Berkeley National Laboratory), Mauro Del Ben (Lawrence Berkeley National Laboratory), Andrew Canning (Lawrence Berkeley National Laboratory)

This poster focuses on a preconditioned conjugate gradient based iterative eigensolver using an unconstrained energy functional minimization scheme. This scheme avoids an explicit reorthogonalization of the trial eigenvectors and becomes an attractive alternative for the solution of very large problems. The unconstrained formulation is implemented in the first-principles materials and chemistry CP2K code, which performs electronic structure calculations based on a density functional theory approximation to the solution of the many-body Schrödinger equation. The systems we use in our studies have a number of atoms ranging from 2,247 to 12,288. We study the convergence of the unconstrained formulation and its scaling on a Cray XC40 (a partition with 9,688 Intel KNL nodes). We show that there is a trade-off between the preconditioner that leads to fast convergence and lower cost preconditioners that lead to best time to solution.
Array management libraries, such as HDF5, Zarr, etc., depend on a complex software stack that consists of parallel I/O middleware (MPI-IO), POSIX-IO, and file systems. Components in the stack are interdependent, such that effort in tuning the parameters in these software libraries for optimal performance is non-trivial. On the other hand, it is challenging to choose an array management library based on the array configuration and access patterns. In this poster, we investigate the performance aspect of two array management libraries, i.e., HDF5 and Zarr, in the context of a neuroscience use case. We highlight the performance variability of HDF5 and Zarr in our preliminary results and discuss potential optimization strategies.
Distributed deep learning using a large mini-batch is a key technology to accelerate training in deep learning. However, it is difficult to achieve a high scalability and maintain validation accuracy in distributed learning on large clusters. We introduce two optimizations, reducing the computation time and overlapping the communication with the computation. By applying the techniques and using 2,048 GPUs, we achieved the world’s fastest ResNet-50 training in MLPerf, which is a de facto standard DNN benchmark (as of July 2019).

Best Poster Finalist: no

**Poster 111: Multiple HPC Environments-Aware Container Image Configuration for Bioinformatics Application**

Kento Aoyama (Tokyo Institute of Technology, National Institute of Advanced Industrial Science and Technology (AIST)), Hiroki Watanabe (Tokyo Institute of Technology, National Institute of Advanced Industrial Science and Technology (AIST)), Masahito Ohue (Tokyo Institute of Technology), Yutaka Akiyama (Tokyo Institute of Technology)

Containers have a considerable advantage for application portability in different environments by isolating process with a small performance overhead; thus it has been rapidly getting popular in a wide range of science fields. However, there are problems in container image configuration when run in multiple HPC environments, and it requires users to have knowledge of systems, container runtimes, container image format, and library compatibilities in HPC environments.

In this study, we introduce our HPC container workflow in multiple supercomputing environments that have different system/library specifications (ABCI, TSUBAME3.0). Our workflow provides custom container image configurations for HPC environments by taking into account differences in container runtime, container image, and library compatibility between the host and inside of the container. We also show the parallel performance of our application in each HPC environment.

Best Poster Finalist: no

**Poster 134: Minimal-Precision Computing for High-Performance, Energy-Efficient, and Reliable Computations**

Daichi Mukunoki (RIKEN Center for Computational Science (R-CCS)), Toshiyuki Imamura (RIKEN Center for Computational Science (R-CCS)), Yiyu Tan (RIKEN Center for Computational Science (R-CCS))
In numerical computations, the precision of floating-point computations is a key factor to determine the performance (speed and energy-efficiency) as well as the reliability (accuracy and reproducibility). However, the precision generally plays a contrary role for both. Therefore, the ultimate concept for maximizing both at the same time is the minimal-precision computation through precision-tuning, which adjusts the optimal precision for each operation and data. Several studies have been already conducted for it so far, but the scope of those studies is limited to the precision-tuning alone. In this study, we propose a more broad concept of the minimal-precision computing with precision-tuning, involving both hardware and software stack.

Best Poster Finalist: no

**Poster 112: Building Complex Software Applications Inside Containers**

*Calvin D. Seamons (Los Alamos National Laboratory)*

High performance computing (HPC) scientific applications require complex dependencies to operate. As user demand for HPC systems increases, it becomes unrealistic to support every unique dependency request. Containers can offer the ability to satisfy the users’ dependency request while simultaneously offering HPC portability across systems. By “containerizing” Model for Prediction Across Scales (MPAS, a large atmospheric simulation suite), we show that it is possible to containerize and run complex software. Furthermore, the container can be run across different HPC systems with nearly identical results (21 bytes difference over 2.1 gigabytes). Containers have the possibility to bring flexibility to code teams in HPC by helping to meet the demand for user defined software stacks (UDSS), and giving teams the ability to choose their software, independently of what is offered by the HPC system.

Best Poster Finalist: no

**Poster 101: Job Performance Overview of Apache Flink and Apache Spark Applications**

*Jan Frenzel (Technical University Dresden), René Jäkel (Technical University Dresden)*

Apache Spark and Apache Flink are two Big Data frameworks used for fast data exploration and
analysis. Both frameworks provide the runtime of program sections and performance metrics, such as the number of bytes read or written, via an integrated dashboard. Performance metrics available in the dashboard lack timely information and are only shown aggregated in a separate part of the dashboard. However, performance investigations and optimizations would benefit from an integrated view with detailed performance metric events. Thus, we propose a system that samples metrics at runtime and collects information about the program sections after the execution finishes. The performance data is stored in an established format independent from Spark and Flink versions and can be viewed with state-of-the-art performance tools, i.e. Vampir. The overhead depends on the sampling interval and was below 10% in our experiments.

Best Poster Finalist: no

**Poster 113: Improvements Toward the Release of the Pavilion 2.0 Test Harness**

Kody J. Everson (Los Alamos National Laboratory, Dakota State University), Maria Francine Lapid (Los Alamos National Laboratory)

High-performance computing production support entails thorough testing in order to evaluate the efficacy of a system for production-grade workloads. There are various phases of a system’s life-cycle to assess, requiring different methods to accomplish effective evaluation of performance and correctness. Due to the unique and distributed nature of an HPC-system, the necessity for sophisticated tools to automatically harness and assess test results, all while interacting with schedulers and programming environment software, requires a customizable, extensible, and lightweight system to manage concurrent testing. Beginning with the recently refactored codebase of Pavilion 1.0, we assisted with the finishing touches on readying this software for open-source release and production usage. Pavilion 2.0 is a Python 3-based testing framework for HPC clusters that facilitates the building, running, and analysis of tests through an easy-to-use, flexible, YAML-based configuration system. This enables users to write their own tests by simply wrapping everything in Pavilion’s well-defined format.

Best Poster Finalist: no

**Poster 119: Toward Lattice QCD on Fugaku: SVE Compiler Studies and Micro-Benchmarks in the RIKEN Fugaku Processor Simulator**

Nils Meyer (University of Regensburg, Bavaria), Tilo Wettig (University of Regensburg, Bavaria), Yuetsu Kodama (RIKEN Center for Computational Science (R-CCS)), Mitsuhisa Sato (RIKEN Center for Computational Science (R-CCS))
The Fugaku supercomputer, successor to the Japanese flagship K-Computer, will start operation in 2021. Fugaku incorporates the Fujitsu A64FX processor, which is the first hardware implementation supporting the Arm SVE instruction set, in this case a 512-bit version. Real hardware is not accessible today, but RIKEN has designed a simulator of the A64FX. We present micro-benchmarks relevant for Lattice QCD obtained in the RIKEN Fugaku processor simulator and compare three different SVE compilers.

Best Poster Finalist: no

Poster 62: Emulating Multi-Pattern Quantum Grover’s Search on a High-Performance Reconfigurable Computer
Naveed Mahmud (University of Kansas), Bennett Haase-Divine (University of Kansas), Bailey K. Srimoungchanh (University of Kansas), Nolan Blankenau (University of Kansas), Annika Kuhnke (University of Kansas), Esam El-Araby (University of Kansas)

Grover’s search (GS) is a widely studied quantum algorithm that can be employed for both single and multi-pattern search problems and potentially provides quadratic speedup over existing classical search algorithms. In this paper, we propose a multi-pattern quantum search methodology based on a modified GS quantum circuit. The proposed method combines classical post-processing permutations with a modified Grover’s circuit to efficiently search for given single/multiple input patterns. Our proposed methodology reduces quantum circuit complexity, realizes space-efficient emulation hardware and improves overall system configurability for dynamic, multi-pattern search. We use a high-performance reconfigurable computer to emulate multi-pattern GS (MGS) and present scalable emulation architectures of a complete multi-pattern search system. We validate the system and provide analysis of experimental results in terms of FPGA resource utilization and emulation time. Our results include a successful hardware architecture that is capable of emulating MGS algorithm up to 32 fully-entangled quantum bits on a single FPGA.

Best Poster Finalist: no

Poster 107: Exploring Interprocess Work Stealing for Balanced MPI Communication
Kaiming Ouyang (University of California, Riverside), Min Si (Argonne National Laboratory), Zizhong Chen (University of California, Riverside)

Workload balance among MPI processes is a critical consideration during the development of HPC applications. However, because of many factors such as complex network interconnections and irregularity of HPC applications, fully achieving workload balance in practice is nearly impossible.
Although interprocess job stealing is a promising solution, existing shared-memory techniques that lack necessary flexibility or cause inefficiency during data access cannot provide an applicable job-stealing implementation. To solve this problem, we propose a new process-in-process (PiP) interprocess job-stealing method to balance communication workload among processes on MPI layers. Our initial experimental results show PiP-based job stealing can efficiently help amortize workload, reduce imbalance, and greatly improve intra- and intersocket ping-pong performance compared with original MPI.

Best Poster Finalist: no

**Poster 127: sFlow Monitoring for Security and Reliability**

Xava A. Grooms (Los Alamos National Laboratory, University of Kentucky), Robert V. Rollins (Los Alamos National Laboratory, Michigan Technological University), Collin T. Rumpca (Los Alamos National Laboratory, Dakota State University)

In the past ten years, High Performance Computing (HPC) has moved far beyond the terascale performance, making petascale systems the new standard. The drastic improvement in performance has been largely unmatched with insignificant improvements in system monitoring. Thus, there is an immediate need for practical and scalable monitoring solutions to ensure the effectiveness of costly compute clusters. This project aims to explore the viability and impact of sFlow enabled switches in cluster network monitoring for security and reliability. A series of tests and exploits were performed to target specific network abnormalities on a nine-node HPC cluster. The results present web-based dashboards that can aid network administrators in improving a cluster’s security and reliability.

Best Poster Finalist: no

**Poster 77: Extreme Scale Phase-Field Simulations of Sintering Processes**

Johannes Hötzer (Karlsruhe University of Applied Sciences), Henrik Hierl (Karlsruhe University of Applied Sciences), Marco Seiz (Karlsruhe Institute of Technology), Andreas Reiter (Karlsruhe Institute of Technology), Britta Nestler (Karlsruhe Institute of Technology)

The sintering process, which turns loose powders into dense materials, is naturally found in the formation of glaciers, but is also the indispensable process to manufacture ceramic materials. This process is described by a dynamically evolving microstructure, which largely influences the resulting material properties.

To investigate this complex three-dimensional, scale-bridging evolution in realistic domain sizes, a
highly optimized and parallelized multiphysics phase-field solver is developed. The solver is optimized in a holistic way, from the application level over the time integration and parallelization, down to the hardware. Optimizations include communication hiding, explicit vectorization, implicit schemes, and local reduction of degrees of freedom.

With this, we are able to investigate large-scale, three-dimensional domains, and long integration times. We have achieved a single-core peak performance of 32.5%, scaled up to 98304 cores on Hazel Hen and SuperMUC-NG, and simulated a multimillion particle system.

Best Poster Finalist: no

**Poster 140: Toward Automatic Function Call Generation for Deep Learning**
Shizhi Tang (Tsinghua University, China), Jidong Zhai (Tsinghua University, China)

Mainstream deep learning frameworks are commonly implemented by invoking underlying high performance tensor libraries on various architectures. However, as these libraries provide increasingly complex semantics including operator fusions, in-place operations, and various memory layouts, the gap between mathematical deep learning models and the underlying libraries becomes larger. In this paper, inspired by the classic problem of Instruction Selection, we design a theorem solver guided exhausted search algorithm to select functions for complex tensor computations. Preliminary results with some micro-benchmarks and a real model show that our approach can outperform both Tensorflow and Tensor Comprehensions at run time.

Best Poster Finalist: no

**Poster 83: ETL: Elastic Training Layer for Deep Learning**
Lei Xie (Tsinghua University, China), Jidong Zhai (Tsinghua University, China)

Due to the rising of deep learning, clusters for deep learning training are widely deployed in production. However, static task configuration and resource fragmentation problems in existing clusters result in low efficiency and poor quality of service. We propose ETL, an elastic training layer for deep learning, to help address them once for all. ETL adopts many novel mechanisms, such as lightweight and configurable report primitive and asynchronous, parallel and IO-free state replication, to achieve both high elasticity and efficiency. The evaluation demonstrates the low overhead and high efficiency of these mechanisms and reveals the advantages of elastic deep learning supported by ETL.
Poster 130: Deep Learning-Based Feature-Aware Data Modeling for Complex Physics Simulations
Qun Liu (Louisiana State University), Subhashis Hazarika (Ohio State University), John M. Patchett (Los Alamos National Laboratory), James P. Ahrens (Los Alamos National Laboratory), Ayan Biswas (Los Alamos National Laboratory)

Data modeling and reduction for in situ is important. Feature-driven methods for in situ data analysis and reduction are a priority for future exascale machines as there are currently very few such methods. We investigate a deep-learning-based workflow that targets in situ data processing using autoencoders. We employ integrated skip connections to obtain higher performance compared to the existing autoencoders. Our experiments demonstrate the initial success of the proposed framework and create optimism for the in situ use case.

Poster 125: Physics Informed Generative Adversarial Networks for Virtual Mechanical Testing
Julian Cuevas (NASA, University of Puerto Rico at Mayaguez), Patrick Leser (NASA), James Warner (NASA), Geoffrey Bomarito (NASA), William Leser (NASA)

Physics-informed generative adversarial networks (PI-GANs) are used to learn the underlying probability distributions of spatially-varying material properties (e.g., microstructure variability in a polycrystalline material). While standard GANs rely solely on data for training, PI-GANs encode physics in the form of stochastic differential equations using automatic differentiation. The goal here is to show that experimental data from a limited number of material tests can be used with PI-GANs to enable unlimited virtual testing for aerospace applications. Preliminary results using synthetically generated data are provided to demonstrate the proposed framework. Deep learning and automatic differentiation capabilities in Tensorflow were implemented on Nvidia Tesla V100 GPUs.

Poster 141: ExaGeoStatR: Harnessing HPC Capabilities for Large Scale Geospatial Modeling Using R
Sameh Abdullah (King Abdullah University of Science and Technology (KAUST)), Yuxiao Li (King Abdullah University of Science and Technology (KAUST)), Jian Cao (King Abdullah University of
Large-scale simulations and parallel computing techniques are becoming essential in Gaussian process calculations to lessen the complexity of geostatistics applications. The log-likelihood function is used in such applications to evaluate the model associated with a given set of measurements in existing n geographic locations. The evaluation of such a function requires $O(n^2)$ memory and $O(n^3)$ computation, which is infeasible for large datasets with existing software tools.

We present ExaGeoStatR, a package for large-scale geostatistics in R that computes the log-likelihood function on shared and distributed-memory, possibly equipped with GPU, using advanced linear algebra techniques. The package provides a high-level abstraction of the underlying architecture while enhancing the R developers' productivity. We demonstrate ExaGeoStatR package by illustrating its implementation details, analyzing its performance on various parallel architectures, and assessing its accuracy using synthetic datasets and a sea surface temperature dataset. The performance evaluation involves spatial datasets with up to 250K observations.

Poster 48: Runtime System for GPU-Based Hierarchical LU Factorization

Qianxiang Ma (Tokyo Institute of Technology), Rio Yokota (Tokyo Institute of Technology, Global Scientific Information and Computing Center; Tokyo Institute of Technology)

Hierarchical low-rank approximation can reduce both the storage and computation costs of dense matrices, but its implementation is challenging. In this research, we tackle one of the most difficult problems of GPU parallelization of the factorization of these hierarchical matrices. To this end, we are developing a new runtime system for GPUs that can schedule all tasks into one GPU kernel. Other existing runtime systems, like cuGraph and Stanford Legion, can only manage streams and kernel-level parallelism. Even without too much tuning, we achieved 4x better performance in H-LU factorization with a single GPU when comparing with a well-tuned CPU-based hierarchical matrix library, HLIBpro, on moderately sized matrices. Additionally, we have significantly less runtime overheads exposed when processing smaller matrices.

Best Poster Finalist: no
Poster 145: Improving Data Compression with Deep Predictive Neural Network for Time Evolutional Data
Rupak Roy (Florida State University), Kento Sato (RIKEN Center for Computational Science (R-CCS)), Jian Guo (RIKEN Center for Computational Science (R-CCS)), Jens Domke (RIKEN Center for Computational Science (R-CCS)), Weikuan Yu (Florida State University), Takaki Hatsui (RIKEN SPring-8 Center), Yasumasa Joti (Japan Synchrotron Radiation Research Institute)

Scientific applications/simulations periodically generate huge intermediate data. Storing or transferring such a large scale of data is critical. Fast I/O is important for making this process faster. One of the approaches to achieve fast I/O is data compression. Our goal is to achieve a delta technique that can improve the performance of existing data compression algorithms for time evolutional intermediate data.

In our approach, we compute the delta values from original data and data predicted by the deep predictive neural network. We pass these delta values through three phases which are preprocessing phase, partitioned entropy coding phase, and density-based spatial delta encoding phase.

In our poster, we present how our predictive delta technique can leverage the time evolutional data to produce highly concentrated small values. We show the improvement in compression ratio when our technique, combined with existing compression algorithms, are applied on the intermediate data for different datasets.

Best Poster Finalist: no

Poster 144: Optimizing Asynchronous Multi-Level Checkpoint/Restart Configurations with Machine Learning
Tonmoy Dey (Florida State University), Kento Sato (RIKEN Center for Computational Science (R-CCS)), Jian Guo (RIKEN Center for Computational Science (R-CCS)), Bogdan Nicolae (Argonne National Laboratory), Jens Domke (RIKEN Center for Computational Science (R-CCS)), Weikuan Yu (Florida State University), Franck Cappello (Argonne National Laboratory), Kathryn Mohror (Lawrence Livermore National Laboratory)

With the emergence of fast local storage, multi-level checkpointing (MLC) has become a common approach for efficient checkpointing. To utilize MLC efficiently, it is important to determine the optimal configuration for the checkpoint/restart (CR). There are mainly two approaches for determining the optimal configuration for CR, namely modeling and simulation approach. However, with MLC, CR becomes more complicated making the modeling approach inaccurate and the simulation approach though accurate, very slow. In this poster, we focus on optimizing the performance of CR by
predicting the optimized checkpoint count and interval. This was achieved by combining the simulation approach with machine learning and neural network to leverage its accuracy without spending time on simulating different CR parameters. We demonstrate that our models can predict the optimized parameter values with minimal error when compared to the simulation approach.

Best Poster Finalist: no

Poster 132: Optimizing Performance at Runtime Using Binary Rewriting
Alexis Engelke (Technical University Munich), David Hildenbrand (Technical University Munich), Martin Schulz (Technical University Munich)

In addition to scalability, performance of sequential code in applications is an important factor in HPC. Typically, programs are compiled once, at which time optimizations are applied, and are then run several times. However, not all information relevant for performance optimizations are available at compile-time, restricting optimization possibilities. The generation of specialized code at runtime allows for further optimizations. Performing such specialization on binary code allows for initial code to be generated at compile-time with only the relevant parts being rewritten at runtime, reducing the optimization overhead. For targeted optimizations and effective use of known runtime information, the rewriting process needs to be guided by the application itself, exploiting information only known to the developer.

We describe three approaches for self-guided binary rewriting explicitly guided by the running application and evaluate the performance of the optimized code as well as the performance of the rewriting process itself.

Best Poster Finalist: no

Poster 53: Unstructured Mesh Technologies for Fusion Simulations
Cameron Smith (Rensselaer Polytechnic Institute (RPI)), Gerrett Diamond (Rensselaer Polytechnic Institute (RPI)), Gopan Perumpilly (Rensselaer Polytechnic Institute (RPI)), Chonglin Zhang (Rensselaer Polytechnic Institute (RPI)), Agnieszka Truszkowska (Rensselaer Polytechnic Institute (RPI)), Morteza Hakimi (Rensselaer Polytechnic Institute (RPI)), Onkar Sahni (Rensselaer Polytechnic Institute (RPI)), Mark Shephard (Rensselaer Polytechnic Institute (RPI)), Eisung Yoon (Ulsan National Institute of Science and Technology, South Korea), Daniel Ibanez (Sandia National Laboratories)

Multiple unstructured mesh technologies are needed to define and execute plasma physics simulations. The domains of interest combine model features defined from physical fields within 3D
CAD of the tokamak vessel with an antenna assembly, and 2D cross sections of the tokamak vessel. Mesh generation technologies must satisfy these geometric constraints and additional constraints imposed by the numerical models. Likewise, fusion simulations over these domains study a range of timescales and physical phenomena within a tokamak.

XGCm studies the development of plasma turbulence in the reactor vessel, GITRm studies impurity transport, and PetraM simulations model RF wave propagation in scrape off layer plasmas. GITRm and XGCm developments are using the PUMIpic infrastructure to manage the storage and access of non-uniform particle distributions in unstructured meshes on GPUs. PetraM combines PUMI adaptive unstructured mesh control with MFEM using CAD models and meshes defined with Simmetrix tools.

Best Poster Finalist: no

**Poster 99: Eithne: A Framework for Benchmarking Micro-Core Accelerators**

*Maurice C. Jamieson (Edinburgh Parallel Computing Centre, University of Edinburgh), Nick Brown (Edinburgh Parallel Computing Centre, University of Edinburgh)*

Running existing HPC benchmarks as-is on micro-core architectures is at best difficult and most often impossible as they have a number of architectural features that makes them significantly different from traditional CPUs: tiny amounts on-chip RAM (c. 32KB), low-level knowledge specific to each device (including the host / device communications interface), limited communications bandwidth and complex or no device debugging environment. In order to compare and contrast different the micro-core architectures, a benchmark framework is required to abstract much of this complexity.

The modular Eithne framework supports the comparison of a number of micro-core architectures. The framework separates the actual benchmark from the details of how this is executed on the different technologies. The framework was evaluated by running the LINPACK benchmark on the Adapteva Epiphany, PicoRV32 and VectorBlox Orca RISC-V soft-cores, NXP RV32M1, ARM Cortex-A9, and Xilinx MicroBlaze soft-core, and comparing resulting performance and power consumption.

Best Poster Finalist: no

**Poster 133: Portable Resilience with Kokkos**

*Jeffery Miles (Sandia National Laboratories), Nicolas Morales (Sandia National Laboratories), Carson Mould (Sandia National Laboratories), Keita Teranishi (Sandia National Laboratories)*
The Kokkos ecosystem is a programming environment that provides performance and portability to many scientific applications that run on DOE supercomputers as well as other smaller scale systems. Leveraging software abstraction concepts within Kokkos, software resilience for end user code is made portable with abstractions and concepts while implementing the most efficient resilience algorithms internally. This addition enables an application to manage hardware failures reducing the cost of interruption without drastically increasing the software maintenance cost. Two main resilience methodologies have been added to the Kokkos ecosystem to validate the resilience abstractions: 1. Checkpointing includes an automatic mode supporting other checkpointing libraries and a manual mode which leverages the data abstraction and memory space concepts. 2. The redundant execution model anticipates failures by replicating data and execution paths. The design and implementation of these additions are illustrated, and appropriate examples are included to demonstrate the simplicity of use.

Best Poster Finalist: no

Poster 79: The HPC PowerStack: A Community-Wide Collaboration Toward an Energy Efficient Software Stack

Siddhartha Jana (Energy Efficient HPC Working Group, Intel Corporation), Stephanie Brink (Lawrence Livermore National Laboratory), Christopher Cantalupo (Intel Corporation), Jonathan Eastep (Intel Corporation), Masaaki Kondo (RIKEN Advanced Institute for Computational Science (AICS), University of Tokyo), Matthias Maiterth (Intel Corporation), Aniruddha Marathe (Lawrence Livermore National Laboratory), Tapasya Patki (Lawrence Livermore National Laboratory), Barry Rountree (Lawrence Livermore National Laboratory), Ryuichi Sakamoto (University of Tokyo), Martin Schulz (Technical University Munich, Leibniz Supercomputing Centre), Carsten Trinitis (Technical University Munich), Josef Weidendorfer (Technical University Munich, Leibniz Supercomputing Centre)

This poster highlights an ongoing community-wide effort among vendors, labs, and academia, to incorporate power-awareness within system-stacks in upcoming exascale machines. HPC PowerStack is the first-and-only community-driven vendor-neutral effort to identify what power optimization software actors are key within the modern-day stack; discuss their interoperability, and work toward gluing together existing open source projects to engineer cost-effective, but cohesive, portable implementations.

This poster disseminates key insights acquired in the project, provides prototyping status updates, highlights open questions, and solicits participation addressing the imminent exascale power challenge.

Best Poster Finalist: no
**Poster 87: Parallelizing Simulations of Large Quantum Circuits**

Michael A. Perlin (University of Colorado, National Institute of Standards and Technology (NIST)), Teague Tomesh (Princeton University), Bradley Pearlman (University of Colorado, National Institute of Standards and Technology (NIST)), Wei Tang (Princeton University), Yuri Alexeev (Argonne National Laboratory), Martin Suchara (Argonne National Laboratory)

We present a parallelization scheme for classical simulations of quantum circuits. Our scheme is based on a recent method to "cut" large quantum circuits into smaller sub-circuits that can be simulated independently, and whose simulation results can in turn be re-combined to infer the output of the original circuit. The exponentially smaller classical computing resources needed to simulate smaller circuits are counterbalanced by exponential overhead in terms of classical post-processing costs. We discuss how this overhead can be massively parallelized to reduce classical computing costs.

Best Poster Finalist: no

---

**Poster 120: ILP-Based Scheduling for Linear-Tape Model Trapped-Ion Quantum Computers**

Xin-Chuan Wu (University of Chicago), Yongshan Ding (University of Chicago), Yunong Shi (University of Chicago), Yuri Alexeev (Argonne National Laboratory), Hal Finkel (Argonne National Laboratory), Kibaek Kim (Argonne National Laboratory), Frederic T. Chong (University of Chicago)

Quantum computing (QC) is emerging as a potential post-Moore high-performance computing (HPC) technology. Trapped-ion quantum bits (qubits) are among the most leading technologies to reach scalable quantum computers that would solve certain problems beyond the capabilities of even the largest classical supercomputers. In trapped-ion QC, qubits can physically move on the ion trap. The state-of-the-art architecture, linear-tape model, only requires a few laser beams to interact with the entire qubits by physically moving the interacting ions to the execution zone. Since the laser beams are limited resources, the ion chain movement and quantum gate scheduling are critical for the circuit latency. To harness the emerging architecture, we present our mathematical model for scheduling the qubit movements and quantum gates in order to minimize the circuit latency. In our experiment, our scheduling reduces 29.47% circuit latency on average. The results suggest classical HPC would further improve the quantum circuit optimization.

Best Poster Finalist: no
Poster 55: MPI+OpenMP Parallelization of DFT Method in GAMESS
Vladimir Mironov (Lomonosov Moscow State University), Yuri Alexeev (Argonne National Laboratory), Dmitry Fedorov (National Institute of Advanced Industrial Science and Technology (AIST))

In this work, the Density Functional Theory (DFT) method is parallelized with MPI-OpenMP in the quantum chemistry package GAMESS. It has been implemented in both regular and Fragment Molecular Orbital (FMO) based DFT codes. The scalability of the FMO-DFT code was demonstrated on Cray XC40 Theta supercomputer. We demonstrated excellent scalability of the code up to 2,048 Intel Xeon Phi nodes (131,072 cores). Moreover, the developed DFT code is about twice as fast as the original code because of our new grid integration algorithm.

Best Poster Finalist: no

Poster 71: AI-Solver: Uncertainty in Prediction and Error Estimation for AI in Engineering
Ahmed Al-Jarro (Fujitsu Laboratories of Europe Ltd), Loic Beheshti (Fujitsu Laboratories of Europe Ltd), Serban Georgescu (Fujitsu Laboratories of Europe Ltd), Koichi Shirahata (Fujitsu Laboratories Ltd), Yasumoto Tomita (Fujitsu Laboratories Ltd), Nakashima Kouta (Fujitsu Laboratories Ltd)

The AI-Solver is a deep learning platform that learns from simulation data to extract general behavior based on physical parameters. The AI-Solver can handle a wide variety of classes of problems including those commonly identified in FEA, CFD and CEM, to name a few, with speedups of up to 250,000X and extremely low error rate of 2-3%. In this work, we build on this recent effort. We first integrate uncertainty quantification, via exploiting the approximation of Bayesian Deep Learning. Second, we develop bespoke error estimation mechanisms capable of processing this uncertainty to provide instant feedback on the confidence in predictions without relying on the availability of ground truth data. To our knowledge, the ability to estimate the discrepancy in predictions without labels is a first in the field of AI for Engineering.

Best Poster Finalist: no

Aaron Scheinberg (Princeton Plasma Physics Laboratory), Guangye Chen (Los Alamos National Laboratory), Stephane Ethier (Princeton Plasma Physics Laboratory), Stuart Slattery (Oak Ridge National Laboratory), Robert Bird (Los Alamos National Laboratory), Pat Worley (PHWorley Consulting), Choong-Seock Chang (Princeton Plasma Physics Laboratory)

Numerical plasma physics models such as the particle-in-cell XGC code are important tools to
understand phenomena encountered in experimental fusion devices. Adequately resolved simulations are computationally expensive, so optimization is essential. To address the need for consistent high performance by cutting-edge scientific software applications, frameworks such as Kokkos have been developed to enable portability as new architectures require hardware-specific coding implementation for best performance. Cabana, a recent extension to Kokkos developed with the ECP-CoPA project, is a library of common kernels and operations typically necessary for particle-based codes. The Kokkos/Cabana framework enables intuitive construction of particle-based codes, while maintaining portability between architectures. Here, we summarize the adoption by XGC of the execution and data layout patterns offered by this framework. We demonstrate a method for Fortran codes to adopt Kokkos and show that it can provide a single, portable code base that performs well on both GPUs and multicore machines.

**Poster 106: Optimizing Hybrid Access Virtual Memory System Using SCM/DRAM Unified Memory Management Unit**

Yusuke Shirota (Toshiba Corporation), Shiyo Yoshimura (Toshiba Corporation), Satoshi Shirai (Toshiba Corporation), Tatsunori Kanai (Toshiba Corporation)

In HPC systems, expectations for storage-class memory (SCM) are increasing in large-scale in-memory processing. While SCM can deliver higher capacity and lower standby power than DRAM, it is slower and the dynamic power is higher. Therefore, in order to realize high-speed, low-power and scalable main memory, it is necessary to build an SCM/DRAM unified memory, and dynamically optimize data placement between the two memories according to the memory access pattern.

In this poster, we describe a new hybrid access type virtual memory method using TLB-extended unified memory management unit which enables collecting and extracting fine-grained memory access locality characteristics. We show that with the proposed method, Hybrid Access control, which is a memory hierarchy control that selectively uses Direct Access to bus attached byte-addressable SCM and low power Aggressive Paging using small DRAM as cache, can be made more accurate, and the efficiency of memory access can be significantly improved.

**Holistic Measurement Driven System Assessment**

Saurabh Jha (University of Illinois), Mike Showerman (National Center for Supercomputing Applications (NCSA), University of Illinois), Aaron Saxton (National Center for Supercomputing Applications)
HPC users deploy a suite of monitors to observe patterns of failures and performance anomalies to improve operational efficiency, achieve higher application performance and inform the design of future systems. However, the promises and the potential of monitoring data have largely been not realized due to various challenges such as inadequacy in monitoring, limited availability of data, lack of methods for fusing monitoring data at time-scales necessary for enabling human-in-the-loop or machine-in-the-loop feedback. To address above challenges, in this work we developed a monitoring fabric Holistic Measurement Driven System Assessment (HMDSA) for large-scale HPC facilities, independent of major component vendor, and within budget constraints of money, space, and power. We accomplish this through development and deployment of scalable, platform-independent, open-source tools and techniques for monitoring, coupled with statistical and machine-learning based runtime analysis and feedback, which enables highly efficient HPC system operation and usage and also informs future system improvements.

Best Poster Finalist: no
Poster 61: Fast 3D Diffeomorphic Image Registration on GPUs
Malte Brunn (University of Stuttgart), Naveen Himthani (University of Texas), George Biros (University of Texas), Miriam Mehl (University of Stuttgart), Andreas Mang (University of Houston)

3D image registration is one of the most fundamental and computationally expensive operations in medical image analysis. Here, we present a mixed-precision, Gauss-Newton-Krylov solver for diffeomorphic registration. Our work extends the publicly available CLAIRE library to GPU architectures. Despite the importance of image registration, only a few implementations of large deformation diffeomorphic registration packages support GPUs. Our contributions are new algorithms and dedicated computational kernels to significantly reduce the runtime of the main computational kernels in CLAIRE: derivatives and interpolation. We deploy (i) highly-optimized, mixed-precision GPU-kernels for the evaluation of scattered-data interpolation, (ii) replace FFT-based first-order derivatives with optimized 8th-order finite differences, and (iii) compare with state-of-the-art CPU and GPU implementations. As a highlight, we demonstrate that we can register 256^3 clinical images in less than 6 seconds on a single NVIDIA Tesla V100. This amounts to over 20x speed-up over CLAIRE and over 30x speed-up over existing GPU implementations.

Best Poster Finalist: no

Poster 138: Across-Stack Profiling and Characterization of State-of-the-Art Machine Learning Models on GPUs
Cheng Li (University of Illinois), Abdul Dakkak (University of Illinois), Wei Wei (Alibaba Inc), Jinjun Xiong (IBM Research), Lingjie Xu (Alibaba Inc), Wei Zhang (Alibaba Inc), Wen-mei Hwu (University of Illinois)

The past few years have seen a surge of using Machine Learning (ML) and Deep Learning (DL) algorithms for traditional HPC tasks such as feature detection, numerical analysis, and graph analytics. While ML and DL enable solving HPC tasks, their adoption has been hampered due to the lack of understanding of how they utilize systems. Optimizing these algorithms requires characterizing their performance across the hardware/software (HW/SW) stack, but the lack of simple tools to automate the process and the reliance on researchers to perform manual characterization is a bottleneck. To alleviate this, we propose an across-stack profiling scheme and integrate it within MLModelScope — a hardware and software agnostic tool for evaluating and benchmarking ML/DL at scale. We demonstrate MLModelScope’s ability to characterize state-of-art ML/DL models and give insights that are only possible obtained by performing across-stack profiling.
Poster 60: Massively Parallel Large-Scale Multi-Model Simulation of Tumor Development
Marco Berghoff (Karlsruhe Institute of Technology), Jakob Rosenbauer (Forschungszentrum Juelich), Alexander Schug (Forschungszentrum Juelich)

The temporal and spatial resolution in the microscopy of tissues has increased significantly within the last years, yielding new insights into the dynamics of tissue development and the role of the single-cell within it. A thorough theoretical description of the connection of single-cell processes to macroscopic tissue reorganizations is still lacking. Especially in tumor development, single cells play a crucial role in advance of tumor properties.

We developed a simulation framework that can model tissue development up to the centimeter scale with micrometer resolution of single cells. Through a full parallelization, it enables the efficient use of HPC systems, therefore enabling detailed simulations on a large scale. We developed a generalized tumor model that respects adhesion driven cell migration, cell-to-cell signaling, and mutation-driven tumor heterogeneity. We scan the response of the tumor development depending on division inhibiting substances such as cytostatic agents.

Poster 114: Optimizing Recommendation System Inference Performance Based on GPU
Xiaowei Shen (Alibaba Inc), Junrui Zhou (Alibaba Inc), Kan Liu (Alibaba Inc), Lingling Jin (Alibaba Inc), Pengfei Fan (Alibaba Inc), Wei Zhang (Alibaba Inc), Jun Yang (University of Pittsburgh)

Neural network-based recommendation models have been widely applied on tracking personalization and recommendation tasks at large Internet companies such as e-commerce companies and social media companies. Alibaba recommendation system deploys WDL (wide and deep learning) models for product recommendation tasks. The WDL model consists of two main parts: embedding lookup and neural network-based feature ranking model that ranks different products for different users. As more and more products and users the model need to rank, the feature length and batch size of the models are increased. The computation of models is also increased so that traditional model inference implementation on CPU cannot meet the requirement of QPS (query per second) and latency of recommendation tasks. In this poster, we develop a GPU based system to speedup recommendation system inference performance. By model quantization and graph transformation, we can achieve 3.9x performance speedup when compared with a baseline GPU implementation.
Poster 59: Accelerating BFS and SSSP on a NUMA Machine for the Graph500 Challenge
Tanuj K. Aasawat (RIKEN), Kazuki Yoshizoe (RIKEN), Tahsin Reza (University of British Columbia), Matei Ripeanu (University of British Columbia)

The NUMA architecture is the design choice for modern multi-CPU shared memory systems. In many ways, a NUMA system resembles a shared-nothing distributed system: memory accesses to remote NUMA domains are more expensive than local accesses.

In this work, we explore how improved data locality and reduced expensive remote communication can be achieved by exploiting "distributed" shared-memory of NUMA machines to develop shared-memory graph processing solutions optimized for NUMA systems. We introduce a novel hybrid design for memory accesses that handles the burst mode in traversal based algorithms, like BFS and SSSP, and reduces the number of remote accesses and updates. We demonstrate that our designs offer up to 84% speedup over our NUMA-oblivious framework Totem and 2.86x over shared-nothing distributed design, for BFS and SSSP algorithms.

Poster 47: Decomposition Algorithms for Scalable Quantum Annealing
Elijah Pelofske (Los Alamos National Laboratory), Georg Hahn (Harvard University), Hristo Djidjev (Los Alamos National Laboratory)

Commercial adiabatic quantum annealers such as D-Wave 2000Q have the potential to solve NP-complete optimization problems efficiently. One of the primary constraints of such devices is the limited number and connectivity of their qubits. This research presents two exact decomposition methods (for the Maximum Clique and the Minimum Vertex Cover problem) that allow us to solve problems of arbitrarily large sizes by splitting them up recursively into a series of arbitrarily small subproblems. Those subproblems are then solved exactly or approximately using a quantum annealer. Whereas some previous approaches are based on heuristics that do not guarantee optimality of their solutions, our decomposition algorithms have the property that the optimal solution of the input problem can be reconstructed given all generated subproblems are solved optimally as well. We investigate various heuristic and exact bounds as well as reduction methods that help to increase the scalability of our approaches.
Poster 92: Nanoporous Flow Simulations on the Summit Supercomputer
Yidong Xia (Idaho National Laboratory), Lixiang Luo (IBM - TJ Watson Research Center), Ansel Blumers (Brown University), Joshua Kane (Idaho National Laboratory), Jan Goral (University of Utah), Yu-Hang Tang (Lawrence Berkeley National Laboratory), Zhen Li (Clemson University, Brown University), Hai Huang (Idaho National Laboratory), Milind Deo (University of Utah)

Fluid flow behaviors in nanoporous materials are distinct from those following the continuum physics. Numerical simulations can be a complement to laboratory experiments. This work presents a dissipative particle dynamics (DPD) package for GPU-accelerated mesoscale flow simulations in nanoporous materials. In an ideal benchmark that minimizes load imbalance, the package delivered nearly perfect strong- and weak-scaling (with up to 4 billion DPD particles) on up to 1,536 V100 GPUs on Oak Ridge National Laboratory’s Summit supercomputer. More remarkably, in a benchmark to measure its usefulness with realistic nanopores in SBA-15 silica, the package exhibited more than 20x speedup over its LAMMPS-based CPU counterpart with the same number nodes (e.g., 384 V100 GPUs vs. 2,688 POWER9 cores). It is worth highlighting that the NVLink2 Host-to-Device interconnects kept the cost of CPU-GPU memory copy as low as only 10% of GPU activity time per rank: 4 times less than their PCIe counterparts.

Best Poster Finalist: yes

Poster 90: You Have to Break It to Make It: How On-Demand, Ephemeral Public Cloud Projects with Alces Flight Compute Resulted in the Open-Source OpenFlightHPC Project
Cristin Merritt (Alces Flight Limited; Alces Software Ltd, UK), Wil Mayers (Alces Flight Limited), Stu Franks (Alces Flight Limited)

Over three years ago the Alces Flight team made a decision to explore on-demand public cloud consumption for High Performance Computing (HPC). Our premise was simple, create a fully-featured, scalable HPC environment for research and scientific computing and provide it as a free subscription service to researchers in order to study how they would approach and consume this then new technology. This tool, Alces Flight Compute, would set out to chart how far away from the traditional bare-metal platforms our subscribers were willing to go. What we didn’t expect was that to get to their destination, our users would proceed to take our tool apart. This deconstruction has resulted in a new approach to HPC environment creation (the open-source OpenFlightHPC project), helped us better understand cloud adoption strategies, and handed over a set of guidelines to help those looking to bring public cloud into their HPC solution.
**Poster 126: Enforcing Crash Consistency of Scientific Applications in Non-Volatile Main Memory Systems**  
Tyler Coy (Washington State University, Vancouver), Xuechen Zhang (Washington State University, Vancouver)  
This poster presents a compiler-assistant technique, NVPath, to automatically generate NVMM-aware persistent data structures which provide the same level of guarantee of crash consistency compared to the baseline code. Compiler-assistant code annotation and transformation is general and can be applied to applications using various data structures. Our experimental results with real-world scientific applications show that the performance of the annotated programs is commensurate with the version using the manual code transformation on the Titan supercomputer.

**Poster 137: Warwick Data Store: A HPC Library for Flexible Data Storage in Multi-Physics Applications**  
Richard O. Kirk (University of Warwick), Timothy R. Law (Atomic Weapons Establishment (AWE), UK), Satheesh Maheswaran (Atomic Weapons Establishment (AWE), UK), Stephen A. Jarvis (University of Warwick)  
With the increasing complexity of memory architectures and multi-physics applications, developing data structures that are performant, portable, scalable, and support developer productivity, is difficult. In order to manage these complexities and allow rapid prototyping of different approaches we are building a lightweight and extensible C++ template library called the Warwick Data Store (WDS). WDS is designed to abstract details of the data structure away from the user, thus easing application development and optimisation. We show that WDS generates minimal performance overhead, via a variety of different scientific benchmarks and proxy-applications.

**Poster 76: HPChain: An MPI-Based Blockchain Framework for High Performance Computing Systems**  
Abdullah Al-Mamun (University of Nevada, Reno; Lawrence Berkeley National Laboratory), Tonglin Li (Lawrence Berkeley National Laboratory), Mohammad Sadoghi (University of California, Davis),
Data fidelity is of prominent importance for scientific experiments and simulations. The state-of-the-art mechanism to ensure data fidelity is through data provenance. However, the provenance data itself may as well exhibit unintentional human errors and malicious data manipulation. To enable a trustworthy and reliable data fidelity service, we advocate achieving the immutability and decentralization of scientific data provenance through blockchains. Specifically, we propose HPChain, a new blockchain framework specially designed for HPC systems. HPChain employs a new consensus protocol compatible with and optimized for HPC systems. Furthermore, HPChain was implemented with MPI and integrated with an off-chain distributed provenance service to tolerate the failures caused by faulty MPI ranks. The HPChain prototype system has been deployed to 500 cores at the University of Nevada’s HPC center and demonstrated strong resilience and scalability while outperforming state-of-the-art blockchains by orders of magnitude.

Best Poster Finalist: no

Poster 104: An Adaptive Checkpoint Model For Large-Scale HPC Systems
Subhendu S. Behera (North Carolina State University), Lipeng Wan (Oak Ridge National Laboratory), Frank Mueller (North Carolina State University), Matthew Wolf (Oak Ridge National Laboratory), Scott Klasky (Oak Ridge National Laboratory)

Checkpoint/Restart is a widely used Fault Tolerance technique for application resilience. However, failures and the overhead of saving application state for future recovery upon failure reduces the application efficiency significantly. This work contributes a failure analysis and prediction model making decisions for checkpoint data placement, recovery, and techniques for reducing checkpoint frequency. We also demonstrate a reduction in application overhead by taking proactive measures guided by failure prediction.

Best Poster Finalist: no

Poster 123: Cloud-Native SmartX Intelligence Cluster for AI-Inspired HPC/HPDA Workloads
Jungsu Han (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science), Jun-Sik Shin (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science), JinCheol Kwon (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science), JongWon Kim (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science)
In this poster, we introduce Cloud-native SmartX Intelligence Cluster for flexibly supporting AI-inspired HPC (high performance computing) / HPDA (high performance data analytics) workloads. This work has been continuously refined from 2013 with a futuristic vision for operating 100 petascale data center. Then, we discuss issues and approaches that come with building a Cloud-native SmartX Intelligence Cluster.

Best Poster Finalist: no

**Poster 86: High-Performance Custom Computing with FPGA Cluster as an Off-Loading Engine**

Takaaki Miyajima (RIKEN Center for Computational Science (R-CCS)), Tomohiro Ueno (RIKEN Center for Computational Science (R-CCS)), Jens Huthmann (RIKEN Center for Computational Science (R-CCS)), Atsushi Koshiba (RIKEN Center for Computational Science (R-CCS)), Kentaro Sano (RIKEN Center for Computational Science (R-CCS)), Mitsuhisa Sato (RIKEN Center for Computational Science (R-CCS))

A heterogeneous system with Field Programmable Gate Arrays (FPGAs) is gathering attention in High-Performance Computing (HPC) area. We have been researching and developing a high-performance custom computing with FPGA cluster as an off-loading engine for an existing HPC system. In this research poster, we describe the motivation of our research and present research topics on a software bridge between the FPGA cluster and existing HPC servers, and dedicated inter-FPGA networks.

Best Poster Finalist: no

**Poster 69: Optimization for Quantum Computer Simulation**

Naoki Yoshioka (RIKEN Center for Computational Science (R-CCS)), Hajime Inaoka (RIKEN Center for Computational Science (R-CCS)), Nobuyasu Ito (RIKEN Center for Computational Science (R-CCS)), Fengping Jin (Forschungszentrum Juelich), Kristel Michielsen (Forschungszentrum Juelich), Hans De Raedt (University of Groningen)

Simulator of quantum circuits is developed for massively parallel classical computers, and it is tested on the K computer in RIKEN R-CCS up to 45 qubits. Two optimization techniques are proposed in order to improve performance of the simulator. The "page method" reduces unnecessary copies in each node. It is found that this method makes approximately 17% speed-up maximum. Initial permutation of qubits is also studied how it affects performance of the simulator. It is found that a simple permutation in ascending order of the number of operations for each qubit is sufficient in the
In the era of data-intensive computing, accessing data with a high-throughput and low-latency is very imperative. Data prefetching is used for hiding read latency by requesting data before it is needed to move it from a high-latency medium to a low-latency one. However, existing solutions do not consider the multi-tiered storage and also suffer from under-utilization of prefetching resources and unnecessary evictions. Additionally, existing approaches implement a client-pull model where understanding the application's I/O behavior drives prefetching decisions. Moving toward exascale, where machines run multiple applications concurrently by accessing files in a workflow, a more data-centric approach resolves challenges such as cache pollution and redundancy. In this study, we present HFetch, a truly hierarchical data prefetcher that adopts a server-push approach to data prefetching. We demonstrate the benefits of such an approach. Results show 10-35% performance gains over existing prefetechers and over 50% when compared to systems with no prefetching.

Poster 52: Design and Specification of Large-Scale Simulations for GPUs Using FFFT
Anuva Kulkarni (Carnegie Mellon University), Daniele Spampinato (Carnegie Mellon University), Franz Franchetti (Carnegie Mellon University)

Large-scale scientific simulations can be ported to heterogeneous environments with GPUs using domain decomposition. However, Fast Fourier Transform (FFT) based simulations require all-to-all communication and large memory, which is beyond the capacity of on-chip GPU memory. To overcome this, domain decomposition solutions are combined with adaptive sampling or pruning around the domain to reduce storage. Expression of such operations is a challenge in existing FFT libraries like FFTW, and thus it is difficult to get a high performance implementation of such methods. We demonstrate algorithm specification for one such simulation (Hooke's law) using FFFT, an emerging API with a SPIRAL-based code generation back-end, and suggest future extensions useful for GPU-based scientific computing.

Best Poster Finalist: no
Poster 136: CHAMELEON: Reactive Load Balancing and Migratable Tasks for Hybrid MPI+OpenMP Applications
Jannis Klinkenberg (RWTH Aachen University), Philipp Samfaß (Technical University Munich), Michael Bader (Technical University Munich), Karl Fürlinger (Ludwig Maximilian University of Munich), Christian Terboven (RWTH Aachen University), Matthias S. Müller (RWTH Aachen University)

Many HPC applications are designed based on underlying performance and execution models. These models could successfully be employed in the past for balancing load within and between compute nodes. However, the increasing complexity of modern software and hardware makes performance predictability and load balancing much more difficult. Tackling these challenges in search for a generic solution, we present a novel library for fine-granular task-based reactive load balancing in distributed memory based on MPI and OpenMP. Our concept allows creating individual migratable tasks that can be executed on any MPI rank. Migration decisions are performed at run time based on online performance or load data. Two fundamental approaches to balance load and at the same time overlap computation and communication are compared. We evaluate our concept under enforced power caps and clock frequency changes using a synthetic benchmark and demonstrate robustness against work-induced imbalances for an AMR application.

Best Poster Finalist: no

Poster 124: Porting Finite State Automata Traversal from GPU to FPGA: Exploring the Implementation Space
Marziyeh Nourian (North Carolina State University), Mostafa Eghbali Zarch (North Carolina State University), Michela Becchi (North Carolina State University)

While FPGAs are traditionally considered hard to program, recently there are efforts to allow using high-level programming models intended for multi-core CPUs and GPUs to program FPGAs. For example, both Intel and Xilinx are now providing OpenCL-to-FPGA toolchains. However, since GPU and FPGA devices offer different parallelism models, OpenCL code optimized for GPU can prove inefficient on FPGA, in terms of both performance and hardware resource utilization.

In this poster, we explore this problem on an emerging workload: finite state automata traversal. Specifically, we explore a set of structural code changes, custom, and best-practice optimizations to retarget an OpenCL NFA engine designed for GPU to FPGA. Our evaluation, which covers traversal throughput and resource utilization, shows that our optimizations lead, on a single execution pipeline, to speedups up to 4x over an already optimized baseline that uses one of the proposed code changes
to fit the original code on FPGA.

Best Poster Finalist: no

**Poster 68: Linking a Next-Gen Remap Library into a Long-Lived Production Code**  
Charles R. Ferenbaugh (Los Alamos National Laboratory), Brendan K. Krueger (Los Alamos National Laboratory)

LANL's long-lived production application xRage contains a remapper capability that maps mesh fields from its native AMR mesh to the GEM mesh format used by some third-party libraries. The current remapper was implemented in a short timeframe and is challenging to maintain. Meanwhile, our next-generation code project has developed a modern remapping library Portage, and the xRage team wanted to link in Portage as an alternate mapper option. But the two codes are very different from each other, and connecting the two required us to deal with a number of challenges. This poster describes the codes, the challenges we worked through, current status, and some initial performance statistics.

Best Poster Finalist: no

**Poster 131: Efficiency of Algorithmic Structures**  
Julian Miller (RWTH Aachen University), Lukas Trümper (RWTH Aachen University), Christian Terboven (RWTH Aachen University), Matthias S. Müller (RWTH Aachen University)

The implementation of high-performance parallel software is challenging and raises issues not seen in serial programs before. It requires a strategy of parallel execution which preserves correctness but maximizes scalability. Efficiently deriving well-scaling solutions remains an unsolved problem especially with the quickly-evolving hardware landscape of high-performance computing (HPC).

This work proposes a framework for classifying the efficiency of parallel programs. It bases on a strict separation between the algorithmic structure of a program and its executed functions. By decomposing parallel programs into a hierarchical structure of parallel patterns, a high-level abstraction is provided which leads to equivalence classes over parallel programs. Each equivalence class possesses efficiency properties, mainly communication and synchronization, dataflow and architecture efficiency. This classification allows for wide application areas and a workflow for structural optimization of parallel algorithms is proposed.

Best Poster Finalist: no
**Poster 98: INSPECT Intranode Stencil Performance Evaluation Collection**

Julian Hammer (University of Erlangen-Nuremberg), Julian Hornich (University of Erlangen-Nuremberg), Georg Hager (University of Erlangen-Nuremberg), Thomas Gruber (University of Erlangen-Nuremberg), Gerhard Wellein (University of Erlangen-Nuremberg)

Modeling and presenting performance data---even for simple kernels such as stencils---is not trivial. We therefore present an overview on how to interpret and what to learn from an INSPECT report, as well as highlighting best practices for performance data reporting.

INSPECT is the "Intranode Stencil Performance Evaluation Collection", which compiles performance benchmarks and reports of various stencil and streaming kernels on a variety of architectures. The goal is to aid performance-aware developers with reference material and a methodology to analyze their own codes.

INSPECT set out to cover these topics and compile a summary of all necessary information to allow reproduction of the performance results, their interpretation and discussion.

Best Poster Finalist: no

**Poster 97: Optimizing Multigrid Poisson Solver of Cartesian CFD Code CUBE**

Kazuto Ando (RIKEN Center for Computational Science (R-CCS)), Rahul Bale (RIKEN), Keiji Onishi (RIKEN Center for Computational Science (R-CCS)), Kiyoshi Kumahata (RIKEN Center for Computational Science (R-CCS)), Kazuo Minami (RIKEN Center for Computational Science (R-CCS)), Makoto Tsubokura (Kobe University, RIKEN Center for Computational Science (R-CCS))

We demonstrate an optimization of multigrid Poisson solver of Cartesian CFD code “CUBE (Complex Unified Building cubE method)”. CUBE is a simulation framework for complex industrial flow problem, such as aerodynamics of vehicles, based on hierarchical Cartesian mesh. In incompressible CFD simulation, solving pressure Poisson equation is the most time-consuming part. In this study, we use a cavity flow simulation as a benchmark problem. With this problem, multigrid Poisson solver dominates 91% of execution time of the time-step loop. Specifically, we evaluate the performance of Gauss-Seidel loop as a computational kernel based on “Byte per Flop” approach. With optimization of the kernel, we achieved 9.8x speedup and peak floating point performance ratio increased from 0.4% to 4.0%. We also measured parallel performance up to 8,192 nodes (65,536 cores) on the K computer. With optimization of the parallel performance, we achieved 2.9x–3.9x sustainable speedup in the time-step loop.
Poster 85: Hybrid Computing Platform for Combinatorial Optimization with the Coherent Ising Machine
Junya Arai (Nippon Telegraph and Telephone Corporation), Yagi Satoshi (Nippon Telegraph and Telephone Corporation), Hiroyuki Uchiyama (Nippon Telegraph and Telephone Corporation), Toshimori Honjo (Nippon Telegraph and Telephone Corporation), Takahiro Inagaki (Nippon Telegraph and Telephone Corporation), Takuya Ikuta (Nippon Telegraph and Telephone Corporation), Kensuke Inaba (Nippon Telegraph and Telephone Corporation), Hiroki Takesue (Nippon Telegraph and Telephone Corporation), Keitaro Horikawa (Nippon Telegraph and Telephone Corporation)

Several institutes are operating cloud platforms that offer Web API access to Ising computers such as quantum annealing machines. Platform users can solve complex combinatorial optimization problems by using hybrid algorithms that utilize both users’ conventional digital computers and remote Ising computers. However, communication via the Internet takes an order of magnitude longer time than optimization on Ising computers. This overheads seriously degrade the performance of hybrid algorithms since they involve frequent communication. In this poster, we first state issues in the design of Ising computing platforms, including communication overheads. Then, we answer the issues by introducing the computing platform for the coherent Ising machine (CIM), an Ising computer based on photonics technologies. Our platform offers efficient CIM-digital communication by allowing users to execute their program on digital computers co-located with the CIM. We have released the platform to our research collaborators in this autumn and started the evaluation.

Best Poster Finalist: no

Poster 117: A New Polymorphic Computing Architecture Based on Fine-Grained Instruction Mobility
David Hentrich (Illinois Institute of Technology), Erdal Oruklu (Illinois Institute of Technology), Jafar Saniie (Illinois Institute of Technology)

This is a summary of the base concepts behind David Hentrich’s May 2018 Ph.D. dissertation in Polymorphic Computing. Polymorphic Computing is the emerging field of changing the computer architecture around the software, rather than vice versa. The main contribution is a new polymorphic computing architecture. The key idea behind the architecture is to create an array of processors where a program’s instructions can be individually and arbitrarily assigned/mobilized to any processor, even during runtime. The key enablers of this architecture are a dataflow instruction set that is conducive to instruction migration, a microarchitectural block called an “operation cell” (op-
细胞），一个处理器，其围绕指令集和“op-cells”构建，并配有多处理器阵列。

Best Poster Finalist: no

**Poster 67: Genie: an MPEG-G Conformant Software to Compress Genomic Data.**

Brian E. Bliss (University of Illinois), Joshua M. Allen (University of Illinois), Saurabh Baheti (Mayo Clinic), Matthew A. Bockol (Mayo Clinic), Shubham Chandak (Stanford University), Jaime Delgado (Polytechnic University of Catalonia), Jan Fostier (Ghent University), Josep L. Gelpi (University of Barcelona), Steven N. Hart (Mayo Clinic), Mikel Hernaez Arrazola (University of Illinois), Matthew E. Hudson (University of Illinois), Michael T. Kalmbach (Mayo Clinic), Eric W. Klee (Mayo Clinic), Liudmila S. Mainzer (University of Illinois), Fabian Müntefering (Leibniz University), Daniel Naro (Barcelona Supercomputing Center), Idoia Ochoa-Alvarez (University of Illinois), Jörn Ostermann (Leibniz University), Tom Paridaens (Ghent University), Christian A. Ross (Mayo Clinic), Jan Voges (Leibniz University), Eric D. Wieben (Mayo Clinic), Mingyu Yang (University of Illinois), Tsachy Weissman (Stanford University), Mathieu Wiepert (Mayo Clinic)

Precision medicine has unprecedented potential for accurate diagnosis and effective treatment. It is supported by an explosion of genomic data, which continues to accumulate at accelerated pace. Yet storage and analysis of petascale genomic data is expensive, and that cost will ultimately be borne by the patients and citizens. The Moving Picture Experts Group (MPEG) has developed MPEG-G, a new open standard to compress, store, transmit and process genomic sequencing data that provides an evolved and superior alternative to currently used genomic file formats. Our poster will showcase software package GENIE, the first open source implementation of an encoder-decoder pair that is compliant with the MPEG-G specifications and delivers all its benefits: efficient compression, selective access, transport and analysis, guarantee of long-term support, and embedded mechanisms for annotation and encryption of compressed information. GENIE will create a step-change in medical genomics by reducing the cost of data storage and analysis.

Best Poster Finalist: no

**Poster 82: A View from the Facility Operations Side on the Water/Air Cooling System of the K Computer**

Jorji Nonaka (RIKEN Center for Computational Science (R-CCS)), Keiji Yamamoto (RIKEN Center for Computational Science (R-CCS)), Akiyoshi Kuroda (RIKEN Center for Computational Science (R-CCS)), Toshiyuki Tsukamoto (RIKEN Center for Computational Science (R-CCS)), Kazuki Koiso (Kobe University, RIKEN Center for Computational Science (R-CCS)), Naohisa Sakamoto (Kobe University, RIKEN Center for Computational Science (R-CCS))
The Operations and Computer Technologies Division at the RIKEN R-CCS is responsible for the operations of the entire K computer facility, which includes the auxiliary subsystems such as the power supply and water/air cooling systems. It is worth noting that part of these subsystems will be reused in the next supercomputer (Fugaku), thus a better understanding of the operational behavior as well as the potential impacts especially on the hardware failure and energy consumption would be greatly beneficial. In this poster, we will present some preliminary impressions of the impact of the water/air cooling system on the K computer system, focusing on the potential benefits of the use of low water/air temperature respectively for the CPU and DRAM memory modules produced by the cooling system. We expect that the obtained knowledge will be helpful for the decision support and/or operation planning of the next supercomputer.

Best Poster Finalist: no

**Poster 135: High-Performance Deep Learning via a Single Building Block**
Evangelos Georganas (Intel Corporation), Kunal Banerjee (Intel Corporation), Dhiraj Kalamkar (Intel Corporation), Sasikanth Avancha (Intel Corporation), Anand Venkat (Intel Corporation), Michael Anderson (Intel Corporation), Greg Henry (Intel Corporation), Hans Pabst (Intel Corporation), Alexander Heinecke (Intel Corporation)

Deep learning (DL) is one of the most prominent branches of machine learning. Due to the immense computational cost of DL workloads, industry and academia have developed DL libraries with highly-specialized kernels for each workload/architecture, leading to numerous, complex code-bases that strive for performance, yet they are hard to maintain and do not generalize. In this work, we introduce the batch-reduce-GEMM kernel and show how the most popular DL algorithms can be formulated with this kernel as basic building-block. Consequently, the DL library-development degenerates to mere (potentially automatic) tuning of loops around this sole optimized kernel. By exploiting our kernel we implement Recurrent Neural Networks, Convolution Neural Networks and Multilayer Perceptron training and inference primitives in just 3K lines of high-level-code. Our primitives outperform vendor-optimized libraries on multi-node CPU-Clusters. We also provide CNN kernels targeting GPUs. Finally, we demonstrate that batch-reduce-GEMM kernel within a tensor compiler yields high-performance CNN primitives.

Best Poster Finalist: no

**Poster 56: Reinforcement Learning for Quantum Approximate Optimization**
Sami Khairy (Illinois Institute of Technology), Ruslan Shaydulin (Clemson University), Lukasz Cincio
The Quantum Approximate Optimization Algorithm (QAOA) is one of the leading candidates for demonstrating quantum advantage. The quality of the solution obtained by QAOA depends on the performance of the classical optimization routine used to optimize the variational parameters. In this work, we propose a Reinforcement Learning (RL) based approach to drastically reduce the number of evaluations needed to find high-quality variational parameters. We train an RL agent on small 8-qubit Max-Cut problem instances on an Intel Xeon Phi supercomputer Bebop, and use (transfer) the learned optimization policy to quickly find high-quality solutions for other larger problem instances coming from different distributions and graph classes. The preliminary results show that our RL based approach is able to improve the quality of the obtained solution by up to 10% within a fixed budget of function evaluations and demonstrate learned optimization policy transferability between different graph classes and sizes.


Accelerating technology disruptions and architectural change create growing opportunities and urgency to reduce the latency in for new architectural innovations to be deployed in extreme scale systems. We are exploring new architectural features that improve memory system performance including word-wise scratchpad memory, a flexible Recode engine, hardware message queues, and the data rearrangement engine (DRE). Performance results are promising yielding as much as 20x benefit. Project 38 is a cross-agency effort undertaken by the US Department of Energy (DOE) and Department of Defense (DoD).

Best Poster Finalist: no
Poster 128: Identifying Time Series Similarity in Large-Scale Earth System Datasets
Payton Linton (Youngstown State University), William Melodia (Youngstown State University), Alina Lazar (Youngstown State University), Deborah Agarwal (Lawrence Berkeley National Laboratory), Ludovico Bianchi (Lawrence Berkeley National Laboratory), Devarshi Ghoshal (Lawrence Berkeley National Laboratory), Kesheng Wu (Lawrence Berkeley National Laboratory), Gilberto Pastorello (Lawrence Berkeley National Laboratory), Lavanya Ramakrishnan (Lawrence Berkeley National Laboratory)

Scientific data volumes are growing every day and instrument configurations, quality control and software updates result in changes to the data. This study focuses on developing algorithms that detect changes in time series datasets in the context of the Deduce project. We propose a combination of methods that include dimensionality reduction and clustering to evaluate similarity measuring algorithms. This methodology can be used to discover existing patterns and correlations within a dataset. The current results indicate that the Euclidean Distance metric provides the best results in terms of internal cluster validity measures for multi-variable analyses of large-scale earth system datasets. The poster will include details on our methodology, results, and future work.

Best Poster Finalist: no

Poster 151: Three-Dimensional Characterization on Edge AI Processors with Object Detection Workloads
Yujie Hui (Ohio State University), Jeffrey Lien (NovuMind Inc), Xiaoyi Lu (Ohio State University)

The Deep Learning inference applications are moving to the edge side, as edge-side AI platforms are cheap and energy-efficient. Different edge AI processors are diversified, since these processors are designed with different approaches. However, it is hard for customers to select an edge AI processor without an overall evaluation of these processors. We propose a three-dimensional characterization (i.e., accuracy, latency, and energy efficiency) approach on three different kinds of edge AI processors (i.e., Edge TPU, NVIDIA Xavier, and NovuTensor). We deploy YOLOv2 and Tiny-YOLO, which are two YOLO-based object detection systems, on these edge AI platforms with Microsoft COCO dataset. I will present our work starting from the problem statement. And then I'll introduce our experiments setup and hardware configuration. Lastly, I'll conclude our experimental results and current work status, as well as the future work.

Best Poster Finalist: no
Poster 148: Unsupervised Clustering of Golden Eagle Telemetry Data

We use a recurrent autoencoder neural network to encode sequential California golden eagle telemetry data. The encoding is followed by an unsupervised clustering technique, Deep Embedded Clustering (DEC), to iteratively cluster the data into a chosen number of behavior classes. We apply the method to simulated movement data sets and telemetry data for a Golden Eagle. The DEC achieves better unsupervised clustering accuracy scores for the simulated data sets as compared to the baseline K-means clustering result.

Best Poster Finalist: no

Poster 66: Hybrid CPU/GPU FE2 Multi-Scale Implementation Coupling Alya and Micropp
Guido Giuntoli (Barcelona Supercomputing Center), Judicaël Grasset (Science and Technology Facilities Council (STFC)), Alejandro Figueroa (George Mason University), Charles Moulinec (Science and Technology Facilities Council (STFC)), Mariano Vázquez (Barcelona Supercomputing Center), Guillaume Houzeaux (Barcelona Supercomputing Center), Stephen Longshaw (Science and Technology Facilities Council (STFC)), Sergio Oller (Polytechnic University of Catalonia)

This poster exposes the results of a new implementation of the FE2 multi-scale algorithm that is achieved by coupling the multi-physics and massively parallel code Alya with the GPU-based code micropp. The coupled code is mainly designed to solve large scale and realistic composite material problems for the aircraft industry. Alya is responsible of solving the macro-scale equations and micropp for solving the representation of fibres at the microscopic level. The poster shows computational performance results that demonstrate that the technique is scalable for real size industrial problems and also how the execution time is dramatically reduced using GPU-based clusters.

Best Poster Finalist: no

Poster 129: Understanding I/O Behavior in Scientific Workflows on High Performance Computing Systems
Fahim Tahmid Chowdhury (Florida State University, Lawrence Livermore National Laboratory), Francesco Di Natale (Lawrence Livermore National Laboratory), Adam Moody (Lawrence Livermore National Laboratory), Elsa Gonsiorowski (Lawrence Livermore National Laboratory), Kathryn Mohror (Lawrence Livermore National Laboratory), Weikuan Yu (Florida State University)
Leadership high performance computing (HPC) systems have the capability to execute workflows of scientific, research or industry applications. Complex HPC workflows can have significant data transfer and I/O requirements. Heterogeneous storage systems in supercomputers equipped with bleeding-edge non-volatile persistent storage devices can be leveraged to handle these data transfer and I/O requirements efficiently.

In this poster, we describe our efforts to extract the I/O characteristics of various HPC workflows and develop strategies to improve I/O performance by leveraging heterogeneous storage systems. We have implemented an emulator to mimic different types of I/O requirements posed by HPC application workflows. We have analyzed the workflow of Cancer Moonshot Pilot 2 (CMP2) project to determine possible I/O inefficiencies. To date, we have performed a systematic characterization and evaluation on the workloads generated by the workflow emulator and a small scale adaptation of the CMP2 workflow.

Best Poster Finalist: no

Poster 116: Advancements in Ultrasound Simulations Enabled by High-Bandwidth GPU Interconnects
Filip Vaverka (Brno University of Technology), Bradley E. Treeby (University College London), Jiri Jaros (Brno University of Technology)

Realistic ultrasound simulations are becoming integral part of many novel medical procedures such as photoacoustic screening and non-invasive treatment planning. The common denominator of all these applications is the need for cheap and relatively large-scale ultrasound simulations with sufficient accuracy. Typical medical applications require full-wave simulations which take frequency-dependent absorption and non-linearity into account.

This poster investigates the benefits of high-bandwidth low-latency interconnects to k-Wave acoustic toolbox in dense multi-GPU environment. The k-Wave multi-GPU code is based on a variant of the local Fourier basis domain decomposition. The poster compares the behavior of the code on a typical PCI-E 3.0 machine with 8 Nvidia Tesla P40 GPUs and a Nvidia DGX-2 server. The performance constraints of PCI-E platforms built around multiple socket servers on multi-GPU applications are deeply explored. Finally, it is shown the k-Wave toolbox can efficiently utilize NVlink 2.0 and achieve over 4x speedup compared to PCI-E systems.

Best Poster Finalist: no
Poster 65: Comparing Granular Dynamics vs. Fluid Dynamics via Large DOF-Count Parallel Simulation on the GPU
Milad Rakhsha (University of Wisconsin), Conlain Kelly (Georgia Institute of Technology), Nicholas Olsen (University of Wisconsin), Lijing Yang (University of Wisconsin), Radu Serban (University of Wisconsin), Dan Negrut (University of Wisconsin)

In understanding granular dynamics, the commonly-used discrete modeling approach that tracks the motion of all particles is computationally demanding, especially with large system size. In such cases, one can contemplate switching to continuum models that are computationally less expensive. In order to assess when such a discrete to continuum switch is justified, we compare granular and fluid dynamics that scales to handle more than 1 billion degrees of freedom (DOFs); i.e., two orders of magnitude higher than the state-of-the-art. On the granular side, we solve the Newton-Euler equations of motion; on the fluid side, we solve the Navier-Stokes equations. Both solvers leverage parallel computing on the GPU, and are publicly available on GitHub as part of an open-source code called Chrono. We report similarities and differences between the dynamics of the discrete, fully-resolved system and the continuum model via numerical experiments including both static and highly transient scenarios.

Best Poster Finalist: no

Poster 78: Understanding HPC Application I/O Behavior Using System Level Statistics
Arnab K. Paul (Virginia Tech), Olaf Faaland (Lawrence Livermore National Laboratory), Adam Moody (Lawrence Livermore National Laboratory), Elsa Gonsiorowski (Lawrence Livermore National Laboratory), Kathryn Mohror (Lawrence Livermore National Laboratory), Ali R. Butt (Virginia Tech)

The processor performance of high performance computing (HPC) systems is increasing at a much higher rate than storage performance. Storage and file system designers therefore require a deep understanding of how HPC application I/O behavior affects current storage system installations in order to improve storage performance. In this work, we contribute to this understanding using application-agnostic file system statistics gathered on compute nodes as well as metadata and object storage file system servers. We analyze file system statistics of more than 4 million jobs over a period of three years on two systems at Lawrence Livermore National Laboratory that include a 15 PiB Lustre file system for storage. Some key observations in our study show that more than 65% HPC users perform significant I/O which are mostly writes; and less than 22% of HPC users who submit write-intensive jobs perform efficient writes to the file system.

Best Poster Finalist: no
Poster 109: A Runtime Approach for Dynamic Load Balancing of OpenMP Parallel Loops in LLVM
Jonas H. Müller Korndörfer (University of Basel, Switzerland), Florina M. Ciorba (University of Basel, Switzerland), Akan Yilmaz (University of Basel, Switzerland), Christian Iwainsky (Technical University Darmstadt), Johannes Doerfert (Argonne National Laboratory), Hal Finkel (Argonne National Laboratory), Vivek Kale (Brookhaven National Laboratory), Michael Klemm (Intel Corporation)

Load imbalance is the major source of performance degradation in computationally-intensive applications that frequently consist of parallel loops. Efficient scheduling can improve the performance of such programs. OpenMP is the de-facto standard for parallel programming on shared-memory systems. The current OpenMP specification provides only three choices for loop scheduling which are insufficient in scenarios with irregular loops, system-induced interference, or both. Therefore, this work augments the LLVM OpenMP runtime library implementation with eleven ready to use scheduling techniques. We tested existing and added scheduling strategies on several applications from NAS, SPEC OMP 2012, and CORAL2 benchmark suites. Experiments show that implemented scheduling techniques outperform others in certain application and system configurations. We measured performance gains of up to 6% compared to the fastest standard scheduling technique. This work aims to be a convincing step toward beyond-standard scheduling options in OpenMP for the benefit of evolving applications executing on multicore architectures.

Best Poster Finalist: no

Poster 143: Quantum Natural Language Processing
Lee James O’Riordan (Irish Centre for High-End Computing (ICHEC), National University of Ireland Galway (NUIG)), Myles Doyle (Irish Centre for High-End Computing (ICHEC), National University of Ireland Galway (NUIG)), Venkatesh Kannan (Irish Centre for High-End Computing (ICHEC), National University of Ireland Galway (NUIG)), Fabio Baruffa (Intel Corporation)

Natural language processing (NLP) algorithms that operate over strings of words are limited since they analyse meanings of the component words in a corpus without information about grammatical rules of the language. Consequently, they often produce unsatisfactory results with increase in problem complexity.

The "distributed compositional semantics" (DisCo) model incorporates grammatical structure of sentences into the algorithms, and offers significant improvements to the quality of results. However, their main challenge is the need for large classical computational resources. The DisCo model presents two quantum algorithms which lower storage and compute requirements compared
to a classic HPC implementation.

In this project, we implement the two DisCo model quantum algorithms on the Intel Quantum Simulator deployed on the Irish national supercomputer. We target corpuses with ~1000 most-common words using up to 36 qubits simulation. The solution will be able to compute the meanings of two sentences and decide if their meanings match.

Best Poster Finalist: no

---

**Poster 152: Deep Domain Adaptation for Runtime Prediction in Dynamic Workload Scheduler**
Hoang H. Nguyen (National Center for Atmospheric Research (NCAR); University of Illinois, Chicago), Ben Matthews (National Center for Atmospheric Research (NCAR)), Irfan Elahi (National Center for Atmospheric Research (NCAR))

In HPC systems, users' requested runtime for submitted jobs plays a crucial role in efficiency. While underestimation of job runtime could terminate jobs before completion, overestimation could result in long queuing of other jobs in HPC systems. In reality, runtime prediction in HPC is challenging due to the complexity and dynamics of running workloads. Most of the current predictive runtime models are trained on static workloads. This poses a risk of over-fitting the predictions with bias from the learned workload distribution. In this work, we propose an adaptation of Correlation Alignment method in our deep neural network architecture (DCORAL) to alleviate the domain shift between workloads for better runtime predictions. Experiments on both standard benchmark workloads and NCAR real-time production workloads reveal that our proposed method results in a more stable training model across different workloads with low accuracy variance as compared to the other state-of-the-art methods.

Best Poster Finalist: no

---

**Poster 75: libCEED - Lightweight High-Order Finite Elements Library with Performance Portability and Extensibility**
Jeremy Thompson (University of Colorado), Valeria Barra (University of Colorado), Yohann Dudouit (Lawrence Livermore National Laboratory), Oana Marin (Argonne National Laboratory), Jed Brown (University of Colorado)

High-order numerical methods are widely used in PDE solvers, but software packages that have provided high-performance implementations have often been special-purpose and intrusive. libCEED is a new library that offers a purely algebraic interface for matrix-free operator
representation and supports run-time selection of implementations tuned for a variety of computational device types, including CPUs and GPUs. We introduce the libCEED API and demonstrate how it can be used in standalone code or integrated with other packages (e.g., PETSc, MFEM, Nek5000) to solve examples of problems that often arise in the scientific computing community, ranging from fast solvers via geometric multigrid methods to Computational Fluid Dynamics (CFD) applications.

Best Poster Finalist: no

**Progress on the Exascale Transition of the VSim Multiphysics PIC code**

Benjamin M. Cowan (Tech-X Corporation), Sergey N. Averkin (Tech-X Corporation), John R. Cary (Tech-X Corporation), Jarrod Leddy (Tech-X Corporation), Scott W. Sides (Tech-X Corporation), Ilya A. Zilberter (Tech-X Corporation)

The highly performant, flexible plasma simulation code VSim was designed nearly 20 years ago (originally as Vorpal), with its first applications roughly four years later. Using object oriented methods, VSim was designed to allow runtime selection from multiple field solvers, particle dynamics, and reactions. It has been successful in modeling for many areas of physics, including fusion plasmas, particle accelerators, microwave devices, and RF and dielectric structures. Now it is critical to move to exascale systems, with their compute accelerator architectures, massive threading, and advanced instruction sets. Here we discuss how we are moving this complex, multiphysics computational application to the new computing paradigm, and how it is done in a way that kept the application producing physics during the move. We present performance results showing significant speedups in all parts of the PIC loop, including field updates, particle pushes, and reactions.

Best Poster Finalist: no

**Poster 58: Lock-Free van Emde Boas Array**

Ziyuan Guo (University of Tokyo)

Lock-based data structures have some potential issues such as deadlock, livelock, and priority inversion, and the progress can be delayed indefinitely if the thread that is holding locks cannot acquire a timeslice from the scheduler. Lock-free data structures, which guarantees the progress of some method call, can be used to avoid these problems. This poster introduces the first lock-free concurrent van Emde Boas Array which is a variant of van Emde Boas Tree. It is linearizable, and the benchmark shows significant performance improvement comparing to other lock-free search
trees when the date set is large and dense enough.

Best Poster Finalist: no

**Poster 63: Adaptive Execution Planning in Biomedical Workflow Management Systems**

*Marta Jaros (Brno University of Technology), Bradley E. Treeby (University College London), Jiri Jaros (Brno University of Technology)*

Biomedical simulations require very powerful computers. Their execution is described by a workflow consisting of a number of different cooperating tasks. The manual execution of individual tasks may be tedious for expert users, but prohibiting for most inexperienced clinicians. k-Dispatch offers a ‘run and forget’ approach where the users are completely screened out from the complexity of HPC systems. k-Dispatch provides task scheduling, execution, monitoring, and fault tolerance. Since the task execution configuration strongly affects the final tasks mapping on the computational resources, the execution planning is of the highest priority. Unlike other tools, k-Dispatch considers a variable amount of computational resources per individual tasks. Since the scaling of the individual HPC codes is never perfect, k-Dispatch may find such a good mapping even an experienced user would miss. The proposed adaptive execution planning is based on collected performance data and the current cluster utilization monitoring.

Best Poster Finalist: no

**Poster 142: Training Deep Neural Networks Directly on Hundred-Million-Pixel Histopathology Images on a Large-Scale GPU Cluster**

*Chi-Chung Chen (AetherAI, Taiwan), Wen-Yu Chuang (Chang-Gung Memorial Hospital, Taiwan), Wei-Hsiang Yu (AetherAI, Taiwan), Hsi-Ching Lin (National Center for High-Performance Computing (NCHC), Taiwan), Shuen-Tai Wang (National Center for High-Performance Computing (NCHC), Taiwan), Fang-An Kuo (National Center for High-Performance Computing (NCHC), Taiwan), Chao-Chun Chuang (National Center for High-Performance Computing (NCHC), Taiwan), Chao-Yuan Yeh (AetherAI, Taiwan)*

Deep learning for digital pathology is challenging because the resolution of whole-slide-images (WSI) is extremely high, often in billions. The most common approach is patch-based method, where WSIs are divided into small patches to train convolutional neural networks (CNN). This approach has significant drawbacks. To have ground truth for individual patches, detailed annotations by pathologists are required. This laborious process has become the major impediment to the development of digital pathology AI. End-to-end WSI training, however, faces the difficulties
of fitting the task into limited GPU memory. In this work, we improved the efficiency of using system memory for GPU compute by 411% through memory optimization and deployed the training pipeline on 8 nodes, totally 32 GPUs distributed system, achieving 147.28x speedup. We demonstrated that CNN is capable of learning features without detailed annotations. The trained CNN can correctly classify cancerous specimen, with performance level closely matching the patch-based methods.

Best Poster Finalist: no

**Poster 96: TSQR on TensorCores**

Hiroyuki Ootomo (Tokyo Institute of Technology), Rio Yokota (Tokyo Institute of Technology)

Tall-Skinny QR (TSQR) is an efficient algorithm for calculating the QR decomposition of $m \times n$ matrices where $m \ll n$, which is done by recursively performing QR decomposition on subdivided blocks of the tall and skinny matrix. Such operations are useful for low-rank approximation methods, which are replacing more and more dense linear algebra in both scientific computing and machine learning fields. The present work focuses on the implementation of this important algorithm on Tensor Cores, which are available on the latest NVIDIA GPUs. We evaluate the speed, accuracy, and stability of TSQR on TensorCores.

Best Poster Finalist: yes

**Poster 95: A Heterogeneous HEVC Video Encoder Based on OpenPOWER Acceleration Platform**

Chenhao Gu (Fudan University, Shanghai, State Key Laboratory of ASIC and System; IBM Corporation), Yang Chen (Fudan University, Shanghai, State Key Laboratory of ASIC and System; IBM Corporation), Yanheng Lu (IBM Corporation), Pengfei Gou (IBM Corporation), Yong Lu (IBM Corporation), Yang Dai (IBM Corporation), Yue Xu (IBM Corporation), Yang Liu (IBM Corporation), Yibo Fan (Fudan University, Shanghai, State Key Laboratory of ASIC and System)

This poster describes a heterogeneous HEVC video encoder system based on the OpenPOWER platform. Our design leverages the Coherent Accelerator Processor Interface (CAPI) on the OpenPOWER, which provides cache-coherent access for FPGA. This technology highly improves CPU-FPGA data communication bandwidth and programming efficiency. X265 is optimized on the OpenPOWER platform to improve its performance with both architecture specific methods and hardware-acceleration methods. For hardware acceleration, frame-level acceleration and functional-unit-level acceleration are introduced and evaluated in this work.
Poster 102: Fast Training of an AI Radiologist: Leveraging Data Pipelining to Efficiently Utilize GPUs
Rakshith Vasudev (Dell EMC), John A. Lockman III (Dell EMC), Lucas A. Wilson (Dell EMC), Srinivas Varadharajan (Dell EMC), Frank Han (Dell EMC), Rengan Xu (Dell EMC), Quy Ta (Dell EMC)

In a distributed deep learning training setting, using accelerators such as GPUs can be challenging to develop a high throughput model. If the accelerators are not utilized effectively, this could mean more time to solution, and thus the model's throughput is low. To use accelerators effectively across multiple nodes, we need to utilize an effective data pipelining mechanism that handles scaling gracefully so GPUs can be exploited of their parallelism. We study the effect of using the correct pipelining mechanism that is followed by tensorflow official models vs a naive pipelining mechanism that doesn't scale well, on two image classification models. Both the models using the optimized data pipeline demonstrate effective linear scaling when GPUs are added. We also show that converting to TF Records is not always necessary.

Poster 94: Multi-GPU Optimization of a Non-Hydrostatic Numerical Ocean Model with Multigrid Preconditioned Conjugate Gradient Method
Takateru Yamagishi (Research Organization for Information Science and Technology, Japan), Yoshimasa Matsumura (University of Tokyo, Atmosphere and Ocean Research Institute), Hiroyasu Hasumi (University of Tokyo, Atmosphere and Ocean Research Institute)

The conjugate gradient method with multigrid preconditioners (MGCG) is used in scientific applications because of its high performance and scalability with many computational nodes. GPUs are thought to be good candidates for accelerating such applications with many meshes where an MGCG solver could show high performance. No previous studies have evaluated and discussed the numerical character of an MGCG solver on GPUs. Consequently, we have implemented and optimized our “kinaco” numerical ocean model with an MGCG solver on GPUs. We evaluated its performance and discussed inter-GPU communications on a coarse grid on which GPUs could be intrinsically problematic. We achieved 3.9 times speedup compared to CPUs and learned how inter-GPU communications depended on the number of GPUs and the aggregation level of information in a multigrid method.
**Poster 108: Power Prediction for High-Performance Computing**
Shigeto Suzuki (Fujitsu Laboratories Ltd), Michiko Hiraoka (Fujitsu Ltd), Takashi Shiraishi (Fujitsu Laboratories Ltd), Enxhi Kreshpa (Fujitsu Laboratories Ltd), Takuji Yamamoto (Fujitsu Laboratories Ltd), Hiroyuki Fukuda (Fujitsu Laboratories Ltd), Shuji Matsui (Fujitsu Ltd), Masahide Fujisaki (Fujitsu Ltd), Atsuya Uno (RIKEN Center for Computational Science (R-CCS))

Exascale computers consume large amounts of power both for computing and cooling-units. As power of the computer varies dynamically corresponding to the load change, cooling-units are desirable to follow it for effective energy management. Because of time lags in cooling-unit operations, advance control is inevitable and an accurate prediction is a key for it. Conventional prediction methods make use of the similarity between job information while in queue. The prediction fails if there is no previously similar job. We developed two models to correct the prediction after queued jobs start running. By taking power histories into account, power-correlated topic model reselects more suitable candidate and recurrent-neural-network model considering variable network sizes predicts power variation from shape features of it. We integrated these into a single algorithm and demonstrated high-precision prediction with an average relative error of 5.7% in K computer as compared to the 18.0% obtained using the conventional method.

**Poster 80: Sharing and Replicability of Notebook-Based Research on Open Testbeds**
Maxine V. King (University of Chicago), Jason Anderson (University of Chicago), Kate Keahey (Argonne National Laboratory, University of Chicago)

We seek to facilitate replicability by creating a way to share experiments easily in and out of notebook-based, open testbed environments and a sharing platform for such experiments in order to allow researchers to combine shareability, consistency of code environment, and well-documented process.

**Poster 121: HFlush: Realtime Flushing for Modern Storage Environments**
Jaime Cernuda (Illinois Institute of Technology), Hugo Trivino (Illinois Institute of Technology), Hariharan Devarajan (Illinois Institute of Technology), Anthony Koukcas (Illinois Institute of
Due to the unparalleled magnitude of data movement in extreme scale computing, I/O has become a central challenge. Modern storage environments have proposed the use of multiple layers between applications and the PFS. Nonetheless, the difference in capacities and speeds between storage layers makes it extremely challenging to evict data from upper layers to lower layers efficiently. However, current solutions are executed in batches, compromising latency; are also push-based implementations, compromising resource utilization. Hence, we propose HFlush, a continuous data eviction mechanism built on a streaming architecture that is pull-based and in which each component is decoupled and executed in parallel. Initial results have shown RFlush to obtain a 7X latency reduction and a 2X bandwidth improvement over a baseline batch-based system. Therefore, RFlush is a promising solution to the growing challenges of extreme scale data generation and eviction shortcomings when archiving data across multiple tiers of storage.

Best Poster Finalist: no

Poster 88: HPC Container Runtime Performance Overhead: At First Order, There Is None
Alfred Torrez (Los Alamos National Laboratory), Reid Priedhorsky (Los Alamos National Laboratory), Timothy Randles (Los Alamos National Laboratory)

Linux containers are an increasingly popular method used by HPC centers to meet increasing demand for greater software flexibility. A common concern is that containers may introduce application performance overhead. Prior work has not tested a broad set of HPC container technologies on a broad set of benchmarks. This poster addresses the gap by comparing performance of the three HPC container implementations (Charliecloud, Shifter, and Singularity) and bare metal on multiple dimensions using industry-standard benchmarks.

We found no meaningful performance differences between the four environments with the possible exception of modest variation in memory usage, which is broadly consistent with prior results. This result suggests that HPC users should feel free to containerize their applications without concern about performance degradation, regardless of the container technology used. It is an encouraging development on the path towards greater adoption of user-defined software stacks to increase the flexibility of HPC.

Best Poster Finalist: no

Poster 64: 416-PFLOPS Fast Scalable Implicit Solver on Low-Ordered Unstructured Finite
Elements Accelerated by 1.10-ExaFLOPS Kernel with Reformulated AI-Like Algorithm: For Equation-Based Earthquake Modeling

Tsuyoshi Ichimura (University of Tokyo, RIKEN), Kohei Fujita (University of Tokyo, RIKEN), Takuma Yamaguchi (University of Tokyo), Akira Naruse (Nvidia Corporation), Jack C. Wells (Oak Ridge National Laboratory), Christopher J. Zimmer (Oak Ridge National Laboratory), Tjerk P. Straatsma (Oak Ridge National Laboratory), Takane Hori (Japan Agency for Marine-Earth Science and Technology), Simone Puel (University of Texas), Thorsten W. Becker (University of Texas), Muneo Hori (Japan Agency for Marine-Earth Science and Technology), Naonori Ueda (RIKEN)

We propose herein an approach for reformulating an equation-based modeling algorithm to an algorithm similar to that of training artificial intelligence (AI) and accelerate this algorithm using high-performance accelerators to reduce the huge computational costs encountered for physics equation-based modeling in earthquake disaster mitigation. A fast scalable equation-based implicit solver on unstructured finite elements is accelerated with a Tensor Core-enabled matrix-vector product kernel. The developed kernel attains 1.10 ExaFLOPS, leading to 416 PFLOPS for the whole solver on full Summit. This corresponds to a 75-fold speedup from a previous state-of-the-art solver running on full Piz Daint. This result could lead to breakthroughs in earthquake disaster mitigation. Our new idea in the HPC algorithm design of combining equation-based modeling with AI is expected to have broad impacts in other earth science and industrial problems.

Best Poster Finalist: no

5:15 pm - 7:00 pm

Poster Reception

Wednesday, November 20

8:30 am - 5:00 pm

Research Posters Display

Poster 74: Enabling Code Portability of a Parallel and Distributed Smooth-Particle Hydrodynamics Application, FleCSPH

Suyash Tandon (University of Michigan), Nicholas Stegmeier (University of Illinois), Vasu Jaganath (University of Wyoming), Jennifer Ranta (Michigan State University), Rathish Ratnasingam
Core-collapse supernovae (CCSNe) are integral to the formation and distribution of heavy elements across the universe. However, CCSNe are highly complex and inherently non-linear phenomena. Large-scale simulations of these cosmic events can provide us a glimpse of their hydrodynamic and nucleosynthetic processes which are difficult to observe. To enable these massive numerical simulations on high-performance computing (HPC) centers, this study uses FleCSPH, a parallel and distributed code, based on the smooth-particle hydrodynamics (SPH) formulation. In the recent years, the HPC architecture has evolved and the next generation of exascale computers are expected to feature heterogenous architecture. Therefore, it is important to maintain code portability across platforms. This work demonstrates code portability of FleCSPH through the incorporation of Kokkos C++ library and containers using Charliecloud.

Best Poster Finalist: no

**Poster 73: Accelerating Large-Scale GW Calculations on Hybrid CPU-GPU Architectures**

Mauro Del Ben (Lawrence Berkeley National Laboratory), Charlene Yang (National Energy Research Scientific Computing Center (NERSC)), Felipe Jornada (University of California, Berkeley; Lawrence Berkeley National Laboratory), Steven G. Louie (University of California, Berkeley; Lawrence Berkeley National Laboratory), Jack Deslippe (National Energy Research Scientific Computing Center (NERSC))

In this poster, we present the strategy, progress, and performance while GPU porting one of the major modules, epsilon, of the electronic structure code BerkeleyGW. Epsilon represents the most time-consuming routines in the BerkeleyGW workflow for large-scale material science simulations. Some of the porting/optimization strategies include, changing our original data layout to efficiently use libraries such as cuBLAS and cuFFT, implementation of specific CUDA kernels to minimize data copies between host/device and keeping data on device, efficient use of data streams to leverage high concurrency on the device, asynchronous memory copies and overlapping (MPI) communication on the host and computation on the device. Preliminary results are presented in terms of the speedup compare to the CPU-only implementation, strong/weak scaling, and power efficiency. Excellent acceleration is demonstrated: up to 30x for specific kernels. Our port also exhibits good scalability and about 16x higher FLOPs/watt efficiency compared to the CPU-only implementation.

Best Poster Finalist: no
**Poster 89: BeeCWL: A CWL Compliant Workflow Management System**
Betis Baheri (Kent State University), Steven Anaya (New Mexico Institute of Mining and Technology), Patricia Grubel (Los Alamos National Laboratory), Qiang Guan (Kent State University), Timothy Randles (Los Alamos National Laboratory)

Scientific workflows are used widely to carry out complex and hierarchical experiments. Although there are many trends to extend the functionality of workflow management systems to cover all possible requirements that may arise from a user community, one unified standard over cloud and HPC systems is still missing. In this paper, we propose a Common Workflow Language (CWL) compliant workflow management system. BeeCWL is a parser to derive meaningful information such as requirements, steps, relationships, etc. from CWL files and to create a graph database from those components. Generated graphs can be passed to an arbitrary scheduler and management system to decide whether there are enough resources to optimize and execute the workflow. Lastly, the user can have control over workflow execution, collecting logs, and restart or rerun some part of a complex workflow.

Best Poster Finalist: no

**Poster 150: A Machine Learning Approach to Understanding HPC Application Performance Variation**
Burak Aksar (Boston University, Sandia National Laboratories), Benjamin Schwaller (Sandia National Laboratories), Omar Aaziz (Sandia National Laboratories), Emre Ates (Boston University), Jim Brandt (Sandia National Laboratories), Ayse K. Coskun (Boston University), Manuel Egele (Boston University), Vitus Leung (Sandia National Laboratories)

Performance anomalies are difficult to detect because often a “healthy system” is vaguely defined, and the ground truth for how a system should be operating is evasive. As we move to exascale, however, detection of performance anomalies will become increasingly important with the increase in size and complexity of systems. There are very few accepted ways of detecting anomalies in the literature, and there are no published and labeled sets of anomalous HPC behavior. In this research, we develop a suite of applications that represent HPC workloads and use data from a lightweight metric collection service to train machine learning models to predict the future behavior of metrics. In the future, this work will be used to predict anomalous runs in compute nodes and determine some root causes of performance issues to help improve the efficiency of HPC system administrators and users.
**Poster 81: Performance of Devito on HPC-Optimised ARM Processors**

Hermes Senger (Federal University of São Carlos, Brazil; University of São Paulo), Jaime Freire de Souza (Federal University of São Carlos, Brazil), Edson Satoshi Gomi (University of São Paulo), Fabio Luporini (Imperial College, London), Gerard Gorman (Imperial College, London)

We evaluate the performance of Devito, a domain specific language (DSL) for finite differences on Arm ThunderX2 processors. Experiments with two common seismic computational kernels demonstrate that Devito can apply automatic code generation and optimization across Arm and Intel platforms. The code transformations include: parallelism, and SIMD vectorization (OpenMP >=4); loop tiling (with best block shape obtained via auto-tuning); domain-specific symbolic optimisation such as common sub-expression elimination and factorisation for Flop reduction, polynomial approximations for trigonometry terms, and heuristic hoisting of time-invariant expressions. Results show that Devito can achieve performance on Arm processors which is competitive to other Intel Xeon processors.

**Poster 103: LIKWID 5: Lightweight Performance Tools**

Thomas Gruber (Erlangen Regional Computing Center), Jan Eitzinger (Erlangen Regional Computing Center), Georg Hager (Erlangen Regional Computing Center), Gerhard Wellein (Erlangen Regional Computing Center)

LIKWID is a tool suite for performance oriented programmers with a worldwide user group. It is developed by the HPC group of the University Erlangen-Nuremberg since 2009 to support them in their daily research and performance engineering of user codes. The HPC landscape has become more and more diverse over the last years with clusters using non-x86 architectures and being equipped with accelerators. With the new major version, the architectural support of LIKWID is extended to ARM and POWER CPUs with the same functionality and features as for x86 architectures. Besides the CPU monitoring, the new version provides access the hardware counting facilities of Nvidia GPUs. This poster introduces the new features and shows the successes of applying LIKWID to identify performance bottlenecks and to test optimizations. Furthermore, the poster gives an overview of how users can integrate the LIKWID tools in their application using a lightweight add-once-and-reuse instrumentation API.
**Poster 149: Solving Phase-Field Equations in Space-Time: Adaptive Space-Time Meshes and Stabilized Variational Formulations**

Kumar Saurabh (Iowa State University), Biswajit Khara (Iowa State University), Milinda Fernando (University of Utah), Masado Ishii (University of Utah), Hari Sundar (University of Utah), Baskar Ganapathysubramanian (Iowa State University)

We seek to efficiently solve a generalized class of partial differential equations called the phase-field equations. These non-linear PDE’s model phase transition (solidification, melting, phase-separation) phenomena which exhibit spatially and temporally localized regions of steep gradients. We consider time as an additional dimension and simultaneously solve for the unknown in large blocks of time (i.e. in space-time), instead of the standard approach of sequential time-stepping. We use variational multiscale (VMS) based finite element approach to solve the ensuing space-time equations. This allows us to (a) exploit parallelism not only in space but also in time, (b) gain high order accuracy in time, and (c) exploit adaptive refinement approaches to locally refine region of interest in both space and time. We illustrate this approach with several canonical problems including melting and solidification of complex snow flake structures.

Best Poster Finalist: no

**Poster 84: ESTEE: A Simulation Toolkit for Distributed Workflow Execution**

Vojtěch Cima (IT4Innovations, Czech Republic), Jakub Beránek (IT4Innovations, Czech Republic), Stanislav Böhm (IT4Innovations, Czech Republic)

Task graphs provide a simple way to describe scientific workflows (sets of tasks with dependencies) that can be executed on both HPC clusters and in the cloud. An important aspect of executing such graphs is the used scheduling algorithm. Many scheduling heuristics have been proposed in existing works; nevertheless, they are often tested in oversimplified environments. We introduce a simulation environment designed for prototyping and benchmarking task schedulers. Our simulation environment, scheduler source codes, and graph datasets are open in order to be fully reproducible. To demonstrate usage of Estee, as an example, we compare the performance of various workflow schedulers in an environment using two different network models.

Best Poster Finalist: no

**Poster 93: Robust Data-Driven Power Simulator for Fast Cooling Control Optimization of a**
Large-Scale Computing System
Takashi Shiraishi (Fujitsu Laboratories Ltd), Hiroshi Endo (Fujitsu Laboratories Ltd), Takaaki Hineno (Fujitsu Laboratories Ltd), Hiroyuki Fukuda (Fujitsu Laboratories Ltd)

Power of large-scale systems such as an HPC or a datacenter is a significant issue. Cooling units consume 30% of the total power. General control policies for cooling units are local and static (manual overall optimization nearly once a week). However, free cooling and IT-load fluctuation may change hourly optimum control variables of the cooling units. In this work, we present a deep neural network (DNN) power simulator that can learn from actual operating logs and can quickly identify the optimum control variables. We demonstrated the power simulator of an actual large-scale system with 4.7-MW-power IT load. Our robust simulator predicted the total power with error of 4.8% without retraining during one year. We achieved optimization by the simulator within 80 seconds that was drastically faster than previous works. The dynamic control optimization each hour showed a 15% power reduction compared to that of conventional policy in the actual system.

Best Poster Finalist: no

Poster 115: sDNA: Software-Defined Network Accelerator Based on Optical Interconnection Architecture
En Shao (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences), Guangming Tan (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences), Zhan Wang (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences), Guojun Yuan (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences), Ninghui Sun (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences)

Software-Defined Network Accelerator (sDNA) is a new accelerated system for the exascale computer. Inspired by the edge forwarding index (EFI), the main contribution of our work is that it presents an extended EFI-based optical interconnection method with slow switching optical device. In our work, we found that sDNA based on extended EFI evaluation is not only able to offload the traffic from an electrical link to an optical link but is also able to avoid congestion inherent to electrical link.

Best Poster Finalist: no

Poster 49: WarpX: Toward Exascale Modeling of Plasma Particle Accelerators on GPU
Maxence Thevenet (Lawrence Berkeley National Laboratory), Jean-Luc Vay (Lawrence Berkeley
Particle accelerators are a vital part of the DOE-supported infrastructure of discovery science and applications, but we need game-changing improvements in the size and cost for future accelerators. Plasma-based particle accelerators stand apart in their potential for these improvements. Turning this from a promising technology into mainstream scientific tools depends critically on high-performance, high-fidelity modeling of complex processes.

WarpX is an open-source particle-in-cell (PIC) code supported by the Exascale Computing Project (ECP) that is combining advanced algorithms with adaptive mesh refinement to allow challenging simulations of a multi-stage plasma-based TeV acceleration relevant for future high-energy physics discoveries. WarpX relies on the ECP co-design center for mesh refinement AMReX, and runs on CPU and GPU-accelerated computers. Production simulation have run on Cori KNL at NERSC and Summit at OLCF. In this poster, recent results and strategies on GPU will be presented, along with recent performance results.

Best Poster Finalist: no

**Poster 50: Implementing an Adaptive Sparse Grid Discretization (ASGarD) for High Dimensional Advection-Diffusion Problems on Exascale Architectures**

M. Graham Lopez (Oak Ridge National Laboratory), David L. Green (Oak Ridge National Laboratory), Lin Mu (University of Georgia), Ed D’Azevedo (Oak Ridge National Laboratory), Wael Elwasif (Oak Ridge National Laboratory), Tyler McDaniel (University of Tennessee), Timothy Younkin (University of Tennessee), Adam McDaniel (Oak Ridge National Laboratory), Diego Del-Castillo-Negrete (Oak Ridge National Laboratory)

Many scientific domains require the solution of high dimensional PDEs. Traditional grid- or mesh-based methods for solving such systems in a noise-free manner quickly become intractable due to the scaling of the degrees of freedom going as $O(N^{d})$ sometimes called "the curse of dimensionality." We are developing an arbitrarily high-order discontinuous-Galerkin finite-element
solver that leverages an adaptive sparse-grid discretization whose degrees of freedom scale as $O(N \log_2 N^{D-1})$. This method and its subsequent reduction in the required resources is being applied to several PDEs including time-domain Maxwell's equations (3D), the Vlasov equation (in up to 6D) and a Fokker-Planck-like problem in ongoing related efforts. Here we present our implementation which is designed to run on multiple accelerated architectures, including distributed systems. Our implementation takes advantage of a system matrix decomposed as the Kronecker product of many smaller matrices which is implemented as batched operations.

Best Poster Finalist: no

Poster 51: SmartK: Efficient, Scalable, and Winning Parallel MCTS
Michael S. Davinroy (Swarthmore College), Shawn Pan (Swarthmore College), Bryce Wiedenbeck (Swarthmore College, Davidson College), Tia Newhall (Swarthmore College)

SmartK is our efficient and scalable parallel algorithm for Monte Carlo Tree Search (MCTS), an approximation technique for game searches. MCTS is also used to solve problems as diverse as planning under uncertainty, combinatorial optimization, and high-energy physics. In these problems, the solution search space is significantly large, necessitating parallel solutions. Shared memory parallel approaches do not scale well beyond the size of a single node's RAM. SmartK is a distributed memory parallelization that takes advantage of both inter-node and intra-node parallelism and a large cumulative RAM found in clusters. SmartK’s novel selection algorithm combined with its ability to efficiently search the solution space, results in better solutions than other MCTS parallel approaches. Results of an MPI implementation of SmartK for the game of Hex, show SmartK yields a better win percentage than other parallel algorithms, and that its performance scales to larger search spaces and high degrees of parallelism.

Best Poster Finalist: no

Poster 70: Numerical Method and Parallelization for the Computation of Coherent Synchrotron Radiation
Boqian Shen (Rice University, Los Alamos National Laboratory)

The purpose of this work is to develop and parallelize an accurate and efficient numerical method for the computation of synchrotron radiation from relativistic electrons in the near field. The high-brilliance electron beam and coherent short-wavelength light source provide a powerful method to understand the microscopic structure and dynamics of materials. Such a method supports a wide range of applications including matter physics, structural biology, and medicine development. To
understand the interaction between the beam and synchrotron radiation, an accurate and efficient numerical simulation is needed. With millions of electrons, the computational cost of the field would be large. Thus, multilevel parallelism and performance portability are desired since modern supercomputers are getting more complex and heterogeneous. The performance model and performance analysis are presented.

Best Poster Finalist: no

**Poster 146: AI Matrix: A Deep Learning Benchmark for Alibaba Data Centers**  
Wei Zhang (Alibaba Inc), Wei Wei (Alibaba Inc), Lingjie Xu (Alibaba Inc), Lingling Jin (Alibaba Inc), Cheng Li (University of Illinois)

This work introduces AI Matrix, an in-house Deep Learning (DL) benchmark suite developed specifically for Alibaba's e-commerce environment. AI Matrix results from a full investigation of the DL applications used inside Alibaba and aims to cover the typical DL applications that account for more than 90% of the GPU usage in Alibaba data centers. This benchmark suite collects DL models that are either directly used or closely resemble the models used in the company's real e-commerce applications. It also collects the real e-commerce applications if no similar DL models are not available. Through the high coverage and close resemblance to real applications, AI Matrix fully represents the DL workloads on Alibaba data centers. The collected benchmarks mainly fall into three categories: computer vision, recommendation, and language processing, which consist of the most majority of DL applications in Alibaba. AI Matrix is made open source, hoping it can benefit the public.

Best Poster Finalist: no

**Poster 100: Comparison of Array Management Library Performance - A Neuroscience Use Case**  
Donghe Kang (Ohio State University), Oliver Rübel (Lawrence Berkeley National Laboratory), Suren Byna (Lawrence Berkeley National Laboratory), Spyros Blanas (Ohio State University)

Array management libraries, such as HDF5, Zarr, etc., depend on a complex software stack that consists of parallel I/O middleware (MPI-IO), POSIX-IO, and file systems. Components in the stack are interdependent, such that effort in tuning the parameters in these software libraries for optimal performance is non-trivial. On the other hand, it is challenging to choose an array management library based on the array configuration and access patterns. In this poster, we investigate the performance aspect of two array management libraries, i.e., HDF5 and Zarr, in the context of a neuroscience use case. We highlight the performance variability of HDF5 and Zarr in our preliminary results and discuss potential optimization strategies.
Poster 118: Self-Driving Reconfigurable Silicon Photonic Interconnects (Flex-LIONS) with Deep Reinforcement Learning
Roberto Proietti (University of California, Davis), Yu Shang (University of California, Davis), Xian Xiao (University of California, Davis), Xiaoliang Chen (University of California, Davis), Yu Zhang (University of California, Davis), SJ Ben Yoo (University of California, Davis)

We propose a self-driving reconfigurable optical interconnect architecture for HPC systems exploiting a deep reinforcement learning (DRL) algorithm and a reconfigurable silicon photonic (SiPh) switching fabric to adapt the interconnect topology to different traffic demands. Preliminary simulation results show that after training, the DRL-based SiPh fabric provides the lowest average end-to-end latency for time-varying traffic patterns.

Poster 147: Extremely Accelerated Deep Learning: ResNet-50 Training in 70.4 Seconds
Akihiro Tabuchi (Fujitsu Laboratories Ltd), Akihiko Kasagi (Fujitsu Laboratories Ltd), Masafumi Yamazaki (Fujitsu Laboratories Ltd), Takumi Honda (Fujitsu Laboratories Ltd), Masahiro Miwa (Fujitsu Laboratories Ltd), Takashi Shiraishi (Fujitsu Laboratories Ltd), Motohiro Kosaki (Fujitsu Laboratories Ltd), Naoto Fukumoto (Fujitsu Laboratories Ltd), Tsuguchika Tabaru (Fujitsu Laboratories Ltd), Atsushi Ike (Fujitsu Laboratories Ltd), Kohta Nakashima (Fujitsu Laboratories Ltd)

Distributed deep learning using a large mini-batch is a key technology to accelerate training in deep learning. However, it is difficult to achieve a high scalability and maintain validation accuracy in distributed learning on large clusters. We introduce two optimizations, reducing the computation time and overlapping the communication with the computation. By applying the techniques and using 2,048 GPUs, we achieved the world's fastest ResNet-50 training in MLPerf, which is a de facto standard DNN benchmark (as of July 2019).

Poster 111: Multiple HPC Environments-Aware Container Image Configuration for Bioinformatics Application
Kento Aoyama (Tokyo Institute of Technology, National Institute of Advanced Industrial Science and
Containers have a considerable advantage for application portability in different environments by isolating process with a small performance overhead; thus it has been rapidly getting popular in a wide range of science fields. However, there are problems in container image configuration when run in multiple HPC environments, and it requires users to have knowledge of systems, container runtimes, container image format, and library compatibilities in HPC environments.

In this study, we introduce our HPC container workflow in multiple supercomputing environments that have different system/library specifications (ABCI, TSUBAME3.0). Our workflow provides custom container image configurations for HPC environments by taking into account differences in container runtime, container image, and library compatibility between the host and inside of the container. We also show the parallel performance of our application in each HPC environment.

Best Poster Finalist: no

**Poster 134: Minimal-Precision Computing for High-Performance, Energy-Efficient, and Reliable Computations**

Daichi Mukunoki (RIKEN Center for Computational Science (R-CCS)), Toshiyuki Imamura (RIKEN Center for Computational Science (R-CCS)), Yiyu Tan (RIKEN Center for Computational Science (R-CCS)), Atsushi Koshiba (RIKEN Center for Computational Science (R-CCS)), Jens Huthmann (RIKEN Center for Computational Science (R-CCS)), Kentaro Sano (RIKEN Center for Computational Science (R-CCS)), Fabienne Jézéquel (Sorbonne University), Stef Graillat (Sorbonne University), Roman Jakymchuk (Sorbonne University), Norihisa Fujita (University of Tsukuba), Taisuke Boku (University of Tsukuba)

In numerical computations, the precision of floating-point computations is a key factor to determine the performance (speed and energy-efficiency) as well as the reliability (accuracy and reproducibility). However, the precision generally plays a contrary role for both. Therefore, the ultimate concept for maximizing both at the same time is the minimal-precision computation through precision-tuning, which adjusts the optimal precision for each operation and data. Several studies have been already conducted for it so far, but the scope of those studies is limited to the precision-tuning alone. In this study, we propose a more broad concept of the minimal-precision computing with precision-tuning, involving both hardware and software stack.

Best Poster Finalist: no
**Poster 112: Building Complex Software Applications Inside Containers**  
*Calvin D. Seamons (Los Alamos National Laboratory)*

High performance computing (HPC) scientific applications require complex dependencies to operate. As user demand for HPC systems increases, it becomes unrealistic to support every unique dependency request. Containers can offer the ability to satisfy the users’ dependency request while simultaneously offering HPC portability across systems. By “containerizing” Model for Prediction Across Scales (MPAS, a large atmospheric simulation suite), we show that it is possible to containerize and run complex software. Furthermore, the container can be run across different HPC systems with nearly identical results (21 bytes difference over 2.1 gigabytes). Containers have the possibility to bring flexibility to code teams in HPC by helping to meet the demand for user defined software stacks (UDSS), and giving teams the ability to choose their software, independently of what is offered by the HPC system.

Best Poster Finalist: no

**Poster 101: Job Performance Overview of Apache Flink and Apache Spark Applications**  
*Jan Frenzel (Technical University Dresden), René Jäkel (Technical University Dresden)*

Apache Spark and Apache Flink are two Big Data frameworks used for fast data exploration and analysis. Both frameworks provide the runtime of program sections and performance metrics, such as the number of bytes read or written, via an integrated dashboard. Performance metrics available in the dashboard lack timely information and are only shown aggregated in a separate part of the dashboard. However, performance investigations and optimizations would benefit from an integrated view with detailed performance metric events. Thus, we propose a system that samples metrics at runtime and collects information about the program sections after the execution finishes. The performance data is stored in an established format independent from Spark and Flink versions and can be viewed with state-of-the-art performance tools, i.e. Vampir. The overhead depends on the sampling interval and was below 10% in our experiments.

Best Poster Finalist: no

**Poster 113: Improvements Toward the Release of the Pavilion 2.0 Test Harness**  
*Kody J. Everson (Los Alamos National Laboratory, Dakota State University), Maria Francine Lapid (Los Alamos National Laboratory)*
High-performance computing production support entails thorough testing in order to evaluate the efficacy of a system for production-grade workloads. There are various phases of a system’s lifecycle to assess, requiring different methods to accomplish effective evaluation of performance and correctness. Due to the unique and distributed nature of an HPC-system, the necessity for sophisticated tools to automatically harness and assess test results, all while interacting with schedulers and programming environment software, requires a customizable, extensible, and lightweight system to manage concurrent testing. Beginning with the recently refactored codebase of Pavilion 1.0, we assisted with the finishing touches on readying this software for open-source release and production usage. Pavilion 2.0 is a Python 3-based testing framework for HPC clusters that facilitates the building, running, and analysis of tests through an easy-to-use, flexible, YAML-based configuration system. This enables users to write their own tests by simply wrapping everything in Pavilion’s well-defined format.

Best Poster Finalist: no

Poster 119: Toward Lattice QCD on Fugaku: SVE Compiler Studies and Micro-Benchmarks in the RIKEN Fugaku Processor Simulator
Nils Meyer (University of Regensburg, Bavaria), Tilo Wettig (University of Regensburg, Bavaria), Yuetsu Kodama (RIKEN Center for Computational Science (R-CCS)), Mitsuhisa Sato (RIKEN Center for Computational Science (R-CCS))

The Fugaku supercomputer, successor to the Japanese flagship K-Computer, will start operation in 2021. Fugaku incorporates the Fujitsu A64FX processor, which is the first hardware implementation supporting the Arm SVE instruction set, in this case a 512-bit version. Real hardware is not accessible today, but RIKEN has designed a simulator of the A64FX. We present micro-benchmarks relevant for Lattice QCD obtained in the RIKEN Fugaku processor simulator and compare three different SVE compilers.

Best Poster Finalist: no

Poster 62: Emulating Multi-Pattern Quantum Grover’s Search on a High-Performance Reconfigurable Computer
Naveed Mahmud (University of Kansas), Bennett Haase-Divine (University of Kansas), Bailey K. Srimoungchanh (University of Kansas), Nolan Blankenau (University of Kansas), Annika Kuhnke (University of Kansas), Esam El-Araby (University of Kansas)

Grover's search(GS) is a widely studied quantum algorithm that can be employed for both single and
multi-pattern search problems and potentially provides quadratic speedup over existing classical search algorithms. In this paper, we propose a multi-pattern quantum search methodology based on a modified GS quantum circuit. The proposed method combines classical post-processing permutations with a modified Grover’s circuit to efficiently search for given single/multiple input patterns. Our proposed methodology reduces quantum circuit complexity, realizes space-efficient emulation hardware and improves overall system configurability for dynamic, multi-pattern search. We use a high-performance reconfigurable computer to emulate multi-pattern GS(MGS) and present scalable emulation architectures of a complete multi-pattern search system. We validate the system and provide analysis of experimental results in terms of FPGA resource utilization and emulation time. Our results include a successful hardware architecture that is capable of emulating MGS algorithm up to 32 fully-entangled quantum bits on a single FPGA.

Best Poster Finalist: no

Poster 107: Exploring Interprocess Work Stealing for Balanced MPI Communication
Kaiming Ouyang (University of California, Riverside), Min Si (Argonne National Laboratory), Zizhong Chen (University of California, Riverside)

Workload balance among MPI processes is a critical consideration during the development of HPC applications. However, because of many factors such as complex network interconnections and irregularity of HPC applications, fully achieving workload balance in practice is nearly impossible. Although interprocess job stealing is a promising solution, existing shared-memory techniques that lack necessary flexibility or cause inefficiency during data access cannot provide an applicable job-stealing implementation. To solve this problem, we propose a new process-in-process (PiP) interprocess job-stealing method to balance communication workload among processes on MPI layers. Our initial experimental results show PiP-based job stealing can efficiently help amortize workload, reduce imbalance, and greatly improve intra- and intersocket ping-pong performance compared with original MPI.

Best Poster Finalist: no

Poster 127: sFlow Monitoring for Security and Reliability
Xava A. Grooms (Los Alamos National Laboratory, University of Kentucky), Robert V. Rollins (Los Alamos National Laboratory, Michigan Technological University), Collin T. Rumpca (Los Alamos National Laboratory, Dakota State University)

In the past ten years, High Performance Computing (HPC) has moved far beyond the terascale
performance, making petascale systems the new standard. The drastic improvement in performance has been largely unmatched with insignificant improvements in system monitoring. Thus, there is an immediate need for practical and scalable monitoring solutions to ensure the effectiveness of costly compute clusters. This project aims to explore the viability and impact of sFlow enabled switches in cluster network monitoring for security and reliability. A series of tests and exploits were performed to target specific network abnormalities on a nine-node HPC cluster. The results present web-based dashboards that can aid network administrators in improving a cluster’s security and reliability.

Best Poster Finalist: no

**Poster 77: Extreme Scale Phase-Field Simulations of Sintering Processes**

Johannes Hötzer (Karlsruhe University of Applied Sciences), Henrik Hierl (Karlsruhe University of Applied Sciences), Marco Seiz (Karlsruhe Institute of Technology), Andreas Reiter (Karlsruhe Institute of Technology), Britta Nestler (Karlsruhe Institute of Technology)

The sintering process, which turns loose powders into dense materials, is naturally found in the formation of glaciers, but is also the indispensable process to manufacture ceramic materials. This process is described by a dynamically evolving microstructure, which largely influences the resulting material properties.

To investigate this complex three-dimensional, scale-bridging evolution in realistic domain sizes, a highly optimized and parallelized multiphysics phase-field solver is developed. The solver is optimized in a holistic way, from the application level over the time integration and parallelization, down to the hardware. Optimizations include communication hiding, explicit vectorization, implicit schemes, and local reduction of degrees of freedom.

With this, we are able to investigate large-scale, three-dimensional domains, and long integration times. We have achieved a single-core peak performance of 32.5%, scaled up to 98304 cores on Hazel Hen and SuperMUC-NG, and simulated a multimillion particle system.

Best Poster Finalist: no

**Poster 140: Toward Automatic Function Call Generation for Deep Learning**

Shizhi Tang (Tsinghua University, China), Jidong Zhai (Tsinghua University, China)

Mainstream deep learning frameworks are commonly implemented by invoking underlying high performance tensor libraries on various architectures. However, as these libraries provide increasingly
complex semantics including operator fusions, in-place operations, and various memory layouts, the gap between mathematical deep learning models and the underlying libraries becomes larger. In this paper, inspired by the classic problem of Instruction Selection, we design a theorem solver guided exhausted search algorithm to select functions for complex tensor computations. Preliminary results with some micro-benchmarks and a real model show that our approach can outperform both Tensorflow and Tensor Comprehensions at run time.

Best Poster Finalist: no

**Poster 83: ETL: Elastic Training Layer for Deep Learning**  
Lei Xie (Tsinghua University, China), Jidong Zhai (Tsinghua University, China)

Due to the rising of deep learning, clusters for deep learning training are widely deployed in production. However, static task configuration and resource fragmentation problems in existing clusters result in low efficiency and poor quality of service. We propose ETL, an elastic training layer for deep learning, to help address them once for all. ETL adopts many novel mechanisms, such as lightweight and configurable report primitive and asynchronous, parallel and IO-free state replication, to achieve both high elasticity and efficiency. The evaluation demonstrates the low overhead and high efficiency of these mechanisms and reveals the advantages of elastic deep learning supported by ETL.

Best Poster Finalist: no

**Poster 130: Deep Learning-Based Feature-Aware Data Modeling for Complex Physics Simulations**  
Qun Liu (Louisiana State University), Subhashis Hazarika (Ohio State University), John M. Patchett (Los Alamos National Laboratory), James P. Ahrens (Los Alamos National Laboratory), Ayan Biswas (Los Alamos National Laboratory)

Data modeling and reduction for in situ is important. Feature-driven methods for in situ data analysis and reduction are a priority for future exascale machines as there are currently very few such methods. We investigate a deep-learning-based workflow that targets in situ data processing using autoencoders. We employ integrated skip connections to obtain higher performance compared to the existing autoencoders. Our experiments demonstrate the initial success of the proposed framework and create optimism for the in situ use case.

Best Poster Finalist: no
Poster 125: Physics Informed Generative Adversarial Networks for Virtual Mechanical Testing
Julian Cuevas (NASA, University of Puerto Rico at Mayaguez), Patrick Leser (NASA), James Warner (NASA), Geoffrey Bomarito (NASA), William Leser (NASA)

Physics-informed generative adversarial networks (PI-GANs) are used to learn the underlying probability distributions of spatially-varying material properties (e.g., microstructure variability in a polycrystalline material). While standard GANs rely solely on data for training, PI-GANs encode physics in the form of stochastic differential equations using automatic differentiation. The goal here is to show that experimental data from a limited number of material tests can be used with PI-GANs to enable unlimited virtual testing for aerospace applications. Preliminary results using synthetically generated data are provided to demonstrate the proposed framework. Deep learning and automatic differentiation capabilities in Tensorflow were implemented on Nvidia Tesla V100 GPUs.

Best Poster Finalist: no

Poster 141: ExaGeoStatR: Harnessing HPC Capabilities for Large Scale Geospatial Modeling Using R
Sameh Abdulah (King Abdullah University of Science and Technology (KAUST)), Yuxiao Li (King Abdullah University of Science and Technology (KAUST)), Jian Cao (King Abdullah University of Science and Technology (KAUST)), Hatem Ltaief (King Abdullah University of Science and Technology (KAUST)), David Keyes (King Abdullah University of Science and Technology (KAUST)), Marc Genton (King Abdullah University of Science and Technology (KAUST)), Ying Sun (King Abdullah University of Science and Technology (KAUST))

Large-scale simulations and parallel computing techniques are becoming essential in Gaussian process calculations to lessen the complexity of geostatistics applications. The log-likelihood function is used in such applications to evaluate the model associated with a given set of measurements in existing n geographic locations. The evaluation of such a function requires $O(n^2)$ memory and $O(n^3)$ computation, which is infeasible for large datasets with existing software tools.

We present ExaGeoStatR, a package for large-scale geostatistics in R that computes the log-likelihood function on shared and distributed-memory, possibly equipped with GPU, using advanced linear algebra techniques. The package provides a high-level abstraction of the underlying architecture while enhancing the R developers' productivity. We demonstrate ExaGeoStatR package by illustrating its implementation details, analyzing its performance on various parallel architectures, and assessing its accuracy using synthetic datasets and a sea surface temperature dataset. The
performance evaluation involves spatial datasets with up to 250K observations.

Best Poster Finalist: no

**Poster 48: Runtime System for GPU-Based Hierarchical LU Factorization**

Qianxiang Ma (Tokyo Institute of Technology), Rio Yokota (Tokyo Institute of Technology, Global Scientific Information and Computing Center; Tokyo Institute of Technology)

Hierarchical low-rank approximation can reduce both the storage and computation costs of dense matrices, but its implementation is challenging. In this research, we tackle one of the most difficult problems of GPU parallelization of the factorization of these hierarchical matrices. To this end, we are developing a new runtime system for GPUs that can schedule all tasks into one GPU kernel. Other existing runtime systems, like cuGraph and Standford Legion, can only manage streams and kernel-level parallelism. Even without too much tuning, we achieved 4x better performance in H-LU factorization with a single GPU when comparing with a well-tuned CPU-based hierarchical matrix library, HLIBpro, on moderately sized matrices. Additionally, we have significantly less runtime overheads exposed when processing smaller matrices.

Best Poster Finalist: no

**Poster 145: Improving Data Compression with Deep Predictive Neural Network for Time Evolitional Data**

Rupak Roy (Florida State University), Kento Sato (RIKEN Center for Computational Science (R-CCS)), Jian Guo (RIKEN Center for Computational Science (R-CCS)), Jens Domke (RIKEN Center for Computational Science (R-CCS)), Weikuan Yu (Florida State University), Takaki Hatsui (RIKEN SPring-8 Center), Yasumasa Joti (Japan Synchrotron Radiation Research Institute)

Scientific applications/simulations periodically generate huge intermediate data. Storing or transferring such a large scale of data is critical. Fast I/O is important for making this process faster. One of the approaches to achieve fast I/O is data compression. Our goal is to achieve a delta technique that can improve the performance of existing data compression algorithms for time evolitional intermediate data.

In our approach, we compute the delta values from original data and data predicted by the deep predictive neural network. We pass these delta values through three phases which are preprocessing phase, partitioned entropy coding phase, and density-based spatial delta encoding phase.
In our poster, we present how our predictive delta technique can leverage the time evolutional data to produce highly concentrated small values. We show the improvement in compression ratio when our technique, combined with existing compression algorithms, are applied on the intermediate data for different datasets.

Best Poster Finalist: no

**Poster 144: Optimizing Asynchronous Multi-Level Checkpoint/Restart Configurations with Machine Learning**

Tonmoy Dey (Florida State University), Kento Sato (RIKEN Center for Computational Science (R-CCS)), Jian Guo (RIKEN Center for Computational Science (R-CCS)), Bogdan Nicolae (Argonne National Laboratory), Jens Domke (RIKEN Center for Computational Science (R-CCS)), Weikuan Yu (Florida State University), Franck Cappello (Argonne National Laboratory), Kathryn Mohror (Lawrence Livermore National Laboratory)

With the emergence of fast local storage, multi-level checkpointing (MLC) has become a common approach for efficient checkpointing. To utilize MLC efficiently, it is important to determine the optimal configuration for the checkpoint/restart (CR). There are mainly two approaches for determining the optimal configuration for CR, namely modeling and simulation approach. However, with MLC, CR becomes more complicated making the modeling approach inaccurate and the simulation approach though accurate, very slow. In this poster, we focus on optimizing the performance of CR by predicting the optimized checkpoint count and interval. This was achieved by combining the simulation approach with machine learning and neural network to leverage its accuracy without spending time on simulating different CR parameters. We demonstrate that our models can predict the optimized parameter values with minimal error when compared to the simulation approach.

Best Poster Finalist: no

**Poster 132: Optimizing Performance at Runtime Using Binary Rewriting**

Alexis Engelke (Technical University Munich), David Hildenbrand (Technical University Munich), Martin Schulz (Technical University Munich)

In addition to scalability, performance of sequential code in applications is an important factor in HPC. Typically, programs are compiled once, at which time optimizations are applied, and are then run several times. However, not all information relevant for performance optimizations are available at compile-time, restricting optimization possibilities. The generation of specialized code at runtime allows for further optimizations. Performing such specialization on binary code allows for initial code
to be generated at compile-time with only the relevant parts being rewritten at runtime, reducing the optimization overhead. For targeted optimizations and effective use of known runtime information, the rewriting process needs to be guided by the application itself, exploiting information only known to the developer.

We describe three approaches for self-guided binary rewriting explicitly guided by the running application and evaluate the performance of the optimized code as well as the performance of the rewriting process itself.

Best Poster Finalist: no

Poster 53: Unstructured Mesh Technologies for Fusion Simulations
Cameron Smith (Rensselaer Polytechnic Institute (RPI)), Gerrett Diamond (Rensselaer Polytechnic Institute (RPI)), Gopan Perumpilly (Rensselaer Polytechnic Institute (RPI)), Chonglin Zhang (Rensselaer Polytechnic Institute (RPI)), Agnieszka Truszkowska (Rensselaer Polytechnic Institute (RPI)), Morteza Hakimi (Rensselaer Polytechnic Institute (RPI)), Onkar Sahni (Rensselaer Polytechnic Institute (RPI)), Mark Shephard (Rensselaer Polytechnic Institute (RPI)), Eisung Yoon (Ulsan National Institute of Science and Technology, South Korea), Daniel Ibanez (Sandia National Laboratories)

Multiple unstructured mesh technologies are needed to define and execute plasma physics simulations. The domains of interest combine model features defined from physical fields within 3D CAD of the tokamak vessel with an antenna assembly, and 2D cross sections of the tokamak vessel. Mesh generation technologies must satisfy these geometric constraints and additional constraints imposed by the numerical models. Likewise, fusion simulations over these domains study a range of timescales and physical phenomena within a tokamak.

XGCm studies the development of plasma turbulence in the reactor vessel, GITRm studies impurity transport, and PetraM simulations model RF wave propagation in scrape off layer plasmas. GITRm and XGCm developments are using the PUMIpic infrastructure to manage the storage and access of non-uniform particle distributions in unstructured meshes on GPUs. PetraM combines PUMI adaptive unstructured mesh control with MFEM using CAD models and meshes defined with Simmetrix tools.

Best Poster Finalist: no

Poster 99: Eithne: A Framework for Benchmarking Micro-Core Accelerators
Maurice C. Jamieson (Edinburgh Parallel Computing Centre, University of Edinburgh), Nick Brown (Edinburgh Parallel Computing Centre, University of Edinburgh)
Running existing HPC benchmarks as-is on micro-core architectures is at best difficult and most often impossible as they have a number of architectural features that makes them significantly different from traditional CPUs: tiny amounts on-chip RAM (c. 32KB), low-level knowledge specific to each device (including the host/device communications interface), limited communications bandwidth and complex or no device debugging environment. In order to compare and contrast different the micro-core architectures, a benchmark framework is required to abstract much of this complexity.

The modular Eithne framework supports the comparison of a number of micro-core architectures. The framework separates the actual benchmark from the details of how this is executed on the different technologies. The framework was evaluated by running the LINPACK benchmark on the Adapteva Epiphany, PicoRV32 and VectorBlox Orca RISC-V soft-cores, NXP RV32M1, ARM Cortex-A9, and Xilinx MicroBlaze soft-core, and comparing resulting performance and power consumption.

**Poster 133: Portable Resilience with Kokkos**

Jeffery Miles (Sandia National Laboratories), Nicolas Morales (Sandia National Laboratories), Carson Mould (Sandia National Laboratories), Keita Teranishi (Sandia National Laboratories)

The Kokkos ecosystem is a programming environment that provides performance and portability to many scientific applications that run on DOE supercomputers as well as other smaller scale systems. Leveraging software abstraction concepts within Kokkos, software resilience for end user code is made portable with abstractions and concepts while implementing the most efficient resilience algorithms internally. This addition enables an application to manage hardware failures reducing the cost of interruption without drastically increasing the software maintenance cost. Two main resilience methodologies have been added to the Kokkos ecosystem to validate the resilience abstractions: 1. Checkpointing includes an automatic mode supporting other checkpointing libraries and a manual mode which leverages the data abstraction and memory space concepts. 2. The redundant execution model anticipates failures by replicating data and execution paths. The design and implementation of these additions are illustrated, and appropriate examples are included to demonstrate the simplicity of use.

Best Poster Finalist: no

**Poster 79: The HPC PowerStack: A Community-Wide Collaboration Toward an Energy Efficient**
Software Stack
Siddhartha Jana (Energy Efficient HPC Working Group, Intel Corporation), Stephanie Brink (Lawrence Livermore National Laboratory), Christopher Cantalupo (Intel Corporation), Jonathan Eastep (Intel Corporation), Masaaki Kondo (RIKEN Advanced Institute for Computational Science (AICS), University of Tokyo), Matthias Maiterth (Intel Corporation), Aniruddha Marathe (Lawrence Livermore National Laboratory), Tapasya Patki (Lawrence Livermore National Laboratory), Barry Rountree (Lawrence Livermore National Laboratory), Ryuichi Sakamoto (University of Tokyo), Martin Schulz (Technical University Munich, Leibniz Supercomputing Centre), Carsten Trinitis (Technical University Munich), Josef Weidendorfer (Technical University Munich, Leibniz Supercomputing Centre)

This poster highlights an ongoing community-wide effort among vendors, labs, and academia, to incorporate power-awareness within system-stacks in upcoming exascale machines. HPC PowerStack is the first-and-only community-driven vendor-neutral effort to identify what power optimization software actors are key within the modern-day stack; discuss their interoperability, and work toward gluing together existing open source projects to engineer cost-effective, but cohesive, portable implementations.

This poster disseminates key insights acquired in the project, provides prototyping status updates, highlights open questions, and solicits participation addressing the imminent exascale power challenge.

Best Poster Finalist: no

Poster 87: Parallelizing Simulations of Large Quantum Circuits
Michael A. Perlin (University of Colorado, National Institute of Standards and Technology (NIST)), Teague Tomesh (Princeton University), Bradley Pearlman (University of Colorado, National Institute of Standards and Technology (NIST)), Wei Tang (Princeton University), Yuri Alexeev (Argonne National Laboratory), Martin Suchara (Argonne National Laboratory)

We present a parallelization scheme for classical simulations of quantum circuits. Our scheme is based on a recent method to “cut” large quantum circuits into smaller sub-circuits that can be simulated independently, and whose simulation results can in turn be re-combined to infer the output of the original circuit. The exponentially smaller classical computing resources needed to simulate smaller circuits are counterbalanced by exponential overhead in terms of classical post-processing costs. We discuss how this overhead can be massively parallelized to reduce classical computing costs.

Best Poster Finalist: no
Poster 120: ILP-Based Scheduling for Linear-Tape Model Trapped-Ion Quantum Computers
Xin-Chuan Wu (University of Chicago), Yongshan Ding (University of Chicago), Yunong Shi (University of Chicago), Yuri Alexeev (Argonne National Laboratory), Hal Finkel (Argonne National Laboratory), Kibaek Kim (Argonne National Laboratory), Frederic T. Chong (University of Chicago)

Quantum computing (QC) is emerging as a potential post-Moore high-performance computing (HPC) technology. Trapped-ion quantum bits (qubits) are among the most leading technologies to reach scalable quantum computers that would solve certain problems beyond the capabilities of even the largest classical supercomputers. In trapped-ion QC, qubits can physically move on the ion trap. The state-of-the-art architecture, linear-tape model, only requires a few laser beams to interact with the entire qubits by physically moving the interacting ions to the execution zone. Since the laser beams are limited resources, the ion chain movement and quantum gate scheduling are critical for the circuit latency. To harness the emerging architecture, we present our mathematical model for scheduling the qubit movements and quantum gates in order to minimize the circuit latency. In our experiment, our scheduling reduces 29.47% circuit latency on average. The results suggest classical HPC would further improve the quantum circuit optimization.

Best Poster Finalist: no

Poster 55: MPI+OpenMP Parallelization of DFT Method in GAMESS
Vladimir Mironov (Lomonosov Moscow State University), Yuri Alexeev (Argonne National Laboratory), Dmitry Fedorov (National Institute of Advanced Industrial Science and Technology (AIST))

In this work, the Density Functional Theory (DFT) method is parallelized with MPI-OpenMP in the quantum chemistry package GAMESS. It has been implemented in both regular and Fragment Molecular Orbital (FMO) based DFT codes. The scalability of the FMO-DFT code was demonstrated on Cray XC40 Theta supercomputer. We demonstrated excellent scalability of the code up 2,048 Intel Xeon Phi nodes (131,072 cores). Moreover, the developed DFT code is about twice as fast as the original code because of our new grid integration algorithm.

Best Poster Finalist: no

Poster 71: AI-Solver: Uncertainty in Prediction and Error Estimation for AI in Engineering
Ahmed Al-Jarro (Fujitsu Laboratories of Europe Ltd), Loic Beheshti (Fujitsu Laboratories of Europe Ltd), Serban Georgescu (Fujitsu Laboratories of Europe Ltd), Koichi Shirahata (Fujitsu Laboratories
The AI-Solver is a deep learning platform that learns from simulation data to extract general behavior based on physical parameters. The AI-Solver can handle a wide variety of classes of problems including those commonly identified in FEA, CFD, and CEM, to name a few, with speedups of up to 250,000X and extremely low error rate of 2-3%. In this work, we build on this recent effort. We first integrate uncertainty quantification, via exploiting the approximation of Bayesian Deep Learning. Second, we develop bespoke error estimation mechanisms capable of processing this uncertainty to provide instant feedback on the confidence in predictions without relying on the availability of ground truth data. To our knowledge, the ability to estimate the discrepancy in predictions without labels is a first in the field of AI for Engineering.

Best Poster Finalist: no

**Poster 72: Kokkos and Fortran in the Exascale Computing Project Plasma Physics Code XGC**

Aaron Scheinberg (Princeton Plasma Physics Laboratory), Guangye Chen (Los Alamos National Laboratory), Stephane Ethier (Princeton Plasma Physics Laboratory), Stuart Slattery (Oak Ridge National Laboratory), Robert Bird (Los Alamos National Laboratory), Pat Worley (PHWorley Consulting), Choong-Seock Chang (Princeton Plasma Physics Laboratory)

Numerical plasma physics models such as the particle-in-cell XGC code are important tools to understand phenomena encountered in experimental fusion devices. Adequately resolved simulations are computationally expensive, so optimization is essential. To address the need for consistent high performance by cutting-edge scientific software applications, frameworks such as Kokkos have been developed to enable portability as new architectures require hardware-specific coding implementation for best performance. Cabana, a recent extension to Kokkos developed with the ECP-CoPA project, is a library of common kernels and operations typically necessary for particle-based codes. The Kokkos/Cabana framework enables intuitive construction of particle-based codes, while maintaining portability between architectures. Here, we summarize the adoption by XGC of the execution and data layout patterns offered by this framework. We demonstrate a method for Fortran codes to adopt Kokkos and show that it can provide a single, portable code base that performs well on both GPUs and multicore machines.

Best Poster Finalist: no

**Poster 106: Optimizing Hybrid Access Virtual Memory System Using SCM/DRAM Unified Memory Management Unit**
Yusuke Shirota (Toshiba Corporation), Shiyo Yoshimura (Toshiba Corporation), Satoshi Shirai (Toshiba Corporation), Tatsunori Kanai (Toshiba Corporation)

In HPC systems, expectations for storage-class memory (SCM) are increasing in large-scale in-memory processing. While SCM can deliver higher capacity and lower standby power than DRAM, it is slower and the dynamic power is higher. Therefore, in order to realize high-speed, low-power and scalable main memory, it is necessary to build an SCM/DRAM unified memory, and dynamically optimize data placement between the two memories according to the memory access pattern.

In this poster, we describe a new hybrid access type virtual memory method using TLB-extended unified memory management unit which enables collecting and extracting fine-grained memory access locality characteristics. We show that with the proposed method, Hybrid Access control, which is a memory hierarchy control that selectively uses Direct Access to bus attached byte-addressable SCM and low power Aggressive Paging using small DRAM as cache, can be made more accurate, and the efficiency of memory access can be significantly improved.

Best Poster Finalist: no

Holistic Measurement Driven System Assessment
Saurabh Jha (University of Illinois), Mike Showerman (National Center for Supercomputing Applications (NCSA), University of Illinois), Aaron Saxton (National Center for Supercomputing Applications (NCSA), University of Illinois), Jeremy Enos (National Center for Supercomputing Applications (NCSA), University of Illinois), Greg Bauer (National Center for Supercomputing Applications (NCSA), University of Illinois), Zbigniew Kalbarczyk (University of Illinois), Ann Gentile (Sandia National Laboratories), Jim Brandt (Sandia National Laboratories), Ravi Iyer (University of Illinois), William T. Kramer (University of Illinois, National Center for Supercomputing Applications (NCSA))

HPC users deploy a suite of monitors to observe patterns of failures and performance anomalies to improve operational efficiency, achieve higher application performance and inform the design of future systems. However, the promises and the potential of monitoring data have largely been not realized due to various challenges such as inadequacy in monitoring, limited availability of data, lack of methods for fusing monitoring data at time-scales necessary for enabling human-in-the-loop or machine-in-the-loop feedback. To address above challenges, in this work we developed a monitoring fabric Holistic Measurement Driven System Assessment (HMDSA) for large-scale HPC facilities, independent of major component vendor, and within budget constraints of money, space, and power. We accomplish this through development and deployment of scalable, platform-independent, open-source tools and techniques for monitoring, coupled with statistical and machine-learning based
runtime analysis and feedback, which enables highly efficient HPC system operation and usage and also informs future system improvements.

Best Poster Finalist: no

**Poster 139: Model Identification of Pressure Drop in Membrane Channels with Multilayer Artificial Neural Networks**

Jiang-hang Gu (Sun Yat-sen University, Zhuhai, School of Chemical Engineering and Technology), Jiu Luo (Sun Yat-sen University, Guangzhou, School of Materials Science and Engineering), Ming-heng Li (California State Polytechnic University, Pomona), Yi Heng (Sun Yat-sen University, Guangzhou, School of Data and Computer Science; Sun Yat-sen University, Guangzhou, China)

This poster presents the work of identifying a data-driven model of pressure drop in spacer-filled reverse osmosis membrane channels and conducting CFD simulations. The established model correlates the pressure drop with a wide range of design objectives, which enables a quantitative description of the geometric structures and operation conditions for improvement. This way, it aims at optimizing the spacer geometry with minimal effort. Furthermore, a high-performance computing strategy is employed to tackle the resulted intractable computational task in the identification procedure and CFD simulations.

Best Poster Finalist: no

**Poster 61: Fast 3D Diffeomorphic Image Registration on GPUs**

Malte Brunn (University of Stuttgart), Naveen Himthani (University of Texas), George Biros (University of Texas), Miriam Mehl (University of Stuttgart), Andreas Mang (University of Houston)

3D image registration is one of the most fundamental and computationally expensive operations in medical image analysis. Here, we present a mixed-precision, Gauss-Newton-Krylov solver for diffeomorphic registration. Our work extends the publicly available CLAIRE library to GPU architectures. Despite the importance of image registration, only a few implementations of large deformation diffeomorphic registration packages support GPUs. Our contributions are new algorithms and dedicated computational kernels to significantly reduce the runtime of the main computational kernels in CLAIRE: derivatives and interpolation. We deploy (i) highly-optimized, mixed-precision GPU-kernels for the evaluation of scattered-data interpolation, (ii) replace FFT-based first-order derivatives with optimized 8th-order finite differences, and (iii) compare with state-of-the-art CPU and GPU implementations. As a highlight, we demonstrate that we can register 256^3 clinical images in less than 6 seconds on a single NVIDIA Tesla V100. This amounts to over 20x
speed-up over CLAIRE and over 30x speed-up over existing GPU implementations.

Best Poster Finalist: no

Poster 138: Across-Stack Profiling and Characterization of State-of-the-Art Machine Learning Models on GPUs
Cheng Li (University of Illinois), Abdul Dakkak (University of Illinois), Wei Wei (Alibaba Inc), Jinjun Xiong (IBM Research), Lingjie Xu (Alibaba Inc), Wei Zhang (Alibaba Inc), Wen-mei Hwu (University of Illinois)

The past few years have seen a surge of using Machine Learning (ML) and Deep Learning (DL) algorithms for traditional HPC tasks such as feature detection, numerical analysis, and graph analytics. While ML and DL enable solving HPC tasks, their adoption has been hampered due to the lack of understanding of how they utilize systems. Optimizing these algorithms requires characterizing their performance across the hardware/software (HW/SW) stack, but the lack of simple tools to automate the process and the reliance on researchers to perform manual characterization is a bottleneck. To alleviate this, we propose an across-stack profiling scheme and integrate it within MLModelScope — a hardware and software agnostic tool for evaluating and benchmarking ML/DL at scale. We demonstrate MLModelScope’s ability to characterize state-of-art ML/DL models and give insights that are only possible obtained by performing across-stack profiling.

Best Poster Finalist: no

Poster 60: Massively Parallel Large-Scale Multi-Model Simulation of Tumor Development
Marco Berghoff (Karlsruhe Institute of Technology), Jakob Rosenbauer (Forschungszentrum Juelich), Alexander Schug (Forschungszentrum Juelich)

The temporal and spatial resolution in the microscopy of tissues has increased significantly within the last years, yielding new insights into the dynamics of tissue development and the role of the single-cell within it. A thorough theoretical description of the connection of single-cell processes to macroscopic tissue reorganizations is still lacking. Especially in tumor development, single cells play a crucial role in advance of tumor properties.

We developed a simulation framework that can model tissue development up to the centimeter scale with micrometer resolution of single cells. Through a full parallelization, it enables the efficient use of HPC systems, therefore enabling detailed simulations on a large scale. We developed a generalized tumor model that respects adhesion driven cell migration, cell-to-cell signaling, and mutation-driven
tumor heterogeneity. We scan the response of the tumor development depending on division inhibiting substances such as cytostatic agents.

Best Poster Finalist: no

Poster 114: Optimizing Recommendation System Inference Performance Based on GPU
Xiaowei Shen (Alibaba Inc), Junrui Zhou (Alibaba Inc), Kan Liu (Alibaba Inc), Lingling Jin (Alibaba Inc), Pengfei Fan (Alibaba Inc), Wei Zhang (Alibaba Inc), Jun Yang (University of Pittsburgh)

Neural network-based recommendation models have been widely applied on tracking personalization and recommendation tasks at large Internet companies such as e-commerce companies and social media companies. Alibaba recommendation system deploys WDL (wide and deep learning) models for product recommendation tasks. The WDL model consists of two main parts: embedding lookup and neural network-based feature ranking model that ranks different products for different users. As more and more products and users the model need to rank, the feature length and batch size of the models are increased. The computation of models is also increased so that traditional model inference implementation on CPU cannot meet the requirement of QPS (query per second) and latency of recommendation tasks. In this poster, we develop a GPU based system to speedup recommendation system inference performance. By model quantization and graph transformation, we can achieve 3.9x performance speedup when compared with a baseline GPU implementation.

Best Poster Finalist: no

Poster 59: Accelerating BFS and SSSP on a NUMA Machine for the Graph500 Challenge
Tanuj K. Aasawat (RIKEN), Kazuki Yoshizoe (RIKEN), Tahsin Reza (University of British Columbia), Matei Ripeanu (University of British Columbia)

The NUMA architecture is the design choice for modern multi-CPU shared memory systems. In many ways, a NUMA system resembles a shared-nothing distributed system: memory accesses to remote NUMA domains are more expensive than local accesses.

In this work, we explore how improved data locality and reduced expensive remote communication can be achieved by exploiting "distributed" shared-memory of NUMA machines to develop shared-memory graph processing solutions optimized for NUMA systems. We introduce a novel hybrid design for memory accesses that handles the burst mode in traversal based algorithms, like BFS and SSSP, and reduces the number of remote accesses and updates. We demonstrate that our designs offer up to 84% speedup over our NUMA-oblivious framework Totem and 2.86x over shared-nothing
Commercial adiabatic quantum annealers such as D-Wave 2000Q have the potential to solve NP-complete optimization problems efficiently. One of the primary constraints of such devices is the limited number and connectivity of their qubits. This research presents two exact decomposition methods (for the Maximum Clique and the Minimum Vertex Cover problem) that allow us to solve problems of arbitrarily large sizes by splitting them up recursively into a series of arbitrarily small subproblems. Those subproblems are then solved exactly or approximately using a quantum annealer. Whereas some previous approaches are based on heuristics that do not guarantee optimality of their solutions, our decomposition algorithms have the property that the optimal solution of the input problem can be reconstructed given all generated subproblems are solved optimally as well. We investigate various heuristic and exact bounds as well as reduction methods that help to increase the scalability of our approaches.

Best Poster Finalist: no

Poster 90: You Have to Break It to Make It: How On-Demand, Ephemeral Public Cloud Projects with Alces Flight Compute Resulted in the Open-Source OpenFlightHPC Project

Cristin Merritt (Alces Flight Limited; Alces Software Ltd, UK), Wil Mayers (Alces Flight Limited), Stu Franks (Alces Flight Limited)

Over three years ago the Alces Flight team made a decision to explore on-demand public cloud consumption for High Performance Computing (HPC). Our premise was simple, create a fully-featured, scalable HPC environment for research and scientific computing and provide it as a free subscription service to researchers in order to study how they would approach and consume this then new technology. This tool, Alces Flight Compute, would set out to chart how far away from the traditional bare-metal platforms our subscribers were willing to go. What we didn’t expect was that to get to their destination, our users would proceed to take our tool apart. This deconstruction has resulted in a new approach to HPC environment creation (the open-source OpenFlightHPC project), helped us better understand cloud adoption strategies, and handed over a set of guidelines to help those looking to bring public cloud into their HPC solution.
**Poster 126: Enforcing Crash Consistency of Scientific Applications in Non-Volatile Main Memory Systems**  
Tyler Coy (Washington State University, Vancouver), Xuechen Zhang (Washington State University, Vancouver)

This poster presents a compiler-assistant technique, NVPath, to automatically generate NVMM-aware persistent data structures which provide the same level of guarantee of crash consistency compared to the baseline code. Compiler-assistant code annotation and transformation is general and can be applied to applications using various data structures. Our experimental results with real-world scientific applications show that the performance of the annotated programs is commensurate with the version using the manual code transformation on the Titan supercomputer.

**Poster 137: Warwick Data Store: A HPC Library for Flexible Data Storage in Multi-Physics Applications**  
Richard O. Kirk (University of Warwick), Timothy R. Law (Atomic Weapons Establishment (AWE), UK), Satheesh Maheswaran (Atomic Weapons Establishment (AWE), UK), Stephen A. Jarvis (University of Warwick)

With the increasing complexity of memory architectures and multi-physics applications, developing data structures that are performant, portable, scalable, and support developer productivity, is difficult. In order to manage these complexities and allow rapid prototyping of different approaches we are building a lightweight and extensible C++ template library called the Warwick Data Store (WDS). WDS is designed to abstract details of the data structure away from the user, thus easing application development and optimisation. We show that WDS generates minimal performance overhead, via a variety of different scientific benchmarks and proxy-applications.

**Poster 76: HPChain: An MPI-Based Blockchain Framework for High Performance Computing Systems**  
Abdullah Al-Mamun (University of Nevada, Reno; Lawrence Berkeley National Laboratory), Tonglin Li
Data fidelity is of prominent importance for scientific experiments and simulations. The state-of-the-art mechanism to ensure data fidelity is through data provenance. However, the provenance data itself may as well exhibit unintentional human errors and malicious data manipulation. To enable a trustworthy and reliable data fidelity service, we advocate achieving the immutability and decentralization of scientific data provenance through blockchains. Specifically, we propose HPChain, a new blockchain framework specially designed for HPC systems. HPChain employs a new consensus protocol compatible with and optimized for HPC systems. Furthermore, HPChain was implemented with MPI and integrated with an off-chain distributed provenance service to tolerate the failures caused by faulty MPI ranks. The HPChain prototype system has been deployed to 500 cores at the University of Nevada's HPC center and demonstrated strong resilience and scalability while outperforming state-of-the-art blockchains by orders of magnitude.

Best Poster Finalist: no

**Poster 104: An Adaptive Checkpoint Model For Large-Scale HPC Systems**
Subhendu S. Behera (North Carolina State University), Lipeng Wan (Oak Ridge National Laboratory), Frank Mueller (North Carolina State University), Matthew Wolf (Oak Ridge National Laboratory), Scott Klasky (Oak Ridge National Laboratory)

Checkpoint/Restart is a widely used Fault Tolerance technique for application resilience. However, failures and the overhead of saving application state for future recovery upon failure reduces the application efficiency significantly. This work contributes a failure analysis and prediction model making decisions for checkpoint data placement, recovery, and techniques for reducing checkpoint frequency. We also demonstrate a reduction in application overhead by taking proactive measures guided by failure prediction.

Best Poster Finalist: no

**Poster 123: Cloud-Native SmartX Intelligence Cluster for AI-Inspired HPC/HPDA Workloads**
Junsu Han (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science), Jun-Sik Shin (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science), JinCheol Kwon (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science), JongWon Kim (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science), Dongfang Zhao (University of Nevada, Reno; University of California, Davis)
In this poster, we introduce Cloud-native SmartX Intelligence Cluster for flexibly supporting AI-inspired HPC (high performance computing) / HPDA (high performance data analytics) workloads. This work has been continuously refined from 2013 with a futuristic vision for operating 100 petascale data center. Then, we discuss issues and approaches that come with building a Cloud-native SmartX Intelligence Cluster.

Best Poster Finalist: no

**Poster 86: High-Performance Custom Computing with FPGA Cluster as an Off-Loading Engine**
Takaaki Miyajima (RIKEN Center for Computational Science (R-CCS)), Tomohiro Ueno (RIKEN Center for Computational Science (R-CCS)), Jens Huthmann (RIKEN Center for Computational Science (R-CCS)), Atsushi Koshiba (RIKEN Center for Computational Science (R-CCS)), Kentaro Sano (RIKEN Center for Computational Science (R-CCS)), Mitsuhashi Sato (RIKEN Center for Computational Science (R-CCS))

A heterogeneous system with Field Programmable Gate Arrays (FPGAs) is gathering attention in High-Performance Computing (HPC) area. We have been researching and developing a high-performance custom computing with FPGA cluster as an off-loading engine for an existing HPC system. In this research poster, we describe the motivation of our research and present research topics on a software bridge between the FPGA cluster and existing HPC servers, and dedicated inter-FPGA networks.

Best Poster Finalist: no

**Poster 69: Optimization for Quantum Computer Simulation**
Naoki Yoshioka (RIKEN Center for Computational Science (R-CCS)), Hajime Inaoka (RIKEN Center for Computational Science (R-CCS)), Nobuyasu Ito (RIKEN Center for Computational Science (R-CCS)), Fengping Jin (Forschungszentrum Juelich), Kristel Michielsen (Forschungszentrum Juelich), Hans De Raedt (University of Groningen)

Simulator of quantum circuits is developed for massively parallel classical computers, and it is tested on the K computer in RIKEN R-CCS up to 45 qubits. Two optimization techniques are proposed in order to improve performance of the simulator. The "page method" reduces unnecessary copies in each node. It is found that this method makes approximately 17% speed-up maximum. Initial permutation of qubits is also studied how it affects performance of the simulator. It is found that a
simple permutation in ascending order of the number of operations for each qubit is sufficient in the case of simulations of quantum adder circuits.

Best Poster Finalist: no

**Poster 52: Design and Specification of Large-Scale Simulations for GPUs Using FTX**

Anuva Kulkarni (Carnegie Mellon University), Daniele Spampinato (Carnegie Mellon University), Franz Franchetti (Carnegie Mellon University)

Large-scale scientific simulations can be ported to heterogeneous environments with GPUs using domain decomposition. However, Fast Fourier Transform (FFT) based simulations require all-to-all communication and large memory, which is beyond the capacity of on-chip GPU memory. To overcome this, domain decomposition solutions are combined with adaptive sampling or pruning around the domain to reduce storage. Expression of such operations is a challenge in existing FFT libraries like FFTW, and thus it is difficult to get a high performance implementation of such methods. We demonstrate algorithm specification for one such simulation (Hooke’s law) using FTX, an emerging API with a SPIRAL-based code generation back-end, and suggest future extensions useful for GPU-based scientific computing.

Best Poster Finalist: no

**Poster 136: CHAMELEON: Reactive Load Balancing and Migratable Tasks for Hybrid MPI+OpenMP Applications**

Jannis Klinkenberg (RWTH Aachen University), Philipp Samfaß (Technical University Munich), Michael Bader (Technical University Munich), Karl Fürlinger (Ludwig Maximilian University of Munich), Christian Terboven (RWTH Aachen University), Matthias S. Müller (RWTH Aachen University)

Many HPC applications are designed based on underlying performance and execution models. These models could successfully be employed in the past for balancing load within and between compute nodes. However, the increasing complexity of modern software and hardware makes performance predictability and load balancing much more difficult. Tackling these challenges in search for a generic solution, we present a novel library for fine-granular task-based reactive load balancing in distributed memory based on MPI and OpenMP. Our concept allows creating individual migratable tasks that can be executed on any MPI rank. Migration decisions are performed at run time based on online performance or load data. Two fundamental approaches to balance load and at the same time overlap computation and communication are compared. We evaluate our concept under enforced power caps and clock frequency changes using a synthetic benchmark and demonstrate robustness...
against work-induced imbalances for an AMR application.

Best Poster Finalist: no

**Poster 124: Porting Finite State Automata Traversal from GPU to FPGA: Exploring the Implementation Space**  
*Marziyeh Nourian (North Carolina State University), Mostafa Eghbali Zarch (North Carolina State University), Michela Becchi (North Carolina State University)*

While FPGAs are traditionally considered hard to program, recently there are efforts to allow using high-level programming models intended for multi-core CPUs and GPUs to program FPGAs. For example, both Intel and Xilinx are now providing OpenCL-to-FPGA toolchains. However, since GPU and FPGA devices offer different parallelism models, OpenCL code optimized for GPU can prove inefficient on FPGA, in terms of both performance and hardware resource utilization.

In this poster, we explore this problem on an emerging workload: finite state automata traversal. Specifically, we explore a set of structural code changes, custom, and best-practice optimizations to retarget an OpenCL NFA engine designed for GPU to FPGA. Our evaluation, which covers traversal throughput and resource utilization, shows that our optimizations lead, on a single execution pipeline, to speedups up to 4x over an already optimized baseline that uses one of the proposed code changes to fit the original code on FPGA.

Best Poster Finalist: no

**Poster 68: Linking a Next-Gen Remap Library into a Long-Lived Production Code**  
*Charles R. Ferenbaugh (Los Alamos National Laboratory), Brendan K. Krueger (Los Alamos National Laboratory)*

LANL’s long-lived production application xRage contains a remapper capability that maps mesh fields from its native AMR mesh to the GEM mesh format used by some third-party libraries. The current remapper was implemented in a short timeframe and is challenging to maintain. Meanwhile, our next-generation code project has developed a modern remapping library Portage, and the xRage team wanted to link in Portage as an alternate mapper option. But the two codes are very different from each other, and connecting the two required us to deal with a number of challenges. This poster describes the codes, the challenges we worked through, current status, and some initial performance statistics.
Best Poster Finalist: no

**Poster 131: Efficiency of Algorithmic Structures**

*Julian Miller (RWTH Aachen University), Lukas Trümper (RWTH Aachen University), Christian Terboven (RWTH Aachen University), Matthias S. Müller (RWTH Aachen University)*

The implementation of high-performance parallel software is challenging and raises issues not seen in serial programs before. It requires a strategy of parallel execution which preserves correctness but maximizes scalability. Efficiently deriving well-scaling solutions remains an unsolved problem especially with the quickly-evolving hardware landscape of high-performance computing (HPC).

This work proposes a framework for classifying the efficiency of parallel programs. It bases on a strict separation between the algorithmic structure of a program and its executed functions. By decomposing parallel programs into a hierarchical structure of parallel patterns, a high-level abstraction is provided which leads to equivalence classes over parallel programs. Each equivalence class possesses efficiency properties, mainly communication and synchronization, dataflow and architecture efficiency. This classification allows for wide application areas and a workflow for structural optimization of parallel algorithms is proposed.

Best Poster Finalist: no

**Poster 98: INSPECT Intranode Stencil Performance Evaluation Collection**

*Julian Hammer (University of Erlangen-Nuremberg), Julian Hornich (University of Erlangen-Nuremberg), Georg Hager (University of Erlangen-Nuremberg), Thomas Gruber (University of Erlangen-Nuremberg), Gerhard Wellein (University of Erlangen-Nuremberg)*

Modeling and presenting performance data---even for simple kernels such as stencils---is not trivial. We therefore present an overview on how to interpret and what to learn from an INSPECT report, as well as highlighting best practices for performance data reporting.

INSPECT is the "Intranode Stencil Performance Evaluation Collection", which compiles performance benchmarks and reports of various stencil and streaming kernels on a variety of architectures. The goal is to aid performance-aware developers with reference material and a methodology to analyze their own codes.

INSPECT set out to cover these topics and compile a summary of all necessary information to allow reproduction of the performance results, their interpretation and discussion.
Poster 97: Optimizing Multigrid Poisson Solver of Cartesian CFD Code CUBE
Kazuto Ando (RIKEN Center for Computational Science (R-CCS)), Rahul Bale (RIKEN), Keiji Onishi (RIKEN Center for Computational Science (R-CCS)), Kiyoshi Kumahata (RIKEN Center for Computational Science (R-CCS)), Kazuo Minami (RIKEN Center for Computational Science (R-CCS)), Makoto Tsubokura (Kobe University, RIKEN Center for Computational Science (R-CCS))

We demonstrate an optimization of multigrid Poisson solver of Cartesian CFD code “CUBE (Complex Unified Building cubE method)”. CUBE is a simulation framework for complex industrial flow problem, such as aerodynamics of vehicles, based on hierarchical Cartesian mesh. In incompressible CFD simulation, solving pressure Poisson equation is the most time-consuming part. In this study, we use a cavity flow simulation as a benchmark problem. With this problem, multigrid Poisson solver dominates 91% of execution time of the time-step loop. Specifically, we evaluate the performance of Gauss-Seidel loop as a computational kernel based on “Byte per Flop” approach. With optimization of the kernel, we achieved 9.8x speedup and peak floating point performance ratio increased from 0.4% to 4.0%. We also measured parallel performance up to 8,192 nodes (65,536 cores) on the K computer. With optimization of the parallel performance, we achieved 2.9x–3.9x sustainable speedup in the time-step loop.

Best Poster Finalist: no

Poster 85: Hybrid Computing Platform for Combinatorial Optimization with the Coherent Ising Machine
Junya Arai (Nippon Telegraph and Telephone Corporation), Yagi Satoshi (Nippon Telegraph and Telephone Corporation), Hiroyuki Uchiyama (Nippon Telegraph and Telephone Corporation), Toshimori Honjo (Nippon Telegraph and Telephone Corporation), Takahiro Inagaki (Nippon Telegraph and Telephone Corporation), Kensuke Inaba (Nippon Telegraph and Telephone Corporation), Takuya Ikuta (Nippon Telegraph and Telephone Corporation), Hiroki Takesue (Nippon Telegraph and Telephone Corporation), Keitaro Horikawa (Nippon Telegraph and Telephone Corporation)

Several institutes are operating cloud platforms that offer Web API access to Ising computers such as quantum annealing machines. Platform users can solve complex combinatorial optimization problems by using hybrid algorithms that utilize both users’ conventional digital computers and remote Ising computers. However, communication via the Internet takes an order of magnitude longer time than optimization on Ising computers. This overheads seriously degrade the performance of hybrid
algorithms since they involve frequent communication. In this poster, we first state issues in the design of Ising computing platforms, including communication overheads. Then, we answer the issues by introducing the computing platform for the coherent Ising machine (CIM), an Ising computer based on photonics technologies. Our platform offers efficient CIM-digital communication by allowing users to execute their program on digital computers co-located with the CIM. We have released the platform to our research collaborators in this autumn and started the evaluation.

Best Poster Finalist: no

**Poster 117: A New Polymorphic Computing Architecture Based on Fine-Grained Instruction Mobility**
David Henrich (Illinois Institute of Technology), Erdal Oruklu (Illinois Institute of Technology), Jafar Saniie (Illinois Institute of Technology)

This is a summary of the base concepts behind David Henrich’s May 2018 Ph.D. dissertation in Polymorphic Computing. Polymorphic Computing is the emerging field of changing the computer architecture around the software, rather than vice versa. The main contribution is a new polymorphic computing architecture. The key idea behind the architecture is to create an array of processors where a program’s instructions can be individually and arbitrarily assigned/mobilized to any processor, even during runtime. The key enablers of this architecture are a dataflow instruction set that is conducive to instruction migration, a microarchitectural block called an “operation cell” (“op-cell”), a processor built around the instruction set and the “op-cells”, and arrays of these processors.

Best Poster Finalist: no

**Poster 67: Genie: an MPEG-G Conformant Software to Compress Genomic Data.**
Brian E. Bliss (University of Illinois), Joshua M. Allen (University of Illinois), Saurabh Baheti (Mayo Clinic), Matthew A. Bockol (Mayo Clinic), Shubham Chandak (Stanford University), Jaime Delgado (Polytechnic University of Catalonia), Jan Fostier (Ghent University), Josep L. Gelpi (University of Barcelona), Steven N. Hart (Mayo Clinic), Michael T. Kalmbach (Mayo Clinic), Eric W. Klee (Mayo Clinic), Liudmila S. Mainzer (University of Illinois), Fabian Müntefering (Leibniz University), Daniel Naro (Barcelona Supercomputing Center), Idoia Ochoa-Alvarez (University of Illinois), Jörn Ostermann (Leibniz University), Tom Paridaens (Ghent University), Christian A. Ross (Mayo Clinic), Jan Voges (Leibniz University), Eric D. Wieben (Mayo Clinic), Mingyu Yang (University of Illinois), Tsachy Weissman (Stanford University), Mathieu Wiepert (Mayo Clinic)
Precision medicine has unprecedented potential for accurate diagnosis and effective treatment. It is supported by an explosion of genomic data, which continues to accumulate at accelerated pace. Yet storage and analysis of petascale genomic data is expensive, and that cost will ultimately be borne by the patients and citizens. The Moving Picture Experts Group (MPEG) has developed MPEG-G, a new open standard to compress, store, transmit and process genomic sequencing data that provides an evolved and superior alternative to currently used genomic file formats. Our poster will showcase software package GENIE, the first open source implementation of an encoder-decoder pair that is compliant with the MPEG-G specifications and delivers all its benefits: efficient compression, selective access, transport and analysis, guarantee of long-term support, and embedded mechanisms for annotation and encryption of compressed information. GENIE will create a step-change in medical genomics by reducing the cost of data storage and analysis.

Best Poster Finalist: no

**Poster 82: A View from the Facility Operations Side on the Water/Air Cooling System of the K Computer**

Jorji Nonaka (RIKEN Center for Computational Science (R-CCS)), Keiji Yamamoto (RIKEN Center for Computational Science (R-CCS)), Akiyoshi Kuroda (RIKEN Center for Computational Science (R-CCS)), Toshiyuki Tsukamoto (RIKEN Center for Computational Science (R-CCS)), Kazuki Koiso (Kobe University, RIKEN Center for Computational Science (R-CCS)), Naohisa Sakamoto (Kobe University, RIKEN Center for Computational Science (R-CCS))

The Operations and Computer Technologies Division at the RIKEN R-CCS is responsible for the operations of the entire K computer facility, which includes the auxiliary subsystems such as the power supply and water/air cooling systems. It is worth noting that part of these subsystems will be reused in the next supercomputer (Fugaku), thus a better understanding of the operational behavior as well as the potential impacts especially on the hardware failure and energy consumption would be greatly beneficial. In this poster, we will present some preliminary impressions of the impact of the water/air cooling system on the K computer system, focusing on the potential benefits of the use of low water/air temperature respectively for the CPU and DRAM memory modules produced by the cooling system. We expect that the obtained knowledge will be helpful for the decision support and/or operation planning of the next supercomputer.

Best Poster Finalist: no

**Poster 135: High-Performance Deep Learning via a Single Building Block**

Evangelos Georganas (Intel Corporation), Kunal Banerjee (Intel Corporation), Dhiraj Kalamkar (Intel
Deep learning (DL) is one of the most prominent branches of machine learning. Due to the immense computational cost of DL workloads, industry and academia have developed DL libraries with highly-specialized kernels for each workload/architecture, leading to numerous, complex code-bases that strive for performance, yet they are hard to maintain and do not generalize. In this work, we introduce the batch-reduce-GEMM kernel and show how the most popular DL algorithms can be formulated with this kernel as basic building-block. Consequently, the DL library-development degenerates to mere (potentially automatic) tuning of loops around this sole optimized kernel. By exploiting our kernel we implement Recurrent Neural Networks, Convolution Neural Networks and Multilayer Perceptron training and inference primitives in just 3K lines of high-level-code. Our primitives outperform vendor-optimized libraries on multi-node CPU-Clusters. We also provide CNN kernels targeting GPUs. Finally, we demonstrate that batch-reduce-GEMM kernel within a tensor compiler yields high-performance CNN primitives.

Best Poster Finalist: no

**Poster 56: Reinforcement Learning for Quantum Approximate Optimization**

Sami Khairy (Illinois Institute of Technology), Ruslan Shaydulin (Clemson University), Lukasz Cincio (Los Alamos National Laboratory), Yuri Alexeev (Argonne National Laboratory), Prasanna Balaprakash (Argonne National Laboratory)

The Quantum Approximate Optimization Algorithm (QAOA) is one of the leading candidates for demonstrating quantum advantage. The quality of the solution obtained by QAOA depends on the performance of the classical optimization routine used to optimize the variational parameters. In this work, we propose a Reinforcement Learning (RL) based approach to drastically reduce the number of evaluations needed to find high-quality variational parameters. We train an RL agent on small 8-qubit Max-Cut problem instances on an Intel Xeon Phi supercomputer Bebop, and use (transfer) the learned optimization policy to quickly find high-quality solutions for other larger problem instances coming from different distributions and graph classes. The preliminary results show that our RL based approach is able to improve the quality of the obtained solution by up to 10% within a fixed budget of function evaluations and demonstrate learned optimization policy transferability between different graph classes and sizes.

Best Poster Finalist: no

John Shalf (Lawrence Berkeley National Laboratory), Dilip Vasudevan (Lawrence Berkeley National Laboratory), David Donofrio (Lawrence Berkeley National Laboratory), Anastasia Butko (Lawrence Berkeley National Laboratory), Andrew Chien (University of Chicago), Yuanwei Fang (University of Chicago), Arjun Rawal (University of Chicago), Chen Zou (University of Chicago), Raymond Bair (Argonne National Laboratory), Kristopher Keipert (Argonne National Laboratory), Arun Rodriguez (Sandia National Laboratories), Maya Gokhale (Lawrence Livermore National Laboratory), Scott Lloyd (Lawrence Livermore National Laboratory), Xiaochen Guo (Lehigh University), Yuan Zeng (Lehigh University)

Accelerating technology disruptions and architectural change create growing opportunities and urgency to reduce the latency in for new architectural innovations to be deployed in extreme scale systems. We are exploring new architectural features that improve memory system performance including word-wise scratchpad memory, a flexible Recode engine, hardware message queues, and the data rearrangement engine (DRE). Performance results are promising yielding as much as 20x benefit. Project 38 is a cross-agency effort undertaken by the US Department of Energy (DOE) and Department of Defense (DoD).

Best Poster Finalist: no

Poster 128: Identifying Time Series Similarity in Large-Scale Earth System Datasets

Payton Linton (Youngstown State University), William Melodia (Youngstown State University), Alina Lazar (Youngstown State University), Deborah Agarwal (Lawrence Berkeley National Laboratory), Ludovico Bianchi (Lawrence Berkeley National Laboratory), Devarshi Ghoshal (Lawrence Berkeley National Laboratory), Kesheng Wu (Lawrence Berkeley National Laboratory), Gilberto Pastorello (Lawrence Berkeley National Laboratory), Lavanya Ramakrishnan (Lawrence Berkeley National Laboratory)

Scientific data volumes are growing every day and instrument configurations, quality control and software updates result in changes to the data. This study focuses on developing algorithms that detect changes in time series datasets in the context of the Deduce project. We propose a combination of methods that include dimensionality reduction and clustering to evaluate similarity measuring algorithms. This methodology can be used to discover existing patterns and correlations within a dataset. The current results indicate that the Euclidean Distance metric provides the best results in terms of internal cluster validity measures for multi-variable analyses of large-scale earth system datasets. The poster will include details on our methodology, results, and future work.
**Poster 151: Three-Dimensional Characterization on Edge AI Processors with Object Detection Workloads**

Yujie Hui (Ohio State University), Jeffrey Lien (NovuMind Inc), Xiaoyi Lu (Ohio State University)

The Deep Learning inference applications are moving to the edge side, as edge-side AI platforms are cheap and energy-efficient. Different edge AI processors are diversified, since these processors are designed with different approaches. However, it is hard for customers to select an edge AI processor without an overall evaluation of these processors. We propose a three-dimensional characterization (i.e., accuracy, latency, and energy efficiency) approach on three different kinds of edge AI processors (i.e., Edge TPU, NVIDIA Xavier, and NovuTensor). We deploy YOLOv2 and Tiny-YOLO, which are two YOLO-based object detection systems, on these edge AI platforms with Microsoft COCO dataset. I will present our work starting from the problem statement. And then I'll introduce our experiments setup and hardware configuration. Lastly, I'll conclude our experimental results and current work status, as well as the future work.

Best Poster Finalist: no

**Poster 148: Unsupervised Clustering of Golden Eagle Telemetry Data**


We use a recurrent autoencoder neural network to encode sequential California golden eagle telemetry data. The encoding is followed by an unsupervised clustering technique, Deep Embedded Clustering (DEC), to iteratively cluster the data into a chosen number of behavior classes. We apply the method to simulated movement data sets and telemetry data for a Golden Eagle. The DEC achieves better unsupervised clustering accuracy scores for the simulated data sets as compared to the baseline K-means clustering result.

Best Poster Finalist: no

**Poster 66: Hybrid CPU/GPU FE2 Multi-Scale Implementation Coupling Alya and Micropip**

Guido Giuntoli (Barcelona Supercomputing Center), Judicaël Grasset (Science and Technology Facilities Council (STFC)), Alejandro Figueroa (George Mason University), Charles Moulinec (Science and Technology Facilities Council (STFC)), Mariano Vázquez (Barcelona Supercomputing Center),
This poster exposes the results of a new implementation of the FE2 multi-scale algorithm that is achieved by coupling the multi-physics and massively parallel code Alya with the GPU-based code micropp. The coupled code is mainly designed to solve large scale and realistic composite material problems for the aircraft industry. Alya is responsible of solving the macro-scale equations and micropp for solving the representation of fibres at the microscopic level. The poster shows computational performance results that demonstrate that the technique is scalable for real size industrial problems and also how the execution time is dramatically reduced using GPU-based clusters.

Best Poster Finalist: no


Fahim Tahmid Chowdhury (Florida State University, Lawrence Livermore National Laboratory), Francesco Di Natale (Lawrence Livermore National Laboratory), Adam Moody (Lawrence Livermore National Laboratory), Elsa Gonsiorowski (Lawrence Livermore National Laboratory), Kathryn Mohror (Lawrence Livermore National Laboratory), Weikuan Yu (Florida State University)

Leadership high performance computing (HPC) systems have the capability to execute workflows of scientific, research or industry applications. Complex HPC workflows can have significant data transfer and I/O requirements. Heterogeneous storage systems in supercomputers equipped with bleeding-edge non-volatile persistent storage devices can be leveraged to handle these data transfer and I/O requirements efficiently.

In this poster, we describe our efforts to extract the I/O characteristics of various HPC workflows and develop strategies to improve I/O performance by leveraging heterogeneous storage systems. We have implemented an emulator to mimic different types of I/O requirements posed by HPC application workflows. We have analyzed the workflow of Cancer Moonshot Pilot 2 (CMP2) project to determine possible I/O inefficiencies. To date, we have performed a systematic characterization and evaluation on the workloads generated by the workflow emulator and a small scale adaptation of the CMP2 workflow.

Best Poster Finalist: no
Poster 116: Advancements in Ultrasound Simulations Enabled by High-Bandwidth GPU Interconnects
Filip Vaverka (Brno University of Technology), Bradley E. Treeby (University College London), Jiri Jaros (Brno University of Technology)

Realistic ultrasound simulations are becoming integral part of many novel medical procedures such as photoacoustic screening and non-invasive treatment planning. The common denominator of all these applications is the need for cheap and relatively large-scale ultrasound simulations with sufficient accuracy. Typical medical applications require full-wave simulations which take frequency-dependent absorption and non-linearity into account.

This poster investigates the benefits of high-bandwidth low-latency interconnects to k-Wave acoustic toolbox in dense multi-GPU environment. The k-Wave multi-GPU code is based on a variant of the local Fourier basis domain decomposition. The poster compares the behavior of the code on a typical PCI-E 3.0 machine with 8 Nvidia Tesla P40 GPUs and a Nvidia DGX-2 server. The performance constraints of PCI-E platforms built around multiple socket servers on multi-GPU applications are deeply explored. Finally, it is shown the k-Wave toolbox can efficiently utilize NVlink 2.0 and achieve over 4x speedup compared to PCI-E systems.

Best Poster Finalist: no

Poster 65: Comparing Granular Dynamics vs. Fluid Dynamics via Large DOF-Count Parallel Simulation on the GPU
Milad Rakhsha (University of Wisconsin), Conlain Kelly (Georgia Institute of Technology), Nicholas Olsen (University of Wisconsin), Lijing Yang (University of Wisconsin), Radu Serban (University of Wisconsin), Dan Negrut (University of Wisconsin)

In understanding granular dynamics, the commonly-used discrete modeling approach that tracks the motion of all particles is computationally demanding, especially with large system size. In such cases, one can contemplate switching to continuum models that are computationally less expensive. In order to assess when such a discrete to continuum switch is justified, we compare granular and fluid dynamics that scales to handle more than 1 billion degrees of freedom (DOFs); i.e., two orders of magnitude higher than the state-of-the-art. On the granular side, we solve the Newton-Euler equations of motion; on the fluid side, we solve the Navier-Stokes equations. Both solvers leverage parallel computing on the GPU, and are publicly available on GitHub as part of an open-source code called Chrono. We report similarities and differences between the dynamics of the discrete, fully-resolved system and the continuum model via numerical experiments including both static and highly transient scenarios.
Poster 78: Understanding HPC Application I/O Behavior Using System Level Statistics
Arnab K. Paul (Virginia Tech), Olaf Faaland (Lawrence Livermore National Laboratory), Adam Moody (Lawrence Livermore National Laboratory), Elsa Gonsiorowski (Lawrence Livermore National Laboratory), Kathryn Mohror (Lawrence Livermore National Laboratory), Ali R. Butt (Virginia Tech)

The processor performance of high performance computing (HPC) systems is increasing at a much higher rate than storage performance. Storage and file system designers therefore require a deep understanding of how HPC application I/O behavior affects current storage system installations in order to improve storage performance. In this work, we contribute to this understanding using application-agnostic file system statistics gathered on compute nodes as well as metadata and object storage file system servers. We analyze file system statistics of more than 4 million jobs over a period of three years on two systems at Lawrence Livermore National Laboratory that include a 15 PiB Lustre file system for storage. Some key observations in our study show that more than 65% HPC users perform significant I/O which are mostly writes; and less than 22% of HPC users who submit write-intensive jobs perform efficient writes to the file system.

Best Poster Finalist: no

Poster 109: A Runtime Approach for Dynamic Load Balancing of OpenMP Parallel Loops in LLVM
Jonas H. Müller Korndörfer (University of Basel, Switzerland), Florina M. Ciorba (University of Basel, Switzerland), Akan Yilmaz (University of Basel, Switzerland), Christian Iwainsky (Technical University Darmstadt), Johannes Doerfert (Argonne National Laboratory), Hal Finkel (Argonne National Laboratory), Vivek Kale (Brookhaven National Laboratory), Michael Klemm (Intel Corporation)

Load imbalance is the major source of performance degradation in computationally-intensive applications that frequently consist of parallel loops. Efficient scheduling can improve the performance of such programs. OpenMP is the de-facto standard for parallel programming on shared-memory systems. The current OpenMP specification provides only three choices for loop scheduling which are insufficient in scenarios with irregular loops, system-induced interference, or both. Therefore, this work augments the LLVM OpenMP runtime library implementation with eleven ready to use scheduling techniques. We tested existing and added scheduling strategies on several applications from NAS, SPEC OMP 2012, and CORAL2 benchmark suites. Experiments show that implemented scheduling techniques outperform others in certain application and system configurations. We
measured performance gains of up to 6% compared to the fastest standard scheduling technique. This work aims to be a convincing step toward beyond-standard scheduling options in OpenMP for the benefit of evolving applications executing on multicore architectures.

Best Poster Finalist: no

**Poster 143: Quantum Natural Language Processing**

Lee James O'Riordan (Irish Centre for High-End Computing (ICHEC), National University of Ireland Galway (NUIG)), Myles Doyle (Irish Centre for High-End Computing (ICHEC), National University of Ireland Galway (NUIG)), Venkatesh Kannan (Irish Centre for High-End Computing (ICHEC), National University of Ireland Galway (NUIG)), Fabio Baruffa (Intel Corporation)

Natural language processing (NLP) algorithms that operate over strings of words are limited since they analyse meanings of the component words in a corpus without information about grammatical rules of the language. Consequently, they often produce unsatisfactory results with increase in problem complexity.

The "distributed compositional semantics" (DisCo) model incorporates grammatical structure of sentences into the algorithms, and offers significant improvements to the quality of results. However, their main challenge is the need for large classical computational resources. The DisCo model presents two quantum algorithms which lower storage and compute requirements compared to a classic HPC implementation.

In this project, we implement the two DisCo model quantum algorithms on the Intel Quantum Simulator deployed on the Irish national supercomputer. We target corpuses with ~1000 most-common words using up to 36 qubits simulation. The solution will be able to compute the meanings of two sentences and decide if their meanings match.

Best Poster Finalist: no

**Poster 152: Deep Domain Adaptation for Runtime Prediction in Dynamic Workload Scheduler**

Hoang H. Nguyen (National Center for Atmospheric Research (NCAR); University of Illinois, Chicago), Ben Matthews (National Center for Atmospheric Research (NCAR)), Irfan Elahi (National Center for Atmospheric Research (NCAR))

In HPC systems, users’ requested runtime for submitted jobs plays a crucial role in efficiency. While underestimation of job runtime could terminate jobs before completion, overestimation could result
in long queuing of other jobs in HPC systems. In reality, runtime prediction in HPC is challenging due to the complexity and dynamics of running workloads. Most of the current predictive runtime models are trained on static workloads. This poses a risk of over-fitting the predictions with bias from the learned workload distribution. In this work, we propose an adaptation of Correlation Alignment method in our deep neural network architecture (DCORAL) to alleviate the domain shift between workloads for better runtime predictions. Experiments on both standard benchmark workloads and NCAR real-time production workloads reveal that our proposed method results in a more stable training model across different workloads with low accuracy variance as compared to the other state-of-the-art methods.

Best Poster Finalist: no

Poster 75: libCEED - Lightweight High-Order Finite Elements Library with Performance Portability and Extensibility
Jeremy Thompson (University of Colorado), Valeria Barra (University of Colorado), Yohann Dudouit (Lawrence Livermore National Laboratory), Oana Marin (Argonne National Laboratory), Jed Brown (University of Colorado)

High-order numerical methods are widely used in PDE solvers, but software packages that have provided high-performance implementations have often been special-purpose and intrusive. libCEED is a new library that offers a purely algebraic interface for matrix-free operator representation and supports run-time selection of implementations tuned for a variety of computational device types, including CPUs and GPUs. We introduce the libCEED API and demonstrate how it can be used in standalone code or integrated with other packages (e.g., PETSc, MFEM, Nek5000) to solve examples of problems that often arise in the scientific computing community, ranging from fast solvers via geometric multigrid methods to Computational Fluid Dynamics (CFD) applications.

Progress on the Exascale Transition of the VSim Multiphysics PIC code
Benjamin M. Cowan (Tech-X Corporation), Sergey N. Averkin (Tech-X Corporation), John R. Cary (Tech-X Corporation), Jarrod Leddy (Tech-X Corporation), Scott W. Sides (Tech-X Corporation), Ilya A. Zilberter (Tech-X Corporation)

The highly performant, flexible plasma simulation code VSim was designed nearly 20 years ago (originally as Vorpal), with its first applications roughly four years later. Using object oriented
methods, VSim was designed to allow runtime selection from multiple field solvers, particle dynamics, and reactions. It has been successful in modeling for many areas of physics, including fusion plasmas, particle accelerators, microwave devices, and RF and dielectric structures. Now it is critical to move to exascale systems, with their compute accelerator architectures, massive threading, and advanced instruction sets. Here we discuss how we are moving this complex, multiphysics computational application to the new computing paradigm, and how it is done in a way that kept the application producing physics during the move. We present performance results showing significant speedups in all parts of the PIC loop, including field updates, particle pushes, and reactions.

Best Poster Finalist: no

**Poster 58: Lock-Free van Emde Boas Array**
Ziyuan Guo (University of Tokyo)

Lock-based data structures have some potential issues such as deadlock, livelock, and priority inversion, and the progress can be delayed indefinitely if the thread that is holding locks cannot acquire a timeslice from the scheduler. Lock-free data structures, which guarantees the progress of some method call, can be used to avoid these problems. This poster introduces the first lock-free concurrent van Emde Boas Array which is a variant of van Emde Boas Tree. It is linearizable, and the benchmark shows significant performance improvement comparing to other lock-free search trees when the date set is large and dense enough.

Best Poster Finalist: no

**Poster 63: Adaptive Execution Planning in Biomedical Workflow Management Systems**
Marta Jaros (Brno University of Technology), Bradley E. Treeby (University College London), Jiri Jaros (Brno University of Technology)

Biomedical simulations require very powerful computers. Their execution is described by a workflow consisting of a number of different cooperating tasks. The manual execution of individual tasks may be tedious for expert users, but prohibiting for most inexperienced clinicians. k-Dispatch offers a 'run and forget' approach where the users are completely screened out from the complexity of HPC systems. k-Dispatch provides task scheduling, execution, monitoring, and fault tolerance. Since the task execution configuration strongly affects the final tasks mapping on the computational resources, the execution planning is of the highest priority. Unlike other tools, k-Dispatch considers a variable amount of computational resources per individual tasks. Since the scaling of the individual
HPC codes is never perfect, k-Dispatch may find such a good mapping even an experienced user would miss. The proposed adaptive execution planning is based on collected performance data and the current cluster utilization monitoring.

Best Poster Finalist: no

**Poster 142: Training Deep Neural Networks Directly on Hundred-Million-Pixel Histopathology Images on a Large-Scale GPU Cluster**
Chi-Chung Chen (AetherAI, Taiwan), Wen-Yu Chuang (Chang-Gung Memorial Hospital, Taiwan), Wei-Hsiang Yu (AetherAI, Taiwan), Hsi-Ching Lin (National Center for High-Performance Computing (NCHC), Taiwan), Shuen-Tai Wang (National Center for High-Performance Computing (NCHC), Taiwan), Fang-An Kuo (National Center for High-Performance Computing (NCHC), Taiwan), Chao-Chun Chuang (National Center for High-Performance Computing (NCHC), Taiwan), Chao-Yuan Yeh (AetherAI, Taiwan)

Deep learning for digital pathology is challenging because the resolution of whole-slide-images (WSI) is extremely high, often in billions. The most common approach is patch-based method, where WSIs are divided into small patches to train convolutional neural networks (CNN). This approach has significant drawbacks. To have ground truth for individual patches, detailed annotations by pathologists are required. This laborious process has become the major impediment to the development of digital pathology AI. End-to-end WSI training, however, faces the difficulties of fitting the task into limited GPU memory. In this work, we improved the efficiency of using system memory for GPU compute by 411% through memory optimization and deployed the training pipeline on 8 nodes, totally 32 GPUs distributed system, achieving 147.28x speedup. We demonstrated that CNN is capable of learning features without detailed annotations. The trained CNN can correctly classify cancerous specimen, with performance level closely matching the patch-based methods.

Best Poster Finalist: no

**Poster 95: A Heterogeneous HEVC Video Encoder Based on OpenPOWER Acceleration Platform**
Chenhao Gu (Fudan University, Shanghai, State Key Laboratory of ASIC and System; IBM Corporation), Yang Chen (Fudan University, Shanghai, State Key Laboratory of ASIC and System; IBM Corporation), Yanheng Lu (IBM Corporation), Pengfei Gou (IBM Corporation), Yong Lu (IBM Corporation), Yang Dai (IBM Corporation), Yue Xu (IBM Corporation), Yang Liu (IBM Corporation), Yibo Fan (Fudan University, Shanghai, State Key Laboratory of ASIC and System)
This poster describes a heterogeneous HEVC video encoder system based on the OpenPOWER platform. Our design leverages the Coherent Accelerator Processor Interface (CAPI) on the OpenPOWER, which provides cache-coherent access for FPGA. This technology highly improves CPU-FPGA data communication bandwidth and programming efficiency. X265 is optimized on the OpenPOWER platform to improve its performance with both architecture specific methods and hardware-acceleration methods. For hardware acceleration, frame-level acceleration and functional-unit-level acceleration are introduced and evaluated in this work.

Best Poster Finalist: no

Poster 102: Fast Training of an AI Radiologist: Leveraging Data Pipelining to Efficiently Utilize GPUs
Rakshith Vasudev (Dell EMC), John A. Lockman III (Dell EMC), Lucas A. Wilson (Dell EMC), Srinivas Varadharajan (Dell EMC), Frank Han (Dell EMC), Rengan Xu (Dell EMC), Quy Ta (Dell EMC)

In a distributed deep learning training setting, using accelerators such as GPUs can be challenging to develop a high throughput model. If the accelerators are not utilized effectively, this could mean more time to solution, and thus the model's throughput is low. To use accelerators effectively across multiple nodes, we need to utilize an effective data pipelining mechanism that handles scaling gracefully so GPUs can be exploited of their parallelism. We study the effect of using the correct pipelining mechanism that is followed by tensorflow official models vs a naive pipelining mechanism that doesn't scale well, on two image classification models. Both the models using the optimized data pipeline demonstrate effective linear scaling when GPUs are added. We also show that converting to TF Records is not always necessary.

Best Poster Finalist: no

Poster 94: Multi-GPU Optimization of a Non-Hydrostatic Numerical Ocean Model with Multigrid Preconditioned Conjugate Gradient Method
Takateru Yamagishi (Research Organization for Information Science and Technology, Japan), Yoshimasa Matsumura (University of Tokyo, Atmosphere and Ocean Research Institute), Hiroyasu Hasumi (University of Tokyo, Atmosphere and Ocean Research Institute)

The conjugate gradient method with multigrid preconditioners (MGCG) is used in scientific applications because of its high performance and scalability with many computational nodes. GPUs are thought to be good candidates for accelerating such applications with many meshes where an
MGCG solver could show high performance. No previous studies have evaluated and discussed the numerical character of an MGCG solver on GPUs. Consequently, we have implemented and optimized our "kinaco" numerical ocean model with an MGCG solver on GPUs. We evaluated its performance and discussed inter-GPU communications on a coarse grid on which GPUs could be intrinsically problematic. We achieved 3.9 times speedup compared to CPUs and learned how inter-GPU communications depended on the number of GPUs and the aggregation level of information in a multigrid method.

Best Poster Finalist: no

**Poster 108: Power Prediction for High-Performance Computing**

Shigeto Suzuki (Fujitsu Laboratories Ltd), Michiko Hiraoka (Fujitsu Ltd), Takashi Shiraishi (Fujitsu Laboratories Ltd), Enxhi Kreshpa (Fujitsu Laboratories Ltd), Takuji Yamamoto (Fujitsu Laboratories Ltd), Hiroyuki Fukuda (Fujitsu Laboratories Ltd), Shuji Matsui (Fujitsu Ltd), Masahide Fujisaki (Fujitsu Ltd), Atsuya Uno (RIKEN Center for Computational Science (R-CCS))

Exascale computers consume large amounts of power both for computing and cooling-units. As power of the computer varies dynamically corresponding to the load change, cooling-units are desirable to follow it for effective energy management. Because of time lags in cooling-unit operations, advance control is inevitable and an accurate prediction is a key for it. Conventional prediction methods make use of the similarity between job information while in queue. The prediction fails if there is no previously similar job. We developed two models to correct the prediction after queued jobs start running. By taking power histories into account, power-correlated topic model reselects more suitable candidate and recurrent-neural-network model considering variable network sizes predicts power variation from shape features of it. We integrated these into a single algorithm and demonstrated high-precision prediction with an average relative error of 5.7% in K computer as compared to the 18.0% obtained using the conventional method.

Best Poster Finalist: no

**Poster 80: Sharing and Replicability of Notebook-Based Research on Open Testbeds**

Maxine V. King (University of Chicago), Jason Anderson (University of Chicago), Kate Keahey (Argonne National Laboratory, University of Chicago)

We seek to facilitate replicability by creating a way to share experiments easily in and out of notebook-based, open testbed environments and a sharing platform for such experiments in order to allow researchers to combine shareability, consistency of code environment, and well-
documented process.

Best Poster Finalist: no

**Poster 121: HFlush: Realtime Flushing for Modern Storage Environments**
Jaime Cernuda (Illinois Institute of Technology), Hugo Trivino (Illinois Institute of Technology), Hariharan Devarajan (Illinois Institute of Technology), Anthony Kougkas (Illinois Institute of Technology), Xian-He Sun (Illinois Institute of Technology)

Due to the unparalleled magnitude of data movement in extreme scale computing, I/O has become a central challenge. Modern storage environments have proposed the use of multiple layers between applications and the PFS. Nonetheless, the difference in capacities and speeds between storage layers makes it extremely challenging to evict data from upper layers to lower layers efficiently. However, current solutions are executed in batches, compromising latency; are also push-based implementations, compromising resource utilization. Hence, we propose HFlush, a continuous data eviction mechanism built on a streaming architecture that is pull-based and in which each component is decoupled and executed in parallel. Initial results have shown RFlush to obtain a 7X latency reduction and a 2X bandwidth improvement over a baseline batch-based system. Therefore, RFlush is a promising solution to the growing challenges of extreme scale data generation and eviction shortcomings when archiving data across multiple tiers of storage.

Best Poster Finalist: no

**Poster 88: HPC Container Runtime Performance Overhead: At First Order, There Is None**
Alfred Torrez (Los Alamos National Laboratory), Reid Priedhorsky (Los Alamos National Laboratory), Timothy Randles (Los Alamos National Laboratory)

Linux containers are an increasingly popular method used by HPC centers to meet increasing demand for greater software flexibility. A common concern is that containers may introduce application performance overhead. Prior work has not tested a broad set of HPC container technologies on a broad set of benchmarks. This poster addresses the gap by comparing performance of the three HPC container implementations (Charliecloud, Shifter, and Singularity) and bare metal on multiple dimensions using industry-standard benchmarks.

We found no meaningful performance differences between the four environments with the possible exception of modest variation in memory usage, which is broadly consistent with prior results. This result suggests that HPC users should feel free to containerize their applications without concern
about performance degradation, regardless of the container technology used. It is an encouraging development on the path towards greater adoption of user-defined software stacks to increase the flexibility of HPC.

Best Poster Finalist: no

**Poster 64: 416-PFLOPS Fast Scalable Implicit Solver on Low-Ordered Unstructured Finite Elements Accelerated by 1.10-ExaFLOPS Kernel with Reformulated AI-Like Algorithm: For Equation-Based Earthquake Modeling**

Tsuyoshi Ichimura (University of Tokyo, RIKEN), Kohei Fujita (University of Tokyo, RIKEN), Takuma Yamaguchi (University of Tokyo), Akira Naruse (Nvidia Corporation), Jack C. Wells (Oak Ridge National Laboratory), Christopher J. Zimmer (Oak Ridge National Laboratory), Tjerk P. Straatsma (Oak Ridge National Laboratory), Takane Hori (Japan Agency for Marine-Earth Science and Technology), Simone Puel (University of Texas), Thorsten W. Becker (University of Texas), Munenori Hori (Japan Agency for Marine-Earth Science and Technology), Naonori Ueda (RIKEN)

We propose herein an approach for reformulating an equation-based modeling algorithm to an algorithm similar to that of training artificial intelligence (AI) and accelerate this algorithm using high-performance accelerators to reduce the huge computational costs encountered for physics equation-based modeling in earthquake disaster mitigation. A fast scalable equation-based implicit solver on unstructured finite elements is accelerated with a Tensor Core-enabled matrix-vector product kernel. The developed kernel attains 1.10 ExaFLOPS, leading to 416 PFLOPS for the whole solver on full Summit. This corresponds to a 75-fold speedup from a previous state-of-the-art solver running on full Piz Daint. This result could lead to breakthroughs in earthquake disaster mitigation. Our new idea in the HPC algorithm design of combining equation-based modeling with AI is expected to have broad impacts in other earth science and industrial problems.

Best Poster Finalist: no

**Poster 91: FreeCompilerCamp: Online Training for Extending Compilers**

Alok Mishra (Stony Brook University, Lawrence Livermore National Laboratory), Anjia Wang (University of North Carolina, Charlotte; Lawrence Livermore National Laboratory), Chunhua Liao (Lawrence Livermore National Laboratory), Yonghong Yan (University of North Carolina, Charlotte), Barbara Chapman (Stony Brook University)

In this presentation, we introduce an ongoing effort of an online training platform aimed to automate the training of developers to quickly extend compilers - FreeCompilerCamp.org. Our free
and open platform allows anyone who is interested in developing compilers to learn the necessary skills. A live training website, built on top of Play-With-Docker, is set up so that anyone with internet access and a web browser will be able to take this training. The entire training system is open-source and developers with relevant skills can contribute new tutorials and deploy it on a private server, workstation or even laptop. We have created some initial tutorials on how to extend the Clang/LLVM or ROSE compilers to support new OpenMP features. Using a web interface consisting of two side-by-side panels, users can follow the tutorials on one side and immediately practice what they learned in a terminal sandbox embedded on the other.

Best Poster Finalist: yes

Poster 92: Nanoporous Flow Simulations on the Summit Supercomputer
Yidong Xia (Idaho National Laboratory), Lixiang Luo (IBM - TJ Watson Research Center), Ansel Blumers (Brown University), Joshua Kane (Idaho National Laboratory), Jan Goral (University of Utah), Yu-Hang Tang (Lawrence Berkeley National Laboratory), Zhen Li (Clemson University, Brown University), Hai Huang (Idaho National Laboratory), Milind Deo (University of Utah)

Fluid flow behaviors in nanoporous materials are distinct from those following the continuum physics. Numerical simulations can be a complement to laboratory experiments. This work presents a dissipative particle dynamics (DPD) package for GPU-accelerated mesoscale flow simulations in nanoporous materials. In an ideal benchmark that minimizes load imbalance, the package delivered nearly perfect strong- and weak-scaling (with up to 4 billion DPD particles) on up to 1,536 V100 GPUs on Oak Ridge National Laboratory’s Summit supercomputer. More remarkably, in a benchmark to measure its usefulness with realistic nanopores in SBA-15 silica, the package exhibited more than 20x speedup over its LAMMPS-based CPU counterpart with the same number of nodes (e.g., 384 V100 GPUs vs. 2,688 POWER9 cores). It is worth highlighting that the NVLink2 Host-to-Device interconnects kept the cost of CPU-GPU memory copy as low as only 10% of GPU activity time per rank: 4 times less than their PCIe counterparts.

Best Poster Finalist: yes

Poster 96: TSQR on TensorCores
Hiroyuki Ootomo (Tokyo Institute of Technology), Rio Yokota (Tokyo Institute of Technology)

Tall-Skinny QR (TSQR) is an efficient algorithm for calculating the QR decomposition of m \times n matrices where m << n, which is done by recursively performing QR decomposition on subdivided blocks of the tall and skinny matrix. Such operations are useful for low-rank approximation methods,
which are replacing more and more dense linear algebra in both scientific computing and machine learning fields. The present work focuses on the implementation of this important algorithm on Tensor Cores, which are available on the latest NVIDIA GPUs. We evaluate the speed, accuracy, and stability of TSQR on TensorCores.

Best Poster Finalist: yes

**Poster 110: Hierarchical Data Prefetching in Multi-Tiered Storage Environments**  
Hariharan Devarajan (Illinois Institute of Technology), Anthony Kougkas (Illinois Institute of Technology), Xian-He Sun (Illinois Institute of Technology)

In the era of data-intensive computing, accessing data with a high-throughput and low-latency is very imperative. Data prefetching is used for hiding read latency by requesting data before it is needed to move it from a high-latency medium to a low-latency one. However, existing solutions do not consider the multi-tiered storage and also suffer from under-utilization of prefetching resources and unnecessary evictions. Additionally, existing approaches implement a client-pull model where understanding the application's I/O behavior drives prefetching decisions. Moving toward exascale, where machines run multiple applications concurrently by accessing files in a workflow, a more data-centric approach resolves challenges such as cache pollution and redundancy. In this study, we present HFetch, a truly hierarchical data prefetcher that adopts a server-push approach to data prefetching. We demonstrate the benefits of such an approach. Results show 10-35% performance gains over existing prefetchers and over 50% when compared to systems with no prefetching.

Best Poster Finalist: yes

**Poster 54: Massively Parallel Eigensolvers Based on Unconstrained Energy Functionals Methods**  
Osni Marques (Lawrence Berkeley National Laboratory), Mauro Del Ben (Lawrence Berkeley National Laboratory), Andrew Canning (Lawrence Berkeley National Laboratory)

This poster focuses on a preconditioned conjugate gradient based iterative eigensolver using an unconstrained energy functional minimization scheme. This scheme avoids an explicit reorthogonalization of the trial eigenvectors and becomes an attractive alternative for the solution of very large problems. The unconstrained formulation is implemented in the first-principles materials and chemistry CP2K code, which performs electronic structure calculations based on a density functional theory approximation to the solution of the many-body Schrödinger equation. The systems we use in our studies have a number of atoms ranging from 2,247 to 12,288. We study
the convergence of the unconstrained formulation and its scaling on a Cray XC40 (a partition with 9,688 Intel KNL nodes). We show that there is a trade-off between the preconditioner that leads to fast convergence and lower cost preconditioners that lead to best time to solution.

Best Poster Finalist: yes

10:30 am - 12:00 pm

Best Research Posters Presentations

Poster 91: FreeCompilerCamp: Online Training for Extending Compilers
Alok Mishra (Stony Brook University, Lawrence Livermore National Laboratory), Anjia Wang (University of North Carolina, Charlotte; Lawrence Livermore National Laboratory), Chunhua Liao (Lawrence Livermore National Laboratory), Yonghong Yan (University of North Carolina, Charlotte), Barbara Chapman (Stony Brook University)

In this presentation, we introduce an ongoing effort of an online training platform aimed to automate the training of developers to quickly extend compilers - FreeCompilerCamp.org. Our free and open platform allows anyone who is interested in developing compilers to learn the necessary skills. A live training website, built on top of Play-With-Docker, is set up so that anyone with internet access and a web browser will be able to take this training. The entire training system is open-source and developers with relevant skills can contribute new tutorials and deploy it on a private server, workstation or even laptop. We have created some initial tutorials on how to extend the Clang/LLVM or ROSE compilers to support new OpenMP features. Using a web interface consisting of two side-by-side panels, users can follow the tutorials on one side and immediately practice what they learned in a terminal sandbox embedded on the other.

Best Poster Finalist: yes

Poster 92: Nanoporous Flow Simulations on the Summit Supercomputer
Yidong Xia (Idaho National Laboratory), Lixiang Luo (IBM - TJ Watson Research Center), Ansel Blumers (Brown University), Joshua Kane (Idaho National Laboratory), Jan Goral (University of Utah), Yu-Hang Tang (Lawrence Berkeley National Laboratory), Zhen Li (Clemson University, Brown University), Hai Huang (Idaho National Laboratory), Milind Deo (University of Utah)

Fluid flow behaviors in nanoporous materials are distinct from those following the continuum
physics. Numerical simulations can be a complement to laboratory experiments. This work presents a dissipative particle dynamics (DPD) package for GPU-accelerated mesoscale flow simulations in nanoporous materials. In an ideal benchmark that minimizes load imbalance, the package delivered nearly perfect strong- and weak-scaling (with up to 4 billion DPD particles) on up to 1,536 V100 GPUs on Oak Ridge National Laboratory’s Summit supercomputer. More remarkably, in a benchmark to measure its usefulness with realistic nanopores in SBA-15 silica, the package exhibited more than 20x speedup over its LAMMPS-based CPU counterpart with the same number nodes (e.g., 384 V100 GPUs vs. 2,688 POWER9 cores). It is worth highlighting that the NVLink2 Host-to-Device interconnects kept the cost of CPU-GPU memory copy as low as only 10% of GPU activity time per rank: 4 times less than their PCIe counterparts.

Best Poster Finalist: yes

**Poster 96: TSQR on TensorCores**

Hirohito Ootomo (Tokyo Institute of Technology), Rio Yokota (Tokyo Institute of Technology)

Tall-Skinny QR (TSQR) is an efficient algorithm for calculating the QR decomposition of $m \times n$ matrices where $m << n$, which is done by recursively performing QR decomposition on subdivided blocks of the tall and skinny matrix. Such operations are useful for low-rank approximation methods, which are replacing more and more dense linear algebra in both scientific computing and machine learning fields. The present work focuses on the implementation of this important algorithm on Tensor Cores, which are available on the latest NVIDIA GPUs. We evaluate the speed, accuracy, and stability of TSQR on TensorCores.

Best Poster Finalist: yes

**Poster 110: Hierarchical Data Prefetching in Multi-Tiered Storage Environments**

Hariharan Devarajan (Illinois Institute of Technology), Anthony Kougkas (Illinois Institute of Technology), Xian-He Sun (Illinois Institute of Technology)

In the era of data-intensive computing, accessing data with a high-throughput and low-latency is very imperative. Data prefetching is used for hiding read latency by requesting data before it is needed to move it from a high-latency medium to a low-latency one. However, existing solutions do not consider the multi-tiered storage and also suffer from under-utilization of prefetching resources and unnecessary evictions. Additionally, existing approaches implement a client-pull model where understanding the application’s I/O behavior drives prefetching decisions. Moving toward exascale, where machines run multiple applications concurrently by accessing files in a workflow, a more
A data-centric approach resolves challenges such as cache pollution and redundancy. In this study, we present HFetch, a truly hierarchical data prefetcher that adopts a server-push approach to data prefetching. We demonstrate the benefits of such an approach. Results show 10-35% performance gains over existing prefetchers and over 50% when compared to systems with no prefetching.

**Poster 54: Massively Parallel Eigensolvers Based on Unconstrained Energy Functionals**

Methods

Osni Marques (Lawrence Berkeley National Laboratory), Mauro Del Ben (Lawrence Berkeley National Laboratory), Andrew Canning (Lawrence Berkeley National Laboratory)

This poster focuses on a preconditioned conjugate gradient based iterative eigensolver using an unconstrained energy functional minimization scheme. This scheme avoids an explicit reorthogonalization of the trial eigenvectors and becomes an attractive alternative for the solution of very large problems. The unconstrained formulation is implemented in the first-principles materials and chemistry CP2K code, which performs electronic structure calculations based on a density functional theory approximation to the solution of the many-body Schrödinger equation. The systems we use in our studies have a number of atoms ranging from 2,247 to 12,288. We study the convergence of the unconstrained formulation and its scaling on a Cray XC40 (a partition with 9,688 Intel KNL nodes). We show that there is a trade-off between the preconditioner that leads to fast convergence and lower cost preconditioners that lead to best time to solution.

**Thursday, November 21**

8:30 am - 5:00 pm

**Research Posters Display**

**Poster 74: Enabling Code Portability of a Parallel and Distributed Smooth-Particle Hydrodynamics Application, FleCSPH**

Suyash Tandon (University of Michigan), Nicholas Stegmeier (University of Illinois), Vasu Jaganath (University of Wyoming), Jennifer Ranta (Michigan State University), Rathish Ratnasingam (Newcastle University), Elizabeth Carlson (University of Nebraska), Julien Loiseau (Los Alamos National Laboratory)
Core-collapse supernovae (CCSNe) are integral to the formation and distribution of heavy elements across the universe. However, CCSNe are highly complex and inherently non-linear phenomena. Large-scale simulations of these cosmic events can provide us a glimpse of their hydrodynamic and nucleosynthetic processes which are difficult to observe. To enable these massive numerical simulations on high-performance computing (HPC) centers, this study uses FleCSPH, a parallel and distributed code, based on the smooth-particle hydrodynamics (SPH) formulation. In the recent years, the HPC architecture has evolved and the next generation of exascale computers are expected to feature heterogenous architecture. Therefore, it is important to maintain code portability across platforms. This work demonstrates code portability of FleCSPH through the incorporation of Kokkos C++ library and containers using Charliecloud.

Best Poster Finalist: no

**Poster 73: Accelerating Large-Scale GW Calculations on Hybrid CPU-GPU Architectures**

Mauro Del Ben (Lawrence Berkeley National Laboratory), Charlene Yang (National Energy Research Scientific Computing Center (NERSC)), Felipe Jornada (University of California, Berkeley; Lawrence Berkeley National Laboratory), Steven G. Louie (University of California, Berkeley; Lawrence Berkeley National Laboratory), Jack Deslippe (National Energy Research Scientific Computing Center (NERSC))

In this poster, we present the strategy, progress, and performance while GPU porting one of the major modules, epsilon, of the electronic structure code BerkeleyGW. Epsilon represents the most time-consuming routines in the BerkeleyGW workflow for large-scale material science simulations. Some of the porting/optimization strategies include, changing our original data layout to efficiently use libraries such as cuBLAS and cuFFT, implementation of specific CUDA kernels to minimize data copies between host/device and keeping data on device, efficient use of data streams to leverage high concurrency on the device, asynchronous memory copies and overlapping (MPI) communication on the host and computation on the device. Preliminary results are presented in terms of the speedup compare to the CPU-only implementation, strong/weak scaling, and power efficiency. Excellent acceleration is demonstrated: up to 30x for specific kernels. Our port also exhibits good scalability and about 16x higher FLOPs/watt efficiency compared to the CPU-only implementation.

Best Poster Finalist: no
Poster 89: BeeCWL: A CWL Compliant Workflow Management System
Betis Baheri (Kent State University), Steven Anaya (New Mexico Institute of Mining and Technology), Patricia Grubel (Los Alamos National Laboratory), Qiang Guan (Kent State University), Timothy Randles (Los Alamos National Laboratory)

Scientific workflows are used widely to carry out complex and hierarchical experiments. Although there are many trends to extend the functionality of workflow management systems to cover all possible requirements that may arise from a user community, one unified standard over cloud and HPC systems is still missing. In this paper, we propose a Common Workflow Language (CWL) compliant workflow management system. BeeCWL is a parser to derive meaningful information such as requirements, steps, relationships, etc. from CWL files and to create a graph database from those components. Generated graphs can be passed to an arbitrary scheduler and management system to decide whether there are enough resources to optimize and execute the workflow. Lastly, the user can have control over workflow execution, collecting logs, and restart or rerun some part of a complex workflow.

Best Poster Finalist: no

Poster 150: A Machine Learning Approach to Understanding HPC Application Performance Variation
Burak Aksar (Boston University, Sandia National Laboratories), Benjamin Schwaller (Sandia National Laboratories), Omar Aaziz (Sandia National Laboratories), Emre Ates (Boston University), Jim Brandt (Sandia National Laboratories), Ayse K. Coskun (Boston University), Manuel Egele (Boston University), Vitus Leung (Sandia National Laboratories)

Performance anomalies are difficult to detect because often a “healthy system” is vaguely defined, and the ground truth for how a system should be operating is evasive. As we move to exascale, however, detection of performance anomalies will become increasingly important with the increase in size and complexity of systems. There are very few accepted ways of detecting anomalies in the literature, and there are no published and labeled sets of anomalous HPC behavior. In this research, we develop a suite of applications that represent HPC workloads and use data from a lightweight metric collection service to train machine learning models to predict the future behavior of metrics. In the future, this work will be used to predict anomalous runs in compute nodes and determine some root causes of performance issues to help improve the efficiency of HPC system administrators and users.

Best Poster Finalist: no
Poster 81: Performance of Devito on HPC-Optimised ARM Processors
Hermes Senger (Federal University of São Carlos, Brazil; University of São Paulo), Jaime Freire de Souza (Federal University of São Carlos, Brazil), Edson Satoshi Gomi (University of São Paulo), Fabio Luporini (Imperial College, London), Gerard Gorman (Imperial College, London)

We evaluate the performance of Devito, a domain specific language (DSL) for finite differences on Arm ThunderX2 processors. Experiments with two common seismic computational kernels demonstrate that Devito can apply automatic code generation and optimization across Arm and Intel platforms. The code transformations include: parallelism, and SIMD vectorization (OpenMP >=4); loop tiling (with best block shape obtained via auto-tuning); domain-specific symbolic optimisation such as common sub-expression elimination and factorisation for Flop reduction, polynomial approximations for trigonometry terms, and heuristic hoisting of time-invariant expressions. Results show that Devito can achieve performance on Arm processors which is competitive to other Intel Xeon processors.

Best Poster Finalist: no

Poster 103: LIKWID 5: Lightweight Performance Tools
Thomas Gruber (Erlangen Regional Computing Center), Jan Eitzinger (Erlangen Regional Computing Center), Georg Hager (Erlangen Regional Computing Center), Gerhard Wellein (Erlangen Regional Computing Center)

LIKWID is a tool suite for performance oriented programmers with a worldwide user group. It is developed by the HPC group of the University Erlangen-Nuremberg since 2009 to support them in their daily research and performance engineering of user codes. The HPC landscape has become more and more diverse over the last years with clusters using non-x86 architectures and being equipped with accelerators. With the new major version, the architectural support of LIKWID is extended to ARM and POWER CPUs with the same functionality and features as for x86 architectures. Besides the CPU monitoring, the new version provides access the hardware counting facilities of Nvidia GPUs. This poster introduces the new features and shows the successes of applying LIKWID to identify performance bottlenecks and to test optimizations. Furthermore, the poster gives an overview of how users can integrate the LIKWID tools in their application using a lightweight add-once-and-reuse instrumentation API.

Best Poster Finalist: no
Poster 149: Solving Phase-Field Equations in Space-Time: Adaptive Space-Time Meshes and Stabilized Variational Formulations
Kumar Saurabh (Iowa State University), Biswajit Khara (Iowa State University), Milinda Fernando (University of Utah), Masado Ishii (University of Utah), Hari Sundar (University of Utah), Baskar Ganapathysubramanian (Iowa State University)

We seek to efficiently solve a generalized class of partial differential equations called the phase-field equations. These non-linear PDE’s model phase transition (solidification, melting, phase-separation) phenomena which exhibit spatially and temporally localized regions of steep gradients. We consider time as an additional dimension and simultaneously solve for the unknown in large blocks of time (i.e. in space-time), instead of the standard approach of sequential time-stepping. We use variational multiscale (VMS) based finite element approach to solve the ensuing space-time equations. This allows us to (a) exploit parallelism not only in space but also in time, (b) gain high order accuracy in time, and (c) exploit adaptive refinement approaches to locally refine region of interest in both space and time. We illustrate this approach with several canonical problems including melting and solidification of complex snowflake structures.

Best Poster Finalist: no

Poster 84: ESTEE: A Simulation Toolkit for Distributed Workflow Execution
Vojtěch Cima (IT4Innovations, Czech Republic), Jakub Beránek (IT4Innovations, Czech Republic), Stanislav Böhm (IT4Innovations, Czech Republic)

Task graphs provide a simple way to describe scientific workflows (sets of tasks with dependencies) that can be executed on both HPC clusters and in the cloud. An important aspect of executing such graphs is the used scheduling algorithm. Many scheduling heuristics have been proposed in existing works; nevertheless, they are often tested in oversimplified environments. We introduce a simulation environment designed for prototyping and benchmarking task schedulers. Our simulation environment, scheduler source codes, and graph datasets are open in order to be fully reproducible. To demonstrate usage of Estee, as an example, we compare the performance of various workflow schedulers in an environment using two different network models.

Best Poster Finalist: no

Poster 93: Robust Data-Driven Power Simulator for Fast Cooling Control Optimization of a Large-Scale Computing System
Takashi Shiraishi (Fujitsu Laboratories Ltd), Hiroshi Endo (Fujitsu Laboratories Ltd), Takaaki Hineno
Power of large-scale systems such as an HPC or a datacenter is a significant issue. Cooling units consume 30% of the total power. General control policies for cooling units are local and static (manual overall optimization nearly once a week). However, free cooling and IT-load fluctuation may change hourly optimum control variables of the cooling units. In this work, we present a deep neural network (DNN) power simulator that can learn from actual operating logs and can quickly identify the optimum control variables. We demonstrated the power simulator of an actual large-scale system with 4.7-MW-power IT load. Our robust simulator predicted the total power with error of 4.8% without retraining during one year. We achieved optimization by the simulator within 80 seconds that was drastically faster than previous works. The dynamic control optimization each hour showed a 15% power reduction compared to that of conventional policy in the actual system.

Best Poster Finalist: no

**Poster 115: sDNA: Software-Defined Network Accelerator Based on Optical Interconnection Architecture**

En Shao (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences), Guangming Tan (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences), Zhan Wang (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences; Chinese Academy of Sciences), Guojun Yuan (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences; Chinese Academy of Sciences), Ninghui Sun (Institute of Computing Technology, Chinese Academy of Sciences; Chinese Academy of Sciences)

Software-Defined Network Accelerator (sDNA) is a new accelerated system for the exascale computer. Inspired by the edge forwarding index (EFI), the main contribution of our work is that it presents an extended EFI-based optical interconnection method with slow switching optical device. In our work, we found that sDNA based on extended EFI evaluation is not only able to offload the traffic from an electrical link to an optical link but is also able to avoid congestion inherent to electrical link.

Best Poster Finalist: no

**Poster 49: WarpX: Toward Exascale Modeling of Plasma Particle Accelerators on GPU**

Maxence Thevenet (Lawrence Berkeley National Laboratory), Jean-Luc Vay (Lawrence Berkeley National Laboratory), Ann Almgren (Lawrence Berkeley National Laboratory), Diana Amorim (Lawrence Berkeley National Laboratory), John Bell (Lawrence Berkeley National Laboratory), Axel
Particle accelerators are a vital part of the DOE-supported infrastructure of discovery science and applications, but we need game-changing improvements in the size and cost for future accelerators. Plasma-based particle accelerators stand apart in their potential for these improvements. Turning this from a promising technology into mainstream scientific tools depends critically on high-performance, high-fidelity modeling of complex processes.

WarpX is an open-source particle-in-cell (PIC) code supported by the Exascale Computing Project (ECP) that is combining advanced algorithms with adaptive mesh refinement to allow challenging simulations of a multi-stage plasma-based TeV acceleration relevant for future high-energy physics discoveries. WarpX relies on the ECP co-design center for mesh refinement AMReX, and runs on CPU and GPU-accelerated computers. Production simulation have run on Cori KNL at NERSC and Summit at OLCF. In this poster, recent results and strategies on GPU will be presented, along with recent performance results.

Best Poster Finalist: no

**Poster 50: Implementing an Adaptive Sparse Grid Discretization (ASGarD) for High Dimensional Advection-Diffusion Problems on Exascale Architectures**

M. Graham Lopez (Oak Ridge National Laboratory), David L. Green (Oak Ridge National Laboratory), Lin Mu (University of Georgia), Ed D’Azevedo (Oak Ridge National Laboratory), Wael Elwasif (Oak Ridge National Laboratory), Tyler McDaniel (University of Tennessee), Timothy Younkin (University of Tennessee), Adam McDaniel (Oak Ridge National Laboratory), Diego Del-Castillo-Negrete (Oak Ridge National Laboratory)

Many scientific domains require the solution of high dimensional PDEs. Traditional grid- or mesh-based methods for solving such systems in a noise-free manner quickly become intractable due to the scaling of the degrees of freedom going as $O(N^d)$ sometimes called "the curse of dimensionality." We are developing an arbitrarily high-order discontinuous-Galerkin finite-element solver that leverages an adaptive sparse-grid discretization whose degrees of freedom scale as $O(N\log 2 N^{D-1})$. This method and its subsequent reduction in the required resources is being
applied to several PDEs including time-domain Maxwell’s equations (3D), the Vlasov equation (in up to 6D) and a Fokker-Planck-like problem in ongoing related efforts. Here we present our implementation which is designed to run on multiple accelerated architectures, including distributed systems. Our implementation takes advantage of a system matrix decomposed as the Kronecker product of many smaller matrices which is implemented as batched operations.

Best Poster Finalist: no

**Poster 91: FreeCompilerCamp: Online Training for Extending Compilers**

Alok Mishra (Stony Brook University, Lawrence Livermore National Laboratory), Anjia Wang (University of North Carolina, Charlotte; Lawrence Livermore National Laboratory), Chunhua Liao (Lawrence Livermore National Laboratory), Yonghong Yan (University of North Carolina, Charlotte), Barbara Chapman (Stony Brook University)

In this presentation, we introduce an ongoing effort of an online training platform aimed to automate the training of developers to quickly extend compilers - FreeCompilerCamp.org. Our free and open platform allows anyone who is interested in developing compilers to learn the necessary skills. A live training website, built on top of Play-With-Docker, is set up so that anyone with internet access and a web browser will be able to take this training. The entire training system is open-source and developers with relevant skills can contribute new tutorials and deploy it on a private server, workstation or even laptop. We have created some initial tutorials on how to extend the Clang/LLVM or ROSE compilers to support new OpenMP features. Using a web interface consisting of two side-by-side panels, users can follow the tutorials on one side and immediately practice what they learned in a terminal sandbox embedded on the other.

Best Poster Finalist: yes

**Poster 51: SmartK: Efficient, Scalable, and Winning Parallel MCTS**

Michael S. Davinroy (Swarthmore College), Shawn Pan (Swarthmore College), Bryce Wiedenbeck (Swarthmore College, Davidson College), Tia Newhall (Swarthmore College)

SmartK is our efficient and scalable parallel algorithm for Monte Carlo Tree Search (MCTS), an approximation technique for game searches. MCTS is also used to solve problems as diverse as planning under uncertainty, combinatorial optimization, and high-energy physics. In these problems, the solution search space is significantly large, necessitating parallel solutions. Shared memory parallel approaches do not scale well beyond the size of a single node's RAM. SmartK is a distributed memory parallelization that takes advantage of both inter-node and intra-node parallelism and a large
cumulative RAM found in clusters. SmartK's novel selection algorithm combined with its ability to efficiently search the solution space, results in better solutions than other MCTS parallel approaches. Results of an MPI implementation of SmartK for the game of Hex, show SmartK yields a better win percentage than other parallel algorithms, and that its performance scales to larger search spaces and high degrees of parallelism.

Best Poster Finalist: no

**Poster 70: Numerical Method and Parallelization for the Computation of Coherent Synchrotron Radiation**

Boqian Shen (Rice University, Los Alamos National Laboratory)

The purpose of this work is to develop and parallelize an accurate and efficient numerical method for the computation of synchrotron radiation from relativistic electrons in the near field. The high-brilliance electron beam and coherent short-wavelength light source provide a powerful method to understand the microscopic structure and dynamics of materials. Such a method supports a wide range of applications including matter physics, structural biology, and medicine development. To understand the interaction between the beam and synchrotron radiation, an accurate and efficient numerical simulation is needed. With millions of electrons, the computational cost of the field would be large. Thus, multilevel parallelism and performance portability are desired since modern supercomputers are getting more complex and heterogeneous. The performance model and performance analysis are presented.

Best Poster Finalist: no

**Poster 146: AI Matrix: A Deep Learning Benchmark for Alibaba Data Centers**

Wei Zhang (Alibaba Inc), Wei Wei (Alibaba Inc), Lingjie Xu (Alibaba Inc), Lingling Jin (Alibaba Inc), Cheng Li (University of Illinois)

This work introduces AI Matrix, an in-house Deep Learning (DL) benchmark suite developed specifically for Alibaba’s e-commerce environment. AI Matrix results from a full investigation of the DL applications used inside Alibaba and aims to cover the typical DL applications that account for more than 90% of the GPU usage in Alibaba data centers. This benchmark suite collects DL models that are either directly used or closely resemble the models used in the company’s real e-commerce applications. It also collects the real e-commerce applications if no similar DL models are not available. Through the high coverage and close resemblance to real applications, AI Matrix fully represents the DL workloads on Alibaba data centers. The collected benchmarks mainly fall into three categories:
computer vision, recommendation, and language processing, which consist of the most majority of DL applications in Alibaba. AI Matrix is made open source, hoping it can benefit the public.

Best Poster Finalist: no

**Poster 54: Massively Parallel Eigensolvers Based on Unconstrained Energy Functionals Methods**  
Osni Marques (Lawrence Berkeley National Laboratory), Mauro Del Ben (Lawrence Berkeley National Laboratory), Andrew Canning (Lawrence Berkeley National Laboratory)

This poster focuses on a preconditioned conjugate gradient based iterative eigensolver using an unconstrained energy functional minimization scheme. This scheme avoids an explicit reorthogonalization of the trial eigenvectors and becomes an attractive alternative for the solution of very large problems. The unconstrained formulation is implemented in the first-principles materials and chemistry CP2K code, which performs electronic structure calculations based on a density functional theory approximation to the solution of the many-body Schrödinger equation. The systems we use in our studies have a number of atoms ranging from 2,247 to 12,288. We study the convergence of the unconstrained formulation and its scaling on a Cray XC40 (a partition with 9,688 Intel KNL nodes). We show that there is a trade-off between the preconditioner that leads to fast convergence and lower cost preconditioners that lead to best time to solution.

Best Poster Finalist: yes

**Poster 100: Comparison of Array Management Library Performance - A Neuroscience Use Case**  
Donghe Kang (Ohio State University), Oliver Rübel (Lawrence Berkeley National Laboratory), Suren Byna (Lawrence Berkeley National Laboratory), Spyros Blanas (Ohio State University)

Array management libraries, such as HDF5, Zarr, etc., depend on a complex software stack that consists of parallel I/O middleware (MPI-IO), POSIX-I/O, and file systems. Components in the stack are interdependent, such that effort in tuning the parameters in these software libraries for optimal performance is non-trivial. On the other hand, it is challenging to choose an array management library based on the array configuration and access patterns. In this poster, we investigate the performance aspect of two array management libraries, i.e., HDF5 and Zarr, in the context of a neuroscience use case. We highlight the performance variability of HDF5 and Zarr in our preliminary results and discuss potential optimization strategies.

Best Poster Finalist: no
Poster 118: Self-Driving Reconfigurable Silicon Photonic Interconnects (Flex-LIONS) with Deep Reinforcement Learning
Roberto Proietti (University of California, Davis), Yu Shang (University of California, Davis), Xian Xiao (University of California, Davis), Xiaoliang Chen (University of California, Davis), Yu Zhang (University of California, Davis), SJ Ben Yoo (University of California, Davis)

We propose a self-driving reconfigurable optical interconnect architecture for HPC systems exploiting a deep reinforcement learning (DRL) algorithm and a reconfigurable silicon photonic (SiPh) switching fabric to adapt the interconnect topology to different traffic demands. Preliminary simulation results show that after training, the DRL-based SiPh fabric provides the lowest average end-to-end latency for time-varying traffic patterns.

Best Poster Finalist: no

Poster 147: Extremely Accelerated Deep Learning: ResNet-50 Training in 70.4 Seconds
Akihiro Tabuchi (Fujitsu Laboratories Ltd), Akihiko Kasagi (Fujitsu Laboratories Ltd), Masafumi Yamazaki (Fujitsu Laboratories Ltd), Takumi Honda (Fujitsu Laboratories Ltd), Masahiro Miwa (Fujitsu Laboratories Ltd), Takashi Shiraishi (Fujitsu Laboratories Ltd), Motohiro Kosaki (Fujitsu Laboratories Ltd), Naoto Fukumoto (Fujitsu Laboratories Ltd), Tsuguchika Tabaru (Fujitsu Laboratories Ltd), Atsushi Ike (Fujitsu Laboratories Ltd), Kohta Nakashima (Fujitsu Laboratories Ltd)

Distributed deep learning using a large mini-batch is a key technology to accelerate training in deep learning. However, it is difficult to achieve a high scalability and maintain validation accuracy in distributed learning on large clusters. We introduce two optimizations, reducing the computation time and overlapping the communication with the computation. By applying the techniques and using 2,048 GPUs, we achieved the world's fastest ResNet-50 training in MLPerf, which is a de facto standard DNN benchmark (as of July 2019).

Best Poster Finalist: no

Poster 111: Multiple HPC Environments-Aware Container Image Configuration for Bioinformatics Application
Kento Aoyama (Tokyo Institute of Technology, National Institute of Advanced Industrial Science and Technology (AIST)), Hiroki Watanabe (Tokyo Institute of Technology, National Institute of Advanced Industrial Science and Technology (AIST)), Masahito Ohue (Tokyo Institute of Technology), Yutaka Akiyama (Tokyo Institute of Technology)
Containers have a considerable advantage for application portability in different environments by isolating process with a small performance overhead; thus it has been rapidly getting popular in a wide range of science fields. However, there are problems in container image configuration when run in multiple HPC environments, and it requires users to have knowledge of systems, container runtimes, container image format, and library compatibilities in HPC environments.

In this study, we introduce our HPC container workflow in multiple supercomputing environments that have different system/library specifications (ABCI, TSUBAME3.0). Our workflow provides custom container image configurations for HPC environments by taking into account differences in container runtime, container image, and library compatibility between the host and inside of the container. We also show the parallel performance of our application in each HPC environment.

Best Poster Finalist: no

Poster 134: Minimal-Precision Computing for High-Performance, Energy-Efficient, and Reliable Computations
Daichi Mukunoki (RIKEN Center for Computational Science (R-CCS)), Toshiyuki Imamura (RIKEN Center for Computational Science (R-CCS)), Yiyu Tan (RIKEN Center for Computational Science (R-CCS)), Atsushi Koshiba (RIKEN Center for Computational Science (R-CCS)), Jens Huthmann (RIKEN Center for Computational Science (R-CCS)), Kentaro Sano (RIKEN Center for Computational Science (R-CCS)), Fabienne Jézéquel (Sorbonne University), Stef Graillat (Sorbonne University), Roman Iakymchuk (Sorbonne University), Norihisa Fujita (University of Tsukuba), Taisuke Boku (University of Tsukuba)

In numerical computations, the precision of floating-point computations is a key factor to determine the performance (speed and energy-efficiency) as well as the reliability (accuracy and reproducibility). However, the precision generally plays a contrary role for both. Therefore, the ultimate concept for maximizing both at the same time is the minimal-precision computation through precision-tuning, which adjusts the optimal precision for each operation and data. Several studies have been already conducted for it so far, but the scope of those studies is limited to the precision-tuning alone. In this study, we propose a more broad concept of the minimal-precision computing with precision-tuning, involving both hardware and software stack.

Best Poster Finalist: no

Poster 112: Building Complex Software Applications Inside Containers
Calvin D. Seamons (Los Alamos National Laboratory)
High performance computing (HPC) scientific applications require complex dependencies to operate. As user demand for HPC systems increases, it becomes unrealistic to support every unique dependency request. Containers can offer the ability to satisfy the users' dependency request while simultaneously offering HPC portability across systems. By "containerizing" Model for Prediction Across Scales (MPAS, a large atmospheric simulation suite), we show that it is possible to containerize and run complex software. Furthermore, the container can be run across different HPC systems with nearly identical results (21 bytes difference over 2.1 gigabytes). Containers have the possibility to bring flexibility to code teams in HPC by helping to meet the demand for user defined software stacks (UDSS), and giving teams the ability to choose their software, independently of what is offered by the HPC system.

Best Poster Finalist: no

**Poster 101: Job Performance Overview of Apache Flink and Apache Spark Applications**
*Jan Frenzel (Technical University Dresden), René Jäkel (Technical University Dresden)*

Apache Spark and Apache Flink are two Big Data frameworks used for fast data exploration and analysis. Both frameworks provide the runtime of program sections and performance metrics, such as the number of bytes read or written, via an integrated dashboard. Performance metrics available in the dashboard lack timely information and are only shown aggregated in a separate part of the dashboard. However, performance investigations and optimizations would benefit from an integrated view with detailed performance metric events. Thus, we propose a system that samples metrics at runtime and collects information about the program sections after the execution finishes. The performance data is stored in an established format independent from Spark and Flink versions and can be viewed with state-of-the-art performance tools, i.e. Vampir. The overhead depends on the sampling interval and was below 10% in our experiments.

Best Poster Finalist: no

**Poster 113: Improvements Toward the Release of the Pavilion 2.0 Test Harness**
*Kody J. Everson (Los Alamos National Laboratory, Dakota State University), Maria Francine Lapid (Los Alamos National Laboratory)*

High-performance computing production support entails thorough testing in order to evaluate the efficacy of a system for production-grade workloads. There are various phases of a system’s life-cycle to assess, requiring different methods to accomplish effective evaluation of performance and
correctness. Due to the unique and distributed nature of an HPC-system, the necessity for sophisticated tools to automatically harness and assess test results, all while interacting with schedulers and programming environment software, requires a customizable, extensible, and lightweight system to manage concurrent testing. Beginning with the recently refactored codebase of Pavilion 1.0, we assisted with the finishing touches on readying this software for open-source release and production usage. Pavilion 2.0 is a Python 3-based testing framework for HPC clusters that facilitates the building, running, and analysis of tests through an easy-to-use, flexible, YAML-based configuration system. This enables users to write their own tests by simply wrapping everything in Pavilion’s well-defined format.

Best Poster Finalist: no

**Poster 119: Toward Lattice QCD on Fugaku: SVE Compiler Studies and Micro-Benchmarks in the RIKEN Fugaku Processor Simulator**

Nils Meyer (University of Regensburg, Bavaria), Tilo Wettig (University of Regensburg, Bavaria), Yuetsu Kodama (RIKEN Center for Computational Science (R-CCS)), Mitsuhisa Sato (RIKEN Center for Computational Science (R-CCS))

The Fugaku supercomputer, successor to the Japanese flagship K-Computer, will start operation in 2021. Fugaku incorporates the Fujitsu A64FX processor, which is the first hardware implementation supporting the Arm SVE instruction set, in this case a 512-bit version. Real hardware is not accessible today, but RIKEN has designed a simulator of the A64FX. We present micro-benchmarks relevant for Lattice QCD obtained in the RIKEN Fugaku processor simulator and compare three different SVE compilers.

Best Poster Finalist: no

**Poster 62: Emulating Multi-Pattern Quantum Grover’s Search on a High-Performance Reconfigurable Computer**

Naveed Mahmud (University of Kansas), Bennett Haase-Divine (University of Kansas), Bailey K. Srimoungchanh (University of Kansas), Nolan Blankenau (University of Kansas), Annika Kuhnke (University of Kansas), Esam El-Araby (University of Kansas)

Grover’s search (GS) is a widely studied quantum algorithm that can be employed for both single and multi-pattern search problems and potentially provides quadratic speedup over existing classical search algorithms. In this paper, we propose a multi-pattern quantum search methodology based on a modified GS quantum circuit. The proposed method combines classical post-processing
permutations with a modified Grover's circuit to efficiently search for given single/multiple input patterns. Our proposed methodology reduces quantum circuit complexity, realizes space-efficient emulation hardware and improves overall system configurability for dynamic, multi-pattern search. We use a high-performance reconfigurable computer to emulate multi-pattern GS(MGS) and present scalable emulation architectures of a complete multi-pattern search system. We validate the system and provide analysis of experimental results in terms of FPGA resource utilization and emulation time. Our results include a successful hardware architecture that is capable of emulating MGS algorithm up to 32 fully-entangled quantum bits on a single FPGA.

Best Poster Finalist: no

**Poster 107: Exploring Interprocess Work Stealing for Balanced MPI Communication**  
Kaiming Ouyang (University of California, Riverside), Min Si (Argonne National Laboratory), Zizhong Chen (University of California, Riverside)

Workload balance among MPI processes is a critical consideration during the development of HPC applications. However, because of many factors such as complex network interconnections and irregularity of HPC applications, fully achieving workload balance in practice is nearly impossible. Although interprocess job stealing is a promising solution, existing shared-memory techniques that lack necessary flexibility or cause inefficiency during data access cannot provide an applicable job-stealing implementation. To solve this problem, we propose a new process-in-process (PiP) interprocess job-stealing method to balance communication workload among processes on MPI layers. Our initial experimental results show PiP-based job stealing can efficiently help amortize workload, reduce imbalance, and greatly improve intra- and intersocket ping-pong performance compared with original MPI.

Best Poster Finalist: no

**Poster 127: sFlow Monitoring for Security and Reliability**  
Xava A. Grooms (Los Alamos National Laboratory, University of Kentucky), Robert V. Rollins (Los Alamos National Laboratory, Michigan Technological University), Collin T. Rumpca (Los Alamos National Laboratory, Dakota State University)

In the past ten years, High Performance Computing (HPC) has moved far beyond the terascale performance, making petascale systems the new standard. The drastic improvement in performance has been largely unmatched with insignificant improvements in system monitoring. Thus, there is an immediate need for practical and scalable monitoring solutions to ensure the effectiveness of costly
compute clusters. This project aims to explore the viability and impact of sFlow enabled switches in cluster network monitoring for security and reliability. A series of tests and exploits were performed to target specific network abnormalities on a nine-node HPC cluster. The results present web-based dashboards that can aid network administrators in improving a cluster’s security and reliability.

Best Poster Finalist: no

**Poster 77: Extreme Scale Phase-Field Simulations of Sintering Processes**  
*Johannes Hötzer (Karlsruhe University of Applied Sciences), Henrik Hierl (Karlsruhe University of Applied Sciences), Marco Seiz (Karlsruhe Institute of Technology), Andreas Reiter (Karlsruhe Institute of Technology), Britta Nestler (Karlsruhe Institute of Technology)*

The sintering process, which turns loose powders into dense materials, is naturally found in the formation of glaciers, but is also the indispensable process to manufacture ceramic materials. This process is described by a dynamically evolving microstructure, which largely influences the resulting material properties.

To investigate this complex three-dimensional, scale-bridging evolution in realistic domain sizes, a highly optimized and parallelized multiphysics phase-field solver is developed. The solver is optimized in a holistic way, from the application level over the time integration and parallelization, down to the hardware. Optimizations include communication hiding, explicit vectorization, implicit schemes, and local reduction of degrees of freedom.

With this, we are able to investigate large-scale, three-dimensional domains, and long integration times. We have achieved a single-core peak performance of 32.5%, scaled up to 98304 cores on Hazel Hen and SuperMUC-NG, and simulated a multimillion particle system.

Best Poster Finalist: no

**Poster 140: Toward Automatic Function Call Generation for Deep Learning**  
*Shizhi Tang (Tsinghua University, China), Jidong Zhai (Tsinghua University, China)*

Mainstream deep learning frameworks are commonly implemented by invoking underlying high performance tensor libraries on various architectures. However, as these libraries provide increasingly complex semantics including operator fusions, in-place operations, and various memory layouts, the gap between mathematical deep learning models and the underlying libraries becomes larger. In this paper, inspired by the classic problem of Instruction Selection, we design a theorem solver guided
exhausted search algorithm to select functions for complex tensor computations. Preliminary results with some micro-benchmarks and a real model show that our approach can outperform both Tensorflow and Tensor Comprehensions at run time.

Best Poster Finalist: no

**Poster 83: ETL: Elastic Training Layer for Deep Learning**
Lei Xie (Tsinghua University, China), Jidong Zhai (Tsinghua University, China)

Due to the rising of deep learning, clusters for deep learning training are widely deployed in production. However, static task configuration and resource fragmentation problems in existing clusters result in low efficiency and poor quality of service. We propose ETL, an elastic training layer for deep learning, to help address them once for all. ETL adopts many novel mechanisms, such as lightweight and configurable report primitive and asynchronous, parallel and IO-free state replication, to achieve both high elasticity and efficiency. The evaluation demonstrates the low overhead and high efficiency of these mechanisms and reveals the advantages of elastic deep learning supported by ETL.

Best Poster Finalist: no

**Poster 130: Deep Learning-Based Feature-Aware Data Modeling for Complex Physics Simulations**
Qun Liu (Louisiana State University), Subhashis Hazarika (Ohio State University), John M. Patchett (Los Alamos National Laboratory), James P. Ahrens (Los Alamos National Laboratory), Ayan Biswas (Los Alamos National Laboratory)

Data modeling and reduction for in situ is important. Feature-driven methods for in situ data analysis and reduction are a priority for future exascale machines as there are currently very few such methods. We investigate a deep-learning-based workflow that targets in situ data processing using autoencoders. We employ integrated skip connections to obtain higher performance compared to the existing autoencoders. Our experiments demonstrate the initial success of the proposed framework and create optimism for the in situ use case.

Best Poster Finalist: no

**Poster 125: Physics Informed Generative Adversarial Networks for Virtual Mechanical Testing**
Physics-informed generative adversarial networks (PI-GANs) are used to learn the underlying probability distributions of spatially-varying material properties (e.g., microstructure variability in a polycrystalline material). While standard GANs rely solely on data for training, PI-GANs encode physics in the form of stochastic differential equations using automatic differentiation. The goal here is to show that experimental data from a limited number of material tests can be used with PI-GANs to enable unlimited virtual testing for aerospace applications. Preliminary results using synthetically generated data are provided to demonstrate the proposed framework. Deep learning and automatic differentiation capabilities in Tensorflow were implemented on Nvidia Tesla V100 GPUs.

Best Poster Finalist: no
Poster 48: Runtime System for GPU-Based Hierarchical LU Factorization
Qianxiang Ma (Tokyo Institute of Technology), Rio Yokota (Tokyo Institute of Technology, Global Scientific Information and Computing Center; Tokyo Institute of Technology)

Hierarchical low-rank approximation can reduce both the storage and computation costs of dense matrices, but its implementation is challenging. In this research, we tackle one of the most difficult problems of GPU parallelization of the factorization of these hierarchical matrices. To this end, we are developing a new runtime system for GPUs that can schedule all tasks into one GPU kernel. Other existing runtime systems, like cuGraph and Standford Legion, can only manage streams and kernel-level parallelism. Even without too much tuning, we achieved 4x better performance in H-LU factorization with a single GPU when comparing with a well-tuned CPU-based hierarchical matrix library, HLIBpro, on moderately sized matrices. Additionally, we have significantly less runtime overheads exposed when processing smaller matrices.

Best Poster Finalist: no

Poster 145: Improving Data Compression with Deep Predictive Neural Network for Time Evolutional Data
Rupak Roy (Florida State University), Kento Sato (RIKEN Center for Computational Science (R-CCS)), Jian Guo (RIKEN Center for Computational Science (R-CCS)), Jens Domke (RIKEN Center for Computational Science (R-CCS)), Weikuan Yu (Florida State University), Takaki Hatsui (RIKEN SPring-8 Center), Yasumasa Joti (Japan Synchrotron Radiation Research Institute)

Scientific applications/simulations periodically generate huge intermediate data. Storing or transferring such a large scale of data is critical. Fast I/O is important for making this process faster. One of the approaches to achieve fast I/O is data compression. Our goal is to achieve a delta technique that can improve the performance of existing data compression algorithms for time evolutional intermediate data.

In our approach, we compute the delta values from original data and data predicted by the deep predictive neural network. We pass these delta values through three phases which are preprocessing phase, partitioned entropy coding phase, and density-based spatial delta encoding phase.

In our poster, we present how our predictive delta technique can leverage the time evolutional data to produce highly concentrated small values. We show the improvement in compression ratio when our technique, combined with existing compression algorithms, are applied on the intermediate data for
With the emergence of fast local storage, multi-level checkpointing (MLC) has become a common approach for efficient checkpointing. To utilize MLC efficiently, it is important to determine the optimal configuration for the checkpoint/restart (CR). There are mainly two approaches for determining the optimal configuration for CR, namely modeling and simulation approach. However, with MLC, CR becomes more complicated making the modeling approach inaccurate and the simulation approach though accurate, very slow. In this poster, we focus on optimizing the performance of CR by predicting the optimized checkpoint count and interval. This was achieved by combining the simulation approach with machine learning and neural network to leverage its accuracy without spending time on simulating different CR parameters. We demonstrate that our models can predict the optimized parameter values with minimal error when compared to the simulation approach.

Poster 132: Optimizing Performance at Runtime Using Binary Rewriting
Alexis Engelke (Technical University Munich), David Hildenbrand (Technical University Munich), Martin Schulz (Technical University Munich)

In addition to scalability, performance of sequential code in applications is an important factor in HPC. Typically, programs are compiled once, at which time optimizations are applied, and are then run several times. However, not all information relevant for performance optimizations are available at compile-time, restricting optimization possibilities. The generation of specialized code at runtime allows for further optimizations. Performing such specialization on binary code allows for initial code to be generated at compile-time with only the relevant parts being rewritten at runtime, reducing the optimization overhead. For targeted optimizations and effective use of known runtime information, the rewriting process needs to be guided by the application itself, exploiting information only known
to the developer.

We describe three approaches for self-guided binary rewriting explicitly guided by the running application and evaluate the performance of the optimized code as well as the performance of the rewriting process itself.

Best Poster Finalist: no

**Poster 53: Unstructured Mesh Technologies for Fusion Simulations**

Cameron Smith (Rensselaer Polytechnic Institute (RPI)), Gerrett Diamond (Rensselaer Polytechnic Institute (RPI)), Gopan Perumpilly (Rensselaer Polytechnic Institute (RPI)), Chonglin Zhang (Rensselaer Polytechnic Institute (RPI)), Agnieszka Truszkowska (Rensselaer Polytechnic Institute (RPI)), Morteza Hakimi (Rensselaer Polytechnic Institute (RPI)), Onkar Sahni (Rensselaer Polytechnic Institute (RPI)), Mark Shephard (Rensselaer Polytechnic Institute (RPI)), Eisung Yoon (Ulsan National Institute of Science and Technology, South Korea), Daniel Ibanez (Sandia National Laboratories)

Multiple unstructured mesh technologies are needed to define and execute plasma physics simulations. The domains of interest combine model features defined from physical fields within 3D CAD of the tokamak vessel with an antenna assembly, and 2D cross sections of the tokamak vessel. Mesh generation technologies must satisfy these geometric constraints and additional constraints imposed by the numerical models. Likewise, fusion simulations over these domains study a range of timescales and physical phenomena within a tokamak.

XGCm studies the development of plasma turbulence in the reactor vessel, GITRm studies impurity transport, and PetraM simulations model RF wave propagation in scrape off layer plasmas. GITRm and XGCm developments are using the PUMIp technology to manage the storage and access of non-uniform particle distributions in unstructured meshes on GPUs. PetraM combines PUMI adaptive unstructured mesh control with MFEM using CAD models and meshes defined with Simmetrix tools.

Best Poster Finalist: no

**Poster 99: Eithne: A Framework for Benchmarking Micro-Core Accelerators**

Maurice C. Jamieson (Edinburgh Parallel Computing Centre, University of Edinburgh), Nick Brown (Edinburgh Parallel Computing Centre, University of Edinburgh)

Running existing HPC benchmarks as-is on micro-core architectures is at best difficult and most often impossible as they have a number of architectural features that makes them significantly
different from traditional CPUs: tiny amounts on-chip RAM (c. 32KB), low-level knowledge specific to each device (including the host/device communications interface), limited communications bandwidth and complex or no device debugging environment. In order to compare and contrast different the micro-core architectures, a benchmark framework is required to abstract much of this complexity.

The modular Eithne framework supports the comparison of a number of micro-core architectures. The framework separates the actual benchmark from the details of how this is executed on the different technologies. The framework was evaluated by running the LINPACK benchmark on the Adapteva Epiphany, PicoRV32 and VectorBlox Orca RISC-V soft-cores, NXP RV32M1, ARM Cortex-A9, and Xilinx MicroBlaze soft-core, and comparing resulting performance and power consumption.

Best Poster Finalist: no

Poster 133: Portable Resilience with Kokkos
Jeffery Miles (Sandia National Laboratories), Nicolas Morales (Sandia National Laboratories), Carson Mould (Sandia National Laboratories), Keita Teranishi (Sandia National Laboratories)

The Kokkos ecosystem is a programming environment that provides performance and portability to many scientific applications that run on DOE supercomputers as well as other smaller scale systems. Leveraging software abstraction concepts within Kokkos, software resilience for end user code is made portable with abstractions and concepts while implementing the most efficient resilience algorithms internally. This addition enables an application to manage hardware failures reducing the cost of interruption without drastically increasing the software maintenance cost. Two main resilience methodologies have been added to the Kokkos ecosystem to validate the resilience abstractions: 1. Checkpointing includes an automatic mode supporting other checkpointing libraries and a manual mode which leverages the data abstraction and memory space concepts. 2. The redundant execution model anticipates failures by replicating data and execution paths. The design and implementation of these additions are illustrated, and appropriate examples are included to demonstrate the simplicity of use.

Best Poster Finalist: no

Poster 79: The HPC PowerStack: A Community-Wide Collaboration Toward an Energy Efficient Software Stack
Siddhartha Jana (Energy Efficient HPC Working Group, Intel Corporation), Stephanie Brink (Lawrence Livermore National Laboratory), Christopher Cantalupo (Intel Corporation), Jonathan Eastep (Intel
This poster highlights an ongoing community-wide effort among vendors, labs, and academia, to incorporate power-awareness within system-stacks in upcoming exascale machines. HPC PowerStack is the first-and-only community-driven vendor-neutral effort to identify what power optimization software actors are key within the modern-day stack; discuss their interoperability, and work toward gluing together existing open source projects to engineer cost-effective, but cohesive, portable implementations.

This poster disseminates key insights acquired in the project, provides prototyping status updates, highlights open questions, and solicits participation addressing the imminent exascale power challenge.

Best Poster Finalist: no

**Poster 87: Parallelizing Simulations of Large Quantum Circuits**

Michael A. Perlin (University of Colorado, National Institute of Standards and Technology (NIST)), Teague Tomesh (Princeton University), Bradley Pearlman (University of Colorado, National Institute of Standards and Technology (NIST)), Wei Tang (Princeton University), Yuri Alexeev (Argonne National Laboratory), Martin Suchara (Argonne National Laboratory)

We present a parallelization scheme for classical simulations of quantum circuits. Our scheme is based on a recent method to "cut" large quantum circuits into smaller sub-circuits that can be simulated independently, and whose simulation results can in turn be recombined to infer the output of the original circuit. The exponentially smaller classical computing resources needed to simulate smaller circuits are counterbalanced by exponential overhead in terms of classical post-processing costs. We discuss how this overhead can be massively parallelized to reduce classical computing costs.

Best Poster Finalist: no

**Poster 120: ILP-Based Scheduling for Linear-Tape Model Trapped-Ion Quantum Computers**
Quantum computing (QC) is emerging as a potential post-Moore high-performance computing (HPC) technology. Trapped-ion quantum bits (qubits) are among the most leading technologies to reach scalable quantum computers that would solve certain problems beyond the capabilities of even the largest classical supercomputers. In trapped-ion QC, qubits can physically move on the ion trap. The state-of-the-art architecture, linear-tape model, only requires a few laser beams to interact with the entire qubits by physically moving the interacting ions to the execution zone. Since the laser beams are limited resources, the ion chain movement and quantum gate scheduling are critical for the circuit latency. To harness the emerging architecture, we present our mathematical model for scheduling the qubit movements and quantum gates in order to minimize the circuit latency. In our experiment, our scheduling reduces 29.47% circuit latency on average. The results suggest classical HPC would further improve the quantum circuit optimization.

Best Poster Finalist: no

**Poster 55: MPI+OpenMP Parallelization of DFT Method in GAMESS**

Vladimir Mironov (Lomonosov Moscow State University), Yuri Alexeev (Argonne National Laboratory), Dmitry Fedorov (National Institute of Advanced Industrial Science and Technology (AIST))

In this work, the Density Functional Theory (DFT) method is parallelized with MPI-OpenMP in the quantum chemistry package GAMESS. It has been implemented in both regular and Fragment Molecular Orbital (FMO) based DFT codes. The scalability of the FMO-DFT code was demonstrated on Cray XC40 Theta supercomputer. We demonstrated excellent scalability of the code up 2,048 Intel Xeon Phi nodes (131,072 cores). Moreover, the developed DFT code is about twice as fast as the original code because of our new grid integration algorithm.

Best Poster Finalist: no

**Poster 71: AI-Solver: Uncertainty in Prediction and Error Estimation for AI in Engineering**

Ahmed Al-Jarro (Fujitsu Laboratories of Europe Ltd), Loic Beheshti (Fujitsu Laboratories of Europe Ltd), Serban Georgescu (Fujitsu Laboratories of Europe Ltd), Koichi Shirahata (Fujitsu Laboratories Ltd), Yasumoto Tomita (Fujitsu Laboratories Ltd), Nakashima Kouta (Fujitsu Laboratories Ltd)

The AI-Solver is a deep learning platform that learns from simulation data to extract general behavior
based on physical parameters. The AI-Solver can handle a wide variety of classes of problems including those commonly identified in FEA, CFD and CEM, to name a few, with speedups of up to 250,000X and extremely low error rate of 2-3%. In this work, we build on this recent effort. We first integrate uncertainty quantification, via exploiting the approximation of Bayesian Deep Learning. Second, we develop bespoke error estimation mechanisms capable of processing this uncertainty to provide instant feedback on the confidence in predictions without relying on the availability of ground truth data. To our knowledge, the ability to estimate the discrepancy in predictions without labels is a first in the field of AI for Engineering.

Best Poster Finalist: no

**Poster 72: Kokkos and Fortran in the Exascale Computing Project Plasma Physics Code XGC**

Aaron Scheinberg (Princeton Plasma Physics Laboratory), Guangye Chen (Los Alamos National Laboratory), Stephane Ethier (Princeton Plasma Physics Laboratory), Stuart Slattery (Oak Ridge National Laboratory), Robert Bird (Los Alamos National Laboratory), Pat Worley (PHWorley Consulting), Choong-Seock Chang (Princeton Plasma Physics Laboratory)

Numerical plasma physics models such as the particle-in-cell XGC code are important tools to understand phenomena encountered in experimental fusion devices. Adequately resolved simulations are computationally expensive, so optimization is essential. To address the need for consistent high performance by cutting-edge scientific software applications, frameworks such as Kokkos have been developed to enable portability as new architectures require hardware-specific coding implementation for best performance. Cabana, a recent extension to Kokkos developed with the ECP-CoPA project, is a library of common kernels and operations typically necessary for particle-based codes. The Kokkos/Cabana framework enables intuitive construction of particle-based codes, while maintaining portability between architectures. Here, we summarize the adoption by XGC of the execution and data layout patterns offered by this framework. We demonstrate a method for Fortran codes to adopt Kokkos and show that it can provide a single, portable code base that performs well on both GPUs and multicore machines.

Best Poster Finalist: no

**Poster 106: Optimizing Hybrid Access Virtual Memory System Using SCM/DRAM Unified Memory Management Unit**

Yusuke Shirota (Toshiba Corporation), Shiyo Yoshimura (Toshiba Corporation), Satoshi Shirai (Toshiba Corporation), Tatsunori Kanai (Toshiba Corporation)
In HPC systems, expectations for storage-class memory (SCM) are increasing in large-scale in-memory processing. While SCM can deliver higher capacity and lower standby power than DRAM, it is slower and the dynamic power is higher. Therefore, in order to realize high-speed, low-power and scalable main memory, it is necessary to build an SCM/DRAM unified memory, and dynamically optimize data placement between the two memories according to the memory access pattern.

In this poster, we describe a new hybrid access type virtual memory method using TLB-extended unified memory management unit which enables collecting and extracting fine-grained memory access locality characteristics. We show that with the proposed method, Hybrid Access control, which is a memory hierarchy control that selectively uses Direct Access to bus attached byte-addressable SCM and low power Aggressive Paging using small DRAM as cache, can be made more accurate, and the efficiency of memory access can be significantly improved.

Best Poster Finalist: no

Holistic Measurement Driven System Assessment
Saurabh Jha (University of Illinois), Mike Showerman (National Center for Supercomputing Applications (NCSA), University of Illinois), Aaron Saxton (National Center for Supercomputing Applications (NCSA), University of Illinois), Jeremy Enos (National Center for Supercomputing Applications (NCSA), University of Illinois), Greg Bauer (National Center for Supercomputing Applications (NCSA), University of Illinois), Zbigniew Kalbarczyk (University of Illinois), Ann Gentile (Sandia National Laboratories), Jim Brandt (Sandia National Laboratories), Ravi Iyer (University of Illinois), William T. Kramer (University of Illinois, National Center for Supercomputing Applications (NCSA))

HPC users deploy a suite of monitors to observe patterns of failures and performance anomalies to improve operational efficiency, achieve higher application performance and inform the design of future systems. However, the promises and the potential of monitoring data have largely been not realized due to various challenges such as inadequacy in monitoring, limited availability of data, lack of methods for fusing monitoring data at time-scales necessary for enabling human-in-the-loop or machine-in-the-loop feedback. To address above challenges, in this work we developed a monitoring fabric Holistic Measurement Driven System Assessment (HMDSA) for large-scale HPC facilities, independent of major component vendor, and within budget constraints of money, space, and power. We accomplish this through development and deployment of scalable, platform-independent, open-source tools and techniques for monitoring, coupled with statistical and machine-learning based runtime analysis and feedback, which enables highly efficient HPC system operation and usage and also informs future system improvements.
Poster 139: Model Identification of Pressure Drop in Membrane Channels with Multilayer Artificial Neural Networks

Jiang-hang Gu (Sun Yat-sen University, Zhuhai, School of Chemical Engineering and Technology), Jiu Luo (Sun Yat-sen University, Guangzhou, School of Materials Science and Engineering), Ming-heng Li (California State Polytechnic University, Pomona), Yi Heng (Sun Yat-sen University, Guangzhou, School of Data and Computer Science; Sun Yat-sen University, Guangzhou, China)

This poster presents the work of identifying a data-driven model of pressure drop in spacer-filled reverse osmosis membrane channels and conducting CFD simulations. The established model correlates the pressure drop with a wide range of design objectives, which enables a quantitative description of the geometric structures and operation conditions for improvement. This way, it aims at optimizing the spacer geometry with minimal effort. Furthermore, a high-performance computing strategy is employed to tackle the resulted intractable computational task in the identification procedure and CFD simulations.

Best Poster Finalist: no

Poster 61: Fast 3D Diffeomorphic Image Registration on GPUs

Malte Brunn (University of Stuttgart), Naveen Himthani (University of Texas), George Biros (University of Texas), Miriam Mehl (University of Stuttgart), Andreas Mang (University of Houston)

3D image registration is one of the most fundamental and computationally expensive operations in medical image analysis. Here, we present a mixed-precision, Gauss-Newton-Krylov solver for diffeomorphic registration. Our work extends the publicly available CLAIRE library to GPU architectures. Despite the importance of image registration, only a few implementations of large deformation diffeomorphic registration packages support GPUs. Our contributions are new algorithms and dedicated computational kernels to significantly reduce the runtime of the main computational kernels in CLAIRE: derivatives and interpolation. We deploy (i) highly-optimized, mixed-precision GPU-kernels for the evaluation of scattered-data interpolation, (ii) replace FFT-based first-order derivatives with optimized 8th-order finite differences, and (iii) compare with state-of-the-art CPU and GPU implementations. As a highlight, we demonstrate that we can register 256^3 clinical images in less than 6 seconds on a single NVIDIA Tesla V100. This amounts to over 20x speed-up over CLAIRE and over 30x speed-up over existing GPU implementations.

Best Poster Finalist: no
Poster 138: Across-Stack Profiling and Characterization of State-of-the-Art Machine Learning Models on GPUs
Cheng Li (University of Illinois), Abdul Dakkak (University of Illinois), Wei Wei (Alibaba Inc), Jinjun Xiong (IBM Research), Lingjie Xu (Alibaba Inc), Wei Zhang (Alibaba Inc), Wen-mei Hwu (University of Illinois)

The past few years have seen a surge of using Machine Learning (ML) and Deep Learning (DL) algorithms for traditional HPC tasks such as feature detection, numerical analysis, and graph analytics. While ML and DL enable solving HPC tasks, their adoption has been hampered due to the lack of understanding of how they utilize systems. Optimizing these algorithms requires characterizing their performance across the hardware/software (HW/SW) stack, but the lack of simple tools to automate the process and the reliance on researchers to perform manual characterization is a bottleneck. To alleviate this, we propose an across-stack profiling scheme and integrate it within MLModelScope — a hardware and software agnostic tool for evaluating and benchmarking ML/DL at scale. We demonstrate MLModelScope’s ability to characterize state-of-art ML/DL models and give insights that are only possible obtained by performing across-stack profiling.

Best Poster Finalist: no

Poster 60: Massively Parallel Large-Scale Multi-Model Simulation of Tumor Development
Marco Berghoff (Karlsruhe Institute of Technology), Jakob Rosenbauer (Forschungszentrum Juelich), Alexander Schug (Forschungszentrum Juelich)

The temporal and spatial resolution in the microscopy of tissues has increased significantly within the last years, yielding new insights into the dynamics of tissue development and the role of the single-cell within it. A thorough theoretical description of the connection of single-cell processes to macroscopic tissue reorganizations is still lacking. Especially in tumor development, single cells play a crucial role in advance of tumor properties.

We developed a simulation framework that can model tissue development up to the centimeter scale with micrometer resolution of single cells. Through a full parallelization, it enables the efficient use of HPC systems, therefore enabling detailed simulations on a large scale. We developed a generalized tumor model that respects adhesion driven cell migration, cell-to-cell signaling, and mutation-driven tumor heterogeneity. We scan the response of the tumor development depending on division inhibiting substances such as cytostatic agents.
**Poster 114: Optimizing Recommendation System Inference Performance Based on GPU**

Xiaowei Shen (Alibaba Inc), Junrui Zhou (Alibaba Inc), Kan Liu (Alibaba Inc), Lingling Jin (Alibaba Inc), Pengfei Fan (Alibaba Inc), Wei Zhang (Alibaba Inc), Jun Yang (University of Pittsburgh)

Neural network-based recommendation models have been widely applied on tracking personalization and recommendation tasks at large Internet companies such as e-commerce companies and social media companies. Alibaba recommendation system deploys WDL (wide and deep learning) models for product recommendation tasks. The WDL model consists of two main parts: embedding lookup and neural network-based feature ranking model that ranks different products for different users. As more and more products and users the model need to rank, the feature length and batch size of the models are increased. The computation of models is also increased so that traditional model inference implementation on CPU cannot meet the requirement of QPS (query per second) and latency of recommendation tasks. In this poster, we develop a GPU based system to speedup recommendation system inference performance. By model quantization and graph transformation, we can achieve 3.9x performance speedup when compared with a baseline GPU implementation.

**Poster 59: Accelerating BFS and SSSP on a NUMA Machine for the Graph500 Challenge**

Tanuj K. Aasawat (RIKEN), Kazuki Yoshizoe (RIKEN), Tahsin Reza (University of British Columbia), Matei Ripeanu (University of British Columbia)

The NUMA architecture is the design choice for modern multi-CPU shared memory systems. In many ways, a NUMA system resembles a shared-nothing distributed system: memory accesses to remote NUMA domains are more expensive than local accesses.

In this work, we explore how improved data locality and reduced expensive remote communication can be achieved by exploiting "distributed" shared-memory of NUMA machines to develop shared-memory graph processing solutions optimized for NUMA systems. We introduce a novel hybrid design for memory accesses that handles the burst mode in traversal based algorithms, like BFS and SSSP, and reduces the number of remote accesses and updates. We demonstrate that our designs offer up to 84% speedup over our NUMA-oblivious framework Totem and 2.86x over shared-nothing distributed design, for BFS and SSSP algorithms.
**Poster 47: Decomposition Algorithms for Scalable Quantum Annealing**

Elijah Pelofske (Los Alamos National Laboratory), Georg Hahn (Harvard University), Hristo Djidjev (Los Alamos National Laboratory)

Commercial adiabatic quantum annealers such as D-Wave 2000Q have the potential to solve NP-complete optimization problems efficiently. One of the primary constraints of such devices is the limited number and connectivity of their qubits. This research presents two exact decomposition methods (for the Maximum Clique and the Minimum Vertex Cover problem) that allow us to solve problems of arbitrarily large sizes by splitting them up recursively into a series of arbitrarily small subproblems. Those subproblems are then solved exactly or approximately using a quantum annealer. Whereas some previous approaches are based on heuristics that do not guarantee optimality of their solutions, our decomposition algorithms have the property that the optimal solution of the input problem can be reconstructed given all generated subproblems are solved optimally as well. We investigate various heuristic and exact bounds as well as reduction methods that help to increase the scalability of our approaches.

Best Poster Finalist: no

**Poster 92: Nanoporous Flow Simulations on the Summit Supercomputer**

Yidong Xia (Idaho National Laboratory), Lixiang Luo (IBM - TJ Watson Research Center), Ansel Blumers (Brown University), Joshua Kane (Idaho National Laboratory), Jan Goral (University of Utah), Yu-Hang Tang (Lawrence Berkeley National Laboratory), Zhen Li (Clemson University, Brown University), Hai Huang (Idaho National Laboratory), Milind Deo (University of Utah)

Fluid flow behaviors in nanoporous materials are distinct from those following the continuum physics. Numerical simulations can be a complement to laboratory experiments. This work presents a dissipative particle dynamics (DPD) package for GPU-accelerated mesoscale flow simulations in nanoporous materials. In an ideal benchmark that minimizes load imbalance, the package delivered nearly perfect strong- and weak-scaling (with up to 4 billion DPD particles) on up to 1,536 V100 GPUs on Oak Ridge National Laboratory’s Summit supercomputer. More remarkably, in a benchmark to measure its usefulness with realistic nanopores in SBA-15 silica, the package exhibited more than 20x speedup over its LAMMPS-based CPU counterpart with the same number nodes (e.g., 384 V100 GPUs vs. 2,688 POWER9 cores). It is worth highlighting that the NVLink2 Host-to-Device interconnects kept the cost of CPU-GPU memory copy as low as only 10% of GPU activity time per rank: 4 times less than their PCIe counterparts.
**Poster 90: You Have to Break It to Make It: How On-Demand, Ephemeral Public Cloud Projects with Alces Flight Compute Resulted in the Open-Source OpenFlightHPC Project**

*Cristin Merritt (Alces Flight Limited; Alces Software Ltd, UK), Wil Mayers (Alces Flight Limited), Stu Franks (Alces Flight Limited)*

Over three years ago the Alces Flight team made a decision to explore on-demand public cloud consumption for High Performance Computing (HPC). Our premise was simple, create a fully-featured, scalable HPC environment for research and scientific computing and provide it as a free subscription service to researchers in order to study how they would approach and consume this then new technology. This tool, Alces Flight Compute, would set out to chart how far away from the traditional bare-metal platforms our subscribers were willing to go. What we didn’t expect was that to get to their destination, our users would proceed to take our tool apart. This deconstruction has resulted in a new approach to HPC environment creation (the open-source OpenFlightHPC project), helped us better understand cloud adoption strategies, and handed over a set of guidelines to help those looking to bring public cloud into their HPC solution.

**Poster 126: Enforcing Crash Consistency of Scientific Applications in Non-Volatile Main Memory Systems**

*Tyler Coy (Washington State University, Vancouver), Xuechen Zhang (Washington State University, Vancouver)*

This poster presents a compiler-assistant technique, NVPath, to automatically generates NVMM-aware persistent data structures which provide the same level of guarantee of crash consistency compared to the baseline code. Compiler-assistant code annotation and transformation is general and can be applied to applications using various data structures. Our experimental results with real-world scientific applications show that the performance of the annotated programs is commensurate with the version using the manual code transformation on the Titan supercomputer.

**Poster 137: Warwick Data Store: A HPC Library for Flexible Data Storage in Multi-Physics Applications**
Richard O. Kirk (University of Warwick), Timothy R. Law (Atomic Weapons Establishment (AWE), UK), Satheesh Maheswaran (Atomic Weapons Establishment (AWE), UK), Stephen A. Jarvis (University of Warwick)

With the increasing complexity of memory architectures and multi-physics applications, developing data structures that are performant, portable, scalable, and support developer productivity, is difficult. In order to manage these complexities and allow rapid prototyping of different approaches we are building a lightweight and extensible C++ template library called the Warwick Data Store (WDS). WDS is designed to abstract details of the data structure away from the user, thus easing application development and optimisation. We show that WDS generates minimal performance overhead, via a variety of different scientific benchmarks and proxy-applications.

Best Poster Finalist: no

Poster 76: HPChain: An MPI-Based Blockchain Framework for High Performance Computing Systems
Abdullah Al-Mamun (University of Nevada, Reno; Lawrence Berkeley National Laboratory), Tonglin Li (Lawrence Berkeley National Laboratory), Mohammad Sadoghi (University of California, Davis), Linhua Jiang (Fudan University, Shanghai), Haoting Shen (University of Nevada, Reno), Dongfang Zhao (University of Nevada, Reno; University of California, Davis)

Data fidelity is of prominent importance for scientific experiments and simulations. The state-of-the-art mechanism to ensure data fidelity is through data provenance. However, the provenance data itself may as well exhibit unintentional human errors and malicious data manipulation. To enable a trustworthy and reliable data fidelity service, we advocate achieving the immutability and decentralization of scientific data provenance through blockchains. Specifically, we propose HPChain, a new blockchain framework specially designed for HPC systems. HPChain employs a new consensus protocol compatible with and optimized for HPC systems. Furthermore, HPChain was implemented with MPI and integrated with an off-chain distributed provenance service to tolerate the failures caused by faulty MPI ranks. The HPChain prototype system has been deployed to 500 cores at the University of Nevada’s HPC center and demonstrated strong resilience and scalability while outperforming state-of-the-art blockchains by orders of magnitude.

Best Poster Finalist: no

Poster 104: An Adaptive Checkpoint Model For Large-Scale HPC Systems
Subhendu S. Behera (North Carolina State University), Lipeng Wan (Oak Ridge National Laboratory),
Checkpoint/Restart is a widely used Fault Tolerance technique for application resilience. However, failures and the overhead of saving application state for future recovery upon failure reduces the application efficiency significantly. This work contributes a failure analysis and prediction model making decisions for checkpoint data placement, recovery, and techniques for reducing checkpoint frequency. We also demonstrate a reduction in application overhead by taking proactive measures guided by failure prediction.

Best Poster Finalist: no

**Poster 123: Cloud-Native SmartX Intelligence Cluster for AI-Inspired HPC/HPDA Workloads**
Jungsu Han (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science), Jun-Sik Shin (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science), JinCheol Kwon (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science), JongWon Kim (Gwangju Institute of Science and Technology, Electrical Engineering and Computer Science)

In this poster, we introduce Cloud-native SmartX Intelligence Cluster for flexibly supporting AI-inspired HPC (high performance computing) / HPDA (high performance data analytics) workloads. This work has been continuously refined from 2013 with a futuristic vision for operating 100 petascale data center. Then, we discuss issues and approaches that come with building a Cloud-native SmartX Intelligence Cluster.

Best Poster Finalist: no

**Poster 86: High-Performance Custom Computing with FPGA Cluster as an Off-Loading Engine**
Takaaki Miyajima (RIKEN Center for Computational Science (R-CCS)), Tomohiro Ueno (RIKEN Center for Computational Science (R-CCS)), Jens Huthmann (RIKEN Center for Computational Science (R-CCS)), Atsushi Koshiba (RIKEN Center for Computational Science (R-CCS)), Kentaro Sano (RIKEN Center for Computational Science (R-CCS)), Mitsuhisa Sato (RIKEN Center for Computational Science (R-CCS))

A heterogeneous system with Field Programmable Gate Arrays (FPGAs) is gathering attention in High-Performance Computing (HPC) area. We have been researching and developing a high-performance custom computing with FPGA cluster as an off-loading engine for an existing HPC
system. In this research poster, we describe the motivation of our research and present research topics on a software bridge between the FPGA cluster and existing HPC servers, and dedicated inter-FPGA networks.

Best Poster Finalist: no

**Poster 69: Optimization for Quantum Computer Simulation**

Naoki Yoshioka (RIKEN Center for Computational Science (R-CCS)), Hajime Inaoka (RIKEN Center for Computational Science (R-CCS)), Nobuyasu Ito (RIKEN Center for Computational Science (R-CCS)), Fengping Jin (Forschungszentrum Juelich), Kristel Michielsen (Forschungszentrum Juelich), Hans De Raedt (University of Groningen)

Simulator of quantum circuits is developed for massively parallel classical computers, and it is tested on the K computer in RIKEN R-CCS up to 45 qubits. Two optimization techniques are proposed in order to improve performance of the simulator. The "page method" reduces unnecessary copies in each node. It is found that this method makes approximately 17% speed-up maximum. Initial permutation of qubits is also studied how it affects performance of the simulator. It is found that a simple permutation in ascending order of the number of operations for each qubit is sufficient in the case of simulations of quantum adder circuits.

Best Poster Finalist: no

**Poster 110: Hierarchical Data Prefetching in Multi-Tiered Storage Environments**

Hariharan Devarajan (Illinois Institute of Technology), Anthony Koukas (Illinois Institute of Technology), Xian-He Sun (Illinois Institute of Technology)

In the era of data-intensive computing, accessing data with a high-throughput and low-latency is very imperative. Data prefetching is used for hiding read latency by requesting data before it is needed to move it from a high-latency medium to a low-latency one. However, existing solutions do not consider the multi-tiered storage and also suffer from under-utilization of prefetching resources and unnecessary evictions. Additionally, existing approaches implement a client-pull model where understanding the application's I/O behavior drives prefetching decisions. Moving toward exascale, where machines run multiple applications concurrently by accessing files in a workflow, a more data-centric approach resolves challenges such as cache pollution and redundancy. In this study, we present HFetch, a truly hierarchical data prefetcher that adopts a server-push approach to data prefetching. We demonstrate the benefits of such an approach. Results show 10-35% performance gains over existing prefetchers and over 50% when compared to systems with no prefetching.
Best Poster Finalist: yes

**Poster 52: Design and Specification of Large-Scale Simulations for GPUs Using FFTX**

Anuva Kulkarni (Carnegie Mellon University), Daniele Spampinato (Carnegie Mellon University), Franz Franchetti (Carnegie Mellon University)

Large-scale scientific simulations can be ported to heterogeneous environments with GPUs using domain decomposition. However, Fast Fourier Transform (FFT) based simulations require all-to-all communication and large memory, which is beyond the capacity of on-chip GPU memory. To overcome this, domain decomposition solutions are combined with adaptive sampling or pruning around the domain to reduce storage. Expression of such operations is a challenge in existing FFT libraries like FFTW, and thus it is difficult to get a high performance implementation of such methods. We demonstrate algorithm specification for one such simulation (Hooke’s law) using FFTX, an emerging API with a SPIRAL-based code generation back-end, and suggest future extensions useful for GPU-based scientific computing.

Best Poster Finalist: no

**Poster 136: CHAMELEON: Reactive Load Balancing and Migratable Tasks for Hybrid MPI+OpenMP Applications**

Jannis Klinkenberg (RWTH Aachen University), Philipp Samfaß (Technical University Munich), Michael Bader (Technical University Munich), Karl Fürlinger (Ludwig Maximilian University of Munich), Christian Terboven (RWTH Aachen University), Matthias S. Müller (RWTH Aachen University)

Many HPC applications are designed based on underlying performance and execution models. These models could successfully be employed in the past for balancing load within and between compute nodes. However, the increasing complexity of modern software and hardware makes performance predictability and load balancing much more difficult. Tackling these challenges in search for a generic solution, we present a novel library for fine-granular task-based reactive load balancing in distributed memory based on MPI and OpenMP. Our concept allows creating individual migratable tasks that can be executed on any MPI rank. Migration decisions are performed at run time based on online performance or load data. Two fundamental approaches to balance load and at the same time overlap computation and communication are compared. We evaluate our concept under enforced power caps and clock frequency changes using a synthetic benchmark and demonstrate robustness against work-induced imbalances for an AMR application.
Poster 124: Porting Finite State Automata Traversal from GPU to FPGA: Exploring the Implementation Space
Marziyeh Nourian (North Carolina State University), Mostafa Eghbali Zarch (North Carolina State University), Michela Becchi (North Carolina State University)

While FPGAs are traditionally considered hard to program, recently there are efforts to allow using high-level programming models intended for multi-core CPUs and GPUs to program FPGAs. For example, both Intel and Xilinx are now providing OpenCL-to-FPGA toolchains. However, since GPU and FPGA devices offer different parallelism models, OpenCL code optimized for GPU can prove inefficient on FPGA, in terms of both performance and hardware resource utilization.

In this poster, we explore this problem on an emerging workload: finite state automata traversal. Specifically, we explore a set of structural code changes, custom, and best-practice optimizations to retarget an OpenCL NFA engine designed for GPU to FPGA. Our evaluation, which covers traversal throughput and resource utilization, shows that our optimizations lead, on a single execution pipeline, to speedups up to 4x over an already optimized baseline that uses one of the proposed code changes to fit the original code on FPGA.

Best Poster Finalist: no

Poster 68: Linking a Next-Gen Remap Library into a Long-Lived Production Code
Charles R. Ferenbaugh (Los Alamos National Laboratory), Brendan K. Krueger (Los Alamos National Laboratory)

LANL’s long-lived production application xRage contains a remapper capability that maps mesh fields from its native AMR mesh to the GEM mesh format used by some third-party libraries. The current remapper was implemented in a short timeframe and is challenging to maintain. Meanwhile, our next-generation code project has developed a modern remapping library Portage, and the xRage team wanted to link in Portage as an alternate mapper option. But the two codes are very different from each other, and connecting the two required us to deal with a number of challenges. This poster describes the codes, the challenges we worked through, current status, and some initial performance statistics.

Best Poster Finalist: no
**Poster 131: Efficiency of Algorithmic Structures**  
*Julian Miller (RWTH Aachen University), Lukas Trümper (RWTH Aachen University), Christian Terboven (RWTH Aachen University), Matthias S. Müller (RWTH Aachen University)*

The implementation of high-performance parallel software is challenging and raises issues not seen in serial programs before. It requires a strategy of parallel execution which preserves correctness but maximizes scalability. Efficiently deriving well-scaling solutions remains an unsolved problem especially with the quickly-evolving hardware landscape of high-performance computing (HPC).

This work proposes a framework for classifying the efficiency of parallel programs. It bases on a strict separation between the algorithmic structure of a program and its executed functions. By decomposing parallel programs into a hierarchical structure of parallel patterns, a high-level abstraction is provided which leads to equivalence classes over parallel programs. Each equivalence class possesses efficiency properties, mainly communication and synchronization, dataflow and architecture efficiency. This classification allows for wide application areas and a workflow for structural optimization of parallel algorithms is proposed.

Best Poster Finalist: no

**Poster 98: INSPECT Intranode Stencil Performance Evaluation Collection**  
*Julian Hammer (University of Erlangen-Nuremberg), Julian Hornich (University of Erlangen-Nuremberg), Georg Hager (University of Erlangen-Nuremberg), Thomas Gruber (University of Erlangen-Nuremberg), Gerhard Wellein (University of Erlangen-Nuremberg)*

Modeling and presenting performance data---even for simple kernels such as stencils---is not trivial. We therefore present an overview on how to interpret and what to learn from an INSPECT report, as well as highlighting best practices for performance data reporting.

INSPECT is the "Intranode Stencil Performance Evaluation Collection", which compiles performance benchmarks and reports of various stencil and streaming kernels on a variety of architectures. The goal is to aid performance-aware developers with reference material and a methodology to analyze their own codes.

INSPECT set out to cover these topics and compile a summary of all necessary information to allow reproduction of the performance results, their interpretation and discussion.

Best Poster Finalist: no
Poster 97: Optimizing Multigrid Poisson Solver of Cartesian CFD Code CUBE
Kazuto Ando (RIKEN Center for Computational Science (R-CCS)), Rahul Bale (RIKEN), Keiji Onishi (RIKEN Center for Computational Science (R-CCS)), Kiyoshi Kumahata (RIKEN Center for Computational Science (R-CCS)), Kazuo Minami (RIKEN Center for Computational Science (R-CCS)), Makoto Tsubokura (Kobe University, RIKEN Center for Computational Science (R-CCS))

We demonstrate an optimization of multigrid Poisson solver of Cartesian CFD code “CUBE (Complex Unified Building cubE method)”. CUBE is a simulation framework for complex industrial flow problem, such as aerodynamics of vehicles, based on hierarchical Cartesian mesh. In incompressible CFD simulation, solving pressure Poisson equation is the most time-consuming part. In this study, we use a cavity flow simulation as a benchmark problem. With this problem, multigrid Poisson solver dominates 91% of execution time of the time-step loop. Specifically, we evaluate the performance of Gauss-Seidel loop as a computational kernel based on “Byte per Flop” approach. With optimization of the kernel, we achieved 9.8x speedup and peak floating point performance ratio increased from 0.4% to 4.0%. We also measured parallel performance up to 8,192 nodes (65,536 cores) on the K computer. With optimization of the parallel performance, we achieved 2.9x–3.9x sustainable speedup in the time-step loop.

Best Poster Finalist: no

Poster 85: Hybrid Computing Platform for Combinatorial Optimization with the Coherent Ising Machine
Junya Arai (Nippon Telegraph and Telephone Corporation), Yagi Satoshi (Nippon Telegraph and Telephone Corporation), Hirohisa Uchiyama (Nippon Telegraph and Telephone Corporation), Toshimori Honjo (Nippon Telegraph and Telephone Corporation), Takahiro Inagaki (Nippon Telegraph and Telephone Corporation), Takuya Ikuta (Nippon Telegraph and Telephone Corporation), Hiroyuki Takesue (Nippon Telegraph and Telephone Corporation), Keitaro Horikawa (Nippon Telegraph and Telephone Corporation)

Several institutes are operating cloud platforms that offer Web API access to Ising computers such as quantum annealing machines. Platform users can solve complex combinatorial optimization problems by using hybrid algorithms that utilize both users’ conventional digital computers and remote Ising computers. However, communication via the Internet takes an order of magnitude longer time than optimization on Ising computers. This overheads seriously degrade the performance of hybrid algorithms since they involve frequent communication. In this poster, we first state issues in the design of Ising computing platforms, including communication overheads. Then, we answer the issues by introducing the computing platform for the coherent Ising machine (CIM), an Ising computer
based on photonics technologies. Our platform offers efficient CIM-digital communication by allowing users to execute their program on digital computers co-located with the CIM. We have released the platform to our research collaborators in this autumn and started the evaluation.

Best Poster Finalist: no

**Poster 117: A New Polymorphic Computing Architecture Based on Fine-Grained Instruction Mobility**

David Hentrich (Illinois Institute of Technology), Erdal Oruklu (Illinois Institute of Technology), Jafar Saniie (Illinois Institute of Technology)

This is a summary of the base concepts behind David Hentrich’s May 2018 Ph.D. dissertation in Polymorphic Computing. Polymorphic Computing is the emerging field of changing the computer architecture around the software, rather than vice versa. The main contribution is a new polymorphic computing architecture. The key idea behind the architecture is to create an array of processors where a program’s instructions can be individually and arbitrarily assigned/mobilized to any processor, even during runtime. The key enablers of this architecture are a dataflow instruction set that is conducive to instruction migration, a microarchitectural block called an “operation cell” (“op-cell”), a processor built around the instruction set and the “op-cells”, and arrays of these processors.

Best Poster Finalist: no

**Poster 67: Genie: an MPEG-G Conformant Software to Compress Genomic Data.**

Brian E. Bliss (University of Illinois), Joshua M. Allen (University of Illinois), Saurabh Baheti (Mayo Clinic), Matthew A. Bockol (Mayo Clinic), Shubham Chandak (Stanford University), Jaime Delgado (Polytechnic University of Catalonia), Jan Fostier (Ghent University), Josep L. Gelpi (University of Barcelona), Steven N. Hart (Mayo Clinic), Mikel Hernaez Arrazola (University of Illinois), Matthew E. Hudson (University of Illinois), Michael T. Kalmbach (Mayo Clinic), Eric W. Klee (Mayo Clinic), Liudmila S. Mainzer (University of Illinois), Fabian Müntefering (Leibniz University), Daniel Naro (Barcelona Supercomputing Center), Idoia Ochoa-Alvarez (University of Illinois), Jörn Ostermann (Leibniz University), Tom Paridaens (Ghent University), Christian A. Ross (Mayo Clinic), Jan Voges (Leibniz University), Eric D. Wieben (Mayo Clinic), Mingyu Yang (University of Illinois), Tsachy Weissman (Stanford University), Mathieu Wiepert (Mayo Clinic)

Precision medicine has unprecedented potential for accurate diagnosis and effective treatment. It is supported by an explosion of genomic data, which continues to accumulate at accelerated pace. Yet storage and analysis of petascale genomic data is expensive, and that cost will ultimately be borne by
the patients and citizens. The Moving Picture Experts Group (MPEG) has developed MPEG-G, a new open standard to compress, store, transmit and process genomic sequencing data that provides an evolved and superior alternative to currently used genomic file formats. Our poster will showcase software package GENIE, the first open source implementation of an encoder-decoder pair that is compliant with the MPEG-G specifications and delivers all its benefits: efficient compression, selective access, transport and analysis, guarantee of long-term support, and embedded mechanisms for annotation and encryption of compressed information. GENIE will create a step-change in medical genomics by reducing the cost of data storage and analysis.

Best Poster Finalist: no

**Poster 82: A View from the Facility Operations Side on the Water/Air Cooling System of the K Computer**

Jorji Nonaka (RIKEN Center for Computational Science (R-CCS)), Keiji Yamamoto (RIKEN Center for Computational Science (R-CCS)), Akiyoshi Kuroda (RIKEN Center for Computational Science (R-CCS)), Toshiyuki Tsukamoto (RIKEN Center for Computational Science (R-CCS)), Kazuki Koiso (Kobe University, RIKEN Center for Computational Science (R-CCS)), Naohisa Sakamoto (Kobe University, RIKEN Center for Computational Science (R-CCS))

The Operations and Computer Technologies Division at the RIKEN R-CCS is responsible for the operations of the entire K computer facility, which includes the auxiliary subsystems such as the power supply and water/air cooling systems. It is worth noting that part of these subsystems will be reused in the next supercomputer (Fugaku), thus a better understanding of the operational behavior as well as the potential impacts especially on the hardware failure and energy consumption would be greatly beneficial. In this poster, we will present some preliminary impressions of the impact of the water/air cooling system on the K computer system, focusing on the potential benefits of the use of low water/air temperature respectively for the CPU and DRAM memory modules produced by the cooling system. We expect that the obtained knowledge will be helpful for the decision support and/or operation planning of the next supercomputer.

Best Poster Finalist: no

**Poster 135: High-Performance Deep Learning via a Single Building Block**

Evangelos Georganas (Intel Corporation), Kunal Banerjee (Intel Corporation), Dhiraj Kalamkar (Intel Corporation), Sasikanth Avancha (Intel Corporation), Anand Venkat (Intel Corporation), Michael Anderson (Intel Corporation), Greg Henry (Intel Corporation), Hans Pabst (Intel Corporation), Alexander Heinecke (Intel Corporation)
Deep learning (DL) is one of the most prominent branches of machine learning. Due to the immense computational cost of DL workloads, industry and academia have developed DL libraries with highly-specialized kernels for each workload/architecture, leading to numerous, complex code-bases that strive for performance, yet they are hard to maintain and do not generalize. In this work, we introduce the batch-reduce-GEMM kernel and show how the most popular DL algorithms can be formulated with this kernel as basic building-block. Consequently, the DL library-development degenerates to mere (potentially automatic) tuning of loops around this sole optimized kernel. By exploiting our kernel we implement Recurrent Neural Networks, Convolution Neural Networks and Multilayer Perceptron training and inference primitives in just 3K lines of high-level-code. Our primitives outperform vendor-optimized libraries on multi-node CPU-Clusters. We also provide CNN kernels targeting GPUs. Finally, we demonstrate that batch-reduce-GEMM kernel within a tensor compiler yields high-performance CNN primitives.

Best Poster Finalist: no

Poster 56: Reinforcement Learning for Quantum Approximate Optimization
Sami Khairy (Illinois Institute of Technology), Ruslan Shaydulin (Clemson University), Lukasz Cincio (Los Alamos National Laboratory), Yuri Alexeev (Argonne National Laboratory), Prasanna Balaprakash (Argonne National Laboratory)

The Quantum Approximate Optimization Algorithm (QAOA) is one of the leading candidates for demonstrating quantum advantage. The quality of the solution obtained by QAOA depends on the performance of the classical optimization routine used to optimize the variational parameters. In this work, we propose a Reinforcement Learning (RL) based approach to drastically reduce the number of evaluations needed to find high-quality variational parameters. We train an RL agent on small 8-qubit Max-Cut problem instances on an Intel Xeon Phi supercomputer Bebop, and use (transfer) the learned optimization policy to quickly find high-quality solutions for other larger problem instances coming from different distributions and graph classes. The preliminary results show that our RL based approach is able to improve the quality of the obtained solution by up to 10% within a fixed budget of function evaluations and demonstrate learned optimization policy transferability between different graph classes and sizes.

Best Poster Finalist: no

John Shalf (Lawrence Berkeley National Laboratory), Dilip Vasudevan (Lawrence Berkeley National Laboratory), David Donofrio (Lawrence Berkeley National Laboratory), Anastasia Butko (Lawrence Berkeley National Laboratory), Andrew Chien (University of Chicago), Yuanwei Fang (University of Chicago), Arjun Rawal (University of Chicago), Chen Zou (University of Chicago), Raymond Bair (Argonne National Laboratory), Kristopher Keipert (Argonne National Laboratory), Arun Rodriguez (Sandia National Laboratories), Maya Gokhale (Lawrence Livermore National Laboratory), Scott Lloyd (Lawrence Livermore National Laboratory), Xiaochen Guo (Lehigh University), Yuan Zeng (Lehigh University)

Accelerating technology disruptions and architectural change create growing opportunities and urgency to reduce the latency in for new architectural innovations to be deployed in extreme scale systems. We are exploring new architectural features that improve memory system performance including word-wise scratchpad memory, a flexible Recode engine, hardware message queues, and the data rearrangement engine (DRE). Performance results are promising yielding as much as 20x benefit. Project 38 is a cross-agency effort undertaken by the US Department of Energy (DOE) and Department of Defense (DoD).

Best Poster Finalist: no

Poster 128: Identifying Time Series Similarity in Large-Scale Earth System Datasets
Payton Linton (Youngstown State University), William Melodia (Youngstown State University), Alina Lazar (Youngstown State University), Deborah Agarwal (Lawrence Berkeley National Laboratory), Ludovico Bianchi (Lawrence Berkeley National Laboratory), Devarshi Ghoshal (Lawrence Berkeley National Laboratory), Kesheng Wu (Lawrence Berkeley National Laboratory), Gilberto Pastorello (Lawrence Berkeley National Laboratory), Lavanya Ramakrishnan (Lawrence Berkeley National Laboratory)

Scientific data volumes are growing every day and instrument configurations, quality control and software updates result in changes to the data. This study focuses on developing algorithms that detect changes in time series datasets in the context of the Deduce project. We propose a combination of methods that include dimensionality reduction and clustering to evaluate similarity measuring algorithms. This methodology can be used to discover existing patterns and correlations within a dataset. The current results indicate that the Euclidean Distance metric provides the best results in terms of internal cluster validity measures for multi-variable analyses of large-scale earth system datasets. The poster will include details on our methodology, results, and future work.

Best Poster Finalist: no
**Poster 151: Three-Dimensional Characterization on Edge AI Processors with Object Detection Workloads**

Yujie Hui (Ohio State University), Jeffrey Lien (NovuMind Inc), Xiaoyi Lu (Ohio State University)

The Deep Learning inference applications are moving to the edge side, as edge-side AI platforms are cheap and energy-efficient. Different edge AI processors are diversified, since these processors are designed with different approaches. However, it is hard for customers to select an edge AI processor without an overall evaluation of these processors. We propose a three-dimensional characterization (i.e., accuracy, latency, and energy efficiency) approach on three different kinds of edge AI processors (i.e., Edge TPU, NVIDIA Xavier, and NovuTensor). We deploy YOLOv2 and Tiny-YOLO, which are two YOLO-based object detection systems, on these edge AI platforms with Microsoft COCO dataset. I will present our work starting from the problem statement. And then I'll introduce our experiments setup and hardware configuration. Lastly, I'll conclude our experimental results and current work status, as well as the future work.

Best Poster Finalist: no

**Poster 148: Unsupervised Clustering of Golden Eagle Telemetry Data**


We use a recurrent autoencoder neural network to encode sequential California golden eagle telemetry data. The encoding is followed by an unsupervised clustering technique, Deep Embedded Clustering (DEC), to iteratively cluster the data into a chosen number of behavior classes. We apply the method to simulated movement data sets and telemetry data for a Golden Eagle. The DEC achieves better unsupervised clustering accuracy scores for the simulated data sets as compared to the baseline K-means clustering result.

Best Poster Finalist: no

**Poster 66: Hybrid CPU/GPU FE2 Multi-Scale Implementation Coupling Alya and Microppt**

Guido Giuntoli (Barcelona Supercomputing Center), Judicaël Grasset (Science and Technology Facilities Council (STFC)), Alejandro Figueroa (George Mason University), Charles Moulinec (Science and Technology Facilities Council (STFC)), Mariano Vázquez (Barcelona Supercomputing Center), Guillaume Houzeaux (Barcelona Supercomputing Center), Stephen Longshaw (Science and Technology Facilities Council (STFC)), Sergio Oller (Polytechnic University of Catalonia)
This poster exposes the results of a new implementation of the FE2 multi-scale algorithm that is achieved by coupling the multi-physics and massively parallel code Alya with the GPU-based code micropp. The coupled code is mainly designed to solve large scale and realistic composite material problems for the aircraft industry. Alya is responsible of solving the macro-scale equations and micropp for solving the representation of fibres at the microscopic level. The poster shows computational performance results that demonstrate that the technique is scalable for real size industrial problems and also how the execution time is dramatically reduced using GPU-based clusters.

Best Poster Finalist: no

Poster 129: Understanding I/O Behavior in Scientific Workflows on High Performance Computing Systems
Fahim Tahmid Chowdhury (Florida State University, Lawrence Livermore National Laboratory), Francesco Di Natale (Lawrence Livermore National Laboratory), Adam Moody (Lawrence Livermore National Laboratory), Elsa Gonsiorowski (Lawrence Livermore National Laboratory), Kathryn Mohror (Lawrence Livermore National Laboratory), Weikuan Yu (Florida State University)

Leadership high performance computing (HPC) systems have the capability to execute workflows of scientific, research or industry applications. Complex HPC workflows can have significant data transfer and I/O requirements. Heterogeneous storage systems in supercomputers equipped with bleeding-edge non-volatile persistent storage devices can be leveraged to handle these data transfer and I/O requirements efficiently.

In this poster, we describe our efforts to extract the I/O characteristics of various HPC workflows and develop strategies to improve I/O performance by leveraging heterogeneous storage systems. We have implemented an emulator to mimic different types of I/O requirements posed by HPC application workflows. We have analyzed the workflow of Cancer Moonshot Pilot 2 (CMP2) project to determine possible I/O inefficiencies. To date, we have performed a systematic characterization and evaluation on the workloads generated by the workflow emulator and a small scale adaptation of the CMP2 workflow.

Best Poster Finalist: no

Poster 116: Advancements in Ultrasound Simulations Enabled by High-Bandwidth GPU Interconnects
Filip Vaverka (Brno University of Technology), Bradley E. Treeby (University College London), Jiri Jaros
Realistic ultrasound simulations are becoming integral part of many novel medical procedures such as photoacoustic screening and non-invasive treatment planning. The common denominator of all these applications is the need for cheap and relatively large-scale ultrasound simulations with sufficient accuracy. Typical medical applications require full-wave simulations which take frequency-dependent absorption and non-linearity into account.

This poster investigates the benefits of high-bandwidth low-latency interconnects to k-Wave acoustic toolbox in dense multi-GPU environment. The k-Wave multi-GPU code is based on a variant of the local Fourier basis domain decomposition. The poster compares the behavior of the code on a typical PCI-E 3.0 machine with 8 Nvidia Tesla P40 GPUs and a Nvidia DGX-2 server. The performance constraints of PCI-E platforms built around multiple socket servers on multi-GPU applications are deeply explored. Finally, it is shown the k-Wave toolbox can efficiently utilize NVlink 2.0 and achieve over 4x speedup compared to PCI-E systems.

Best Poster Finalist: no

**Poster 65: Comparing Granular Dynamics vs. Fluid Dynamics via Large DOF-Count Parallel Simulation on the GPU**

*Milad Rakhsha (University of Wisconsin), Conlain Kelly (Georgia Institute of Technology), Nicholas Olsen (University of Wisconsin), Lijing Yang (University of Wisconsin), Radu Serban (University of Wisconsin), Dan Negrut (University of Wisconsin)*

In understanding granular dynamics, the commonly-used discrete modeling approach that tracks the motion of all particles is computationally demanding, especially with large system size. In such cases, one can contemplate switching to continuum models that are computationally less expensive. In order to assess when such a discrete to continuum switch is justified, we compare granular and fluid dynamics that scales to handle more than 1 billion degrees of freedom (DOFs); i.e., two orders of magnitude higher than the state-of-the-art. On the granular side, we solve the Newton-Euler equations of motion; on the fluid side, we solve the Navier-Stokes equations. Both solvers leverage parallel computing on the GPU, and are publicly available on GitHub as part of an open-source code called Chrono. We report similarities and differences between the dynamics of the discrete, fully-resolved system and the continuum model via numerical experiments including both static and highly transient scenarios.

Best Poster Finalist: no
Poster 78: Understanding HPC Application I/O Behavior Using System Level Statistics
Arnab K. Paul (Virginia Tech), Olaf Faaland (Lawrence Livermore National Laboratory), Adam Moody (Lawrence Livermore National Laboratory), Elsa Gonsiorowski (Lawrence Livermore National Laboratory), Kathryn Mohror (Lawrence Livermore National Laboratory), Ali R. Butt (Virginia Tech)

The processor performance of high performance computing (HPC) systems is increasing at a much higher rate than storage performance. Storage and file system designers therefore require a deep understanding of how HPC application I/O behavior affects current storage system installations in order to improve storage performance. In this work, we contribute to this understanding using application-agnostic file system statistics gathered on compute nodes as well as metadata and object storage file system servers. We analyze file system statistics of more than 4 million jobs over a period of three years on two systems at Lawrence Livermore National Laboratory that include a 15 PiB Lustre file system for storage. Some key observations in our study show that more than 65% HPC users perform significant I/O which are mostly writes; and less than 22% of HPC users who submit write-intensive jobs perform efficient writes to the file system.

Best Poster Finalist: no

Poster 109: A Runtime Approach for Dynamic Load Balancing of OpenMP Parallel Loops in LLVM
Jonas H. Müller Korndörfer (University of Basel, Switzerland), Florina M. Ciorba (University of Basel, Switzerland), Akan Yilmaz (University of Basel, Switzerland), Christian Iwainsky (Technical University Darmstadt), Johannes Doerfert (Argonne National Laboratory), Hal Finkel (Argonne National Laboratory), Vivek Kale (Brookhaven National Laboratory), Michael Klemm (Intel Corporation)

Load imbalance is the major source of performance degradation in computationally-intensive applications that frequently consist of parallel loops. Efficient scheduling can improve the performance of such programs. OpenMP is the de-facto standard for parallel programming on shared-memory systems. The current OpenMP specification provides only three choices for loop scheduling which are insufficient in scenarios with irregular loops, system-induced interference, or both. Therefore, this work augments the LLVM OpenMP runtime library implementation with eleven ready to use scheduling techniques. We tested existing and added scheduling strategies on several applications from NAS, SPEC OMP 2012, and CORAL2 benchmark suites. Experiments show that implemented scheduling techniques outperform others in certain application and system configurations. We measured performance gains of up to 6% compared to the fastest standard scheduling technique. This work aims to be a convincing step toward beyond-standard scheduling options in OpenMP for the benefit of evolving applications executing on multicore architectures.
Poster 143: Quantum Natural Language Processing
Lee James O’Riordan (Irish Centre for High-End Computing (ICHEC), National University of Ireland Galway (NUIG)), Myles Doyle (Irish Centre for High-End Computing (ICHEC), National University of Ireland Galway (NUIG)), Venkatesh Kannan (Irish Centre for High-End Computing (ICHEC), National University of Ireland Galway (NUIG)), Fabio Baruffa (Intel Corporation)

Natural language processing (NLP) algorithms that operate over strings of words are limited since they analyse meanings of the component words in a corpus without information about grammatical rules of the language. Consequently, they often produce unsatisfactory results with increase in problem complexity.

The "distributed compositional semantics" (DisCo) model incorporates grammatical structure of sentences into the algorithms, and offers significant improvements to the quality of results. However, their main challenge is the need for large classical computational resources. The DisCo model presents two quantum algorithms which lower storage and compute requirements compared to a classic HPC implementation.

In this project, we implement the two DisCo model quantum algorithms on the Intel Quantum Simulator deployed on the Irish national supercomputer. We target corpuses with ~1000 most-common words using up to 36 qubits simulation. The solution will be able to compute the meanings of two sentences and decide if their meanings match.

Best Poster Finalist: no

Poster 152: Deep Domain Adaptation for Runtime Prediction in Dynamic Workload Scheduler
Hoang H. Nguyen (National Center for Atmospheric Research (NCAR); University of Illinois, Chicago), Ben Matthews (National Center for Atmospheric Research (NCAR)), Irfan Elahi (National Center for Atmospheric Research (NCAR))

In HPC systems, users’ requested runtime for submitted jobs plays a crucial role in efficiency. While underestimation of job runtime could terminate jobs before completion, overestimation could result in long queuing of other jobs in HPC systems. In reality, runtime prediction in HPC is challenging due to the complexity and dynamics of running workloads. Most of the current predictive runtime models are trained on static workloads. This poses a risk of over-fitting the predictions with bias from the learned workload distribution. In this work, we propose an adaptation of Correlation Alignment
method in our deep neural network architecture (DCORAL) to alleviate the domain shift between workloads for better runtime predictions. Experiments on both standard benchmark workloads and NCAR real-time production workloads reveal that our proposed method results in a more stable training model across different workloads with low accuracy variance as compared to the other state-of-the-art methods.

Best Poster Finalist: no

Poster 75: libCEED - Lightweight High-Order Finite Elements Library with Performance Portability and Extensibility
Jeremy Thompson (University of Colorado), Valeria Barra (University of Colorado), Yohann Dudouit (Lawrence Livermore National Laboratory), Oana Marin (Argonne National Laboratory), Jed Brown (University of Colorado)

High-order numerical methods are widely used in PDE solvers, but software packages that have provided high-performance implementations have often been special-purpose and intrusive. libCEED is a new library that offers a purely algebraic interface for matrix-free operator representation and supports run-time selection of implementations tuned for a variety of computational device types, including CPUs and GPUs. We introduce the libCEED API and demonstrate how it can be used in standalone code or integrated with other packages (e.g., PETSc, MFEM, Nek5000) to solve examples of problems that often arise in the scientific computing community, ranging from fast solvers via geometric multigrid methods to Computational Fluid Dynamics (CFD) applications.

Best Poster Finalist: no

Progress on the Exascale Transition of the VSim Multiphysics PIC code
Benjamin M. Cowan (Tech-X Corporation), Sergey N. Averkin (Tech-X Corporation), John R. Cary (Tech-X Corporation), Jarrod Leddy (Tech-X Corporation), Scott W. Sides (Tech-X Corporation), Ilya A. Zilberter (Tech-X Corporation)

The highly performant, flexible plasma simulation code VSim was designed nearly 20 years ago (originally as Vorpal), with its first applications roughly four years later. Using object oriented methods, VSim was designed to allow runtime selection from multiple field solvers, particle dynamics, and reactions. It has been successful in modeling for many areas of physics, including fusion plasmas, particle accelerators, microwave devices, and RF and dielectric structures. Now it is critical to move to exascale systems, with their compute accelerator architectures, massive
threading, and advanced instruction sets. Here we discuss how we are moving this complex, multiphysics computational application to the new computing paradigm, and how it is done in a way that kept the application producing physics during the move. We present performance results showing significant speedups in all parts of the PIC loop, including field updates, particle pushes, and reactions.

Best Poster Finalist: no

**Poster 58: Lock-Free van Emde Boas Array**  
Ziyuan Guo (University of Tokyo)

Lock-based data structures have some potential issues such as deadlock, livelock, and priority inversion, and the progress can be delayed indefinitely if the thread that is holding locks cannot acquire a timeslice from the scheduler. Lock-free data structures, which guarantees the progress of some method call, can be used to avoid these problems. This poster introduces the first lock-free concurrent van Emde Boas Array which is a variant of van Emde Boas Tree. It is linearizable, and the benchmark shows significant performance improvement comparing to other lock-free search trees when the date set is large and dense enough.

Best Poster Finalist: no

**Poster 63: Adaptive Execution Planning in Biomedical Workflow Management Systems**  
Marta Jaros (Brno University of Technology), Bradley E. Treeby (University College London), Jiri Jaros (Brno University of Technology)

Biomedical simulations require very powerful computers. Their execution is described by a workflow consisting of a number of different cooperating tasks. The manual execution of individual tasks may be tedious for expert users, but prohibiting for most inexperienced clinicians. k-Dispatch offers a ‘run and forget’ approach where the users are completely screened out from the complexity of HPC systems. k-Dispatch provides task scheduling, execution, monitoring, and fault tolerance. Since the task execution configuration strongly affects the final tasks mapping on the computational resources, the execution planning is of the highest priority. Unlike other tools, k-Dispatch considers a variable amount of computational resources per individual tasks. Since the scaling of the individual HPC codes is never perfect, k-Dispatch may find such a good mapping even an experienced user would miss. The proposed adaptive execution planning is based on collected performance data and the current cluster utilization monitoring.
**Poster 142: Training Deep Neural Networks Directly on Hundred-Million-Pixel Histopathology Images on a Large-Scale GPU Cluster**

Chi-Chung Chen (AetherAI, Taiwan), Wen-Yu Chuang (Chang-Gung Memorial Hospital, Taiwan), Wei-Hsiang Yu (AetherAI, Taiwan), Hsi-Ching Lin (National Center for High-Performance Computing (NCHC), Taiwan), Shuen-Tai Wang (National Center for High-Performance Computing (NCHC), Taiwan), Fang-An Kuo (National Center for High-Performance Computing (NCHC), Taiwan), Chao-Chun Chuang (National Center for High-Performance Computing (NCHC), Taiwan), Chao-Yuan Yeh (AetherAI, Taiwan)

Deep learning for digital pathology is challenging because the resolution of whole-slide-images (WSI) is extremely high, often in billions. The most common approach is patch-based method, where WSIs are divided into small patches to train convolutional neural networks (CNN). This approach has significant drawbacks. To have ground truth for individual patches, detailed annotations by pathologists are required. This laborious process has become the major impediment to the development of digital pathology AI. End-to-end WSI training, however, faces the difficulties of fitting the task into limited GPU memory. In this work, we improved the efficiency of using system memory for GPU compute by 411% through memory optimization and deployed the training pipeline on 8 nodes, totally 32 GPUs distributed system, achieving 147.28x speedup. We demonstrated that CNN is capable of learning features without detailed annotations. The trained CNN can correctly classify cancerous specimen, with performance level closely matching the patch-based methods.

Best Poster Finalist: no

**Poster 96: TSQR on TensorCores**

Hiroyuki Ootomo (Tokyo Institute of Technology), Rio Yokota (Tokyo Institute of Technology)

Tall-Skinny QR (TSQR) is an efficient algorithm for calculating the QR decomposition of m x n matrices where m << n, which is done by recursively performing QR decomposition on subdivided blocks of the tall and skinny matrix. Such operations are useful for low-rank approximation methods, which are replacing more and more dense linear algebra in both scientific computing and machine learning fields. The present work focuses on the implementation of this important algorithm on Tensor Cores, which are available on the latest NVIDIA GPUs. We evaluate the speed, accuracy, and stability of TSQR on TensorCores.
**Poster 95: A Heterogeneous HEVC Video Encoder Based on OpenPOWER Acceleration Platform**

Chenhao Gu (Fudan University, Shanghai, State Key Laboratory of ASIC and System; IBM Corporation), Yang Chen (Fudan University, Shanghai, State Key Laboratory of ASIC and System; IBM Corporation), Yanheng Lu (IBM Corporation), Pengfei Gou (IBM Corporation), Yong Lu (IBM Corporation), Yang Dai (IBM Corporation), Yue Xu (IBM Corporation), Yang Liu (IBM Corporation), Yibo Fan (Fudan University, Shanghai, State Key Laboratory of ASIC and System)

This poster describes a heterogeneous HEVC video encoder system based on the OpenPOWER platform. Our design leverages the Coherent Accelerator Processor Interface (CAPI) on the OpenPOWER, which provides cache-coherent access for FPGA. This technology highly improves CPU-FPGA data communication bandwidth and programming efficiency. X265 is optimized on the OpenPOWER platform to improve its performance with both architecture specific methods and hardware-acceleration methods. For hardware acceleration, frame-level acceleration and functional-unit-level acceleration are introduced and evaluated in this work.

**Poster 102: Fast Training of an AI Radiologist: Leveraging Data Pipelining to Efficiently Utilize GPUs**

Rakshith Vasudev (Dell EMC), John A. Lockman III (Dell EMC), Lucas A. Wilson (Dell EMC), Srinivas Varadharajan (Dell EMC), Frank Han (Dell EMC), Rengan Xu (Dell EMC), Quy Ta (Dell EMC)

In a distributed deep learning training setting, using accelerators such as GPUs can be challenging to develop a high throughput model. If the accelerators are not utilized effectively, this could mean more time to solution, and thus the model's throughput is low. To use accelerators effectively across multiple nodes, we need to utilize an effective data pipelining mechanism that handles scaling gracefully so GPUs can be exploited of their parallelism. We study the effect of using the correct pipelining mechanism that is followed by tensorflow official models vs a naive pipelining mechanism that doesn't scale well, on two image classification models. Both the models using the optimized data pipeline demonstrate effective linear scaling when GPUs are added. We also show that converting to TF Records is not always necessary.
Poster 94: Multi-GPU Optimization of a Non-Hydrostatic Numerical Ocean Model with Multigrid Preconditioned Conjugate Gradient Method

Takateru Yamagishi (Research Organization for Information Science and Technology, Japan), Yoshimasa Matsumura (University of Tokyo, Atmosphere and Ocean Research Institute), Hiroyasu Hasumi (University of Tokyo, Atmosphere and Ocean Research Institute)

The conjugate gradient method with multigrid preconditioners (MGCG) is used in scientific applications because of its high performance and scalability with many computational nodes. GPUs are thought to be good candidates for accelerating such applications with many meshes where an MGCG solver could show high performance. No previous studies have evaluated and discussed the numerical character of an MGCG solver on GPUs. Consequently, we have implemented and optimized our “kinaco” numerical ocean model with an MGCG solver on GPUs. We evaluated its performance and discussed inter-GPU communications on a coarse grid on which GPUs could be intrinsically problematic. We achieved 3.9 times speedup compared to CPUs and learned how inter-GPU communications depended on the number of GPUs and the aggregation level of information in a multigrid method.

Best Poster Finalist: no

Poster 108: Power Prediction for High-Performance Computing

Shigeto Suzuki (Fujitsu Laboratories Ltd), Michiko Hiraoka (Fujitsu Ltd), Takashi Shiraishi (Fujitsu Laboratories Ltd), Enxhi Kreshpa (Fujitsu Laboratories Ltd), Takuji Yamamoto (Fujitsu Laboratories Ltd), Hiroyuki Fukuda (Fujitsu Laboratories Ltd), Shuji Matsui (Fujitsu Ltd), Masahide Fujisaki (Fujitsu Ltd), Atsuya Uno (RIKEN Center for Computational Science (R-CCS))

Exascale computers consume large amounts of power both for computing and cooling-units. As power of the computer varies dynamically corresponding to the load change, cooling-units are desirable to follow it for effective energy management. Because of time lags in cooling-unit operations, advance control is inevitable and an accurate prediction is a key for it. Conventional prediction methods make use of the similarity between job information while in queue. The prediction fails if there is no previously similar job. We developed two models to correct the prediction after queued jobs start running. By taking power histories into account, power-correlated topic model reselects more suitable candidate and recurrent-neural-network model considering variable network sizes predicts power variation from shape features of it. We integrated these into a single algorithm and demonstrated high-precision prediction with an average relative error of 5.7% in K computer as compared to the 18.0% obtained using the conventional method.
Poster 80: Sharing and Replicability of Notebook-Based Research on Open Testbeds
Maxine V. King (University of Chicago), Jason Anderson (University of Chicago), Kate Keahey (Argonne National Laboratory, University of Chicago)

We seek to facilitate replicability by creating a way to share experiments easily in and out of notebook-based, open testbed environments and a sharing platform for such experiments in order to allow researchers to combine shareability, consistency of code environment, and well-documented process.

Poster 121: HFlush: Realtime Flushing for Modern Storage Environments
Jaime Cernuda (Illinois Institute of Technology), Hugo Trivino (Illinois Institute of Technology), Hariharan Devarajan (Illinois Institute of Technology), Anthony Kougkas (Illinois Institute of Technology), Xian-He Sun (Illinois Institute of Technology)

Due to the unparalleled magnitude of data movement in extreme scale computing, I/O has become a central challenge. Modern storage environments have proposed the use of multiple layers between applications and the PFS. Nonetheless, the difference in capacities and speeds between storage layers makes it extremely challenging to evict data from upper layers to lower layers efficiently. However, current solutions are executed in batches, compromising latency; are also push-based implementations, compromising resource utilization. Hence, we propose HFlush, a continuous data eviction mechanism built on a streaming architecture that is pull-based and in which each component is decoupled and executed in parallel. Initial results have shown RFlush to obtain a 7X latency reduction and a 2X bandwidth improvement over a baseline batch-based system. Therefore, RFlush is a promising solution to the growing challenges of extreme scale data generation and eviction shortcomings when archiving data across multiple tiers of storage.

Poster 88: HPC Container Runtime Performance Overhead: At First Order, There Is None
Alfred Torrez (Los Alamos National Laboratory), Reid Priedhorsky (Los Alamos National Laboratory), Timothy Randles (Los Alamos National Laboratory)
Linux containers are an increasingly popular method used by HPC centers to meet increasing demand for greater software flexibility. A common concern is that containers may introduce application performance overhead. Prior work has not tested a broad set of HPC container technologies on a broad set of benchmarks. This poster addresses the gap by comparing performance of the three HPC container implementations (Charliecloud, Shifter, and Singularity) and bare metal on multiple dimensions using industry-standard benchmarks.

We found no meaningful performance differences between the four environments with the possible exception of modest variation in memory usage, which is broadly consistent with prior results. This result suggests that HPC users should feel free to containerize their applications without concern about performance degradation, regardless of the container technology used. It is an encouraging development on the path towards greater adoption of user-defined software stacks to increase the flexibility of HPC.

Best Poster Finalist: no

Tsuyoshi Ichimura (University of Tokyo, RIKEN), Kohei Fujita (University of Tokyo, RIKEN), Takuma Yamaguchi (University of Tokyo), Akira Naruse (Nvidia Corporation), Jack C. Wells (Oak Ridge National Laboratory), Christopher J. Zimmer (Oak Ridge National Laboratory), Tjerk P. Straatsma (Oak Ridge National Laboratory), Takane Hori (Japan Agency for Marine-Earth Science and Technology), Simone Puel (University of Texas), Thorsten W. Becker (University of Texas), Muneo Hori (Japan Agency for Marine-Earth Science and Technology), Naonori Ueda (RIKEN)

We propose herein an approach for reformulating an equation-based modeling algorithm to an algorithm similar to that of training artificial intelligence (AI) and accelerate this algorithm using high-performance accelerators to reduce the huge computational costs encountered for physics equation-based modeling in earthquake disaster mitigation. A fast scalable equation-based implicit solver on unstructured finite elements is accelerated with a Tensor Core-enabled matrix-vector product kernel. The developed kernel attains 1.10 ExaFLOPS, leading to 416 PFLOPS for the whole solver on full Summit. This corresponds to a 75-fold speedup from a previous state-of-the-art solver running on full Piz Daint. This result could lead to breakthroughs in earthquake disaster mitigation. Our new idea in the HPC algorithm design of combining equation-based modeling with AI is expected to have broad impacts in other earth science and industrial problems.

Best Poster Finalist: no
SC Theater

Tuesday, November 19

10:00 am - 6:00 pm

SC Theater

Hybrid Quantum-Classical Algorithms for Graph Problems: Forging a Path to Near-Term Applications
Ruslan Shaydulin (Clemson University)

Technology Challenge Introduction and Jury Introduction

Technology Challenge UEN Demo

Technology Challenge RENCI Demo Prep

Technology Challenge RENCI Demo

Break

SCinet Architecture

Technology Challenge ANL Demo Prep

Technology Challenge ANL Demo

Technology Challenge Jury Closing Remarks

Student Cluster Challenge "How We Got Here"

Student Cluster Challenge "How We Got Here"

Student Cluster Challenge "How We Got Here"
Student Cluster Challenge "How We Got Here"

SCinet Routing

Live Interview Video Broadcast

Open

Modern HPC approaches to solve time dependent Quantum Mechanical equations
Konstantin Rygol (University of Bergen)

Wednesday, November 20

10:00 am - 6:00 pm

SC Theater

2:30 pm - 4:30 pm

Computing4Change Lightning Talks

Session Description: Computing4Change is an undergraduate student competition sponsored by SIGHPC and co-located with SC19. Students work in assigned teams on a data intensive, socially relevant topic assigned at SC19. Each student will present their technical approach to the given problem and their results in this lightning talk session. Come support these undergraduate students from across the country and hear their innovative approaches for this competition.
Thursday, November 21

10:00 am - 3:00 pm

SC Theater
SC20

Tuesday, November 19
10:00 am - 5:00 pm
SC20 Preview Booth

Wednesday, November 20
10:00 am - 5:00 pm
SC20 Preview Booth

Thursday, November 21
8:00 am - 8:35 am
SC20 Conference Preview

10:00 am - 3:00 pm
SC20 Preview Booth
Visualization of Entrainment and Mixing Phenomena at Cloud Edges

Bipin Kumar (Indian Institute of Tropical Meteorology), Matt Rehme (National Center for Atmospheric Research (NCAR)), Neethi Suresh (Indian Institute of Tropical Meteorology), Nihanth Cherukuru (National Center for Atmospheric Research (NCAR)), Stanislaw Jaroszynski (National Center for Atmospheric Research (NCAR)), Shaomeng Li (National Center for Atmospheric Research (NCAR)), Scott Pearse (National Center for Atmospheric Research (NCAR)), Tim Scheitlin (National Center for Atmospheric Research (NCAR))

Entrainment and mixing processes occur during the entire life of a cloud. These processes change the droplet size distribution, which determines rain formation and radiative properties. Since it is a microphysical process, it cannot be resolved in large scale weather forecasting models. Small scale simulations such as Direct Numerical Simulations (DNS) are required to resolve the smallest scale of these processes. However, it has been a challenge to visualize these processes in a 3D domain as it generates petabytes of data. Visualization plays a vital role in analyzing and understanding these huge data outputs. Here, we present different techniques for 3D visualization of data obtained from DNS carried out at Indian Institute of Tropical Meteorology (IITM) to understand cloud microphysical properties more closely.

Multiple tools were used to conduct a visual analysis of this data. Two of these tools are well established and tested technologies: ParaView and VAPOR. The others are emergent technologies created at the National Center for Atmospheric Research (NCAR) and are in the development phase. This data simulation, in addition to exploring DNS as mentioned above, provided an opportunity to test and improve development of several tools and methods.
Visualizing Deep Learning at Scale for the Construction of Galaxy Catalogs in the Dark Energy Survey

Janet Y. K. Knowles (Argonne National Laboratory), Joseph A. Insley (Argonne National Laboratory, Northern Illinois University), Silvio Rizzi (Argonne National Laboratory), Elise Jennings (Argonne National Laboratory), Asad Khan (University of Illinois), Eliu Huerta (University of Illinois), Sibo Wang (University of Illinois), Robert Gruendl (University of Illinois), Huihuo Zheng (Argonne National Laboratory)

The advent of machine and deep learning algorithms on petascale supercomputers is accelerating the pace of discovery in astrophysics and poses significant challenges to the interpretability of these deep neural networks. We present a novel visualization of a deep neural network output during training as it is learning to classify galaxy images as either spiral or elliptical. The network is trained using labeled datasets from the citizen science campaign, Galaxy Zoo, adopted by the Sloan Digital Sky Survey. These trained neural network models can then be used to classify galaxies in the Dark Energy Survey that overlap the footprint of both surveys. Visualizing a reduced representation of the network output, projected into 3-D parameter space, reveals how the network has discovered two distinct clusters of features which allows it to classify galaxies into two groups. These visualizations of the neural network during training aid in the interpretability of the black box of deep learning and reveal how the network responds to the input images at various stages of training. Finally, it allows a wider net to be cast to a general audience, thereby generating interest in and visibility to an otherwise highly specialized field.

NVIDIA IndeX Accelerated Computing for Visualizing Cholla's Galactic Winds

Christopher Lux (Nvidia Corporation), Alexander Kuhn (Nvidia Corporation), Evan Schneider (Princeton University), Brant Robertson (University of California, Santa Cruz), Marc Nienhaus (Nvidia Corporation)

Galactic winds – outflows of gas driven out of galaxies by the combined effects of thousands of supernovae – are a crucial feature of galaxy evolution. Despite their importance, a complete theoretical picture of these winds has been elusive. Simulating the complicated interaction between the hot, high pressure gas created by supernovae and the cooler, high density gas in the galaxy disk requires massive computational resources and sophisticated software. For this purpose, Computational Hydrodynamics On Parallel Architectures (Cholla) has been demonstrated to be a scalable and efficient tool that operates in large, distributed multi-GPU environments at high levels of performance. This additional performance results in unprecedented resolution for this type of simulation and massive amounts of generated data. This raises the requirements for analysis tools that can cope with scale and complexity of the simulated physical processes. To address those
challenges, we utilize NVIDIA IndeX as a scalable framework to visualize the simulation output. NVIDIA IndeX features a streaming-based architecture to interactively explore simulation results in large-scale, multi-GPU environments. We utilize customized sampling programs for multi-volume and surface rendering to address analysis questions of galactic wind simulations. This combination of massively parallel simulation and analysis allows us to utilize recent supercomputer capabilities and to speed up the exploration of galactic wind simulations.

Visualizing the World’s Largest Turbulence Simulation
Salvatore Cielo (Leibniz Supercomputing Centre), Luigi Iapichino (Leibniz Supercomputing Centre), Johannes Günther (Intel Corporation), Christoph Federrath (Australian National University, Research School of Astronomy and Astrophysics), Elisabeth Mayer (Leibniz Supercomputing Centre), Markus Wiedemann (Leibniz Supercomputing Centre)

We present the visualization of the largest interstellar turbulence simulations ever performed, unravelling key astrophysical processes concerning the formation of stars and the relative role of magnetic fields. The simulations, including pure hydrodynamical (HD) and magneto-hydrodynamical (MHD) runs, up to a size of 10048^3 grid elements, were produced on the supercomputers of the Leibniz Supercomputing Centre and visualized using the hybrid parallel (MPI + TBB) ray-tracing engine OSPRay associated with VisIt. Besides revealing features of turbulence with an unprecedented resolution, the visualizations brilliantly showcase the stretching-and-folding mechanisms through which astrophysical processes such as supernova explosions drive turbulence and amplify the magnetic field in the interstellar gas, and how the first structures, the seeds of newborn stars are shaped by this process.

An Accessible Visual Narrative for the Primary Energy Source of Life from the Fulldome Show Birth of Planet Earth
Melih Sener (University of Illinois), Stuart Levy (University of Illinois), AJ Christensen (University of Illinois), Robert Patterson (University of Illinois), Kalina Borkiewicz (University of Illinois), John E. Stone (University of Illinois), Barry Isralewitz (University of Illinois), Jeffrey Carpenter (University of Illinois), Donna Cox (University of Illinois)

Conversion of sunlight into chemical energy, namely photosynthesis, is the primary energy source of life on Earth. An explanatory visualization depicting this process is presented in the form of an excerpt from the fulldome show Birth of Planet Earth. This accessible visual narrative shows a lay audience, especially children, how the energy of sunlight is captured, converted, and stored through a chain of proteins to power living cells. The visualization is the result of a multi-year collaboration among biophysicists, visualization scientists, and artists, which, in turn, is based on a decade-long
experimental-computational collaboration on structural and functional modeling that produced an atomic detail description of a bacterial bioenergetic organelle, the chromatophore. The energy conversion steps depicted feature an integration of function from electronic to cell levels, spanning nearly 12 orders of magnitude in time scales modeled with multi-scale computational approaches. This atomic detail description uniquely enables a modern retelling of one of humanity’s earliest stories—the interplay between light and life.

**Visualizing Supersonic Retropropulsion for Mars: The Summit Early Science Results**
Patrick Moran (NASA Ames Research Center), Timothy Sandstrom (NASA Ames Research Center), Ashley Korzun (NASA Langley Research Center), Eric Nielsen (NASA Langley Research Center), Aaron Walden (NASA Langley Research Center)

As part of the Oak Ridge Leadership Computing Facility (OLCF) Early Science program, NASA ran a series of simulations on the new, GPU-accelerated system Summit, using the flow solver FUN3D to produce solutions for a set of representative descent conditions. The raw output from these simulations was over a petabyte of data. We produced visualizations at the NASA Advanced Supercomputing Division at Ames Research Center using parallel techniques where feasible to accelerate the process. Included with this summary is an animation with some of the highlights of our visualizations so far. We discuss some of our visualization design choices, and plans for future work.

5:15 pm - 7:00 pm

**Poster Reception**

Wednesday, November 20

8:30 am - 5:00 pm

**Scientific Visualization & Data Analytics Showcase Posters Display**

**Visualization of Entrainment and Mixing Phenomena at Cloud Edges**
Bipin Kumar (Indian Institute of Tropical Meteorology), Matt Rehme (National Center for Atmospheric Research (NCAR)), Neethi Suresh (Indian Institute of Tropical Meteorology), Nihan Cherukuru (National Center for Atmospheric Research (NCAR)), Stanislaw Jaroszynski (National
Entrainment and mixing processes occur during the entire life of a cloud. These processes change the droplet size distribution, which determines rain formation and radiative properties. Since it is a microphysical process, it cannot be resolved in large scale weather forecasting models. Small scale simulations such as Direct Numerical Simulations (DNS) are required to resolve the smallest scale of these processes. However, it has been a challenge to visualize these processes in a 3D domain as it generates petabytes of data. Visualization plays a vital role in analyzing and understanding these huge data outputs. Here, we present different techniques for 3D visualization of data obtained from DNS carried out at Indian Institute of Tropical Meteorology (IITM) to understand cloud microphysical properties more closely.

Multiple tools were used to conduct a visual analysis of this data. Two of these tools are well established and tested technologies: ParaView and VAPOR. The others are emergent technologies created at the National Center for Atmospheric Research (NCAR) and are in the development phase. This data simulation, in addition to exploring DNS as mentioned above, provided an opportunity to test and improve development of several tools and methods.

**Visualizing Deep Learning at Scale for the Construction of Galaxy Catalogs in the Dark Energy Survey**

Janet Y. K. Knowles (Argonne National Laboratory), Joseph A. Insley (Argonne National Laboratory, Northern Illinois University), Silvio Rizzi (Argonne National Laboratory), Elise Jennings (Argonne National Laboratory), Asad Khan (University of Illinois), Eliu Huerta (University of Illinois), Sibo Wang (University of Illinois), Robert Gruendl (University of Illinois), Huihuo Zheng (Argonne National Laboratory)

The advent of machine and deep learning algorithms on petascale supercomputers is accelerating the pace of discovery in astrophysics and poses significant challenges to the interpretability of these deep neural networks. We present a novel visualization of a deep neural network output during training as it is learning to classify galaxy images as either spiral or elliptical. The network is trained using labeled datasets from the citizen science campaign, Galaxy Zoo, adopted by the Sloan Digital Sky Survey. These trained neural network models can then be used to classify galaxies in the Dark Energy Survey that overlap the footprint of both surveys. Visualizing a reduced representation of the network output, projected into 3-D parameter space, reveals how the network has discovered two distinct clusters of features which allows it to classify galaxies into two groups. These visualizations of the neural network during training aid in the interpretability of the black box of
deep learning and reveal how the network responds to the input images at various stages of training. Finally, it allows a wider net to be cast to a general audience, thereby generating interest in and visibility to an otherwise highly specialized field.

**NVIDIA IndeX Accelerated Computing for Visualizing Cholla's Galactic Winds**

Christopher Lux (Nvidia Corporation), Alexander Kuhn (Nvidia Corporation), Evan Schneider (Princeton University), Brant Robertson (University of California, Santa Cruz), Marc Nienhaus (Nvidia Corporation)

Galactic winds – outflows of gas driven out of galaxies by the combined effects of thousands of supernovae – are a crucial feature of galaxy evolution. Despite their importance, a complete theoretical picture of these winds has been elusive. Simulating the complicated interaction between the hot, high pressure gas created by supernovae and the cooler, high density gas in the galaxy disk requires massive computational resources and sophisticated software. For this purpose, Computational Hydrodynamics On Parallel Architectures (Cholla) has been demonstrated to be a scalable and efficient tool that operates in large, distributed multi-GPU environments at high levels of performance. This additional performance results in unprecedented resolution for this type of simulation and massive amounts of generated data. This raises the requirements for analysis tools that can cope with scale and complexity of the simulated physical processes. To address those challenges, we utilize NVIDIA IndeX as a scalable framework to visualize the simulation output: NVIDIA IndeX features a streaming-based architecture to interactively explore simulation results in large-scale, multi-GPU environments. We utilize customized sampling programs for multi-volume and surface rendering to address analysis questions of galactic wind simulations. This combination of massively parallel simulation and analysis allows us to utilize recent supercomputer capabilities and to speed up the exploration of galactic wind simulations.

**Visualizing the World's Largest Turbulence Simulation**

Salvatore Cielo (Leibniz Supercomputing Centre), Luigi Iapichino (Leibniz Supercomputing Centre), Johannes Günther (Intel Corporation), Christoph Federrath (Australian National University, Research School of Astronomy and Astrophysics), Elisabeth Mayer (Leibniz Supercomputing Centre), Markus Wiedemann (Leibniz Supercomputing Centre)

We present the visualization of the largest interstellar turbulence simulations ever performed, unravelling key astrophysical processes concerning the formation of stars and the relative role of magnetic fields. The simulations, including pure hydrodynamical (HD) and magneto-hydrodynamical (MHD) runs, up to a size of 10048^3 grid elements, were produced on the supercomputers of the Leibniz Supercomputing Centre and visualized using the hybrid parallel (MPI + TBB) ray-tracing
engine OSPRay associated with VisIt. Besides revealing features of turbulence with an unprecedented resolution, the visualizations brilliantly showcase the stretching-and-folding mechanisms through which astrophysical processes such as supernova explosions drive turbulence and amplify the magnetic field in the interstellar gas, and how the first structures, the seeds of newborn stars are shaped by this process.

An Accessible Visual Narrative for the Primary Energy Source of Life from the Fulldome Show Birth of Planet Earth

Melih Sener (University of Illinois), Stuart Levy (University of Illinois), AJ Christensen (University of Illinois), Robert Patterson (University of Illinois), Kalina Borkiewicz (University of Illinois), John E. Stone (University of Illinois), Barry Isralewitz (University of Illinois), Jeffrey Carpenter (University of Illinois), Donna Cox (University of Illinois)

Conversion of sunlight into chemical energy, namely photosynthesis, is the primary energy source of life on Earth. An explanatory visualization depicting this process is presented in the form of an excerpt from the fulldome show Birth of Planet Earth. This accessible visual narrative shows a lay audience, especially children, how the energy of sunlight is captured, converted, and stored through a chain of proteins to power living cells. The visualization is the result of a multi-year collaboration among biophysicists, visualization scientists, and artists, which, in turn, is based on a decade-long experimental-computational collaboration on structural and functional modeling that produced an atomic detail description of a bacterial bioenergetic organelle, the chromatophore. The energy conversion steps depicted feature an integration of function from electronic to cell levels, spanning nearly 12 orders of magnitude in time scales modeled with multi-scale computational approaches. This atomic detail description uniquely enables a modern retelling of one of humanity’s earliest stories—the interplay between light and life.

Visualizing Supersonic Retropropulsion for Mars: The Summit Early Science Results

Patrick Moran (NASA Ames Research Center), Timothy Sandstrom (NASA Ames Research Center), Ashley Korzun (NASA Langley Research Center), Eric Nielsen (NASA Langley Research Center), Aaron Walden (NASA Langley Research Center)

As part of the Oak Ridge Leadership Computing Facility (OLCF) Early Science program, NASA ran a series of simulations on the new, GPU-accelerated system Summit, using the flow solver FUN3D to produce solutions for a set of representative descent conditions. The raw output from these simulations was over a petabyte of data. We produced visualizations at the NASA Advanced Supercomputing Division at Ames Research Center using parallel techniques where feasible to accelerate the process. Included with this summary is an animation with some of the highlights of
our visualizations so far. We discuss some of our visualization design choices, and plans for future work.

1:30 pm - 3:00 pm

Scientific Visualization & Data Analytics Showcase Posters Presentations

Visualization of Entrainment and Mixing Phenomena at Cloud Edges
Bipin Kumar (Indian Institute of Tropical Meteorology), Matt Rehme (National Center for Atmospheric Research (NCAR)), Neethi Suresh (Indian Institute of Tropical Meteorology), Nihanth Cherukuru (National Center for Atmospheric Research (NCAR)), Stanislaw Jaroszynski (National Center for Atmospheric Research (NCAR)), Shaomeng Li (National Center for Atmospheric Research (NCAR)), Scott Pearse (National Center for Atmospheric Research (NCAR)), Tim Scheitlin (National Center for Atmospheric Research (NCAR))

Entrainment and mixing processes occur during the entire life of a cloud. These processes change the droplet size distribution, which determines rain formation and radiative properties. Since it is a microphysical process, it cannot be resolved in large scale weather forecasting models. Small scale simulations such as Direct Numerical Simulations (DNS) are required to resolve the smallest scale of these processes. However, it has been a challenge to visualize these processes in a 3D domain as it generates petabytes of data. Visualization plays a vital role in analyzing and understanding these huge data outputs. Here, we present different techniques for 3D visualization of data obtained from DNS carried out at Indian Institute of Tropical Meteorology (IITM) to understand cloud microphysical properties more closely.

Multiple tools were used to conduct a visual analysis of this data. Two of these tools are well established and tested technologies: ParaView and VAPOR. The others are emergent technologies created at the National Center for Atmospheric Research (NCAR) and are in the development phase. This data simulation, in addition to exploring DNS as mentioned above, provided an opportunity to test and improve development of several tools and methods.

Visualizing Deep Learning at Scale for the Construction of Galaxy Catalogs in the Dark Energy Survey
Janet Y. K. Knowles (Argonne National Laboratory), Joseph A. Insley (Argonne National Laboratory, Northern Illinois University), Silvio Rizzi (Argonne National Laboratory), Elise Jennings (Argonne National Laboratory), Asad Khan (University of Illinois), Eliu Huerta (University of Illinois), Sibo Wang
The advent of machine and deep learning algorithms on petascale supercomputers is accelerating the pace of discovery in astrophysics and poses significant challenges to the interpretability of these deep neural networks. We present a novel visualization of a deep neural network output during training as it is learning to classify galaxy images as either spiral or elliptical. The network is trained using labeled datasets from the citizen science campaign, Galaxy Zoo, adopted by the Sloan Digital Sky Survey. These trained neural network models can then be used to classify galaxies in the Dark Energy Survey that overlap the footprint of both surveys. Visualizing a reduced representation of the network output, projected into 3-D parameter space, reveals how the network has discovered two distinct clusters of features which allows it to classify galaxies into two groups. These visualizations of the neural network during training aid in the interpretability of the black box of deep learning and reveal how the network responds to the input images at various stages of training. Finally, it allows a wider net to be cast to a general audience, thereby generating interest in and visibility to an otherwise highly specialized field.

**NVIDIA IndeX Accelerated Computing for Visualizing Cholla's Galactic Winds**

Christopher Lux (Nvidia Corporation), Alexander Kuhn (Nvidia Corporation), Evan Schneider (Princeton University), Brant Robertson (University of California, Santa Cruz), Marc Nienhaus (Nvidia Corporation)

Galactic winds – outflows of gas driven out of galaxies by the combined effects of thousands of supernovae – are a crucial feature of galaxy evolution. Despite their importance, a complete theoretical picture of these winds has been elusive. Simulating the complicated interaction between the hot, high pressure gas created by supernovae and the cooler, high density gas in the galaxy disk requires massive computational resources and sophisticated software. For this purpose, Computational Hydrodynamics On Parallel Architectures (Cholla) has been demonstrated to be a scalable and efficient tool that operates in large, distributed multi-GPU environments at high levels of performance. This additional performance results in unprecedented resolution for this type of simulation and massive amounts of generated data. This raises the requirements for analysis tools that can cope with scale and complexity of the simulated physical processes. To address those challenges, we utilize NVIDIA IndeX as a scalable framework to visualize the simulation output: NVIDIA IndeX features a streaming-based architecture to interactively explore simulation results in large-scale, multi-GPU environments. We utilize customized sampling programs for multi-volume and surface rendering to address analysis questions of galactic wind simulations. This combination of massively parallel simulation and analysis allows us to utilize recent supercomputer capabilities and to speed up the exploration of galactic wind simulations.
Visualizing the World’s Largest Turbulence Simulation

Salvatore Cielo (Leibniz Supercomputing Centre), Luigi Iapichino (Leibniz Supercomputing Centre), Johannes Günther (Intel Corporation), Christoph Federrath (Australian National University, Research School of Astronomy and Astrophysics), Elisabeth Mayer (Leibniz Supercomputing Centre), Markus Wiedemann (Leibniz Supercomputing Centre)

We present the visualization of the largest interstellar turbulence simulations ever performed, unravelling key astrophysical processes concerning the formation of stars and the relative role of magnetic fields. The simulations, including pure hydrodynamical (HD) and magneto-hydrodynamical (MHD) runs, up to a size of 10048^3 grid elements, were produced on the supercomputers of the Leibniz Supercomputing Centre and visualized using the hybrid parallel (MPI + TBB) ray-tracing engine OSPRay associated with VisIt. Besides revealing features of turbulence with an unprecedented resolution, the visualizations brilliantly showcase the stretching-and-folding mechanisms through which astrophysical processes such as supernova explosions drive turbulence and amplify the magnetic field in the interstellar gas, and how the first structures, the seeds of newborn stars are shaped by this process.

An Accessible Visual Narrative for the Primary Energy Source of Life from the Fulldome Show Birth of Planet Earth

Melih Sener (University of Illinois), Stuart Levy (University of Illinois), AJ Christensen (University of Illinois), Robert Patterson (University of Illinois), Kalina Borkiewicz (University of Illinois), John E. Stone (University of Illinois), Barry Isralewitz (University of Illinois), Jeffrey Carpenter (University of Illinois), Donna Cox (University of Illinois)

Conversion of sunlight into chemical energy, namely photosynthesis, is the primary energy source of life on Earth. An explanatory visualization depicting this process is presented in the form of an excerpt from the fulldome show Birth of Planet Earth. This accessible visual narrative shows a lay audience, especially children, how the energy of sunlight is captured, converted, and stored through a chain of proteins to power living cells. The visualization is the result of a multi-year collaboration among biophysicists, visualization scientists, and artists, which, in turn, is based on a decade-long experimental-computational collaboration on structural and functional modeling that produced an atomic detail description of a bacterial bioenergetic organelle, the chromatophore. The energy conversion steps depicted feature an integration of function from electronic to cell levels, spanning nearly 12 orders of magnitude in time scales modeled with multi-scale computational approaches. This atomic detail description uniquely enables a modern retelling of one of humanity’s earliest stories---the interplay between light and life.
Visualizing Supersonic Retropropulsion for Mars: The Summit Early Science Results

Patrick Moran (NASA Ames Research Center), Timothy Sandstrom (NASA Ames Research Center), Ashley Korzun (NASA Langley Research Center), Eric Nielsen (NASA Langley Research Center), Aaron Walden (NASA Langley Research Center)

As part of the Oak Ridge Leadership Computing Facility (OLCF) Early Science program, NASA ran a series of simulations on the new, GPU-accelerated system Summit, using the flow solver FUN3D to produce solutions for a set of representative descent conditions. The raw output from these simulations was over a petabyte of data. We produced visualizations at the NASA Advanced Supercomputing Division at Ames Research Center using parallel techniques where feasible to accelerate the process. Included with this summary is an animation with some of the highlights of our visualizations so far. We discuss some of our visualization design choices, and plans for future work.

Thursday, November 21

8:30 am - 5:00 pm

Scientific Visualization & Data Analytics Showcase Posters Display

Visualization of Entrainment and Mixing Phenomena at Cloud Edges

Bipin Kumar (Indian Institute of Tropical Meteorology), Matt Rehme (National Center for Atmospheric Research (NCAR)), Neethi Suresh (Indian Institute of Tropical Meteorology), Nihanth Cherukuru (National Center for Atmospheric Research (NCAR)), Stanislaw Jaroszynski (National Center for Atmospheric Research (NCAR)), Shaomeng Li (National Center for Atmospheric Research (NCAR)), Scott Pearse (National Center for Atmospheric Research (NCAR)), Tim Scheitlin (National Center for Atmospheric Research (NCAR))

Entrainment and mixing processes occur during the entire life of a cloud. These processes change the droplet size distribution, which determines rain formation and radiative properties. Since it is a microphysical process, it cannot be resolved in large scale weather forecasting models. Small scale simulations such as Direct Numerical Simulations (DNS) are required to resolve the smallest scale of these processes. However, it has been a challenge to visualize these processes in a 3D domain as it generates petabytes of data. Visualization plays a vital role in analyzing and understanding these huge data outputs. Here, we present different techniques for 3D visualization of data obtained from DNS carried out at Indian Institute of Tropical Meteorology (IITM) to understand cloud microphysical
properties more closely.

Multiple tools were used to conduct a visual analysis of this data. Two of these tools are well established and tested technologies: ParaView and VAPOR. The others are emergent technologies created at the National Center for Atmospheric Research (NCAR) and are in the development phase. This data simulation, in addition to exploring DNS as mentioned above, provided an opportunity to test and improve development of several tools and methods.

**Visualizing Deep Learning at Scale for the Construction of Galaxy Catalogs in the Dark Energy Survey**

Janet Y. K. Knowles (Argonne National Laboratory), Joseph A. Insley (Argonne National Laboratory, Northern Illinois University), Silvio Rizzi (Argonne National Laboratory), Elise Jennings (Argonne National Laboratory), Asad Khan (University of Illinois), Eliu Huerta (University of Illinois), Sibo Wang (University of Illinois), Robert Gruendl (University of Illinois), Huihuo Zheng (Argonne National Laboratory)

The advent of machine and deep learning algorithms on petascale supercomputers is accelerating the pace of discovery in astrophysics and poses significant challenges to the interpretability of these deep neural networks. We present a novel visualization of a deep neural network output during training as it is learning to classify galaxy images as either spiral or elliptical. The network is trained using labeled datasets from the citizen science campaign, Galaxy Zoo, adopted by the Sloan Digital Sky Survey. These trained neural network models can then be used to classify galaxies in the Dark Energy Survey that overlap the footprint of both surveys. Visualizing a reduced representation of the network output, projected into 3-D parameter space, reveals how the network has discovered two distinct clusters of features which allows it to classify galaxies into two groups. These visualizations of the neural network during training aid in the interpretability of the black box of deep learning and reveal how the network responds to the input images at various stages of training. Finally, it allows a wider net to be cast to a general audience, thereby generating interest in and visibility to an otherwise highly specialized field.

**NVIDIA IndeX Accelerated Computing for Visualizing Cholla's Galactic Winds**

Christopher Lux (Nvidia Corporation), Alexander Kuhn (Nvidia Corporation), Evan Schneider (Princeton University), Brant Robertson (University of California, Santa Cruz), Marc Nienhaus (Nvidia Corporation)

Galactic winds – outflows of gas driven out of galaxies by the combined effects of thousands of supernovae – are a crucial feature of galaxy evolution. Despite their importance, a complete
Visualizing the World’s Largest Turbulence Simulation

Salvatore Cielo (Leibniz Supercomputing Centre), Luigi Iapichino (Leibniz Supercomputing Centre), Johannes Günther (Intel Corporation), Christoph Federrath (Australian National University, Research School of Astronomy and Astrophysics), Elisabeth Mayer (Leibniz Supercomputing Centre), Markus Wiedemann (Leibniz Supercomputing Centre)

We present the visualization of the largest interstellar turbulence simulations ever performed, unravelling key astrophysical processes concerning the formation of stars and the relative role of magnetic fields. The simulations, including pure hydrodynamical (HD) and magneto-hydrodynamical (MHD) runs, up to a size of 10048^3 grid elements, were produced on the supercomputers of the Leibniz Supercomputing Centre and visualized using the hybrid parallel (MPI + TBB) ray-tracing engine OSPRay associated with VisIt. Besides revealing features of turbulence with an unprecedented resolution, the visualizations brilliantly showcase the stretching-and-folding mechanisms through which astrophysical processes such as supernova explosions drive turbulence and amplify the magnetic field in the interstellar gas, and how the first structures, the seeds of newborn stars are shaped by this process.

An Accessible Visual Narrative for the Primary Energy Source of Life from the Fulldome Show Birth of Planet Earth

Melih Sener (University of Illinois), Stuart Levy (University of Illinois), AJ Christensen (University of Illinois), Robert Patterson (University of Illinois), Kalina Borkiewicz (University of Illinois), John E. Stone (University of Illinois), Barry Isralewitz (University of Illinois), Jeffrey Carpenter (University of Illinois)
Conversion of sunlight into chemical energy, namely photosynthesis, is the primary energy source of life on Earth. An explanatory visualization depicting this process is presented in the form of an excerpt from the fulldome show Birth of Planet Earth. This accessible visual narrative shows a lay audience, especially children, how the energy of sunlight is captured, converted, and stored through a chain of proteins to power living cells. The visualization is the result of a multi-year collaboration among biophysicists, visualization scientists, and artists, which, in turn, is based on a decade-long experimental-computational collaboration on structural and functional modeling that produced an atomic detail description of a bacterial bioenergetic organelle, the chromatophore. The energy conversion steps depicted feature an integration of function from electronic to cell levels, spanning nearly 12 orders of magnitude in time scales modeled with multi-scale computational approaches. This atomic detail description uniquely enables a modern retelling of one of humanity’s earliest stories---the interplay between light and life.

Visualizing Supersonic Retropropulsion for Mars: The Summit Early Science Results
Patrick Moran (NASA Ames Research Center), Timothy Sandstrom (NASA Ames Research Center), Ashley Korzun (NASA Langley Research Center), Eric Nielsen (NASA Langley Research Center), Aaron Walden (NASA Langley Research Center)

As part of the Oak Ridge Leadership Computing Facility (OLCF) Early Science program, NASA ran a series of simulations on the new, GPU-accelerated system Summit, using the flow solver FUN3D to produce solutions for a set of representative descent conditions. The raw output from these simulations was over a petabyte of data. We produced visualizations at the NASA Advanced Supercomputing Division at Ames Research Center using parallel techniques where feasible to accelerate the process. Included with this summary is an animation with some of the highlights of our visualizations so far. We discuss some of our visualization design choices, and plans for future work.

Friday, November 22
8:30 am - 12:00 pm

Scientific Visualization & Data Analytics Showcase Posters Display

Visualization of Entrainment and Mixing Phenomena at Cloud Edges
Entrainment and mixing processes occur during the entire life of a cloud. These processes change the droplet size distribution, which determines rain formation and radiative properties. Since it is a microphysical process, it cannot be resolved in large scale weather forecasting models. Small scale simulations such as Direct Numerical Simulations (DNS) are required to resolve the smallest scale of these processes. However, it has been a challenge to visualize these processes in a 3D domain as it generates petabytes of data. Visualization plays a vital role in analyzing and understanding these huge data outputs. Here, we present different techniques for 3D visualization of data obtained from DNS carried out at Indian Institute of Tropical Meteorology (IITM) to understand cloud microphysical properties more closely.

Multiple tools were used to conduct a visual analysis of this data. Two of these tools are well established and tested technologies: ParaView and VAPOR. The others are emergent technologies created at the National Center for Atmospheric Research (NCAR) and are in the development phase. This data simulation, in addition to exploring DNS as mentioned above, provided an opportunity to test and improve development of several tools and methods.

Visualizing Deep Learning at Scale for the Construction of Galaxy Catalogs in the Dark Energy Survey
Janet Y. K. Knowles (Argonne National Laboratory), Joseph A. Insley (Argonne National Laboratory, Northern Illinois University), Silvio Rizzi (Argonne National Laboratory), Elise Jennings (Argonne National Laboratory), Asad Khan (University of Illinois), Eliu Huerta (University of Illinois), Sibo Wang (University of Illinois), Robert Gruendl (University of Illinois), Huihuo Zheng (Argonne National Laboratory)

The advent of machine and deep learning algorithms on petascale supercomputers is accelerating the pace of discovery in astrophysics and poses significant challenges to the interpretability of these deep neural networks. We present a novel visualization of a deep neural network output during training as it is learning to classify galaxy images as either spiral or elliptical. The network is trained using labeled datasets from the citizen science campaign, Galaxy Zoo, adopted by the Sloan Digital Sky Survey. These trained neural network models can then be used to classify galaxies in the Dark Energy Survey that overlap the footprint of both surveys. Visualizing a reduced representation of
the network output, projected into 3-D parameter space, reveals how the network has discovered two distinct clusters of features which allows it to classify galaxies into two groups. These visualizations of the neural network during training aid in the interpretability of the black box of deep learning and reveal how the network responds to the input images at various stages of training. Finally, it allows a wider net to be cast to a general audience, thereby generating interest in and visibility to an otherwise highly specialized field.

**NVIDIA IndeX Accelerated Computing for Visualizing Cholla's Galactic Winds**
Christopher Lux (Nvidia Corporation), Alexander Kuhn (Nvidia Corporation), Evan Schneider (Princeton University), Brant Robertson (University of California, Santa Cruz), Marc Nienhaus (Nvidia Corporation)

Galactic winds – outflows of gas driven out of galaxies by the combined effects of thousands of supernovae – are a crucial feature of galaxy evolution. Despite their importance, a complete theoretical picture of these winds has been elusive. Simulating the complicated interaction between the hot, high pressure gas created by supernovae and the cooler, high density gas in the galaxy disk requires massive computational resources and sophisticated software. For this purpose, Computational Hydrodynamics On Parallel Architectures (Cholla) has been demonstrated to be a scalable and efficient tool that operates in large, distributed multi-GPU environments at high levels of performance. This additional performance results in unprecedented resolution for this type of simulation and massive amounts of generated data. This raises the requirements for analysis tools that can cope with scale and complexity of the simulated physical processes. To address those challenges, we utilize NVIDIA IndeX as a scalable framework to visualize the simulation output: NVIDIA IndeX features a streaming-based architecture to interactively explore simulation results in large-scale, multi-GPU environments. We utilize customized sampling programs for multi-volume and surface rendering to address analysis questions of galactic wind simulations. This combination of massively parallel simulation and analysis allows us to utilize recent supercomputer capabilities and to speed up the exploration of galactic wind simulations.

**Visualizing the World’s Largest Turbulence Simulation**
Salvatore Cielo (Leibniz Supercomputing Centre), Luigi Iapichino (Leibniz Supercomputing Centre), Johannes Günther (Intel Corporation), Christoph Federrath (Australian National University, Research School of Astronomy and Astrophysics), Elisabeth Mayer (Leibniz Supercomputing Centre), Markus Wiedemann (Leibniz Supercomputing Centre)

We present the visualization of the largest interstellar turbulence simulations ever performed, unravelling key astrophysical processes concerning the formation of stars and the relative role of
magnetic fields. The simulations, including pure hydrodynamical (HD) and magneto-hydrodynamical (MHD) runs, up to a size of $10048^3$ grid elements, were produced on the supercomputers of the Leibniz Supercomputing Centre and visualized using the hybrid parallel (MPI + TBB) ray-tracing engine OSPRay associated with VisIt. Besides revealing features of turbulence with an unprecedented resolution, the visualizations brilliantly showcase the stretching-and-folding mechanisms through which astrophysical processes such as supernova explosions drive turbulence and amplify the magnetic field in the interstellar gas, and how the first structures, the seeds of newborn stars are shaped by this process.

An Accessible Visual Narrative for the Primary Energy Source of Life from the Fulldome Show Birth of Planet Earth
Melih Sener (University of Illinois), Stuart Levy (University of Illinois), AJ Christensen (University of Illinois), Robert Patterson (University of Illinois), Kalina Borkiewicz (University of Illinois), John E. Stone (University of Illinois), Barry Isralewitz (University of Illinois), Jeffrey Carpenter (University of Illinois), Donna Cox (University of Illinois)

Conversion of sunlight into chemical energy, namely photosynthesis, is the primary energy source of life on Earth. An explanatory visualization depicting this process is presented in the form of an excerpt from the fulldome show Birth of Planet Earth. This accessible visual narrative shows a lay audience, especially children, how the energy of sunlight is captured, converted, and stored through a chain of proteins to power living cells. The visualization is the result of a multi-year collaboration among biophysicists, visualization scientists, and artists, which, in turn, is based on a decade-long experimental-computational collaboration on structural and functional modeling that produced an atomic detail description of a bacterial bioenergetic organelle, the chromatophore. The energy conversion steps depicted feature an integration of function from electronic to cell levels, spanning nearly 12 orders of magnitude in time scales modeled with multi-scale computational approaches. This atomic detail description uniquely enables a modern retelling of one of humanity’s earliest stories---the interplay between light and life.

Visualizing Supersonic Retropropulsion for Mars: The Summit Early Science Results
Patrick Moran (NASA Ames Research Center), Timothy Sandstrom (NASA Ames Research Center), Ashley Korzun (NASA Langley Research Center), Eric Nielsen (NASA Langley Research Center), Aaron Walden (NASA Langley Research Center)

As part of the Oak Ridge Leadership Computing Facility (OLCF) Early Science program, NASA ran a series of simulations on the new, GPU-accelerated system Summit, using the flow solver FUN3D to produce solutions for a set of representative descent conditions. The raw output from these
simulations was over a petabyte of data. We produced visualizations at the NASA Advanced Supercomputing Division at Ames Research Center using parallel techniques where feasible to accelerate the process. Included with this summary is an animation with some of the highlights of our visualizations so far. We discuss some of our visualization design choices, and plans for future work.
Tuesday, November 19

10:00 am - 6:00 pm

SC Theater

Hybrid Quantum-Classical Algorithms for Graph Problems: Forging a Path to Near-Term Applications
Ruslan Shaydulin (Clemson University)

Technology Challenge Introduction and Jury Introduction

Technology Challenge UEN Demo

Technology Challenge RENCI Demo Prep

Technology Challenge RENCI Demo

Break

SCinet Architecture

Technology Challenge ANL Demo Prep

Technology Challenge ANL Demo

Technology Challenge Jury Closing Remarks

Student Cluster Challenge "How We Got Here"

Student Cluster Challenge "How We Got Here"

Student Cluster Challenge "How We Got Here"
Modern HPC approaches to solve time dependent Quantum Mechanical equations

Konstantin Rygol (University of Bergen)

Wednesday, November 20
10:00 am - 6:00 pm
SC Theater

Thursday, November 21
10:00 am - 3:00 pm
SC Theater
Student Cluster Competition

(back to top)

Monday, November 18
7:30 pm - 8:00 pm
Student Cluster Competition Kick-Off

Session Description: Join us for the kick-off of the Student Cluster Competition at Booth 1299 in the exhibit hall.

Tuesday, November 19
10:00 am - 6:00 pm
Student Cluster Competition

Session Description: View the progress of the Student Cluster Competition at Booth 1299 in the exhibit hall.

Wednesday, November 20
10:00 am - 6:00 pm
Student Cluster Competition

Session Description: It's the final day of the Student Cluster Competition at Booth 1299 in the exhibit hall! Come see how the teams are doing, as the time ticks down.
Thursday, November 21

10:00 am - 3:00 pm

Student Cluster Competition

Session Description: Talk to the Student Cluster Competition teams at Booth 1299 in the exhibit hall about how the competition went.
Student Job/Opportunity Fair

Wednesday, November 20

10:00 am - 3:00 pm

Student/Postdoc Job Fair

All students attending SC19 are invited to attend the Student/Postdoc Job Fair. Hiring organizations will be present from 10am-3pm. Students may stop by at any time.
Students@SC

(back to top)

Sunday, November 17

8:30 am - 8:45 am

Decoding SC - How to Maximize Your Time at the Conference

**Session Description:** This session is open to all students attending the SC19 conference. During this session, we will highlight aspects of the conference that we think will be of particular interest and benefit to our student attendees. Please join us to learn how to maximize your time at the conference.

8:45 am - 9:15 am

Students@SC: Keynote Address "Delivering on the Exascale Computing Project Mission for the U.S. Department of Energy"

Students@SC: Keynote Address "Delivering on the Exascale Computing Project Mission for the U.S. Department of Energy"

Doug Kothe (Oak Ridge National Laboratory, US Department of Energy Exascale Computing Project)

We invite all students to attend the Students@SC Keynote talk given by Doug Kothe, the Director of the US Department of Energy (DOE) Exascale Computing Project (ECP). The Exascale Computing Project was initiated in 2016 to accelerate innovation with exascale simulation and data science solutions that enhance U.S. economic competitiveness, strengthen our national security, and change our quality of life.

ECP’s mission is to deliver exascale-ready applications and solutions that address currently intractable problems of strategic importance and national interest; create and deploy an expanded and vertically integrated software stack on DOE HPC exascale and pre-exascale systems, thereby defining the enduring US exascale ecosystem; and leverage U.S. HPC vendor research activities
and products into DOE HPC exascale systems.

Join us for this exciting talk to hear about DOE’s path to exascale.

9:15 am - 10:30 am

Careers in HPC Panel

Careers in HPC Panel
Eric Nielsen (NASA), Fernanda Foertter (Nvidia Corporation), Doug Kothe (Oak Ridge National Laboratory), Kate Evans (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. We invite all students attendees to join us for this lively discussion.

10:30 am - 12:30 pm

Parallel Algorithm Design

Parallel Algorithm Design
Rebecca Hartman-Baker (Lawrence Berkeley National Laboratory)
We invite all SC19 student attendees to join us to learn more about designing scalable parallel algorithms for high-performance computing. We will begin with an overview of parallelization concepts and parallel computer architectures, followed by a discussion of the elements of parallel algorithms and strategies for designing them. In addition to discussing concepts, we will present real-life problems and solutions and the implications of future architectures on parallel algorithm design.

1:30 pm - 3:00 pm

Perfecting Your Elevator Speech and Growing Your Professional Network

Session Description:
Perfecting Your Elevator Speech and Growing Your Professional Network
Betsy Hillery (Purdue University)
All SC19 student attendees are invited to this session that is designed to help attendees develop and perfect your elevator speech. Introducing yourself well sets the stage for a professional conversation, whether that’s at a networking event, with a colleague or at the beginning of an interview. In this session, you will learn how to use one tool, called an elevator speech, that you can utilize to help make these introductions both simple and effective. A personal elevator speech is a quick summary of yourself. It’s named for the time it takes to ride an elevator from bottom to top of a building. An elevator speech is useful to have ready at conferences such as SC where you have many opportunities to network as well as in future job interviews. From the phone screen to in-person interviews, you’ll be asked to provide a summary of who you are, your background and what you want from your next job. Come learn how to deliver an effective elevator speech.

*Students are highly encouraged to bring a copy of your resume/CV to assist in the development of your elevator speech.*

3:00 pm - 5:00 pm
Resume Workshop

**Session Description:** Students will have an opportunity to consult with professionals from industry, national labs, and academia to get feedback on their résumés and curriculum vitae in advance of the Student/Postdoc Job Fair that takes place on Wednesday, November 20, from 10:00 am to 3:00 pm. To get the most out of the session, we recommend students bring a copy of your résumé or curriculum vitae to share with a mentor. Students can also use this session to consult with mentors to learn more about the Student/Postdoc Job Fair to help maximize students’ experience at this event.

Monday, November 18

9:00 am - 10:00 am
How to Create and Deliver Effective Research Posters

**How to Create and Deliver Effective Research Posters**
Seetharami Seelam (IBM Corporation), Kathryn Mohror (Lawrence Livermore National Laboratory)
The ability to create effective poster presentations is an important skill. Academic posters, when one effectively, are a succinct and attractive way to showcase your work at conferences and meetings. Successful posters can generate discussion amongst the audience members and therefore, it is important to have a clear plan of what to say when presenting your poster. This interactive session will help attendees better understand how to both create and present an effective research poster. The SC19 Poster Chair, Seetharami Seelam, and Research Posters Track Chair, Kathryn Mohror will lead this interactive session. All SC19 student attendees are invited to participate.

11:00 am - 1:00 pm

Resume Workshop

Session Description: Students will have an opportunity to consult with professionals from industry, national labs, and academia to get feedback on their résumés and curriculum vitae in advance of the Student/Postdoc Job Fair that takes place on Wednesday, November 20, from 10:00 am to 3:00 pm. To get the most out of the session, we recommend students bring a copy of your résumé or curriculum vitae to share with a mentor. Students can also use this session to consult with mentors to learn more about the Student/Postdoc Job Fair to help maximize students’ experience at this event.

2:30 pm - 5:00 pm

Modern Software Design, Tools, and Practices

Modern Software Design, Tools, and Practices
Elsa Gonsiorowski (Lawrence Livermore National Laboratory), Reed Milewicz (Sandia National Laboratories)
Interested in a career in supercomputing? If yes, the chances are very good that you will interact in some way with software. Software has become a foundation of discovery in computational science and engineering and faces increasing complexity in computational models and computer architectures. This session is open to all students attending SC and will introduce attendees to modern software design, tools, and practices that will be applicable to students across many different disciplines.
The featured speakers include Elsa Gonsiorowski of Lawrence Livermore National Laboratory and Reed Milewicz of Sandia National Laboratories. Both are members of the Interoperable Design of Extreme-scale Application Software (IDEAS) project, which is part of the Exascale Computing Project (ECP). IDEAS-ECP serves to help ease the challenges of software development in this environment, and to help the development teams ensure that DOE investment in the exascale software ecosystem is as productive and sustainable as possible.

**Tuesday, November 19**

10:00 am - 6:00 pm

**SC Theater**

Hybrid Quantum-Classical Algorithms for Graph Problems: Forging a Path to Near-Term Applications  
Ruslan Shaydulin (Clemson University)

Technology Challenge Introduction and Jury Introduction

Technology Challenge UEN Demo

Technology Challenge RENCI Demo Prep

Technology Challenge RENCI Demo

Break

SCinet Architecture

Technology Challenge ANL Demo Prep

Technology Challenge ANL Demo

Technology Challenge Jury Closing Remarks

Student Cluster Challenge "How We Got Here"

Student Cluster Challenge "How We Got Here"
Student Cluster Challenge "How We Got Here"

Student Cluster Challenge "How We Got Here"

Student Cluster Challenge "How We Got Here"

Student Cluster Challenge "How We Got Here"

Student Cluster Challenge "How We Got Here"

Student Cluster Challenge "How We Got Here"

SCinet Routing

Live Interview Video Broadcast

Live Interview Video Broadcast

Open

Modern HPC approaches to solve time dependent Quantum Mechanical equations
Konstantin Rygol (University of Bergen)

Wednesday, November 20

10:00 am - 3:00 pm

Student/Postdoc Job Fair

Student/Postdoc Job Fair

All students attending SC19 are invited to attend the Student/Postdoc Job Fair. Hiring organizations will be present from 10am-3pm. Students may stop by at any time.
10:00 am - 6:00 pm

SC Theater

Thursday, November 21
10:00 am - 3:00 pm
SC Theater

Friday, November 22
8:00 am - 12:00 pm

Hands-On with the Summit Supercomputer

Hands-On with the Summit Supercomputer
Thomas Papatheodore (Oak Ridge National Laboratory), Jack Morrison (Oak Ridge National Laboratory), Ashley Barker (Oak Ridge National Laboratory)
This session will provide an introduction to the Summit supercomputer at Oak Ridge National Laboratory and give SC19 student attendees the opportunity to explore its use. Attendees of this session will be provided accounts on the machine, led through hands-on material to become familiar with the computing and programming environment, then challenged to apply their skills in competition with one another. Come discover what it’s like to use one of the most powerful computers ever built!

*Note: Students need to bring a laptop to participate.*
Technology Challenge

Tuesday, November 19

10:00 am - 6:00 pm

SC Theater

Hybrid Quantum-Classical Algorithms for Graph Problems: Forging a Path to Near-Term Applications

Ruslan Shaydulin (Clemson University)

Technology Challenge Introduction and Jury Introduction

Technology Challenge UEN Demo

Technology Challenge RENCI Demo Prep

Technology Challenge RENCI Demo

Break

SCinet Architecture

Technology Challenge ANL Demo Prep

Technology Challenge ANL Demo

Technology Challenge Jury Closing Remarks

Student Cluster Challenge "How We Got Here"

Student Cluster Challenge "How We Got Here"

Student Cluster Challenge "How We Got Here"
SCinet Routing

Live Interview Video Broadcast

Open

Modern HPC approaches to solve time dependent Quantum Mechanical equations
Konstantin Rygol (University of Bergen)

Wednesday, November 20
10:00 am - 6:00 pm
SC Theater

Thursday, November 21
10:00 am - 3:00 pm
SC Theater
Test of Time

(back to top)

Tuesday, November 19

3:30 pm - 4:15 pm

Test of Time Award Presentation

From Dense Linear Algebra to GPU Occupancy
Vasily Volkov (Nvidia Corporation), James Demmel (University of California, Berkeley)

The CUDA programming model was introduced in 2007 and featured a number of new concepts, such as occupancy and shared memory. In this work, we considered performance implications of these concepts in application to dense matrix factorizations. Our findings were contrarian to the widely accepted recommendations of the day. (i) We found a performance optimization pattern that leads to lower occupancy, whereas it was recommended to maximize occupancy in order to hide memory latencies. (ii) We found that instruction-level parallelism contributes to latency hiding on GPUs, which was believed to be not the case. (iii) We found that performance can be improved by using massive register blocking, whereas it was recommended to minimize register use to maximize occupancy. (iv) We found that shared memory is slower than registers and the use of the latter should be favored over the former, if possible. These novel insights led us to a design of the matrix multiply routine that substantially outperformed the state-of-the-art vendor BLAS library. The optimization pattern we pointed out is found today in many high-performance GPU codes.
Containers in HPC

Brian Skjerven (Pawsey Supercomputing Center), Mark Gray (Pawsey Supercomputing Center)

No longer an experimental topic, containers are here to stay in HPC. They offer software portability, improved collaboration, and data reproducibility. A variety of tools (e.g. Docker, Shifter, Singularity, Podman) exist for users who want to incorporate containers into their workflows, but oftentimes they may not know where to start.

This tutorial will cover the basics of creating and using containers in an HPC environment. We will make use of hands-on demonstrations from a range of disciplines to highlight how containers can be used in scientific workflows. These examples will draw from Bioinformatics, Machine Learning, Computational Fluid Dynamics and other areas.

Through this discussion, attendees will learn how to run GPU- and MPI-enabled applications with containers. We will also show how containers can be used to improve performance in Python workflows and I/O-intensive jobs.

Lastly, we will discuss best practices for container management and administration. These practices include how to incorporate good software engineering principles, such as the use of revision control and continuous integration tools.


Dhabaleswar Panda (Ohio State University), Ammar Ahmad Awan (Ohio State University), Hari Subramoni (Ohio State University), Ching-Hsiang Chu (Ohio State University)
The recent advances in Deep Learning (DL) has led to many exciting challenges and opportunities for Computer Science and AI researchers alike. Modern DL frameworks like TensorFlow, PyTorch, Cognitive Toolkit, Caffe2, and several others have emerged that offer ease of use and flexibility to train, and deploy various types of Deep Neural Networks (DNNs).

In this tutorial, we will provide an overview of interesting trends in DNN design and how cutting-edge hardware architectures are playing a key role in moving the field forward. We will also present an overview of different DNN architectures and DL frameworks. Most DL frameworks started with a single-node/single-GPU design. However, approaches to parallelize the process of DNN training are also being actively explored. The DL community has moved along different distributed training designs that exploit communication runtimes like gRPC, MPI, and NCCL. We will highlight new challenges and opportunities for communication runtimes to efficiently support distributed DNN training. We also highlight some of our co-design efforts to utilize CUDA-aware MPI for large-scale DNN training on modern GPU clusters. Finally, we include hands-on exercises to enable the attendees to gain first-hand experience of running distributed DNN training experiments on a modern GPU cluster.

8:30 am - 12:00 pm

Using the SPEC HPG Benchmarks for Better Analysis and Evaluation of Current and Future HPC Systems

Robert Henschel (Indiana University), Junjie Li (Indiana University), Verónica G. Melesse Vergara (Oak Ridge National Laboratory), Mayara Gimenes (University of Delaware), Sandra Wienke (RWTH Aachen University)

The High Performance Group (HPG) of the Standard Performance Evaluation Corporation (SPEC) develops benchmark methodologies for High Performance Computing systems. The group also releases production quality benchmark suites like SPEC MPI2007, SPEC OMP2012, and SPEC ACCEL, that can evaluate all dimensions of parallelism. These benchmarks are used in academia and industry to conduct research in HPC systems, facilitate procurement, testing, and tuning of HPC systems. In order to make these benchmarks more available, SPEC HPG offers these benchmark suites free of charge to non-commercial users. In this half-day tutorial, participants will learn how to leverage SPEC benchmarks for performance evaluation, tuning of system parameters, comparison of systems (e.g., for procurement), and get an outlook on its power measurement capabilities. The presenters will provide demos and hands-on guidance on how to install, compile, and run the benchmarks on HPC systems provided. The presenters will also show how the results will be interpreted, discuss various use cases, and present the publication process of results. An
SSH-capable device is required for the hands-on sessions.

8:30 am - 12:00 pm

**Boosting Power Efficiency of HPC Applications with GEOPM**

Siddhartha Jana (Energy Efficient HPC Working Group), Martin Schulz (Technical University Munich, Leibniz Supercomputing Centre), Tapasya Patki (Lawrence Livermore National Laboratory)

Power and energy are critical constraints for exascale supercomputing. Optimizing for application performance under such constraints is becoming challenging due to the dynamic phase behavior of scientific codes, increasing variation in processor power efficiency resulting from manufacturing, and due to complexities arising from upcoming heterogeneous architectures.

In order to address some of these challenges, this half-day hands-on tutorial will discuss GEOPM (Global Extensible Open Power Manager) as an emerging, cross-community, cross-platform, open source (BSD 3-clause), job-level power management framework that is capable of providing feedback to job schedulers and resource managers in order to drive system power efficiency improvements. This framework is part of OpenHPC and a key pillar of the software stack in the upcoming first exascale system at Argonne National Laboratory in the US. It optimizes for time-to-solution by leveraging techniques from learning and control systems. GEOPM leverages hardware monitoring and control knobs provided by modern processors.

In this hands-on tutorial, we will discuss the state-of-the-art power management techniques and show the areas where GEOPM can provide benefit over and above those techniques. A high-level overview of the GEOPM architecture, a walkthrough of the GEOPM plugin infrastructure, and example exercises for hands-on practice will be demoed.

8:30 am - 5:00 pm

**Best Practices for HPC in the Cloud**

Pierre-Yves Aquilanti (Amazon Web Services), Nina Vogl (Amazon Web Services), Arthur Petitpierre (Amazon Web Services), Sean Smith (Amazon Web Services), Matt Koop (Amazon Web Services)

Cloud computing technologies have matured to the point that nearly all HPC production workloads can be efficiently accommodated in a cloud environment. Cloud computing presents a nearly
unrivaled completeness of services and the ability to accommodate the most demanding computational workloads at scale. However, the complexity and scale that comes with such an environment also can make the first experience a daunting proposition. This is where this tutorial offers “solution through training” explanations and hands-on experience in a safe environment.

In the first part, we will provide an overview of computational, storage, and network technologies offered by AWS and demonstrate how to best utilize them to run HPC workflows. In the second part, we will review the different computational methods such as auto-scaling, serverless computing with an emphasis on optimizing applications for a cloud environment. In the third part, we will cover specific use cases, techniques, and optimizations to run application at extreme scale.

This tutorial will be composed of presentations and hands-on sessions where attendees will have the opportunity to put into practice their learnings on the AWS cloud. A laptop/tablet and a shell with Bash are required for the hands-on portions of this tutorial.

8:30 am - 5:00 pm

Introduction to Quantum Computing

Scott Pakin (Los Alamos National Laboratory), Eleanor G. Rieffel (NASA Ames Research Center)

Quantum computing offers the potential to revolutionize high-performance computing by providing a means to solve certain computational problems asymptotically faster than any classical computer. Relatively recently, quantum computing has advanced from merely a theoretical possibility to engineered reality, including commercial entities offering early prototype quantum processors, both special-purpose quantum annealers and general-purpose gate-model processors. The media has been showcasing each new development and implicitly conveying the message that quantum-computing ubiquity is nigh. Here, we will respond to this hype and provide an overview of the exciting but still early state of the field.

In this tutorial, we introduce participants to the computational model that gives quantum computing its immense computational power. We examine the thought processes that programmers need to map problems onto the two dominant quantum-computing architectures. And we discuss the current hardware and algorithmic challenges that must be overcome before quantum computing becomes a component of the HPC developer’s repertoire.

8:30 am - 5:00 pm
Parallel I/O in Practice

Robert Latham (Argonne National Laboratory, Mathematics and Computer Science Division), Robert Ross (Argonne National Laboratory), Brent Welch (Google LLC), Glenn Lockwood (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 HDF5). 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, HDFS, 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.

8:30 am - 5:00 pm

OpenMP Common Core: A “Hands-On” Exploration

Tim Mattson (Intel Corporation), Alice Koniges (University of Hawaii), Yun (Helen) He (National Energy Research Scientific Computing Center (NERSC)), David Eder (Maui High Performance Computing Center)

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 in the market. It’s also very complicated. We created OpenMP to be the “easy API” for application programmers. 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 20 items from OpenMP. We call this subset 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 OpenMP Common Core. We utilize active learning through a carefully selected set of exercises, so students master the common core and learn to apply it to their own problems. The exercises use a simple subset of C: Fortran programmers will have no problem keeping up.

Students will use their own laptops 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.

8:30 am - 5:00 pm

**Node-Level Performance Engineering**

Georg Hager (University of Erlangen-Nuremberg, Erlangen Regional Computing Center), Gerhard Wellein (University of Erlangen-Nuremberg, Department of Computer Science; Erlangen Regional Computing Center)

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.

8:30 am - 5:00 pm

**Secure Coding Practices and Automated Assessment Tools**

Elisa Heymann (University of Wisconsin), Barton Miller (University of Wisconsin)
High performance computing increasingly involves the development and deployment of network and cloud services to access resources for computation, communication, data, instruments, and analytics. Unique to the HPC field is the large amount of software that we develop to drive these services. These services must assure data integrity and availability, while providing access to a global scientific and engineering community.

Securing your network is not enough. Every service that you deploy is a window into your data center from the outside world, and a window that could be exploited by an attacker. This tutorial is relevant to anyone wanting to learn about minimizing security flaws in the software they develop or manage. You will learn skills critical for software developers and analysts concerned with security.

Software assurance tools – tools that scan the source or binary code of a program to find weaknesses – are the first line of defense in assessing the security of a software project. These tools can catch flaws in a program that affect both the correctness and safety of the code. This tutorial is also relevant to anyone wanting to learn how to use these automated assessment tools to minimize security flaws.

8:30 am - 5:00 pm

In Situ Analysis and Visualization with SENSEI and Ascent

Hank Childs (University of Oregon), Cyrus Harrison (Lawrence Livermore National Laboratory), Matthew Larsen (Lawrence Livermore National Laboratory), Burlen Loring (Lawrence Berkeley National Laboratory), Silvio Rizzi (Argonne National Laboratory), Brad Whitlock (Intelligent Light), David Rogers (Los Alamos National Laboratory), David Thompson (Kitware Inc)

A key trend facing extreme-scale computational science is the widening gap between computational and I/O rates. The challenge 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 the idea of in situ processing, where one performs visualization and analysis processing while data is still resident in memory. The HPC visualization community provides several production tools for in situ processing and is actively researching infrastructures and approaches to help with in situ use cases on current and future HPC platforms.

This tutorial blends lectures and hands-on examples to introduce attendees to in situ processing concepts and provide practical experience using state-of-the-art and -practice in situ tools, including: SENSEI, ParaView Catalyst, VisIt Libsim, ADIOS, and ALPINE Ascent. Attendees will learn how to use the SENSEI and Ascent in situ interfaces to couple simulation codes to multiple
endpoints. This tutorial is ideal for those who would like to learn how to analyze, visualize, or process data on HPC platforms while it is still resident in memory.

8:30 am - 5:00 pm

Tools and Techniques for Programming GPUs

Christian Trott (Sandia National Laboratories), Mark Gates (University of Tennessee), Siva Rajamanickam (Sandia National Laboratories), Jakub Kurzak (University of Tennessee)

At this point, GPUs are an integral part of high-performance computing, and GPU acceleration seems to be the only viable path for reaching exascale. Under such circumstances, it is critical for computational scientists to invest in GPU programming skills if they want to be competitive in their field. In this tutorial, we attempt to cover the full spectrum of GPU kernel-development techniques. We start by introducing the basics of GPU programming with CUDA and HIP and then shift the focus to solutions aimed at portability: OpenMP, OpenACC, Kokkos, and KokkosKernels. We also include an extensive presentation of automated performance tuning methodologies. Our objective is twofold: (1) prevent younger researchers from reinventing the wheel when it comes to producing GPU-accelerated numerical software and (2) steer the attendees away from one-off solutions and motivate them to invest their efforts in more portable and more sustainable approaches.

8:30 am - 5:00 pm

Parallel Computing 101

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 machine learning problems. These examples illustrate using MPI on distributed memory systems; OpenMP on shared memory systems; MPI+OpenMP on hybrid systems; and CUDA and compiler directives on GPUs and accelerators. 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.

8:30 am - 5:00 pm

Floating-Point Analysis and Reproducibility Tools for Scientific Software

Ignacio Laguna (Lawrence Livermore National Laboratory), Ganesh Gopalakrishnan (University of Utah), Michael Bentley (University of Utah), Ian Briggs (University of Utah), Michael O. Lam (James Madison University), Cindy Rubio-Gonzalez (University of California, Davis)

While scientific software is widely used in several science and engineering disciplines, developing accurate and reliable scientific software is notoriously difficult. One of the most serious difficulties comes from dealing with floating-point arithmetic to perform numerical computations. Round-off errors occur and accumulate at all levels of computation, and compiler optimizations and low precision arithmetic can significantly affect the final computational results. With accelerators dominating high-performance computing systems, computational scientists are faced with even bigger challenges, given that ensuring numerical reproducibility in these systems pose a very difficult problem.

This tutorial will demonstrate tools that are available today to analyze floating-point scientific software. We focus on tools that allow programmers to get insight about how different aspects of floating-point arithmetic affects their code and how to fix potential bugs. Some of the floating-point analysis areas that we cover in the tutorial are compiler optimizations, floating-point exceptions on GPUs, precision tuning, sensitivity to rounding errors, non-determinism, and data races.

8:30 am - 5:00 pm

Mastering Tasking with OpenMP

Christian Terboven (RWTH Aachen University), Michael Klemm (Intel Corporation), 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. We will include hands-on sessions. Throughout all topics, we present the features of OpenMP 4.5 and the additions that appeared in OpenMP 5.0, released during SC18.

1:30 pm - 5:00 pm

**Container Computing for HPC and Scientific Workflows**

Richard Shane Canon (Lawrence Berkeley National Laboratory), Andrew J. Younge (Sandia National Laboratories), Sameer Shende (University of Oregon), Eduardo Arango (Sylabs Inc)

Container computing has revolutionized 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, Shifter, Singularity and Charliecloud 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 various container runtimes. 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 updated version of the highly successful tutorial presented at SC16, SC17, SC18. It was attended by more than 100 people each year. The 2018 tutorial was very highly rated with 2.8 / 3 stars for "would recommend" and 4.3 / 5 stars for overall quality.

1:30 pm - 5:00 pm
Tools and Best Practices for Distributed Deep Learning on Supercomputers

Xu Weijia (Texas Advanced Computing Center (TACC)), Zhao Zhang (University of Texas), David Walling (University of Texas)

This tutorial is a practical guide on how to run distributed deep learning over multiple compute nodes effectively. Deep Learning (DL) has emerged as an effective analysis method and has adapted quickly across many scientific domains in recent years. Domain scientists are embracing DL as both a standalone data science method and an effective approach to reducing dimensionality in the traditional simulation. However, due to its inherent high computational requirement, application of DL is limited by the available computational resources. Recently, we have seen the fusion of DL and high-performance computing (HPC): supercomputers show an unparalleled capacity to reduce DL training time from days to minutes; HPC techniques have been used to speed up parallel DL training. Therefore, distributed deep learning has great potential to augment DL applications by leveraging existing high performance computing cluster.

This tutorial consists of three sessions. First, we will give an overview of the state-of-art approaches to enabling deep learning at scale. The second session is an interactive hands-on session to help attendees running distributed deep learning on Frontera at the Texas Advanced Computing Center. In the last session, we will focus on the best practices on how to scale, evaluate, and tune up performance.

1:30 pm - 5:00 pm

High Performance Computing (HPC) Data Center Planning and TCO: A Case Study and Roadmap

Michael P. Thomas (kW Mission Critical Engineering), Anna Maria Bailey (Lawrence Livermore National Laboratory, Energy Efficient HPC Working Group), David Lambiaso (Environmental Systems Design (ESD))

Building upon the “Data Center Planning and Design” Tutorials presented at SC18, SC17 and SC16 and attendees’ feedback, this tutorial provides best practices and lessons learned gleaned from multiple HPC data center infrastructure capacity planning, total cost of ownership (TCO) analysis and business case justification initiatives over the last 15 years and, significantly, provides a deep-dive case study and roadmap around Lawrence Livermore National Laboratory’s (LLNL) methodology to accommodate future HPC systems. The presenters understand that the HPC data center mechanical, electrical and structural infrastructure can be a significant enabler or roadblock to the timely deployment and optimized performance of HPC systems. We are encouraged to see
more TCO/business case dialogue in the HPC community than in years past and will share firsthand experience about how participants can develop/enhance their organizations’ HPC facility infrastructure to match their deployment methodology. Topics covered include identifying and managing stakeholders, defining the mission, developing a facilities infrastructure roadmap, performing HPC vendor surveys, developing program requirements and key performance parameters, following project processes, developing alternative studies, and establishing scope and budget expectations. This tutorial will improve your technical knowledge, project management skills, and financial confidence to navigate this process and drive successful outcomes.

1:30 pm - 5:00 pm

**Productive Parallel Programming for FPGA with High-Level Synthesis**

*Johannes de Fine Licht (ETH Zurich), Torsten Hoefler (ETH Zurich)*

Energy efficiency has become a first class citizen in the design of large computing systems. While GPUs and custom processors have shown merit in this regard, reconfigurable architectures, such as FPGAs, promise another major step in energy efficiency, constituting a middle ground between fixed hardware architectures and custom-built ASICs. Programming FPGAs has traditionally been done in hardware description languages, requiring extensive hardware knowledge and significant engineering effort. This tutorial shows how high-level synthesis (HLS) can be harnessed to productively achieve scalable pipeline parallelism on FPGAs. Attendees will learn how to target FPGA resources from high-level C++ or OpenCL code, guiding the mapping from imperative code to hardware, enabling them to develop massively parallel designs with real performance benefits. We treat well-known examples from the software world, relating traditional code optimizations to both corresponding and new transformations for hardware, building on existing knowledge when introducing new topics. By bridging the gap between software and hardware optimization, our tutorial aims to enable developers from a larger set of backgrounds to start tapping into the potential of FPGAs with real high performance codes.

**Monday, November 18**

8:30 am - 12:00 pm

**InfiniBand, Omni-Path, and High-Speed Ethernet for Beginners**

*Dhabaleswar Panda (Ohio State University), Hari Subramoni (Ohio State University),*
Mohammadreza Bayatpour (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 being widely deployed.

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 libfabrics stack will also be provided. An overview of the emerging NVLink, NVLink2, and NVSwitch architectures will also be given. 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.

8:30 am - 12:00 pm

Fast Parallel Direct Linear Solvers and Preconditioners

Pieter Ghysels (Lawrence Berkeley National Laboratory), Xiaoye Sherry Li (Lawrence Berkeley National Laboratory), Yang Liu (Lawrence Berkeley National Laboratory)

Matrix factorizations and the accompanying solution algorithms (e.g., triangular solution associated with the LU factorization) are often the most robust algorithmic choices for linear systems from multi-physics and multi-scale simulations. They are indispensable tools for building various algebraic equation solvers. They can be used as direct solvers, as coarse-grid solvers in multigrid, or as preconditioners for iterative solvers. As we are approaching the exascale computing era, demand for algorithm innovation is increasingly high. It is imperative to develop optimal-complexity scalable algorithms both in flop count and more importantly in data movement, such as, in the form of communication-avoiding formulations, and low-rank and butterfly compressions. On the software and implementation side, it is imperative to exploit multiple levels of parallelism presented by the heterogeneous node architectures through well orchestrated use of MPI, OpenMP and GPU programming like CUDA.
In this tutorial, we will present our recently developed novel techniques to address scalability gaps. We will demonstrate their efficacies through three solver libraries: SuperLU, STRUMPACK and ButterflyPACK, with representative use cases from simulations and data analytics. Through hands-on exercises, the participants will learn how to use each solver most effectively for their target problems and parallel machines.

8:30 am - 12:00 pm

**Practical Persistent Memory Programming**

Adrian Jackson *(University of Edinburgh)*, Javier Conejero *(Barcelona Supercomputing Center)*

Persistent memory, such as Intel’s Optane DCPMM, is now available for use in systems and will be included in future exascale deployments such as the DoE Aurora system. This new form of memory requires both different programming approaches to exploit the persistent functionality and storage performance and redesign of some applications to benefit from the full performance of the hardware.

This tutorial aims to educate attendees on the persistent memory hardware currently available, the software methods to exploit such hardware, and the choices that users of systems and system designers have when deciding what persistent memory functionality and configurations to utilize.

The tutorial will provide hands-on experience on programming persistent memory along with a wealth of information on the hardware and software ecosystem and potential performance and functionality benefits. We will be using an HPC system that has compute nodes with Optane memory for the tutorial practicals.

8:30 am - 5:00 pm

**Delivering HPC: Procurement, Cost Models, Metrics, Value, and More**


HPC leadership and management skills are essential to the success of HPC. This includes securing funding, procuring the right technology, building effective support teams, ensuring value for money, and delivering a high-quality service to users.
This tutorial will provide practical, experience-based training on delivering HPC. This includes stakeholder management, requirements capture, market engagement, hardware procurement, benchmarking, bid scoring, acceptance testing, total cost of ownership, cost recovery models, metrics, and value.

The presenters have been involved in numerous major HPC procurements in several countries, over three decades, as HPC managers or advisors. The tutorial is applicable to HPC procurements and service delivery in most countries, public or private sector, and is based on experiences from a diversity of real-world cases.

The lead author (Jones) has become the de-facto international leader in delivering training on these topics, with a desire to improve the best practice of the community, and without a sales focus or product to favor.

The SC tutorials by these authors have been strongly attended and highly rated by attendees for several years. For SC19, we have combined our previous multiple half-day tutorials into a single full-day program to reduce duplication and allow for more coherent scheduling of topics.

8:30 am - 5:00 pm

Unified Cyber Infrastructure with Kubernetes

Igor Sfiligoi (San Diego Supercomputer Center), Dmitry Yurievich Mishin (San Diego Supercomputer Center)

Kubernetes has emerged as the leading container orchestration solution over the past few years. Developed at Google, now maintained by Cloud Native Foundation, it sports a very diverse and active development community. While initially developed for the cloud, it can also be deployed on-prem. It also supports federating several independent clusters into much larger logical resource pools, which can span multiple administrative and geographical regions.

One of Kubernetes advantages is its ability to effectively and securely co-schedule service containers alongside user containers. Service applications run with high privileges, with functionality and performance typical of bare metal deployments. User applications are given only basic privileges and limited resources, allowing for both a safe operating environment and fair resource sharing. The Kubernetes framework provides the glue to link them together allowing for a fully functional system.
In this tutorial, the attendees will learn how Kubernetes was used to create a nationwide cyber infrastructure that is serving hundreds of scientific groups through NFS-funded projects like OSG, PRP, TNRP and CHASE-CI. The program includes a Kubernetes architectural overview, hands on sessions operating on the PRP production Kubernetes cluster, and examples of how OSG and PRP are leveraging the Kubernetes cluster for facilitating scientific computing.

8:30 am - 5:00 pm

Deep Learning at Scale

Steven Farrell (Lawrence Berkeley National Laboratory), Mustafa Mustafa (Lawrence Berkeley National Laboratory), Wahid Bhimji (Lawrence Berkeley National Laboratory), Mr Prabhat (Lawrence Berkeley National Laboratory), Michael Ringenburg (Cray Inc), Victor Lee (Intel Corporation), Laurie Stephey (Lawrence Berkeley National Laboratory)

Deep learning is rapidly and fundamentally transforming the way science and industry use data to solve problems. Deep neural network models have been shown to be powerful tools for extracting insights from data across a large number of domains. As these models grow in complexity to solve increasingly challenging problems with larger and larger datasets, the need for scalable methods and software to train them grows accordingly.

The Deep Learning at Scale tutorial aims to provide attendees with a working knowledge on deep learning on HPC class systems, including core concepts, scientific applications, and techniques for scaling. We will provide training accounts, example code, and datasets to allow attendees to experiment hands-on with scalable distributed training and hyperparameter optimization of deep neural network machine learning models.

8:30 am - 5:00 pm

Managing HPC Software Complexity with Spack

Todd Gamblin (Lawrence Livermore National Laboratory), Gregory Becker (Lawrence Livermore National Laboratory), Massimiliano Culpo (Sylabs Inc), Mario Melara (Lawrence Berkeley National Laboratory), Peter Scheibel (Lawrence Livermore National Laboratory), Adam J. Stewart (University of Illinois at Urbana-Champaign)

The modern scientific software stack includes thousands of packages, from C, C++, and Fortran libraries, to packages written in interpreted languages like Python and R. HPC applications may
Spack is an open source tool for HPC package management that simplifies building, installing, customizing, and sharing HPC software stacks. In recent years, its adoption has grown rapidly: by end-users, by HPC developers, and by the world’s largest HPC centers. Spack provides a powerful and flexible dependency model, a simple Python syntax for writing package build recipes, and a repository of over 3,200 community-maintained packages. This tutorial provides a thorough introduction to Spack’s capabilities: installing and authoring packages, integrating Spack with development workflows, and using Spack for deployment at HPC facilities. Attendees will leave with foundational skills for using Spack to automate day-to-day tasks, along with deeper knowledge for applying Spack to advanced use cases.

8:30 am - 5:00 pm

Programming your GPU with OpenMP: A Hands-On Introduction

Tim Mattson (Intel Corporation), Simon McIntosh-Smith (University of Bristol), Eric Stotzer (Mythic)

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 at SC19 the day before) 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. You may also want to load an OpenMP compiler onto their laptops before the tutorial. Information about OpenMP compilers is available at www.openmp.org.
Advanced MPI Programming

William Gropp (University of Illinois), Yanfei Guo (Argonne National Laboratory), 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.

Advanced OpenMP: Host Performance and 5.0 Features

Christian Terboven (RWTH Aachen University), Michael Klemm (Intel Corporation), Ruud van der Pas (Oracle), 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. Based on the feedback from previous years, we have taken out the part on directives for attached compute
accelerators, to focus solely on performance programming for multi-core architectures. Throughout all topics, we present the recent additions of OpenMP 5.0.

8:30 am - 5:00 pm

**Better Scientific Software**

David E. Bernholdt (Oak Ridge National Laboratory), Anshu Dubey (Argonne National Laboratory), Michael A. Heroux (Sandia National Laboratories, St. John’s University), Jared O’Neal (Argonne National Laboratory), Patricia Grubel (Los Alamos National Laboratory), Rinku Gupta (Argonne National Laboratory)

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).

8:30 am - 5:00 pm

**Hands-On Practical Hybrid Parallel Application Performance Engineering**

Markus Geimer (Forschungszentrum Juelich), Sameer Shende (University of Oregon), Matthias Weber (Technical University Dresden), Brian Wylie (Forschungszentrum Juelich)
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 Score-P, 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 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 a contemporary HPC system where remote access 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.

The morning sessions will focus on the tools’ complementary execution analysis capabilities, while the common approach to application instrumentation and measurement will be covered in the afternoon.

8:30 am - 5:00 pm

Fault-Tolerance for High Performance and Big Data Applications: Theory and Practice

George Bosilca (University of Tennessee), Aurelien Bouteiller (University of Tennessee), Thomas Herault (University of Tennessee), Yves Robert (ENS Lyon, University of Tennessee)

Resilience is a critical issue for large-scale platforms. This tutorial provides a comprehensive survey of fault-tolerant techniques for high-performance and big data applications, with a fair balance between theory and practice. 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, silent error detection and correction;
(iii) Application-specific techniques, such as user-level in-memory checkpointing, data replication (map-reduce) or fixed-point convergence for iterative applications (backpropagation); and

(iv) Practical deployment of fault tolerance techniques with User Level Fault Mitigation (a proposed MPI standard extension).

Relevant examples will include widely used routines such as map-reduce and backpropagation in neural networks. A step-by-step approach will show how to protect these routines and make them fault-tolerant, using a variety of techniques, in a hands-on session.

The tutorial is open to all SC19 attendees who are interested in the current status and expected promise of fault-tolerant approaches for scientific and big data 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.

8:30 am - 5:00 pm

**Application Porting and Optimization on GPU-Accelerated POWER Architectures**

Dirk Pleiter (Forschungszentrum Juelich), Christoph Hagleitner (IBM Zurich Research Laboratory), Andreas Herten (Forschungszentrum Juelich), Tom Papatheodore (Oak Ridge National Laboratory), Archana Ravindar (IBM India), Mathias Wagner (Nvidia Corporation)

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 NVLink technology. Two major sites in the US, ORNL and LLNL, deployed their pre-exascale systems based on this new architecture (Summit and Sierra, respectively).

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 NVLink-based node architectures, lectures on first-hand experience in porting to this architecture, and exercises using tools to focus on performance.

8:30 am - 5:00 pm

**High Performance I/O Frameworks 101**
As concurrency and complexity continue to increase on high-end machines, I/O performance is rapidly becoming a fundamental challenge to achieving exascale computing. Wider adoption of higher-level I/O abstractions will be critically important to address this challenge. Modern I/O libraries provide data models, portable APIs, storage abstractions, and self-describing data containers. They achieve high performance and scalability, allow data to be managed more effectively throughout the data lifecycle, and enable reproducible science.

Part I of this tutorial will provide an overview of parallel I/O systems and summarize the key techniques for obtaining high performance I/O on high-performance computing (HPC) resources at scale. Part II introduces ADIOS and HDF5 libraries, delving through their usage models and examples, showing how to achieve high performance scalable I/O. Part III explains data compression and shows how to use it with I/O libraries. Part IV covers techniques for creating in situ analytics and teaches how to generate visualization services using VTK-M. Finally, Part V will explain data indexing/querying and how to use the libraries to query data both in situ and on files. Over one half of this tutorial will be hands-on sessions, where we provide access to the software, and go through live examples.

1:30 pm - 5:00 pm

InfiniBand, Omni-Path, and High-Speed Ethernet: Advanced Features, Challenges in Designing HEC Systems and Usage

Dhabaleswar Panda (Ohio State University), Hari Subramoni (Ohio State University), Jahanzeb Hashmi (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 supporting MPI, Storage and Parallel File Systems, Cloud Computing systems with SR-IOV Virtualization, Grid Computing systems, and Deep Learning systems. These systems are ringing 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 Libfabrics 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. Case studies focusing on domain-specific challenges in designing these systems, their solutions, and sample performance numbers will be presented. Finally, hands-on exercises will be carried out with Open Fabrics and Libfabrics software stacks and Network Management tools.

1:30 pm - 5:00 pm

Compression for Scientific Data

Franck Cappello (Argonne National Laboratory), Peter Lindstrom (Lawrence Livermore National Laboratory), Sheng Di (Argonne National Laboratory)

Large-scale numerical simulations, observations, and experiments are generating very large datasets that are difficult to analyze, store, and transfer. Data compression is an attractive and efficient technique to significantly reduce the size of scientific datasets. This tutorial reviews the state of the art in lossy compression of scientific datasets, discusses in detail two lossy compressors (SZ and ZFP), introduces compression error assessment metrics and the Z-checker tool to analyze the difference between initial and decompressed datasets. The tutorial addresses the following questions: Why lossless and lossy compression? How does compression work? How to measure and control compression error? The tutorial uses examples of real-world compressors and scientific datasets to illustrate the different compression techniques and their performance. Participants will also have the opportunity to learn how to use SZ, ZFP and Z-checker for their own datasets. The tutorial is given by two of the leading teams in this domain and targets primarily beginners interested in learning about lossy compression for scientific data. This half-day tutorial is improved from the evaluations of the highly rated tutorials given on this topic at ISC17, SC17 and SC18.

1:30 pm - 5:00 pm

Performance Tuning with the Roofline Model on GPUs and CPUs

Samuel Williams (Lawrence Berkeley National Laboratory), Charlene Yang (Lawrence Berkeley National Laboratory), Aleksandar Ilic (INESC-ID, Portugal), Kirill Rogozhin (Intel Corporation)
The Roofline performance model offers an insightful and intuitive method for extracting the key execution characteristics of HPC applications and comparing them against the performance bounds of modern CPUs and GPUs. Its capability to abstract the complexity of memory hierarchies and identify the most profitable optimization techniques have made Roofline-based analysis increasingly popular in the HPC community. Although different flavors of the Roofline model have been developed to deal with various definitions of memory data movement, there remains a need for a systematic methodology when applying them to analyze the efficiency of applications running on multicore, manycore, and accelerated systems.

The tutorial aims to bridge this gap on both CPUs and GPUs by exposing the fundamental aspects behind different Roofline modeling principles and providing several practical use case scenarios to highlight their efficacy 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 practice-oriented, hands-on topics and labs are given by the lead methodology researchers and the main designer of Intel’s Roofline automation tools. The tutorial presenters have a long history of working with the Roofline model and have presented several Roofline-based tutorials.
Workshop

Sunday, November 17

9:00 am - 12:30 pm

Computational Reproducibility at Exascale 2019 (CRE2019)

Session Description: 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. The workshop addresses issues in reproducibility that arise when computing at exascale. It will include issues of numerical reproducibility as well as approaches and best practices to sharing and running code and the reproducible dissemination of computational results. The workshop is meant to address the scope of the problems of computational reproducibility in HPC in general, and those anticipated as we scale up to exascale machines in the next decade. The participants of this workshop will include government, academic, and industry stakeholders; the goals of this workshop are to understand the current state of the problems that arise, what work is being done to deal with this issues, and what the community thinks the possible approaches to these problem are.


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.

The workshop addresses issues in reproducibility that arise when computing at exascale. It will include issues of numerical reproducibility as well as approaches and best practices to sharing and running code and the reproducible dissemination of computational results. The workshop is meant to address the scope of the problems of computational reproducibility in HPC in general, and those anticipated as we scale up to exascale machines in the next decade. The participants of this workshop will include government, academic, and industry stakeholders; the goals of this workshop are to understand the current state of the problems that arise, what work is being done to deal with this issues, and what the community thinks the possible approaches to these problem are.

**Reproducibility, Computability, and the Scientific Method**
Peter Coveney (University College London)
Plenary I

**Comparing Perturbation Models for Evaluating Stability of Post-Processing Pipelines in Neuroimaging**
Gregory Kiar (McGill University), Pablo de Oliveira Castro (University of Versailles), Pierre Rioux (McGill University), Eric Petit (Intel Corporation), Shawn T. Brown (McGill University), Alan C. Evans (McGill University), Tristan Glatard (Concordia University)

**Fast, Good, and Repeatable: Summations, Vectorization, and Reproducibility**
Brett Neuman (Los Alamos National Laboratory), Andy Dubois (Los Alamos National Laboratory), Laura Monroe (Los Alamos National Laboratory), Robert W. Robey (Los Alamos National Laboratory)

CRE2019 Morning Break
Reproducibility and Provenance of James Webb Space Telescope Data Products
Howard Bushouse (Space Telescope Science Institute)

Numerical Reproducibility Based on Minimal-Precision Validation
Toshiyuki Imamura (RIKEN Center for Computational Science (R-CCS)), Daichi Mukunoki (RIKEN Center for Computational Science (R-CCS)), Roman Iakymchuk (Sorbonne University), Fabienne Jézéquel (Sorbonne University), Stef Graillat (Sorbonne University)

Reproducibility and Variable Precision Computing
David H. Bailey (Lawrence Berkeley National Laboratory; University of California, Davis)

Panel Discussion
Howard Bushouse (Space Telescope Science Institute), Peter Coveney (University College London), Thorsten Hoefler (ETH Zurich), Thomas Ludwig (German Climate Computing Center), Line Pouchard (Brookhaven National Laboratory)

9:00 am - 5:30 pm

Innovating the Network for Data Intensive Science (INDIS)

Session Description: 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. [https://scinet.supercomputing.org/workshop/tags/indis2019](https://scinet.supercomputing.org/workshop/tags/indis2019)
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.

SCinet DTN-as-a-Service Framework
Se-young YU (Northwestern University, International Center for Advanced Internet Research (iCAIR))

Transferring big data over Wide Area Networks (WANs) is challenging because optimization is dependent on the specifics of multiple parameters. Network services, paths, and technologies have different characteristics, including loss rate, latency, and available capacity. Yet, frameworks currently used to configure and orchestrate transfer systems, measure performance, and analyze results have limited capabilities. We propose a framework, DTN-as-a-Service (DaaS), for high-performance network data transfers using and integration of techniques, including virtualization, network provisioning, and performance data analysis. This framework has a modular design for supporting multiple transfer tools, optimizers and orchestrators for the data transfer environment, including Docker and Kubernetes. We present a Jupyter based workflow for high-speed network data transfer in data-intensive science and evaluate the performance of the transfer with a simple programmable visualizer implemented in the framework. With the increase in the number and the capacity of WAN links at the conferences (multiple 100 Gbps WAN circuits), the challenges involved in setting up, testing, debugging, verifying and running applications on high-performance systems connecting to the conference SCinet WAN circuits also increase. The SCinet implementation of the DaaS framework for the conference community allowed users to control hardware, software, and network infrastructure for high-speed network data transfer, primarily for large scale applications. Through the evaluation of the framework in our test setup, we demonstrated that NVMe over Fabrics with TCP is twice as efficient compared to using conventional TCP in high-speed NVMe-to-NVMe transfers. We also implemented a 400 Gbps LAN experiment to evaluate the DaaS framework.
Daniel Nägele (University of Stuttgart, BelWü Koordination), Stefan Wesner (Ulm University, Institute of Information Resource Management)

In times of increasing bandwidth, network operators strive for increased visibility of their network’s utilization as well as an indication of the legitimacy of traffic processed across network nodes. Additionally, the detection and mitigation of illegitimate traffic such as denial of service attacks remains a current and persistently active field of research.

Flow-based network monitoring can provide this information live from any network interface. This paper introduces a flow processing platform meant to receive flow information from border interfaces and distribute the acquired information to specialized applications. Transit providers deploying our platform can use this information directly, but also provide all interested customers or network entities with the specific subset concerning them.

Between collecting and redistributing the flow information, our platform offers different methods of enrichment using a variety of sources, allowing for high-level views incorporating additional data compared to plain Netflow records. However, a provided tool can reencode and reexport standard Netflow to ensure compatibility and allow for seamless integration of customer-specific streams into preexisting setups.

This platform’s components allow the enrichment, division and anonymization of flow data to a number of highly customized streams for any type of application, either on a customer-specific or a network-wide provider level. Applications include the conversion of flow data for time-series databases and the accompanying dashboards, the detection of DDoS attacks or other high-traffic situation on any network level, the identification of faulty network routing policies, or any other use case conceivable on regular flow data, but within an arbitrary network scope.

INDIS Morning Break

INDIS Showcases Panel: SCinet NRE, XNET and Architecture
Jim Stewart (Utah Education and Telehealth Network (UETN)), Marc Lyonnais (Ciena Corporation)
INDIS Keynote: The Relevance of Software Defined Exchange Points in the Path of Scientific Workflows

Julio Ibarra (Florida International University)

International research network connections are increasing, providing new paths for scientific workflows to move data and work across continents: In the north Atlantic between North America and Europe, the Advanced North Atlantic (ANA) collaboration operates 740G of distributed bandwidth capacity; between the U.S. and South America, the AmLight Express and Protect (AmLight-ExP) project added three 200G optical waves between Florida and Brazil, for a total capacity of 630Gbps; the AmLight-SACS project will be activating a 100G optical wave between Brazil and Angola, establishing a new south Atlantic route between the Americas, Africa and Europe; and Bella, a project led by RedCLARA and GEANT, will provide new high-capacity network paths between South America and Europe. While this phenomenon provides the opportunity for network operators to add resiliency to the global R&E fabric, it also adds complexity to network management and traffic engineering. Software Defined Exchange points (SDX) provide enhanced capabilities to simplify flow-management and enable new network services. Unlike traditional exchange points, an SDX controller has global visibility of the network topology, making it possible to maintain related states for the full path of a flow. This presentation will describe what an SDX is and why SDXs are relevant to scientific workflows in the global R&E network fabric. The AtlanticWave-SDX project and its SDX controller will be presented as a novel distributed programmable controller along with use cases for two science drivers. Results from extensive experiments in a testbed will be presented that demonstrate the AtlanticWave-SDX controller’s high performance in terms of scalability, response time, and resource utilization, which are extremely important to support advanced scientific applications. Finally, future work will be described to clarify where SDXs can have an impact on the global R&E network fabric.

Estimation of RTT and Loss Rate of Wide-Area Connections Using MPI Measurements

Nageswara Rao (Oak Ridge National Laboratory), Neena Imam (Oak Ridge National Laboratory), Raj Kettimuthu (Argonne National Laboratory)

Scientific computations are expected to be increasingly distributed across wide-area networks, and Message Passing Interface (MPI) has been shown to scale to support their communications over long distances. The execution times of MPI basic operations over long distance connections reflect the connection length and losses, which should be accounted for by the applications, for example, by rolling back to a single site under high network loss conditions. We utilize execution time measurements of MPI_Sendrecv operations collected over emulated 10Gbps connections with 0-366ms round-trip times, wherein the longest connection spans the globe, under up to 20% periodic losses. We describe five machine leaning methods to estimate the connection RTT and loss rate from these MPI execution times. They provide disparate, namely, linear and non-linear, and smooth and non-smooth, estimators of RTT and loss rate. Our results show that accurate estimates...
can be generated at low loss rates but become inaccurate at loss rates 10% and higher. Overall, these results constitute a case study of the strengths and limitations of machine learning methods in inferring network-level parameters using application-level measurements.

**INDIS Lunch Break**

**Co-Scheduling of Advance and Immediate Bandwidth Reservations for Inter-Data Center Transfer**

*Aiqin Hou (Northwest University, China)*

As scientific applications and business services increasingly migrate to clouds, big data of various types with different priorities need to be transferred between geographically distributed cloud-based data centers. It has become a critical task for Cloud Service Providers (CSP) to fully utilize the expensive bandwidth resources of the links connecting such data centers while guaranteeing users’ Quality of Experience (QoE). Most high-performance networks based on software-defined networking (SDN) provide the capability of advance bandwidth reservation. This paper focuses on the scheduling of multiple user requests of two different types with different priorities, namely, advance bandwidth reservation with a lower priority and immediate bandwidth reservation with a higher priority, to maximize the total satisfaction of user requests. We formulate this co-scheduling problem as a generic optimization problem, which is shown to be NP-complete. We design a heuristic algorithm to maximize the number of successfully scheduled requests and minimize the number of preempted advance reservation requests, while minimizing the completion time of each request. Extensive simulation results show that our scheduling scheme significantly outperforms greedy approaches in terms of user satisfaction degree.

**INDIS Short Break**

**Hop Recording and Forwarding State Logging: Two Implementations for Path Tracking in P4**

*Joseph Hill (University of Amsterdam)*

Full information on the path travelled by packets is extremely important for network management and network security. We implemented two path tracking methods in hardware with P4. The first approach tracks a packet’s path by recording each node along the path of a packet (hop recording).
The complete path a packet took can be extracted from the packet in the last node of the path. The second approach tracks a packet’s path by logging the forwarding state of a network (forwarding state logging). The complete path can be reconstructed based on the node where the packet entered a network. We conducted experiments with the two implemented approaches and showed that the paths of the packets are reconstructed correctly. The advantage of using P4 is that the control plane only gets involved when the path of a packet is reconstructed. We finally show how our work provides a working tool in P4 networks that can be used to gain deep insights in traffic patterns.

INDIS Afternoon Break

G2: A Network Optimization Framework for High-Precision Analysis of Bottleneck and Flow Performance
Jordi Ros-Giralt (Reservoir Labs Inc)

Congestion control algorithms for data networks have been the subject of intense research for the last three decades. While most of the work has focused around the characterization of a flow’s bottleneck link, understanding the interactions amongst links and the ripple effects that perturbations in a link can cause on the rest of the network has remained much less understood. The Theory of Bottleneck Ordering is a recently developed mathematical framework that reveals the bottleneck structure of a network and provides a model to understand such effects. In this paper we present G2, the first operational network optimization framework that utilizes this new theoretical framework to characterize with high-precision the performance of bottlenecks and flows. G2 generates an interactive graph structure that describes how perturbations in links and flows propagate, providing operators new optimization insights and traffic engineering recommendations to help improve network performance. We provide a description of the G2 implementation and a set of experiments using real TCP/IP code to demonstrate its operational efficacy.

INDIS Keynote: Applications of ML and AI in Next Generation Wired and Wireless Networks
Awanish Verma (Xilinx Inc)

The emergence of new network technologies such as Intent based networks, coherent core optical networks, flow-based firewalls and 5G wireless has caused exponential increase in complexity and throughout of edge, access, metro and core networks. In such complex flow-based networks, manual and reactive adjustment of network parameters is not efficient and cannot guarantee reliable and efficient network deployment and operation. The monitoring and control of network
parameters such as QoS policy assignment in access networks, modulation and symbol rate adjustments in optical networks, packet filtering and anomaly detection in security networks, beam-forming and scheduling in 5G wireless networks is extremely complex and challenging. Machine learning (ML) and Artificial Intelligence (AI) can be used to perform proactive prediction and adjustment from learned behavior in next generation networks to achieve high efficiency and reliability in networks. This session discusses the applications of machine learning in next generation networks and provides an overview of how machine learning models (supervised, unsupervised or recurrent) deployed on a programmable hardware can help to build automated, efficient and resilient next generation networks.

**Training Classifiers to Identify TCP Signatures in Scientific Workflows**

*George Papadimitriou (University of Southern California), Cong Wang (Renaissance Computing Institute (RenCI))*

Identifying network anomalies is an important measure to ensure reliability and quality of data transfers among facilities. Scientific workflows in particular heavily rely on good network performance to ensure their smooth executions. In this paper, we present a lightweight classifier system that is able to recognize anomalous TCP transfers. Using random forest trees and labeled data sets, we evaluate the classifier with real workflow transfers for ground truth data. Our studies reveal that various TCP congestion algorithms behave differently in anomalous conditions. We show that training classifiers on these separately can aid detection in network performance deterioration. Results reveal that our classifiers are able to better predict anomalous flows for TCP Reno and Hamilton compared to Cubic and BBR, due to the manner how their congestion control algorithms handle the anomalies.

**Sample Transfer Optimization with Adaptive Deep Neural Network**

Transfer configurations play a crucial role in achieving desirable performance in high-speed networks where suboptimal settings could lead to poor transfer throughput. However, discovering the optimal configuration for a given transfer task is a difficult problem as it depends on various factors including dataset characteristics and network settings. The state-of-the-art transfer tuning solutions rely on sample transfers and evaluate different transfer configurations in attempt to discover the optimal one in real-time. Yet, current approaches to run sample transfers incur significant delay and measurement errors, thus limit the gain offered by tuning algorithms. In this paper, we take advantage of feed forward deep neural network (DNN) to minimize execution time of sample transfers without sacrificing measurement accuracy. To achieve this goal, we collected 115K data transfer logs in four networks and trained multiple DNNs that can predict convergence time of transfers by analyzing real-time throughput metrics. The results gathered in various
networks with rich set of transfer configurations indicate that DNN can reduce error rate by up to 50% compared to the state-of-the-art solution while achieving similar sample transfer execution time in most cases. Moreover, by tuning its hyperparameters and model settings, one can achieve low execution time and error rate based on the specific needs of the user or application.

INDIS Closing Remarks

9:00 am - 5:30 pm

Deep Learning on Supercomputers

Session Description: The Deep Learning (DL) on Supercomputers workshop provides a forum for practitioners working on any and all aspects of DL for science and engineering in the High Performance Computing (HPC) context to present their latest research results and development, deployment, and application experiences. The general theme of this workshop series is the intersection of DL and HPC. Its scope encompasses application development in scientific scenarios using HPC platforms; DL methods applied to numerical simulation; fundamental algorithms, enhanced procedures, and software development methods to enable scalable training and inference; hardware changes with impact on future supercomputer design; and machine deployment, performance evaluation, and reproducibility practices for DL applications with an emphasis on scientific usage. This workshop will be centered around published papers. Submissions will be peer-reviewed, and accepted papers will be published as part of the Joint Workshop Proceedings by Springer. https://dlonsc19.github.io/
future supercomputer design; and machine deployment, performance evaluation, and reproducibility practices for DL applications with an emphasis on scientific usage. This workshop will be centered around published papers. Submissions will be peer-reviewed, and accepted papers will be published as part of the Joint Workshop Proceedings by Springer.

Keynote
Satoshi Matsuoka (Tokyo Institute of Technology)

Deep Learning on Supercomputers Morning Break

DeepDriveMD: Deep-Learning Driven Adaptive Molecular Simulations for Protein Folding

Simulations of biological macromolecules are important in understanding the physical basis of complex processes such as protein folding. However, even with increasing computational capacity and specialized architectures, the ability to simulate protein folding at atomistic scales still remains challenging. This stems from the dual aspects of high dimensionality of protein conformational landscapes, and the inability of atomistic molecular dynamics (MD) simulations to sufficiently sample these landscapes to observe folding events. Machine learning/deep learning (ML/DL) techniques, when combined with atomistic MD simulations offer the opportunity to potentially overcome these limitations by: (1) effectively reducing the dimensionality of MD simulations to automatically build latent representations that correspond to biophysically relevant reaction coordinates (RCs), and (2) driving MD simulations to automatically sample potentially novel conformational states based on these RCs. We examine how coupling DL approaches with MD simulations can lead to effective approaches to fold small proteins on supercomputers. In particular, we study the computational costs and effectiveness of scaling DL-coupled MD workflows implemented using RADICAL-Cybertools in folding two prototypical systems, namely Fs-peptide and the fast-folding variant of the villin head piece protein. We demonstrate that a DL-coupled MD workflow is able to effectively learn latent representations and drive adaptive simulations. Compared to traditional MD-based approaches, our approach achieves an effective performance gain in sampling the folded states by at least 2.3x. Together, our study provides quantitative basis to understand how coupling DL approaches to MD simulations, can lead to effective performance gains and reduced times to solution on supercomputing resources.
Deep Facial Recognition Using Tensorflow

Facial recognition is a tractable problem today because of the prevalence of Deep Learning implementations. Approaches for creating structured datasets from unstructured web data are more easily accessible as are GPUs that deep learning frameworks can use to learn from this data. In DARPA’s MEMEX effort, which sought to create better search capabilities for law enforcement to scan the deep and dark web, we are interested in leveraging the Tensorflow framework to reproduce a seminal Deep Learning facial recognition model called VGG-Face. On MEMEX we desired to build the VGG-Face model and to train feature extraction for use in prioritization of leads for possible law enforcement follow-up. We describe our efforts to recreate the VGG-Face dataset, along with our efforts to create the Deep Learning network implementation for it using Tensorflow. Though other implementations of VGG-Face on Tensorflow exist, none of them fully reproduce as much of the dataset as we do today (~ 48% of the data still exists), nor have detailed documentation and steps for reproducing each step in the workflow. We contribute those instructions and leverage Texas Advanced Computing Center’s Maverick2 supercomputer to perform the work. We report experimental results on building the dataset, and training the network to achieve a 77.99% validation accuracy on the 2,622 celebrity use case from VGG-Face. This paper can be a useful recipe in building new Tensorflow facial recognition.

Deep Learning Accelerated Light Source Experiments

Experimental protocols at synchrotron light sources typically process and validate data only after an experiment has completed, which can lead to undetected errors and cannot enable online steering. Real-time data analysis can enable both detection of, and recovery from, errors, and optimization of data acquisition. However, modern scientific instruments, such as detectors at synchrotron light sources, can generate data at GBs/sec rates. Data processing methods such as the widely used computational tomography usually require considerable computational resources, and yield poor quality reconstructions in the early stages of data acquisition when available views are sparse.

We describe here how a deep convolutional neural network can be integrated into the real-time streaming tomography pipeline to enable better-quality images in the early stages of data acquisition. Compared with conventional streaming tomography processing, our method can significantly improve tomography image quality, deliver comparable images using only 32% of the data needed for conventional streaming processing, and save 68% experiment time for data acquisition.

DC-S3GD: Delay-Compensated Stale-Synchronous SGD for Large-Scale Decentralized Neural
Network Training

Data parallelism has become the de facto standard for training Deep Neural Network on multiple processing units. In this work we propose DC-S3GD, a decentralized (without Parameter Server) stale-synchronous version of the Delay-Compensated Asynchronous Stochastic Gradient Descent (DC-ASGD) algorithm. In our approach, we allow for the overlap of computation and communication, by averaging in parameter space and compensating the inherent error with a first-order correction of the locally computed gradients. We prove the effectiveness of our approach by training Convolutional Neural Network with large batches and achieving state-of-the-art results.

Aggregating Local Storage for Scalable Deep Learning I/O

Deep learning applications introduce heavy I/O loads on computer systems. The inherently long-running, highly concurrent, and random file accesses can easily saturate traditional shared file systems and negatively impact other users. We investigate here a solution to these problems based on leveraging local storage and the interconnect to serve training datasets at scale. We present FanStore, a user-level transient object store that provides low-latency and scalable POSIX-compliant file access by integrating the function interception technique and various metadata/data placement strategies. On a single node, FanStore provides performance similar to that of the XFS journaling file system. On many nodes, our experiments with real applications show that FanStore achieves over 90% scaling efficiency.

Deep Learning on Supercomputers Lunch Break

Highly-Scalable, Physics-Informed GANs for Learning Solutions of Stochastic PDEs

Uncertainty quantification for forward and inverse problems is a central challenge across physical and biomedical disciplines. We address this challenge for the problem of modeling subsurface flow at the Hanford Site by combining stochastic computational models with observational data using physics-informed GAN models. The geographic extent, spatial heterogeneity, and multiple correlation length scales of the Hanford Site require training a computationally intensive GAN model to thousands of dimensions. We develop a highly optimized implementation that scales to 27,500 NVIDIA Volta GPUs. We develop a hierarchical scheme based on a multi-player game-theoretic
approach for exploiting domain parallelism, map discriminators and generators to multiple GPUs, and employ efficient communication schemes to ensure training stability and convergence. Our implementation scales to 4584 nodes on the Summit supercomputer with a 93.1% scaling efficiency, achieving peak and sustained half-precision rates of 1228 PF/s and 1207 PF/s.

**Deep Learning for Gap Crossing Ability of Ground Vehicles**

In this work we present our results designing a deep neural network (DNN) to act as a surrogate model for costly HPC simulations. In order to determine a ground vehicle's gap crossing ability in extreme weather scenarios, several HPC simulations are currently used. Hydrologic models are first run to determine the environmental conditions over an area of interest. Once these conditions are known they are given, along with the terrain data, to a vehicle simulation which determines if a particular vehicle can cross a stream at a given point. Every point of interest must be evaluated independently, which quickly becomes infeasible for a large numbers of crossing points. In order to accelerate this phase of the process, we have created a DNN that acts as a surrogate model for the vehicle simulator. Despite several challenges converting irregular data into a form that can be used with a DNN, and incorporating scalars into the models, we were able to produce DNN models that predicted the gap crossing ability of all vehicle types with over 95% accuracy.

**Deep Learning on Supercomputers Afternoon Break**

**Scaling Distributed Training of Flood-Filling Networks on HPC Infrastructure for Brain Mapping**

Mapping all the neurons in the brain requires automatic reconstruction of entire cells from volume electron microscopy data. The flood-filling network (FFN) architecture has demonstrated leading performance for segmenting structures from this data. However, the training of the network is computationally expensive. In order to reduce the training time, we implemented synchronous and data-parallel distributed training using the Horovod library, which is different from the asynchronous training scheme used in the published FFN code. We demonstrated that our distributed training scaled well up to 2048 Intel Knights Landing (KNL) nodes on the Theta supercomputer. Our trained models achieved similar level of inference performance, but took less training time compared to previous methods. Our study on the effects of different batch sizes on FFN training suggests ways to further improve training efficiency. Our findings on optimal learning
rate and batch sizes agree with previous works.

**Evolving Larger Convolutional Layer Kernel Sizes for a Settlement Detection Deep-Learner on Summit**

Deep-learner hyper-parameters, such as kernel sizes, batch sizes, and learning rates, can significantly influence the quality of trained models. The state of the art for finding optimal hyper-parameters generally uses a brute force, grid search approach, random search, or Bayesian-based optimization among other techniques. We applied an evolutionary algorithm to optimize kernel sizes for a convolutional neural network used to detect settlements in satellite imagery. Usually convolutional layer kernel sizes are small – typically one, three, or five – but we found that the system converged at, or near, kernel sizes of nine for the last convolutional layer, and that this occurred for multiple runs using two different datasets. Moreover, the larger kernel sizes had fewer false positives than the 3x3 kernel sizes found as optimal via a brute force uniform grid search. This suggests that this large kernel size may be leveraging patterns found in larger areal features in the source imagery, and that this may be generalized as possible guidance for similar remote sensing deep-learning tasks.

**Scaling TensorFlow, PyTorch, and MXNet Using MVAPICH2 for High-Performance Deep Learning on Frontera**

Frontera is the largest NSF-funded cluster in the US and comprises of 8,008 nodes equipped with the latest Intel Xeon processors (Cascade-Lake). In this paper, we explore the potential of Frontera for training state-of-the-art Deep Learning (DL) models at scale. Most DL studies present performance data from large-scale GPU clusters that are equipped with NVIDIA GPUs. However, our earlier performance characterization studies have helped us achieve comparable performance with CPU-only clusters as well. Based on this, we configure three important DL frameworks; 1) TensorFlow, 2) PyTorch, and 3) MXNet, using Horovod and two Message Passing Interface (MPI) libraries on Frontera: 1) MVAPICH2 and 2) Intel MPI. We provide a systematic performance comparison for TensorFlow using MVAPICH2 and Intel MPI on 2,048 Frontera nodes. Using a four process per-node configuration, we observe near-linear scaling for ResNet-50 training for TensorFlow up to 8,192 MPI processes (on 2,048 nodes) offering a sustained performance of 250,000 images/second. In addition, we provide insights into process per node and batch size configurations for TensorFlow as well as for PyTorch and MXNet. Based on single-node performance behavior, we scale all three DL frameworks up to 1,024 processes (256 nodes) for various models like ResNet-50/101/152 and Inception-v3/v4.
Strategies to Deploy and Scale Deep Learning on the Summit Supercomputer

The rapid growth and wide applicability of Deep Learning (DL) frameworks poses challenges to computing centers which need to deploy and support the software, and also to domain scientists who have to keep up with the system environment and scale up scientific exploration through DL. We offer recommendations for deploying and scaling DL frameworks on the Summit supercomputer, currently atop the Top500 list, at the Oak Ridge National Laboratory Leadership Computing Facility (OLCF). We discuss DL software deployment in the form of containers, and compare performance of native-built frameworks and containerized deployment. Software containers show no noticeable negative performance impact and exhibit faster Python loading times and promise easier maintenance. To explore strategies for scaling up DL model training campaigns, we assess DL compute kernel performance, discuss and recommend I/O data formats and staging, and identify communication needs for scalable message exchange for DL runs at scale. We recommend that users take a step-wise tuning approach beginning with algorithmic kernel choice, node I/O configuration, and communications tuning as best-practice. We present baseline examples of scaling efficiency 87% for a DL run of ResNet50 running on 1024 nodes (6144 V100 GPUs).

9:00 am - 5:30 pm

Sixth SC Workshop on Best Practices for HPC Training and Education

Session Description: High-performance computing has become central for empowering scientific progress in the most fundamental research in various academic and business domains. It is remarkable to observe that the rapid advancement in the mainstream computing technology has facilitated the ability to solve complex, large-scale scientific applications that perform advanced simulations of the implementation of various numerical models pertaining to diverse fields. However, the inherent wide distribution, heterogeneity, and dynamism of the current and emerging computing and software environments increasingly challenge cyberinfrastructure facilitators, trainers and educators. The challenge is how to support and train the diverse current users and prepare the future educators, researchers, developers and policymakers to keep pace with the rapidly evolving HPC environments to advance discovery and economic competitiveness for many generations. Since 2014, the Best Practices for HPC Training Workshop at SC has been successful in creating a global platform for addressing common challenges for enhancing HPC training and education and for numerous sharing and collaborating opportunities across the globe. The sixth annual full-day workshop on HPC training and education is an ACM SIGHPC Education Chapter coordinated effort, aimed at fostering more collaborations among the practitioners from traditional
High-performance computing has become central for empowering scientific progress in the most fundamental research in various academic and business domains. It is remarkable to observe that the rapid advancement in the mainstream computing technology has facilitated the ability to solve complex, large-scale scientific applications that perform advanced simulations of the implementation of various numerical models pertaining to diverse fields. However, the inherent wide distribution, heterogeneity, and dynamism of the current and emerging computing and software environments increasingly challenge cyberinfrastructure facilitators, trainers, and educators. The challenge is how to support and train the diverse current users and prepare the future educators, researchers, developers, and policymakers to keep pace with the rapidly evolving HPC environments to advance discovery and economic competitiveness for many generations.

Since 2014, the Best Practices for HPC Training Workshop at SC has been successful in creating a global platform for addressing common challenges for enhancing HPC training and education and for numerous sharing and collaborating opportunities across the globe. The sixth annual full-day workshop on HPC training and education is an ACM SIGHPEDducation Chapter coordinated effort, aimed at fostering more collaborations among the practitioners from traditional and emerging fields to explore educational needs in HPC, to develop and deploy HPC training, and to identify new challenges and opportunities for latest HPC platforms. The workshop will also be a platform for disseminating results and lessons learned in these areas and will be captured in a Special Edition of the Journal of Computational Science Education.

FreeCompilerCamp.org: Training for OpenMP Compiler Development from Cloud

Anjia Wang (University of North Carolina, Charlotte; Lawrence Livermore National Laboratory), Alok Mishra (Stony Brook University, Lawrence Livermore National Laboratory)

OpenMP is one of the most popular programming models to exploit node-level parallelism of supercomputers. Many researchers are interested in developing OpenMP compilers or extending existing standard for new capabilities. However, there is a lack of training resources for researchers who are involved in the compiler and language development around OpenMP, making learning
curve in this area steep.

In this paper, we introduce an ongoing effort, FreeCompilerCamp.org, a free and open online learning platform aimed to train researchers to quickly develop OpenMP compilers. The platform is built on top of Play-With-Docker, a docker playground for users to conduct experiments in an online terminal sandbox. It provides a live training website that is set up on cloud, so anyone with internet access and a web browser will be able to take the training. It also enables developers with relevant skills to contribute new tutorials. The entire training system is open-source and can be deployed on a private server, workstation or even laptop for personal use. We have created some initial tutorials to train users to learn how to extend the Clang/LLVM and ROSE compiler to support new OpenMP features. We welcome anyone to try out our system, give us feedback, contribute new training courses, or enhance the training platform to make it an effective learning resource for the HPC community.

**Introducing Novices to Scientific Parallel Computing**
Betsy Hillery (Purdue University), Xiao Zhu (Purdue University)

HPC and Scientific Computing are integral tools for sustaining the growth of scientific research. Additionally, educating future domain scientists and research-focused IT staff about the use of computation to support research is as important as capital expenditures on new resources. In this paper we describe the parallel computing portion of our HPC seminar series which we use as a tool to introduce students from many non-traditional disciplines to scientific, parallel and high-performance computing.

**Best Practices for HPC Training and Education Morning Break**

**A Modular Course on Developing Research Software**
Kyle Niemeyer (Oregon State University)

This talk describes the motivation and design of learning modules for teaching best practices for developing research software and practical computational science. These modules have been offered twice as a 10-week graduate course in an engineering program, although the content applies broadly to any field of scientific and/or engineering research where software may be developed. Topics taught include local and remote version control, licensing and copyright, structuring Python modules, testing and test coverage, continuous integration, packaging and distribution, open science, introduction to parallelism, software citation, and reproducibility basics,
among others. Lectures are supplemented by in-class activities and discussions, and all course material is shared openly via GitHub. In the 10-week course, student work is heavily based on a single, term-long project where students individually develop a software package targeted at their own research topic; all contributions must be submitted as pull requests and reviewed/merged by other students. The course was offered in 2018 and 2019 to 24 students total, and shorter-length workshops will also be offered.

Teaching HPC Systems Administrators
Alex Younts (Purdue University)

The ability to grow and teach systems professionals relies on having the ability to let students interact with supercomputers at levels not given to normal users. In this paper we explore the teaching methods and hardware platforms used by Purdue Research Computing to train undergraduates for HPC system roles. From Raspberry Pi clusters to the LittleFe project, tremendous previous work has focused on providing miniature hardware platforms and developing curriculum for teaching. We have come up and employed a method using virtual machines to reach a wider audience and remove barriers as well as some best practices for approaching coursework. This paper outlines the system we have designed, expands on the benefits and drawbacks over hardware systems, and discusses the failures and successes we have had teaching HPC System Administrators.

The Supercomputing Institute: A Systems-Focused Approach to HPC Training and Education
J. Lowell Wofford (Los Alamos National Laboratory), Cory Lueninghoener (Los Alamos National Laboratory)

For the past thirteen years, Los Alamos National Laboratory HPC Division has hosted the Computer System, Cluster and Networking Summer Institute summer internship program (recently renamed "The Supercomputing Institute") to provide a basis is cluster computing for undergraduate and graduate students. The institute invites 12 students each year to participate in a 10-week internship program. This program has been a strong educational experience for many students through this time and has been an important recruitment tool for HPC Division. In this paper, we describe the institute as a whole and dive into individual components that were changed this year to keep the program up to date. We also provide some qualitative and quantitative results that indicate that these changes have improved the program over recent years.

Computational Mathematics, Science and Engineering (CMSE): Establishing an Academic Department Dedicated to Scientific Computation as a Discipline
Dirk Colbry (Michigan State University)
The Computational Mathematics Science and Engineering (CMSE) department is one of the newest units at Michigan State University (MSU). Founded in 2015, CMSE recognizes computation as the "triple junction" of algorithm development and analysis, high performance computing, and applications to scientific and engineering modeling and data science (as illustrated in Figure 1). This approach is designed to engage with computation as a new integrated discipline, rather than a series of decentralized, isolated sub-specialties. In the four years since its inception, the department has grown and flourished; however, the pathway was sometimes arduous. This paper shares lessons learned during the department’s development and the initiatives it has taken on to support computational research and education across the university. By sharing these lessons, we hope to encourage and support the foundation of similar departments at other universities and grow this integrated style discipline as a new profession.

An Informal Introduction to HPC Via a Numerical Weather Model and Low-Cost Hardware

Computing power and paradigms are perpetually evolving with time and technology. A corresponding growth in the volume and availability of data has driven the implementation of high performance computing (HPC) across a broader spectrum of disciplines, which in turn is driving demand for new HPC professionals. However, financial, logistical, and educational barriers exist which inhibit students from beginning HPC career pathways.

As a means of circumventing these challenges of learning foundational HPC concepts, viable low-cost methods can be implemented. The aim of this work is to create a portable low-cost cluster where students can be introduced to computational science and its applications within an informal environment. This is done by creating a graphical user interface based application that allows a user to run a modern numerical weather model on a cluster of Raspberry Pi computers.

This work helps introduce fundamental HPC concepts to a broader audience that may not have any prior exposure to computational science or possess technical skills. By creating an informal and simplistic supercomputing environment, the learning process for such students can be accelerated.

Computational Biology as a Compelling Pedagogical Tool in Computer Science Education
Vijayalakshmi Saravanan (Rochester Institute of Technology)

High-performance computing (HPC) and parallel and distributed computing (PDC) are widely discussed topics in computer science (CS) and computer engineering (CE) education. In the past decade, high-performance computing has also contributed significantly to addressing complex problems in bioengineering, healthcare and systems biology. Therefore, computational biology applications provide several compelling examples that can be potent pedagogical tools in teaching
high-performance computing. In this paper, we introduce a novel course curriculum to teach high-performance, parallel and distributed computing to senior graduate students (PhD) in a hands-on setup through examples drawn from a wealth of areas in computational biology. We introduce the concepts of parallel programming, algorithms and architectures and implementations via carefully chosen examples from computational biology. We believe that this course curriculum will provide students an engaging and refreshing introduction to these well-established domains.

Best Practices for HPC Training and Education Lunch Break

Directed Internship and Apprenticeship Toward a Sustainable HPC Workforce Development Pipeline
Elizabeth Bautista (Lawrence Berkeley National Laboratory)

In high performance computing (HPC), staff who manage the computational center can consistently turn over and recruiting new staff can be a challenge. The skill sets required to perform this function involve understanding of system administration, storage administration, wide area network engineering and management of the facility environment. As such, this area intersects with concepts in computing science, electrical engineering and mechanical engineering, which is currently not taught in any educational system in this combination. This talk provides information on a training model that has been adopted at the National Energy Research Scientific Computing Center (NERSC) at Lawrence Berkeley National Laboratory (LBNL). As a computational facility that have onsite staff 24/7, staff are trained with required technical understanding and independent thought to resolve problems to bring the facility resources back to normal operation.

Most education systems focus on one or two of these areas. The model involves working with the school and curriculum to provide the basic education required. Students are placed in an internship to demonstrate their knowledge of these areas but also direct their training toward the HPC area. Completion of an internship can lead to the one-year apprenticeship program which is milestone based and meant to direct the apprentice to master each of the skill areas to prepare to meet the minimum qualifications of a Site Reliability Engineer at a computational facility. NERSC has seen success from this program and has created a workforce pipeline from schools into a potential career position.

Contributing HPC Skills to the HPC Certification Forum
Julian Kunkel (University of Reading)
The International HPC Certification Program has been officially launched over a year ago at ISC’18 and since then made significant progress in categorising and defining the skills required to proficiently use a variety of HPC systems. The program reached the stage when the support and input from the HPC community is essential. For the certification to be recognised widely, it needs to capture skills required by majority of HPC users, regardless of their level. This cannot be achieved without contributions from the community. This extended abstract briefly presents the current state of the developed Skill Tree and explains how contributors can extend it. In the talk, we focus on the contribution aspects.

Creating a Relevant, Application-Based Curriculum for High Performance Computing in High School
Vincent C. Betro (Baylor School, Chattanooga, Tennessee), Mary E. Loveless (Baylor School, Chattanooga, Tennessee)

While strides have been made to improve science and math readiness at a college-preparatory level, some key fundamentals have been left unaddressed that can cause students to turn away from the STEM disciplines before they find their niche. Introducing collegiate level research and project-based, group-centered learning at a high school level has a multi-faceted effect; in addition to elevated learning outcomes in science and math, students exhibit improved critical thinking and communication skills, leading to improved preparedness for subsequent academic endeavors. The work presented here outlines the development of a STEM ecosystem where both the science department and math department have implemented an interdisciplinary approach to introduce a spectrum of laboratory and computing research skills. This takes the form of both “in situ,” micro-curricular elements and stand-alone research and computer science classes which integrate the language-independent concepts of abstraction and object-oriented programming, distributed and high-performance computing, and high and low level language control applications. This pipeline has been an effective tool that has allowed several driven and interested students to participated in collegiate-level and joint-collegiate projects involving virtual reality, robotics and systems controls, and modeling. The willingness of the departments to cross-pollinate, hire faculty well-versed in research, and support students and faculty with the proper resources are critical factors in readying the next generation of computing leaders.

Self-paced Learning in HPC Lab Courses
Christian Terboven (RWTH Aachen University)

In a software lab, groups of students develop parallel code using modern tools, document the results and present their solutions. The learning objectives include the foundations of High-Performance Computing (HPC), such as the understanding of modern architectures, the development of parallel programming skills, and course-specific topics, like accelerator
programming or cluster set-up.

In order to execute the labs successfully with limited personnel resources and still provide students with access to world-class HPC architectures, we developed a set of concepts to motivate students and to track their progress. This includes the learning status survey and the developer diary, which are presented in this work. We also report on our experiences with using innovative teaching concepts to incentivize students to optimize their codes, such as using competition among the groups. Our concepts enable us to track the effectiveness of our labs and to steer them for increasing sizes of diverse students.

We conclude that software labs are effective in adding practical experiences to HPC education. Our approach to hand out open tasks and to leave creative freedom in implementing the solutions enables the students to self-pace their learning process and to vary their investment of effort during the semester. Our effort and progress tracking ensures the achieving of the extensive learning objectives and enables our research on HPC programming productivity.

Evaluating the Effectiveness of an Online Learning Platform in Transitioning from High Performance Computing to a Commercial Cloud Computing Environment
Dhruva Chakravorty (Texas A&M University)

Developments in large scale computing environments have led to design of workflows that rely on containers and analytics platform that are well supported by the commercial cloud. The National Science Foundation also envisions a future in science and engineering that includes commercial cloud service providers (CSPs) such as Amazon Web Services, Azure and Google Cloud. These win forces have made researchers consider the commercial cloud as an alternative option to current high performance computing (HPC) environments. Training and knowledge on how to migrate workflows, cost-control, data-management, and system administration remain some of the commonly listed concerns with adoption of cloud computing. In an effort to ameliorate this situation, CSPs have developed online and in-person training platforms to help address this problem. Scalability, ability to impart knowledge, evaluating knowledge gain, and accreditation are the core concepts that have driven this approach. Here, we present a review of our experience using Google’s Qwiklabs online platform for remote and in-person training from the perspective of a HPC user. For this study, we completed over 50 online courses, earned five badges and attended a one-day session. Here, we identify the strengths of the approach, identify avenues to refine them, and consider means to further community engagement. We further evaluate the readiness of these resources for a cloud-curious researcher who is familiar with HPC. Finally, we present recommendations on how the large scale computing community can leverage these opportunities to work with CSPs to assist researchers nationally and at their home institutions.
Leveraging SC to Learn High-Performance Communication  
Rohit Zambre (University of California, Irvine)

Best Practices for HPC Training and Education Afternoon Break

Panel: Building the Future  
Linda McIlver (Australian Data Science Education Institute), Aaron Weeden (Shodor Education Foundation), Maria Ribera Sancho (Barcelona Supercomputing Center), Christine Harvey (MITRE Corporation), Tim Powell (Daresbury Laboratory)

Open Discussion  
Susan Mehringer (Cornell University)

9:00 am - 5:30 pm

The 5th International Workshop on Data Analysis and Reduction for Big Scientific Data (DRBSD-5) in Conjunction with SC19

Session Description: A growing disparity between simulation speeds and I/O rates makes it increasingly infeasible for applications to save all results for analysis. In this new world, applications must increasingly perform online data analysis and reduction—tasks that introduce algorithmic, implementation, and programming model challenges that are unfamiliar to many scientists and that have major implications for the design of various elements of exascale systems. This trend has spurred interest in 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
A growing disparity between simulation speeds and I/O rates makes it increasingly infeasible for applications to save all results for analysis. In this new world, applications must increasingly perform online data analysis and reduction—tasks that introduce algorithmic, implementation, and programming model challenges that are unfamiliar to many scientists and that have major implications for the design of various elements of exascale systems.

This trend has spurred interest in 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.

**DRBSD-5 Keynote Talk**


**Understanding Performance-Quality Trade-offs in Scientific Visualization Workflows with**
Lossy Compression
Jieyang Chen (Oak Ridge National Laboratory)

The cost of I/O is a significant challenge on current supercomputers, and the trend is likely to continue into the foreseeable future. This challenge is amplified in scientific visualization because of the requirement to consume large amounts of data before processing can begin. Lossy compression has become an important technique in reducing the cost of performing I/O. In this paper we consider the implications of using compressed data for visualization within a scientific workflow. We use visualization operations on simulation data that is reduced using three different state-of-the-art compression techniques. We study the storage efficiency and preservation of visualization features on the resulting compressed data, and draw comparisons between the three techniques used. Our contributions can help inform both scientists and researchers in the use and design of compression techniques for preservation of important visualization details.

DRBSD-5 Morning Break

DRBSD-5 Invited Talk 1: Data Challenges for Nuclear Femtography
Amber Boehnlein (Thomas Jefferson National Accelerator Facility)

A Collaborative Effort to Improve Lossy Compression Methods for Climate Data
Dorit M. Hammerling (Colorado School of Mines)

Climate model simulations produce large volumes of data, and reducing the storage burden with data compression is increasingly of interest to climate scientists. A key concern to the climate community, though, is ensuring that any data loss due to compression does not in any way affect their scientific analysis. For this reason, the climate community is taking a cautious approach to adopting lossy compression by carefully investigating the potential existence of artifacts due to compression in a wide variety of analysis settings. Spatio-temporal statistical analysis in particular can highlight compression-induced features that would go unnoticed by the standard metrics common to the data compression community. Communicating such findings to the algorithm developers in the context of a collaborative improvement cycle is one – in our view productive – way to foster trust within the climate community and pave the way for eventual adoption of lossy compression. In this work, we report on the initial results of a successful and mutually beneficial collaboration between the two communities that led to improvements in a well regarded
Machine learning (ML) has emerged as a tool for understanding data at scale. However, this new methodology comes at a cost because ML requires the use of even more HPC resources to generate ML algorithms. In addition to the compute resources required to develop ML algorithms, ML does not sidestep one of the biggest challenges on leading-edge HPC systems: the increasing gap between compute performance and I/O bandwidth. This has led to a strong push towards in situ, processing the data as it is generated, strategies to mitigate the I/O bottleneck. Unfortunately, there are no in situ frameworks dedicated to coupling scientific visualization and ML at scale to develop ML algorithms for scientific visualization.

To address the ML and in situ visualization gap, we introduce PAVE. PAVE is an in situ framework which addresses the data management needs between visualisation and machine learning tasks. We demonstrate our framework with a case study that accelerates physically-based light rendering, path-tracing, through the use of a conditional Generative Adversarial neural Network (cGAN). PAVE couples the training over path-traced images resulting in a generative model able to produce scene renderings with accurate light transport and global illumination of a quality comparable to offline approaches in a more efficient manner.

Complex workflows consisting of multiple simulation and analysis codes running concurrently through in-memory coupling is becoming popular due to inherent advantages in online management of large-scale data, resilience, and the code development process. However, orchestrating such a multi-application workflow to efficiently utilize resources on a heterogeneous architecture is challenging.

In this paper, we present our results with running the Fusion Whole Device Modeling benchmark workflow on Summit, a pre-exascale supercomputer at Oak Ridge National Laboratory. We explore various resource distribution and process placement mechanisms, including sharing compute nodes...
between processes from separate applications. We show that fine-grained process placement can have a significant impact towards efficient utilization of the compute power of a node on Summit, and conclude that sophisticated tools for performing co-design studies of multi-application workflows can play an important role towards efficient orchestration of such workflows.

DRBSD-5 Lunch Break

DRBSD Keynote Talk 2

DRBSD-5 Invited Talk 3: Scientific Data at Exascale: Architecting Systems for Performance, Productivity, and Parallelism
Rangan Sukumar (Cray Inc)

DRBSD-5 Afternoon Break

DRBSD-5 Invited Talk 4
Keith Gray (BP plc)

DRBSD-5 Invited Talk: In Situ Data Analytics for Next Generation Molecular Dynamics Workflows
Michela Taufer (University of Tennessee)

Analyzing the Performance and Accuracy of Lossy Checkpointing on Sub-Iteration of NWChem
Future exascale systems are expected to be characterized by more frequent failures than current petascale systems. This places increased importance on the application to minimize the amount of time wasted due to recomputation when recovering from a checkpoint. Typically HPC application checkpoint at iteration boundaries. However, for applications that have a high per-iteration cost, checkpointing inside the iteration limits the amount of re-computation. This paper analyzes the performance and accuracy of using lossy compressed check-pointing in the computational chemistry application NWChem. Our results indicate that lossy compression is an effective tool for reducing the sub-iteration checkpoint size. Moreover, compression error tolerances that yield acceptable deviation in accuracy and iteration count are quantified.

**Using Machine Learning to Reduce Ensembles of Geological Models for Oil and Gas Exploration**

Oliver Brown (Edinburgh Parallel Computing Centre)

Exploration using borehole drilling is a key activity in determining the most appropriate locations for the petroleum industry to develop oil fields. However estimating the amount of Oil In Place (OIP) relies on computing with a very significant number of geological models which, due to the ever increasing capability to capture and refine data, is becoming infeasible. As such data reduction techniques are required to reduce this set down to a much smaller yet still fully representative ensemble. In this paper we explore different approaches to identifying the key grouping of models, based on their most important features, and then using this information select a reduced set which we can be confident fully represent the overall model space. The result of this work is an approach which enables us to describe the entire state space using only 0.5% of the models, along with a series of lessons learnt. The techniques that we describe are not only applicable to oil and gas exploration, but also more generally to the HPC community as we are forced to work with reduced data-sets due to the rapid increase in data collection capability.

**Exploring Lossy Compression of Gene Expression Matrices**

Coleman B. McKnight (Clemson University)

Gene Expression Matrices (GEMs) are a fundamental data type in the genomics domain. As the size and scope of genomics experiments increase, researchers are struggling to process large GEMs through downstream workflows with currently accepted practices. In this paper, we propose a methodology to reduce the size of GEMs using multiple approaches. Our method partitions data into discrete fields based on data type and employs state-of-the-art lossless and lossy compression algorithms to reduce the input data size. This work explores a variety of lossless and lossy compression methods to determine which methods work the best for each component of a GEM. We evaluate the accuracy of the compressed GEMs by running them through the Knowledge
Independent Network Construction (KINC) workflow and comparing the quality of the resulting gene co-expression network with a lossless control to verify result fidelity. Results show that utilizing a combination of lossy and lossless compression results in compression ratios up to 9.77× on a Yeast GEM, while still preserving the biological integrity of the data. Usage of the compression methodology on the Cancer Cell Line Encyclopedia (CCLE) GEM resulted in compression ratios up to 9.26×. By using this methodology, researchers in the Genomics domain may be able to process previously inaccessible GEMs while realizing significant reduction in computational costs.

9:00 am - 5:30 pm

Parallel Applications Workshop, Alternatives to MPI+X

Session Description: Supercomputers are becoming increasingly complex due to the prevalence of hierarchy and heterogeneity in emerging node and system architectures. As a result of these trends, users of conventional programming models for scalable high-performance applications increasingly find themselves writing applications using a mix of distinct programming models—such as Fortran90, C, C++, MPI, OpenMP, and CUDA—which are also often becoming more complex and detail-oriented themselves. These trends negatively impact the costs of developing, porting, and maintaining HPC applications. Meanwhile, new programming models and languages are being developed that strive to improve upon the status quo by unifying the expression of parallelism and locality across the system, raising the level of abstraction, making use of modern language design features, and/or leveraging the respective strengths of programmers, compilers, runtimes, and operating systems. These alternatives may take the form of parallel programming languages (e.g., Chapel, Fortran 2018, Julia, UPC), frameworks for large-scale data processing and analytics (e.g., Spark, Tensorflow, Dask), or libraries and embedded DSLs that extend existing languages (e.g., Legion, COMPSs, SHMEM, HPX, Charm++, UPC++, Coarray C++, Global Arrays). The PAW-ATM workshop is designed to explore the expression of applications in scalable parallel programming models that serve as an alternative to the status quo. It is designed to bring together applications experts and proponents of high-level programming models to present concrete and practical examples of using such alternative models and to illustrate the benefits of high-level approaches to scalable programming. [http://sourceryinstitute.github.io/PAW/](http://sourceryinstitute.github.io/PAW/)

Supercomputers are becoming increasingly complex due to the prevalence of hierarchy and heterogeneity in emerging node and system architectures. As a result of these trends, users of
conventional programming models for scalable high-performance applications increasingly find themselves writing applications using a mix of distinct programming models—such as Fortran90, C, C++, MPI, OpenMP, and CUDA—which are also often becoming more complex and detail-oriented themselves. These trends negatively impact the costs of developing, porting, and maintaining HPC applications.

Meanwhile, new programming models and languages are being developed that strive to improve upon the status quo by unifying the expression of parallelism and locality across the system, raising the level of abstraction, making use of modern language design features, and/or leveraging the respective strengths of programmers, compilers, runtimes, and operating systems. These alternatives may take the form of parallel programming languages (e.g., Chapel, Fortran 2018, Julia, UPC), frameworks for large-scale data processing and analytics (e.g., Spark, Tensorflow, Dask), or libraries and embedded DSLs that extend existing languages (e.g., Legion, COMPSs, SHMEM, HPX, Charm++, UPC++, Coarray C++, Global Arrays).

The PAW-ATM workshop is designed to explore the expression of applications in scalable parallel programming models that serve as an alternative to the status quo. It is designed to bring together applications experts and proponents of high-level programming models to present concrete and practical examples of using such alternative models and to illustrate the benefits of high-level approaches to scalable programming.

**Computer Science Challenges to Imaging the Universe with the SKA Radio-Telescope**  
Peter J. Braam (University of Oxford)  
The SKA radio telescope will be a massive scientific instrument entering service in the late 2020s. The conversion of its antenna signals to images and the detection of transient phenomena is a truly massive computational undertaking, requiring 200PB/sec of memory bandwidth, involving domain-specific data. In this lecture, we will give an overview of the data processing in the telescope and the process that has been followed to design suitable algorithms and systems. We will highlight the difficulties observed in utilizing existing programming language approaches and discuss opportunities for innovation from the computer science perspective.

**PAW-ATM Morning Break**

**Soleil-X: Turbulence, Particles, and Radiation in the Regent Programming Language**  
Hilario C. Torres (Stanford University)
The Predictive Science Academic Alliance Program (PSAAP) II at Stanford University is developing an Exascale-ready multi-physics solver to investigate particle-laden turbulent flows in a radiation environment for solar energy receiver applications. In order to simulate the proposed concentrated particle-based receiver design three distinct but coupled physical phenomena must be modeled: fluid flows, Lagrangian particle dynamics, and the transport of thermal radiation. Therefore, three different physics solvers (fluid, particles, and radiation) must run concurrently with significant cross-communication in an integrated multi-physics simulation. However, each solver uses substantially different algorithms and data access patterns. Coordinating the overall data communication, computational load balancing, and scaling these different physics solvers together on modern massively parallel, heterogeneous high performance computing systems presents several major challenges. We have adopted the Legion programming system, via the Regent programming language, and its task parallel programming model to address these challenges. Our multi-physics solver Soleil-X is written entirely in the high level Regent programming language and is one of the largest and most complex applications written in Regent to date. At this workshop we will give an overview of the software architecture of Soleil-X as well as discuss how our multi-physics solver was designed to use the task parallel programming model provided by Legion. We will also discuss the development experience, scaling, performance, portability, and multi-physics simulation results.

Exploring the Use of Novel Programming Models in Land Surface Models
Ethan T. Coon (Oak Ridge National Laboratory)

A wide range of programming models are currently under rapid development to meet the needs of application developers looking to work on more complex machines. These models fill a variety of roles. Some look to abstract supercomputer architecture, including both processors and memory, to present a strategy for portable performance across a wide range of machines. Others look to expose concurrency by explicitly constructing task-driven dependency graphs that allow a scheduler to find parallelism. Here we explore the implications for application codes of adopting two such programming models, Kokkos and Legion, one from each class of models. We specifically focus on the software design implications on refactoring existing applications, rather than the performance and performance tuning of these models. We identify a strategy for refactoring the Energy Exascale Earth System Model’s Land Surface Model, an extremely complex code for climate applications, and prototype a series of mini-apps that explore the adoption of Kokkos and Legion. In doing this, we identify commonalities across the models, leading to a series of conclusions about application software design and refactoring for the adoption of novel programming models.

Simulating Ultralight Dark Matter with Chapel: An Experience Report
Nikhil Padmanabhan (Yale University)
We describe our implementation of an astrophysical code to simulate the dynamics of ultralight dark matter in Chapel. We focus on the programmability of Chapel, highlighting the relative ease of translating the physics of this system into a code that can run efficiently on distributed systems. We also demonstrate that this code can scale well from small problem sizes that can be run on laptops to large problem sizes run across hundreds of processors. We finally present the results from a few simulations of astrophysical interest. An interesting by-product of broader interest is a distributed FFT routine written in Chapel; we summarize its implementation and performance.

**A UPC++ Actor Library and Its Evaluation On a Shallow Water Proxy Application**

*Alexander Pöppl (Technical University of Munich)*

Programmability is one of the key challenges of Exascale Computing. Using the actor model for distributed computations may be one solution. The actor model separates computation from communication while still enabling their overlap. Each actor possesses specified communication endpoints to publish and receive information. Computations are undertaken based on the data available on these channels. We present a library that implements this programming model using UPC++, and evaluate three different parallelization strategies, one based on rank-sequential execution, one based on multiple threads in a rank, and one based on OpenMP tasks. In an evaluation of our library using shallow water proxy applications, our solution compares favorably against an earlier implementation based on X10, and a BSP-based approach.

**Evaluation of Programming Models to Address Load Imbalance on Distributed Multi-Core CPUs: A Case Study with Block Low-Rank Factorization**

*Yu Pei (University of Tennessee)*

To minimize data movement, many parallel applications statically distribute computational tasks among the processes. However, modern simulations often encounters irregular computational tasks. As a result, load imbalance among the processes must be dealt with at the programming level.

One critical application for many domains is the LU factorization of a large dense matrix stored in the Block Low-Rank (BLR) format. Using the low-rank format can significantly reduce the cost of factorization in many scientific applications, including the boundary element analysis of electrostatic field. However, the partitioning of the matrix based on underlying geometry leads to different sizes of the matrix, thus load imbalance among the processes at each step of factorization.

We use BLR LU factorization as a test case to study the programmability and performance of five different programming approaches: (1) flat MPI, (2) Adaptive MPI (Charm++), (3) MPI + OpenMP, (4) parameterized task graph (PTG), and (5) dynamic task discovery (DTD). The last two versions use a
task-based paradigm to express the algorithm; we rely on the PaRSEC runtime system to execute the tasks. We first point out programming features needed to efficiently solve this category of problems, hinting at possible alternatives to the MPI+X programming paradigm. We then evaluate the programmability of the different approaches. Finally, we show the performance result on the Intel Haswell--based Bridges system and analyze the effectiveness of the implementations to address the load imbalance.

**Scalable Machine Learning with OpenSHMEM**
Gerard Taylor (Intel Corporation)

Deep convolutional neural networks (DNNs) have had a significant, and lasting impact across the computing industry. Training these large neural networks is computationally intensive and is often parallelized to shorten training times that could otherwise range from days to weeks. The Message Passing Interface (MPI) communication model has been commonly used to facilitate the data exchange and synchronization required for parallel DNN training. We observe that OpenSHMEM supports many of the same communication operations as MPI — in particular, the all-reduce operation needed to support data parallelism — and that OpenSHMEM may further provide a unique solution to fine-grain model parallel computation. In this work, we present an initial evaluation of OpenSHMEM’s suitability for use in DNN training and compare its performance with MPI. Results indicate that OpenSHMEM data-parallel performance is comparable with MPI. The usage of OpenSHMEM to support model parallelism will be explored in our future work.

**PAW-ATM Lunch Break**

**Designing, Implementing, and Evaluating the Upcoming OpenSHMEM Teams API**
David Ozog (Intel Corporation)

For many years, the OpenSHMEM parallel programming interface has provided a high-performance alternative to MPI that emphasizes one-sided messaging, simplifies communication across a global memory space, and bolsters the capabilities of rapidly evolving fabric interconnect technologies. The OpenSHMEM specification standardizes the library interfaces, prioritizing a performant and portable API. The specification continues to mature with vigorous support from several authoritative vendors and researchers. For example, the OpenSHMEM specification committee is actively standardizing a unique teams API that enables user-defined subsets of application processes to efficiently and productively perform communication operations, such as collectives routines, remote memory accesses, and remote atomic operations.
This paper describes the OpenSHMEM teams interface and several interesting aspects and challenges in implementing the API, as well as possible extensions that could improve the programmability and/or performance. We evaluate the performance of a preliminary implementation and show that using teams effectively can facilitate impressive improvements of collective operations at scale, even while simplifying the underlying programming model.

Enabling Low-Overhead Communication in Multi-threaded OpenSHMEM Applications using Contexts
Wenbin Lu (Stony Brook University)

As the number of shared-memory cores per node in modern High Performance Computing (HPC) machines continues to grow, hybrid programming models like MPI+threads are becoming a preferred choice for scientific applications. While being able to utilize computation resources efficiently, threads in hybrid applications often compete with each other for communication resources, resulting in a negative impact on performance. The OpenSHMEM distributed programming model provides communication context objects that can be used to provide threads with isolated access to the network, thus reducing contention. In this work, we discuss a design for OpenSHMEM contexts and an implementation of the context construct to support hybrid multi-threaded applications and evaluate the performance of the implementation. In all our micro-benchmarks, threads show nearly identical communication performance compared to single-threaded OpenSHMEM processes. By using contexts in hybrid benchmarks, we have achieved up to 43.1% performance improvement for 3D halo exchange, 339% improvement for all-to-all communication, and 35.4% improvement for inter-node load balancing.

Efficient Active Message RMA in GASNet Using a Target-Side Reassembly Protocol (Extended Abstract)
Paul H. Hargrove (Lawrence Berkeley National Laboratory)

GASNet is a portable, open-source, high-performance communication library designed to efficiently support the networking requirements of PGAS runtime systems and other alternative models on future exascale machines. This paper investigates strategies for efficient implementation of GASNet’s “AM Long” API that couples an RMA transfer with an Active Message (AM) delivery.

We discuss several network-level protocols for AM Long and propose a new target-side reassembly protocol. We present a microbenchmark evaluation on the Cray XC Aries network hardware. The target-side reassembly protocol on this network improves AM Long end-to-end latency by up to 33%, and the effective bandwidth by up to 49%, while also enabling asynchronous source completion that drastically reduces injection overheads.
The improved AM Long implementation for Aries is available in GASNet-EX release v2019.9.0 and later.

PAW-ATM Afternoon Break

Pygion: Flexible, Scalable Task-Based Parallelism with Python
Elliott Slaughter (SLAC National Accelerator Laboratory)

Dynamic languages provide the flexibility needed to implement expressive support for task-based parallel programming constructs. We present Pygion, a Python interface for the Legion task-based programming system, and show that it can provide features comparable to Regent, a statically typed programming language with dedicated support for the Legion programming model. Furthermore, we show that the dynamic nature of Python permits the implementation of several key optimizations (index launches, futures, mapping) currently implemented in the Regent compiler. Together these features enable Pygion code that is comparable in expressiveness to Regent, but more flexible, and substantially more concise, less error prone, and easier to use than C++ Legion code. We show that, in combination with high-performance kernels written in the Regent programming language, Pygion is able to achieve efficient, scalable execution on up to 1024 nodes of the heterogeneous supercomputer Piz Daint.

Arkouda: NumPy-like arrays at massive scale backed by Chapel
Michael H. Merrill (US Department of Defense)

Exploratory data analysis (EDA) is a prerequisite for all data science, as illustrated by the ubiquity of Jupyter notebooks, the preferred interface for EDA among data scientists. In order to expand the scale of EDA, we have designed and built a software package for using NumPy-like arrays and Pandas-like data frames at massive scale. We have also integrated this software package into our data science workflow. This software package, called Arkouda, is primarily implemented in the Chapel programming language. Arkouda is currently in process to be open-sourced so our work can be shared with a broader community. In the proposed talk we will cover the why, how, and what of Arkouda.

Panel Discussion: Applications in Alternatives to MPI+X

Panel Chair: Patrick McCormick — Los Alamos National Laboratory
Panelists: Alex
Women in HPC: Diversifying the HPC Community and Engaging Male Allies

Session Description: The eleventh international Women in HPC workshop will be held at SC19, Denver, USA with the goal of increasing the recruitment, retention, and success of women in the international HPC workforce. The WHPC workshop series has become the leading international event on improving equity, diversity and inclusion, providing all attendees, irrespective of their gender, with the skills to thrive in the workplace and the information to build frameworks for diverse recruitment and retention policies and actions. Discussions provide benefits to both the individual, by assisting in personal development and also the employer as the discussions help improve productivity and workforce retention. WHPC@SC19 will focus on the following topics: - Surviving difficult events and how to minimize the impact on your career - Managing and resolving imposter syndrome - Building an effective network - How to get a new job or promotion - Behaviors for inclusion: coping strategies for unconscious bias and micro-aggression - Being a parent, guardian and caregiver: dealing with the guilt - Pointers on making and engaging male allies at workplace. We will also provide opportunities aimed at promoting and providing women with the skills to thrive in HPC including short lightning talks by women working in HPC. [https://womeninhpc.org/whpc-sc19/workshop/](https://womeninhpc.org/whpc-sc19/workshop/)

The eleventh international Women in HPC workshop will be held at SC19, Denver, USA with the goal of increasing the recruitment, retention, and success of women in the international HPC workforce.

The WHPC workshop series has become the leading international event on improving equity, diversity and inclusion, providing all attendees, irrespective of their gender, with the skills to thrive in the workplace and the information to build frameworks for diverse recruitment and retention policies and actions. Discussions provide benefits to both the individual, by assisting in personal development and also the employer as the discussions help improve productivity and workforce retention.
WHPC@SC19 will focus on the following topics: - Surviving difficult events and how to minimize the impact on your career - Managing and resolving imposter syndrome - Building an effective network - How to get a new job or promotion - Behaviors for inclusion: coping strategies for unconscious bias and micro-aggression - Being a parent, guardian and caregiver: dealing with the guilt - Pointers on making and engaging male allies at workplace.

We will also provide opportunities aimed at promoting and providing women with the skills to thrive in HPC including short lightning talks by women working in HPC.

**Keynote: The Butterfly Effect of Inclusive Leadership**
Bev Crair (Lenovo)
Keynote talk

**Women in HPC Morning Break**

**Thriving at Workplace (Short Talks)**
Hai Ah Nam (Los Alamos National Laboratory), Elizabeth Bautista (Lawrence Berkeley National Laboratory), Jo Adegbola (Amazon Web Services), AJ Lauer (National Center for Atmospheric Research (NCAR))
This session will have short talks on the following topics: - Surviving difficult events & minimizing the impact on your career - Managing and resolving imposter syndrome during a career change - Coping strategies for unconscious bias -- Elizabeth Bautista, NERSC - Being a parent, guardian and caregiver: dealing with guilt - Dealing with sexism in the workplace

**Panel Discussion: Engaging Male Allies**
Cory Snavely (National Energy Research Scientific Computing Center (NERSC)), Ian Foster (Argonne National Laboratory), Brendan Bouffler (Amazon Web Services), Patty Lopez (Intel Corporation), Misbah Mubarak (Amazon Web Services)
Research shows that engaging male allies in gender inclusion programs has lead to significant progress where 96% of organizations see progress when males are engaged vs. 30% in cases where males are not engaged. The panel would discuss strategies on engaging males allies and the impact it is having on diversity and inclusion efforts.
Women in HPC Lunch Break

Early/Mid Career Lightning Talks Session
Mariam Umar (Intel Corporation)
The lightning talks session will have presentations from early/mid career women about their work in the area of HPC, diversity and inclusion.

Women in HPC Afternoon Break

Panel Discussion: Mentoring: Why We Need It, Best Practice, and the Business Case for Great Mentoring Programs
Toni Collis (Women-in-HPC), Mariam Umar (Intel Corporation), Sarvani Chadalapaka (University of California, Merced), Bronis Supinski (Lawrence Livermore National Laboratory)
Our primary goal is to talk about effective mentoring, explain that it goes beyond the traditional technical guidance and that it can make/break someone's career when they have access to good (and bad!) mentoring. We want to encourage the leaders in the room to create valuable mentoring programmes and offer mentorship that really works, as well as help the younger audience to understand how to get a mentor even when there aren't official programmes to apply to.

Women-in-HPC Workshop Closing
Toni Collis (Women-in-HPC), Misbah Mubarak (Amazon Web Services)
Workshop closing remarks and outcomes

9:00 am - 5:30 pm

Workshop on Exascale MPI (ExaMPI)
Session Description: 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.

https://sites.google.com/site/workshopexampi/

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.

ExaMPI Keynote

ExaMPI Morning Break

Multirate: A Flexible MPI Benchmark for Fast Assessment of Multithreaded Communication Performance
George Bosilca (University of Tennessee)

As the modern hardware landscape continues to drastically change, the degree of parallelism required to maintain a high occupancy of resources has substantially increased. These hardware changes have highlighted the limitations of the traditional method of using one process per processing unit, which indicates that a more flexible programming paradigm is necessary. In the context of the message passing paradigm, MPI needs a significant improvement in threaded performance in order to fully utilize all hardware capabilities. However, for developers to know what needs to be improved, and for users to know what performance to expect, benchmarks are needed
to quickly assess the capabilities and performance of MPI implementations. This paper introduces a new communication benchmark designed to replicate typical application communication patterns and assess their performance with a varied amount of resources. We evaluate three MPI implementations with our benchmark suite and assess their strengths and weaknesses.

**Impacts of Multi-GPU MPI Collective Communications on Large FFT Computation**

*Alan Ayala (University of Tennessee)*

Most applications targeting exascale, such as those part of the Exascale Computing Project (ECP), are designed for heterogeneous architectures and rely on the Message Passing Interface (MPI) as their underlying parallel programming model. In this paper we analyze the limitations of collective MPI communication for the computation of fast Fourier transforms (FFTs), which are relied on heavily for large-scale particle simulations. We present experiments made at one of the largest heterogeneous platforms, the Summit supercomputer at ORNL. We discuss communication models from state-of-the-art FFT libraries, and propose a new FFT library, named HEFFTE (Highly Efficient FFTs for Exascale), which supports heterogeneous architectures and yields considerable speedups compared with CPU libraries, while maintaining good weak as well as strong scalability.

**Node-Aware Improvements to Allreduce**

*Amanda J. Bienz (University of Illinois)*

The MPI_Allreduce collective operation is a core kernel of many parallel codebases, particularly for reductions over a single value per process. The commonly used allreduce recursive-doubling algorithm obtains the lower bound message count, yielding optimality for small reduction sizes based on node-agnostic performance models. However, this algorithm yields duplicate messages between sets of nodes. Node-aware optimizations in MPICH remove duplicate messages through use of a single master process per node, yielding a large number of inactive processes at each inter-node step. In this paper, we present an algorithm that uses the multiple processes available per node to reduce the maximum number of inter-node messages communicated by a single process, improving the performance of allreduce operations, particularly for small message sizes.

**Accelerating the Global Arrays ComEx Runtime Using Multiple Progress Ranks**

*Nitin Gawande (Pacific Northwest National Laboratory (PNNL)), Bruce Palmer (Pacific Northwest National Laboratory (PNNL)), Sriram Krishnamoorthy (Pacific Northwest National Laboratory (PNNL)), Joseph Manzano (Pacific Northwest National Laboratory (PNNL))*

Abstract—Partitioned Global Address Space (PGAS) models are a part of system software that is being designed to support communication runtimes for exascale applications. MPI has been shown to be a viable option to develop a scalable PGAS communication subsystem and has the
advantages of its standardization and higher performance. We used MPI two-sided semantics with a combination of automatic and user defined splitting of MPI communicators to achieve asynchronous progress. Our implementation can make use of multiple asynchronous progress ranks (PR) per node that can be mapped to the computing architecture of a node in a distributed cluster. We are able to show significant speed up of over 2.0X and scaling of a communication bound computational chemistry application distributed over 1024 nodes of state-of-the-art HPC clusters. Our results show that while running a communication bound application workload on a certain number of cluster nodes, an optimum number of ranks dedicated for communication can be found to achieve asynchronous communication progress and obtain highest performance.

ExaMPI Lunch Break

RDMA-Based Library for Collective Operations in MPI
Alexander Margolin (Hebrew University of Jerusalem)

In most MPI implementations, abstraction layers separate the collective operation algorithms from the communication primitives, thus hindering its optimization with network acceleration technologies, such as RDMA. Open UCX is an RDMA-based point-to-point communication library, that can reduce the latency between processes in MPI applications, particularly in large-scale system. This paper presents a design and implementation of a library for MPI collective operations, by extending Open UCX. Our approach is transparent to MPI applications, and can reduce the latency of repeated calls to such operations by an average of 8% for relatively small message sizes and as much as 90% for larger messages.

Using MPI-3 RMA for Active Messages
Joseph Schuchart (High Performance Computing Center Stuttgart, University of Stuttgart)

Distributed asynchronous programming systems require a scalable, low-latency way to exchange information on the state of units of execution (tasks) across process boundaries. These interactions can be modeled as active messages, i.e., a message that, upon reception, triggers an action on the receiving side. Such actions, generally of short duration, include the posting of the reception for a task input data or the transition of a task into the runnable state. Traditionally, such messages have been implemented using MPI two-sided communication primitives. However, processing reception queues for numerous small messages may incur significant overheads in the progress of the MPI library, notably when handling unexpected messages and iterating over a number of pre-posted requests. In this work, we investigate a different approach for implementing an active message
queue using MPI-3 RMA primitives and compare the resulting performance in a task runtime in terms of scalable performance and latency against two-sided implementations. We discuss the latency of basic RMA operations and the resulting performance in terms of latency and scalable throughput of the RMA-based queues. Using a benchmark application, we show that using RMA-based queues may provide notable benefits compared to using MPI two-sided communication for inter-scheduler communication.

ExaMPI Afternoon Break

ExaMPI Invited Talk #1 - Evaluating MPI Message Size Summary Statistics
Kurt Ferreira (Sandia National Laboratories)

ExaMPI Invited Talk #2 - The Case for Modular Generalizable Proxy Applications for Systems Software Research
William Marts (University of New Mexico)

ExaMPI Vendor Panel - MPI for Exascale: Challenges and Opportunities

9:00 am - 5:30 pm

H2RC 2019: Fifth International Workshop on Heterogeneous High-Performance Reconfigurable Computing

Session Description: As in the previous four years, this workshop will bring together application experts, software developers, and hardware engineers, both from industry and academia, to share experiences and best practices to leverage the practical application of reconfigurable logic to Scientific Computing, Machine/Deep Learning, 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, SYCL, C, C++, and other high-level design flows, which enable and improve cross-platform functional and performance portability while also improving productivity. Particular emphasis is given to cross-platform comparisons and combinations 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, GPU, Many-cores and hybrid devices, ASICs), and on how to most effectively achieve cross-platform compatibility.  

http://h2rc.cse.sc.edu/

As in the previous four years, this workshop will bring together application experts, software developers, and hardware engineers, both from industry and academia, to share experiences and best practices to leverage the practical application of reconfigurable logic to Scientific Computing, Machine/Deep Learning, 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, SYCL, C, C++, and other high-level design flows, which enable and improve cross-platform functional and performance portability while also improving productivity. Particular emphasis is given to cross-platform comparisons and combinations 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, GPU, Many-cores and hybrid devices, ASICs), and on how to most effectively achieve cross-platform compatibility.

SYCL: A Single-Source C++ Standard for Heterogeneous Computing  
Ronan Keryell (Xilinx Inc)

hlslib: Software Engineering for Hardware Design  
Johannes de Fine Licht (ETH Zurich)

High-level synthesis (HLS) tools have brought FPGA development into the mainstream, by allowing programmers to design architectures using familiar languages such as C, C++, and OpenCL. While the move to these languages has brought significant benefits, many aspects of traditional software engineering are still unsupported, or not exploited by developers in practice. Furthermore, designing reconfigurable architectures requires support for hardware constructs and workflows that are not covered by CPU-oriented tools and languages. To address this gap, we have developed hlslib, a
collection of software tools, plug-in hardware modules, and code samples, designed to enhance the productivity of HLS developers. The goal of hlslib is two-fold: first, create a community-driven arena of bleeding edge development, which can move quicker, and provides more powerful abstractions than what is provided by vendors; and second, collect a wide range of example codes, both minimal proofs of concept, and larger, real-world applications, that can be reused directly or inspire other work. hlslib is offered as an open source library, containing CMake files, C++ headers, convenience scripts, and examples codes, and is receptive to any contribution that can benefit HLS developers, through general functionality or examples.

Morning Coffee Break

2GRVI Phalanx: A 1332-Core RISC-V RV64I Processor Cluster Array with an HBM2 High Bandwidth Memory System, and an OpenCL-like Programming Model, in a Xilinx VU37P FPGA [WIP Report]
Jan Gray (Gray Research LLC)

2GRVI (and its predecessor, GRVI) are FPGA-efficient 64b (resp. 32b) RISC-V processing element cores. Phalanx is a parallel processor and accelerator array overlay framework. Groups of PEs and accelerator cores form shared memory compute clusters. Clusters, DRAM, NICs and other I/O controllers communicate by message passing on an FPGA-optimal Hoplite torus soft NoC. This extended abstract summarizes work-in-progress to redesign the 2017 GRVI Phalanx to take advantage of new Xilinx FPGAs with 460 GB/s dual stack HBM2 DRAM-in-package, and to provide a familiar parallel programming experience via an OpenCL-like programming model and tools. The new system is the first kilocore RV64I SoC and the first RISC-V multiprocessor with an HBM2 memory system.

CFD Acceleration with FPGA
Viraj Paropkari (Xilinx Inc; byteLAKE, Poland), Marcin Rojek (byteLAKE, Poland), Krzysztof Rojek (byteLAKE, Poland; Czestochowa University of Technology, Poland)

The goal of this work is to adapt 4 CFD kernels to the Xilinx ALVEO U250 FPGA, including first-order step of the non-linear iterative upwind advection MPDATA schemes (non-oscillatory forward in time), the divergence part of the matrix-free linear operator formulation in the iterative Krylov scheme, tridiagonal Thomas algorithm for vertical matrix inversion inside preconditioner for the iterative solver, and computation of the pseudovelocity for the second pass of upwind algorithm in MPDATA. All the kernels use 3-dimensional compute domain consisted from 7 to 11 arrays. Since
all the kernels belong to the group of memory bound algorithms, our main challenge is to provide the highest utilization of global memory bandwidth. In this work we present FPGA as a powerful device to accelerate HPC codes. The proposed adaptation is compared with a CPU implementation that was strongly optimized in order to provide realistic and objective benchmarks. Our adaptation allows us to reduce the execution time up to 40% and the energy consumption up to 80% compared to the CPU processors.

Data Flow Pipes: A SYCL Extension for Spatial Architectures
Michael Kinsner (Intel Corporation), John Freeman (Intel Corporation)

FIFOs are a common construct in design for spatial and data flow architectures. OpenCL 2.0 defined a “pipe” feature to expose the FIFO construct, but the design didn’t meet all needs of spatial architectures. This talk describes a pipes extension to the Khronos SYCL single-source, C++-based programming framework, that exposes a pipe abstraction which closes the gaps in the OpenCL design, while also offering a more usable interface. The C++ type system is leveraged to provide static connectivity guarantees without extensive compiler implementation effort, and to provide well-defined interaction with C++ features. The described extension provides a usable interface that can also act as a substrate for additional abstractions to be built on top. This talk will motivate the utility of FIFOs/pipes in high level language FPGA design, describe the SYCL pipes extension and its mapping to SPIR-V and OpenCL, and provide examples of use in common spatial design patterns.

Morning 5-Minute Break

It's All about Data Movement: Optimizing FPGA Data Access to Boost Performance
Nick Brown (Edinburgh Parallel Computing Centre, University of Edinburgh)

The use of reconfigurable computing, and FPGAs in particular, to accelerate computational kernels has the potential to be of great benefit to scientific codes and the HPC community in general. However, whilst recent advanced in FPGA tooling have made the physical act of programming reconfigurable architectures much more accessible, in order to gain good performance the entire algorithm must be rethought and recast in a dataflow style. Reducing the cost of data movement for all computing devices is critically important, and in this paper we explore the most appropriate techniques for FPGAs. We do this by describing the optimisation of an existing FPGA implementation of an atmospheric model's advection scheme. By taking an FPGA code that was over four times slower than running on the CPU, mainly due to data movement overhead, we
describe the profiling and optimisation strategies adopted to significantly reduce the runtime and bring the performance of our FPGA kernels to a much more practical level for real-world use. The result of this work is a set of techniques, steps, and lessons learnt that we have found significantly improves the performance of FPGA based HPC codes and that others can adopt in their own codes to achieve similar results.

The Memory Controller Wall: Benchmarking the Intel FPGA SDK for OpenCL Memory Interface
Hamid Reza Zohouri (Tokyo Institute of Technology; EdgeCortex Inc, Japan)

Supported by their high power efficiency and recent advancements in High Level Synthesis (HLS), FPGAs are quickly finding their way into HPC and cloud systems. Large amounts of work have been done so far on loop and area optimizations for different applications on FPGAs using HLS. However, a comprehensive analysis of the behavior and efficiency of the memory controller of FPGAs is missing in literature, which becomes even more crucial when the limited memory bandwidth of modern FPGAs compared to their GPU counterparts is taken into account. In this work, we will analyze the memory interface generated by Intel FPGA SDK for OpenCL with different configurations for input/output arrays, vector size, interleaving, kernel programming model, on-chip channels, operating frequency, padding, and multiple types of overlapped blocking. Our results point to multiple shortcomings in the memory controller of Intel FPGAs, especially with respect to memory access alignment, that can hinder the programmer’s ability in maximizing memory performance in their design. For some of these cases, we will provide work-arounds to improve memory bandwidth efficiency; however, a general solution will require major changes in the memory controller itself.

Accelerating Large Garbled Circuits on an FPGA-Enabled Cloud
Miriam Leeser (Northeastern University)

Garbled Circuits (GC) is a technique for ensuring the privacy of inputs from users and is particularly well suited for FPGA implementations in the cloud where data analytics is frequently run. Secure Function Evaluation, such as that enabled by GC, is orders of magnitude slower than processing in the clear. We present our best implementation of GC on Amazon Web Services (AWS) that implements garbling on Amazon’s FPGA enabled F1 instances. In this paper we present the largest problems garbled to date on FPGA instances, which includes problems that are represented by over four million gates. Our implementation speeds up garbling 20 times over software over a range of different circuit sizes.

H2RC’19 Lunch Break
ML Acceleration with Heterogeneous Computing for Big Data Physics Experiments
Philip Harris (Massachusetts Institute of Technology (MIT))

Testbed for the Research Community Exploring Next-Generation Cloud Platforms
Miriam Leeser (Northeastern University)

Afternoon Break

High-Throughput Multi-Threaded Sum-Product Network Inference in the Reconfigurable Cloud
Lukas Sommer (Technical University Darmstadt)

Large cloud providers have started to make powerful FPGAs available as part of their public cloud offers. One promising application area for this kind of instances is the acceleration of machine learning tasks.

This work presents an accelerator architecture that uses multiple accelerator cores for the inference in so-called Sum-Product Networks and complements it with a host software interface that overlaps data-transfer and actual computation.

The evaluation shows that, the proposed architecture deployed to Amazon AWS F1 instances is able to outperform a 12-core Xeon processor by a factor of up to 1.9x and a Nvidia Tesla V100 GPU by a factor of up to 6.6x.

Implementation and Impact of an Ultra-Compact Multi-FPGA Board for Large System Prototyping
Fabien Chaix (Institute of Computer Science, Foundation for Research and Technology - Hellas)

Efficient prototyping of a large complex system can be significantly facilitated by the use of a flexible and versatile physical platform where both new hardware and software components can readily be
implemented and tightly integrated in a timely manner. Towards this end, we have developed the 120 130 mm QFDB board and associated firmware, including the system software environment. We developed a large system based on this advanced dense and modular building block. The QFDB features 4 interconnected Xilinx Zynq Ultrascale+ devices, each one consisting of an ARM-based subsystem tightly coupled with reconfigurable logic. Each Zynq Ultrascale+ is connected to 16 GB of DDR4 memory. In addition, one Zynq provides storage through an M.2 Solid State Disk (SSD). In this paper, we present the design and the implementation of this board, as well as the software environment for board operation. Moreover, we describe a 10 Gb Ethernet communication infrastructure for interconnecting multiple boards together. Finally, we highlight the impact of this board on a number of ongoing research activities that leverage the QFDB versatility, both as a large scale prototyping system for HPC solutions, and as a host for the development of FPGA integration techniques.

Afternoon 5-Minute Break

FBLAS: Streaming Linear Algebra Kernels on FPGA
Tiziano De Matteis (ETH Zurich)

Reconfigurable hardware represents an attractive alternative to load-store architectures, as it allows eliminating expensive control and data movement overheads in computations. In practice, these devices are often not considered in the high-performance computing community, due to the steep learning curve and low productivity of hardware design, and the lack of available library support for fundamental operations. We present FBLAS, an open source implementation of Basic Linear Algebra Subroutines (BLAS) for FPGAs. The library is implemented with a modern HLS tool to promote productivity, reusability, and maintainability. Numerical routines are designed to be easily composed exploiting on-chip connections, to reduce off-chip communication resulting in lower communication volume.

Combining Perfect Shuffle and Bitonic Networks for Efficient Quantum Sorting
Naveed Mahmud (University of Kansas)

The emergence of quantum computers in the last decade has generated research interest in applications such as quantum sorting. Quantum sorting plays a critical role in creating ordered sets of data that can be better utilized, e.g., quantum ordered search or quantum network switching. In this paper, we propose a quantum sorting algorithm that combines highly parallelizable bitonic merge networks with perfect shuffle permutations (PSP), for sorting data represented in the
quantum domain. The combination of bitonic networks with PSP improves the temporal complexity of bitonic merge sorting which is critical for reducing decoherence effects for quantum processing. We present space-efficient quantum circuits that can be used for quantum bit comparison and permutation. We also present a reconfigurable hardware quantum emulator for prototyping the proposed quantum algorithm. The emulator has a fully-pipelined architecture and supports double-precision floating-point computations, resulting in high throughput and accuracy. The proposed hardware architectures are implemented on a high-performance reconfigurable computer (HPRC). In our experiments, we emulated quantum sorting circuits of up to 31 fully-entangled quantum bits on a single FPGA node of the HPRC platform. To the best of our knowledge, our effort is the first to investigate a reconfigurable hardware emulation of quantum sorting using bitonic networks and perfect shuffle.

Performance and Energy Efficiency Analysis of Reverse Time Migration on a FPGA Platform
Adhvan Furtado (SENAI CIMATEC)

Reverse time migration (RTM) modeling is a computationally intensive component in the seismic processing workflow of oil and gas exploration, often demanding the manipulation of terabytes of data. Therefore, the computational kernels of the RTM algorithms need to access a large range of memory locations. However, most of these accesses result in cache misses, degrading the overall system performance. GPGPUs and FPGAs are the two endpoints in the spectrum of acceleration platforms, since both can achieve better performance in comparison to CPU on several high-performance applications. Recent literature highlights FPGA better energy efficiency when compared to GPGPU. The present work proposes a FPGA accelerated platform prototype targeting the computation of the RTM algorithm on an HPC environment. Experimental results highlight that speedups of 112x can be achieved, when compared to a sequential execution on CPU. When compared to a GPU, the power consumption has been reduced up to 55%.

9:00 am - 5:30 pm

Workshop on Education for High Performance Computing (EduHPC)

Session Description: 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). Data science curriculum (e.g. for new degree programs and within data science centers) and topics related to Internet of Things are also in the workshop scope. 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
Welcome: Workshop on Education for High Performance Computing (EduHPC)
Sushil Prasad (University of Texas, San Antonio), Debzani Deb (Winston-Salem State University), Trilce Estrada (University of New Mexico)

Keynote: Preparing HPC Education for a Post-Moore Future
Vivek Sarkar (Georgia Institute of Technology)
Though considerable progress has been made on integrating parallel and distributed computing into Computer Science curricula, many challenges still remain especially earlier in the undergraduate curriculum and later in continuing education for computing professionals. The first part of this talk will summarize experiences with HPC education in a sophomore-level course titled "Fundamentals of Parallel Programming" (COMP 322) introduced at Rice University in 2011. The learning outcomes for COMP 322 fall into three course modules -- Parallelism, Concurrency, and Distribution. We will also present experiences with a Coursera specialization that was introduced in 2017 based on COMP 322 material, but adapted into a continuing education format for computing professionals. This specialization consists of three courses, derived from the three modules in COMP 322.

In the second part of the talk, we will summarize future computing trends as we approach the end of Moore's Law, and identify pedagogic concepts that will be relevant for the post-Moore era. These "extreme heterogeneity" disruptions include new kinds of heterogenous processors and accelerators, heterogeneous memories, near/in-memory computation structures, and even non von Neumann computing elements. In our opinion, selected concepts from current pedagogy can
provide a valuable starting point for creating new pedagogical materials for the post-Moore era. These concepts include task parallelism, data flow execution, and locality/distribution control as first-class primitives, which can enable strong semantic guarantees, as well as a promising foundation for post-Moore systems and also modern heterogeneous processors including CPUs, GPUs, DSPs, FPGAs, and NMPs.

**Morning Break**

**Teaching Parallel and Distributed Computing Concepts in Simulation with WRENCH**
Ryan Tanaka (University of Hawaii at Manoa)

Teaching topics related to high performance computing and parallel and distributed computing in a hands-on manner is challenging, especially at introductory, undergraduate levels. There is a participation challenge due to the need to secure access to a platform on which students can learn via hands-on activities, which is not always possible. There are also pedagogic challenges. For instance, any particular platform provided to students imposes constraints on which learning objectives can be achieved. These challenges become steeper as the topics being taught target more heterogeneous, more distributed, and/or larger platforms, as needed to prepare students for using and developing cyberinfrastructure.

To address the above challenges, we have developed a set of pedagogic activities that can be integrated piecemeal in university courses, starting at freshman levels. These activities use simulation so that students can experience hands-on any relevant application and platform scenarios. This is achieved by capitalizing on the capabilities of the WRENCH and SimGrid simulation frameworks. After describing our approach and the pedagogic activities currently available, we present results from an evaluation performed in an undergraduate university course.

**A Gentle Introduction to Heterogeneous Computing for CS1 Students**
Apan Qasem (Texas State University)

Heterogeneous architectures have emerged as a dominant platform, not only in high performance computing but also in mobile processing, cloud computing, and the Internet of Things (IoTs). Because the curriculum is over-crowded in its current state, it is difficult to include a new course as a required part of the curriculum without increasing the number of hours to graduation. Integration of heterogeneous computing content requires a module-based approach, such as those undertaken for introducing parallel and distributed computing
In this paper, we present a teaching module that introduces CS1 students to some of the fundamental concepts in heterogeneous computing. The goal of this module is not to teach students how to program heterogeneous systems but rather expose them to this emerging trend and prepare them for material they are expected to see in future classes. Although concepts are covered at a high-level, the module emphasizes active learning and includes a lab assignment that provides students hands-on experience with respect to task mapping and performance evaluation of a heterogeneous system. The module was implemented at our home institution in Fall 2018. Initial evaluation results are quite encouraging both in terms of learning outcomes and student engagement and interest.

Assessing the Integration of Parallel and Distributed Computing in Early Undergraduate Computer Science Curriculum Using Unplugged Activities
Srishti Srivastava (University of Southern Indiana), Mary Smith (Hawai‘i Pacific University)

Due to the pervasive presence of parallel and distributed computing (PDC) in most of the application domains, the need for PDC skills in computing professionals is necessary. This has inspired much interest in including PDC topics in existing computer science (CS) undergraduate curriculum, as evidenced in the ACM/IEEE joint curriculum recommendations. We address this gap by developing new active learning style PDC activities. The application and evaluation of the unplugged activities are described in this paper. The activities highlight key benefits of parallel computing, and the appropriate application of important PDC concepts. An assessment of student engagement has been utilized to measure the effectiveness of our active learning PDC modules. We determine whether the engagement differs by gender, age, and class standing (freshman, sophomore, junior, senior). Results indicated there were slight differences with respect to age and class standing. This research contributes to helping other practitioners in the CS community by providing PDC lesson modules to be integrated into their existing courses. The developed activities and the related lesson plans are intended to provide easier learning of the fundamental PDC concepts in an early undergraduate CS curriculum.

Paper Session A: Discussions

Lightning Talk Session
Trilce Estrada (University of New Mexico)
Parallelism for Beginners with Fun Examples

The presentation deals with the following:

- Assembling in parallel a 3D jigsaw puzzle ball.
- Parallel sorting of binary coded numbered cards.
- Using the twenty one card trick and its generalizations to motivate beginning students of Computer Science and Mathematics.

Actionable Guidance for Fresh HPC Researchers
Hang Liu (Stevens Institute of Technology)

High Performance Computing (HPC) and, in general, Parallel and Distributed Computing (PDC) has become pervasive, so does the pertinent research. While a collection of research advice write ups has surged in the community, very few of them are actionable and specific for HPC researchers. This lightning talk aims to provide such guidelines. As an early success, we have posted these guidelines on GitHub (https://github.com/asherliu/researchHOWTO), which draws around 30 stars and positive feedbacks. In addition, our graduate and undergraduate students find the advice to be constructive.

Hearing Program Behavior with TSAL
Joel C. Adams (Calvin University)

Much work has been done in the area of real-time algorithm visualization, in which a program produces a graphical representation of its behavior as it executes. In this lightning talk, we examine the relatively uncharted territory of real-time audialization, in which a program produces a sonic representation of its behavior as it executes. Such work seems apt to be beneficial for auditory learners, especially those with visual disabilities. To support this exploration, we have created the Thread Safe Audio Library (TSAL), a platform-independent, object-oriented C++ library that provides thread-safe classes for mixing, synthesizing, and playing sounds. Using TSAL, we can create an audialization by taking a working program and adding library calls that generate behavior-representing sounds in real time. If a program is multithreaded, each thread can play distinct sounds, allowing us to hear the multithreaded behavior. This lightning talk provides an overview of TSAL and demonstrates several audializations, including the Producers-Consumers Problem, Parallel MergeSort, and others.
A Jupyter Notebook Based Tool for Building Skills in Computational Statistical Mechanics  
Benjamin H. Glick (Lewis & Clark College)

When learning about High-Performance Computing, almost of equal importance to developing computational skills is building an understanding of the applications and implications computational research can have. One field in which high-performance computing is particularly important is computational physics. Many physicists depend on high-performance hardware and software to carry out their research. This project describes an open-source, Jupyter notebook-based web app which helps students both develop skills relating to and appreciation for HPC techniques and builds understanding of complex physical models in the domain of thermodynamics/quantum field theory. This project is designed to help students understand, appreciate, and build skills relating to simulations of statistical systems, which are too complex to describe analytically and instead need to be described by complex monte-carlo simulations.

Teaching and Advising HPC at NTNU  
Anne C. Elster (Norwegian University of Science and Technology)

Since the world was forced into parallelism in the mid 2000s as we hit the Power Wall, Parallel and Distributed Computing (PDC) and High Performance Computing (HPC) in particular, is no longer a specialty seen at the US government labs and weather forecasting centers, but is now everywhere from embedded systems in smartphones that typically have one or more multicore processor and 3 or more GPUs, to laptops, workstations, large server and server farms to supercomputers. This has created a huge demand for technical talent with HPC skills. Big Data, AI, and associated Data Analytics has put even higher pressure on the demand.

In this short-paper/lightning talk, I will highlight some of my personal experiences in attracting students into my parallel computing class, my experience with advising 80+ graduate students (most of them master students), and current efforts and ideas for introducing HPC to first-year students. This includes why I believe in teaching MPI before OpenMP, threading or CUDA, why optimizing code/performance engineering is so important, the importance of group meeting and team building, and how and why parallel computing concepts should also be taught in the very first introduction to computers classed.

I will also mention the need to have good mentors early in the curriculum. Finally, I will highlight the need to have good mentors early in the curriculum and do out-reach to not only high school students, but also grade schools and middle schools.

The Data Science Journal Club  
William Monroe (University of Alabama, Birmingham), Ravi Tripathi (University of Alabama,
In 2012, the Harvard Business Review termed data science "The Sexiest Job of the 21st century." Since that time, with the rise of GPU computation and the concurrent rise deep learning, "data science" skills are sought after more than ever. In early 2019, job postings for data science on Indeed.com have increased by 256% since December 2013. As such, many academic disciplines are searching for ways of integrating data science techniques into their programs. As many data science techniques are computationally intensive, it is natural to pair data science training with training in proper use of HPC resources. At UAB, Research Computing has initiated a weekly Data Science Journal club. By training members in data science, we naturally touch on HPC topics as well, making this an ideal frontier for HPC education. Similarly to traditional academic journal clubs, new topics are discussed each week; however, rather than reading papers, the club focuses on working through practical data science applications with published code.

Excellence in Computer Engineering Education (EXCEED): Integrating PDC Topics in the ECE Courses at UIUC
Yuting Chen (University of Illinois)

The widespread presence of multicore and general-purpose graphics processing units (GPUs) in PCs, laptops, and now even handhelds has changed the computing landscape and empowered users to make valued, innovative contributions to the technology of computing. However, the current undergraduate Electrical and Computer Engineering (ECE) curriculum of the University of Illinois at Urbana-Champaign (UIUC) at the introductory level has not yet adequately addressed the needs for this rapid change in computing hardware platforms, devices, languages, and supporting programming environments. Recognizing this urgency, the principal goal of our EXCEED project is to expose our freshmen and sophomore students to the concepts of parallel and distributed computing (PDC) by integrating modules of PDC topics in the existing curriculum. The main focus is to allow our students to develop parallel computational thinking while they learn the core topics on von-Neumann-model-based sequential computational techniques.

Increasing Undergraduate Research in Computational Sciences
Maria Pantoja (California Polytechnic State University)

High Performance Computing (HPC) is used everywhere from e-commerce to machine learning. The main problems solved by HPC are related to computational sciences and affect almost all scientific disciplines. Training the students that will have sufficient knowledge in HPC is a crucial task of computer science departments in all educational institutions. This paper describes the steps taken to increase the undergraduate engagement in HPC disciplines in our undergraduate institution. The increased exposure allowed our computer science students to collaborate with other
departments research projects by accelerating different computational intensive problems. In order to increase students knowledge in HPC we decided to offer classes at the undergraduate level in parallel programming and distributed systems; and to properly support these classes we invested in a lab specifically designed for parallel programming, with machines equipped with the best multicores cpus and gpus so students can experience the effects of massively parallel architectures in their code.

Using detailed enrollment, master thesis and publications from the last few years, since the last has been available; we can see a substantial increase in undergraduate students interest in the HPC area, and a great success in the number of research master thesis as well as the number of peer reviewed undergraduate and graduate student-authored publications.

**Building a Data Science Education Ecosystem Resource Collection**
Gladys K. Andino (Purdue University, University of Virginia)

Purdue University is creating a Data Science Education Ecosystem (DSEE) to prepare students for an increasingly data-driven world. Projects include The Data Mine, an undergraduate community of learning communities from many disciplines united by an interest in data science; disciplinary credit courses in digital humanities, microbiome analysis, and GIS; and modular courses and workshops on data science skills. These activities have created a need for innovative instructional material to support and disseminate these projects.

Our project was funded to create a Data Science Education Ecosystem Resource Collection (DSEERC) leveraging the expertise of faculty in the Purdue University Libraries and School of Information Studies (PULSIS), the expertise and infrastructure of Research Computing, and the students and instructors of The Data Mine. The scope of the collection will be curated instructional material produced by and for DSEE efforts on campus. Formats will include interactive tutorials, videos, lessons and datasets with an emphasis on modular content that can be used in courses or as needed by students. The DSEERC will lower the barrier to data science education because learners will have access to material even if they are not currently enrolled in class. Instructors can use the resource to explore new ways of teaching data science, minimize duplication of effort by reusing resources, and to grasp the DSEE landscape. This resource is critical for Purdue if we hope to build data literacy and fluency for all learners and to promote best practices in data science education.

**Lightning Talk Discussions**
Lunch Break

PDC Curriculum Update
Charles Weems (University of Massachusetts, Amherst), Sushil Prasad (University of Texas, San Antonio), Alan Sussman (University of Maryland)

Teaching on Demand: an HPC Experience
Rocío Carratalá-Sáez (Jaume I University)

In this work, we present the experience of the course "Build your own supercomputer with Raspberry Pi", offered as a non-mandatory workshop with the purpose of bringing High Performance Computing (HPC) closer to bachelor students of Universitat Jaume I (UJI, Spain). The intention of the course is twofold; on the one hand, we target toward increasing the knowledge of Computer Science and Engineering students about the labor performed by the HPC community; on the other hand, we aim to create a personalized experience for each student by fulfilling their curiosity about the topics presented and discussed in the class. In order to evaluate the impact and learning, we analyze two surveys filled out by the students respectively before and after the course, where HPC interest and knowledge are exposed.

Successful Systems in Production Graduate Teaching
Ali Shoker (HASLab, INESC TEC and Minho Univ.)

We present our experience in coordinating and teaching a novel graduate systems and computing course named “Successful Systems in Production” (SSP). The course targets graduate students of different research interests in Computer Science. The course aims at giving a breadth knowledge on cutting-edge well-known systems in production, and exploring the potential synergies across different areas of research. Having its roots in Distributed Computing, SSP addresses those systems that overlap with other research areas like Computational Systems, Parallel Computing, Databases, Cloud Computing, Artificial Intelligence, Security, etc. SSP exhibits an agile topic selection model that fits several students’ backgrounds in each academic year. The topics focus on the practical aspects of each selected system that is considered “successful”, i.e., based on its worldwide impact and technical significance. This is important for graduate students to acquire best practices in industry and academia, necessary to build practical computing systems. In the same
vein, the assessment method includes a project that is based on one of the presented systems and also intersects with the student’s own research plan. Based on our teaching experience and the excellent feedback of the students, we strongly recommend this graduate course to be taught at other universities.

**Paper Session B: Discussions**

Katharine Cahill (Ohio Supercomputer Center)

**Afternoon Break**

**Teaching Concurrent and Distributed Programming With Concepts Over Mathematical Proofs**

David Marchant (University of Copenhagen, Niels Bohr Institute; University of Copenhagen)

We describe how a concept based approach to teaching was used to update how concurrent and distributed systems were taught at the University of Copenhagen. This approach focuses on discussion to drive student engagement whilst fostering a deeper understanding of the presented topics compared to more traditional displays of crude facts. The course is split into three sections: local concurrency, networked concurrency, and concurrency in hardware. This allows for an easier student journey through the course, as they are introduced to all core concepts in the first section, then have them reinforced in greater detail in the subsequent sections. Finally, the experience gained in updating this course is presented so others attempting to do similar may learn from it.

**Measuring the Impact of HPC Training**

Julian Miller (RWTH Aachen University)

The demand for computational power is rising rapidly in science and engineering and is now expanding into areas from the social sciences and humanities. This is met with the use of large-scale, and increasingly complex, High-Performance Computing (HPC) systems. Programming such heterogeneous and quickly evolving systems efficiently has always been challenging, particularly for novice HPC programmers, but is continually increasing in complexity. HPC service providers are therefore growing their training offerings as part of their overall service provision. However, there is no standard for understanding the impact and effectiveness of this training. Traditional training and education metrics, such as examination, are typically not applicable. Additionally, the key criteria for service providers are not only knowledge gained but the impact that the knowledge has on HPC
usage. To capture and evaluate the effectiveness of training, a novel methodology with two key components is proposed: progress productivity and training productivity. The methodology includes a process for deriving an activity-specific weighted selection of the metrics that comprise the productivity measures. The proposed productivity criteria could then be applied to improve the training of HPC application programmers taking the participant's pre-knowledge and focus areas into account.

**Toward Improving Collaborative Behavior During Competitive Programming Assignments**

Arturo Gonzalez-Escribano (University of Valladolid, Spain)

Competitive gamification has been successfully used to improve the learning experience in computer science courses. Assignments based on high-performance and parallel programming contests have been shown to be highly effective to promote the interest and involvement of students. Nevertheless, during programming-contest assignments the competitive approach sometimes diminishes the usual collaboration between students to solve basic questions and common difficulties; for example, those related to the usage of the programming languages or tools. In this paper, we present an approach to integrate collaborative gamification techniques with the competitive elements of programming contests, in order to balance both aspects and to enhance the overall learning experience. The results show that this approach effectively promotes participation and collaborative behavior during a programming contest. They also reveal potential improvements that can be considered for this approach.

**Paper Session C: Discussions**

**Peachy Parallel Assignment Session**

David Bunde (Knox College)

**Parallel Computing in the Cloud for All Levels**

Steven A. Bogaerts (DePauw University)

This document describes a project, adaptable for all undergraduate levels, in which students access cloud computing resources and leverage them in a parallel application.

**Pin Finder**
Daniel A. Ellsworth (Colorado College)

A basic assignment that connects to parallel programming, security, and performance topics.

Dask Processing and Analytics for Large Datasets
Alina Lazar (Youngstown State University)

This paper describes the assignment titled "Dask Analytics" that is used for student evaluation as part of a graduate data science and data mining course. For this assignment students are required to read, process and answer queries using a large dataset that does not fit in the RAM memory of a commodity laptop. Using the Python framework Dask, which extends a small set of Pandas's operations, students can become familiar with parallel and distributed processing. In addition, the assignment teaches students about the basics operations implemented in Dask in a very interesting and applied way, as well as operations and algorithms that are harder to parallelize.

Agent-Based Simulation of Fire Extinguishing: an Assignment for OpenMP, MPI, and CUDA/OpenCL
Arturo Gonzalez-Escribano (University of Valladolid, Spain)

We present a new assignment used in a parallel computing course to teach the approaches to the same problem in different parallel programming models. It targets concepts of shared-memory programming with OpenMP, distributed-memory programming with MPI, and/or GPU programming with CUDA or OpenCL. This assignment is based on a simplified agent-based simulation where teams of firefighters aim to extinguish a set of fire focal points in a dynamically evolving scenario. The program is designed to be simple, easy to understand by students, and to include specific parallelization and optimization opportunities. Although there is a quite direct parallel solution in the three programming models, the program has plenty of opportunities for further improvements. It extends the ideas of a previously presented assignment, in order to use more interesting data structures, load balancing techniques, and code optimizations. It has been successfully used in parallel programming contests during a real course, using the performance obtained by the students’ code as a measure of success.

Thread-to-Core Optimization Problem
Muhammad Alfian Amrizal (Tohoku University)

Introducing Parallel Prefix-Sums Algorithm and Its Applications in an Undergraduate Course
Sukhamay Kundu (Louisiana State University)
The parallel prefix-sums algorithm and its variations are often used as key substeps in more complex parallel computations. We have successfully introduced the parallel prefix-sum algorithm and its key ideas in a senior level undergraduate Computer Science course. To improve the understanding of these ideas, students were required to solve several problems using simple variations of the basic prefix-sums algorithm. They were also required to solve several non-trivial problems, including their 2-dimensional generalization. We briefly discuss four problems covered in the second category.

Peachy Parallel Assignment Discussions

Best Paper Announcement and Closing Remarks

9:00 am - 5:30 pm

Workshop on Programming and Performance Visualization Tools (ProTools)

Session Description: Understanding program behavior is critical to overcome the expected architectural and programming complexities, such as limited power budgets, heterogeneity, hierarchical memories, shrinking I/O bandwidths, and performance variability, that arise on modern HPC platforms. To do so, HPC software developers need intuitive support tools for debugging, performance measurement, analysis, and tuning of large-scale HPC applications. Moreover, data collected from these tools such as hardware counters, communication traces, and network traffic can be far too large and too complex to be analyzed in a straightforward manner. We need new automatic analysis and visualization approaches to help application developers intuitively understand the multiple, interdependent effects that algorithmic choices have on application correctness or performance. The Workshop on Programming and Performance Visualization Tools (ProTools) intends to bring together HPC application developers, tool developers, and researchers from the visualization, performance, and program analysis fields for an exchange of new approaches to assist developers in analyzing, understanding, and optimizing programs for extreme-scale platforms. Covering all aspects of HPC program and performance analysis - data collection, visualization, and tools - the ProTools workshop intends to attract a broad set of research communities, and facilitate better interactions between these communities. The Virtual Institute -
High Productivity Supercomputing (VI-HPS) will provide support for the ProTools workshop. VI-HPS is an international initiative of HPC researchers and developers focused on programming and performance tools for parallel systems. [https://protools19.github.io](https://protools19.github.io)

Understanding program behavior is critical to overcome the expected architectural and programming complexities, such as limited power budgets, heterogeneity, hierarchical memories, shrinking I/O bandwidths, and performance variability, that arise on modern HPC platforms. To do so, HPC software developers need intuitive support tools for debugging, performance measurement, analysis, and tuning of large-scale HPC applications. Moreover, data collected from these tools such as hardware counters, communication traces, and network traffic can be far too large and too complex to be analyzed in a straightforward manner. We need new automatic analysis and visualization approaches to help application developers intuitively understand the multiple, interdependent effects that algorithmic choices have on application correctness or performance. The Workshop on Programming and Performance Visualization Tools (ProTools) intends to bring together HPC application developers, tool developers, and researchers from the visualization, performance, and program analysis fields for an exchange of new approaches to assist developers in analyzing, understanding, and optimizing programs for extreme-scale platforms.

Covering all aspects of HPC program and performance analysis - data collection, visualization, and tools - the ProTools workshop intends to attract a broad set of research communities, and facilitate better interactions between these communities. The Virtual Institute - High Productivity Supercomputing (VI-HPS) will provide support for the ProTools workshop. VI-HPS is an international initiative of HPC researchers and developers focused on programming and performance tools for parallel systems.

**Keynote**

Stephane Eranian (Google LLC)

**ProTools Morning Break**
Understanding the Performance of GPGPU Applications from a Data-Centric View

Using a CPU-GPU hybrid computing framework is becoming a common configuration for supercomputers. The wide deployment of GPUs (as well as other hardware accelerators) brings to the HPC community a big question: Are we using them effectively? Inappropriate use of GPUs can generate incorrect results in certain cases, but more often, will slow down the program instead of speeding it up. This paper describes a tool that satisfies the needs of programmers to analyze the runtime performance of kernels and obtain insights for better GPU utilization. Compared to existing GPU performance tools, ours provides some unique features: data-centric profiling and generating complete GPU call stacks. With the guidance of the tool, we were able to improve the kernel performance of three widely-studied GPU benchmarks by a factor of up to 46.6x with minor code modification.

Asvie: A Timing-Agnostic SVE Optimization Methodology

As we are quickly approaching exascale and moving onwards towards the next challenge, we are exploring a wider range of technologies and architectures. The further out the timeframes considered, the less likely prototype hardware is available. A popular method of exploring new architectural extensions is to emulate them on existing platforms. The Arm Instruction Emulator (ArmIE) is such a tool, which we use on existing Armv8 platforms to run Arm's latest vector architecture, the Scalable Vector Extension (SVE).

To aid with porting applications towards SVE, we developed an application optimization methodology based on ArmIE that uses timing-agnostic metrics to assess application quality. We show how we have successfully optimized the High Performance Conjugate Gradient (HPCG) High Performance Computing benchmark to SVE by using our methodology, resulting in a hand-optimized intrinsics-based version.

Designing Efficient Parallel Software via Compositional Performance Modeling

Performance models are powerful instruments for understanding the performance of parallel systems and uncovering their bottlenecks. Already during system design, performance models can help ponder alternatives. However, creating a performance model - whether theoretically or empirically - for an entire application that does not exist yet is challenging unless the interactions between all system components are well understood, which is often not the case during design. In
this paper, we propose to generate performance models of full programs from performance models of their components using formal composition operators derived from parallel design patterns such as pipeline or task pool. As long as the design of the overall system follows such a pattern, its performance model can be predicted with reasonable accuracy without an actual implementation.

**Performance Analysis of Tile Low-Rank Cholesky Factorization Using PaRSEC Instrumentation Tools**

This paper highlights the necessary development of new instrumentation tools within the PaRSE task-based runtime system to leverage the performance of low-rank matrix computations. In particular, the tile low-rank (TLR) Cholesky factorization represents one of the most critical matrix operations toward solving challenging large-scale scientific applications. The challenge resides in the heterogeneous arithmetic intensity of the various computational kernels, which stresses PaRSE’s dynamic engine when orchestrating the task executions at runtime. Such irregular workload imposes the deployment of new scheduling heuristics to privilege the critical path, while exposing task parallelism to maximize hardware occupancy. To measure the effectiveness of PaRSE’s engine and its various scheduling strategies for tackling such workloads, it becomes paramount to implement adequate performance analysis and profiling tools tailored to fine-grained and heterogeneous task execution. This permits us not only to provide insights from PaRSE, but also to identify potential applications’ performance bottlenecks. These instrumentation tools may actually foster synergism between applications and PaRSE developers for productivity as well as high-performance computing purposes. We demonstrate the benefits of these amenable tools, while assessing the performance of TLR Cholesky factorization from data distribution, communication-reducing and synchronization-reducing perspectives. This tool-assisted performance analysis results in three major contributions: a new hybrid data distribution, a new hierarchical TLR Cholesky algorithm, and a new performance model for tuning the tile size. The new TLR Cholesky factorization achieves an 8X performance speedup over existing implementations on massively parallel supercomputers, toward solving large-scale 3D climate and weather prediction applications.

**ProTools Lunch Break**

**The Case for a Common Instrumentation Interface for HPC Codes**
Lightweight timekeeping functionality for basic performance logging, regression testing, and anomaly detection is essential in HPC codes. We present the Caliper, TiMemory, and PerfStubs libraries that have recently been developed as common solutions for these tasks. Lightweight, always-on profiling solutions are typically built around user-defined instrumentation points, which can benefit a variety of use cases beyond application timekeeping. We argue for the creation of a tool-agnostic adapter layer to make these instrumentation points available to third-party tools, runtime systems, and system software.

Keynote / Panel [TBD]

ProTools Afternoon Break

Automatic Instrumentation Refinement for Empirical Performance Modeling

The analysis of runtime performance is important during the development and throughout the life cycle of HPC applications. One important objective in performance analysis is to identify regions in the code that show significant runtime increase with larger problem sizes or more processes. One approach to identify such regions is to use empirical performance modeling, i.e., building performance models based on measurements. While the modeling itself has already been streamlined and automated, the generation of the required measurements is time consuming and tedious. In this paper, we propose an approach to automatically adjust the instrumentation to reduce overhead and focus the measurements to relevant regions, i.e., such that show increasing runtime with larger input parameters or increasing number of MPI ranks. Our approach employs Extra-P to generate performance models, which it then uses to extrapolate runtime and, finally, decide which functions should be kept for measurement. Also, the analysis expands the instrumentation, by heuristically adding functions based on static source-code features. We evaluate our approach using benchmarks from SPEC CPU 2006, SU2, and parallel MILC. The evaluation shows that our approach can filter functions of little interest and generate profiles that contain mostly relevant regions. For example, the overhead for SU2 can be improved automatically from 200% to 11% compared to filtered Score-P measurements.

Multi-Level Performance Instrumentation for Kokkos Applications Using TAU
The TAU Performance System® provides a multi-level instrumentation strategy for instrumentation of Kokkos applications. Kokkos provides a performance portable API for expressing parallelism at the node level. TAU uses the Kokkos profiling system to expose performance factors using user-specified parallel kernel names for lambda functions or C++ functors. It can also use instrumentation at the OpenMP, CUDA, pthread, or other runtime levels to expose the implementation details giving a dual focus of higher-level abstractions as well as low-level execution dynamics. This multi-level instrumentation strategy adopted by TAU can highlight performance problems across multiple layers of the runtime system without modifying the application binary.

CHAMPVis: Comparative Hierarchical Analysis of Microarchitectural Performance

Performance analysis and optimization are essential tasks for hardware and software engineers. In the age of datacenter-scale computing, it is particularly important to conduct comparative performance analysis to understand discrepancies and limitations among different hardware systems and applications. However, there is a distinct lack of productive visualization tools for these comparisons.

We present CHAMPVis, a web-based, interactive visualization tool that leverages the hierarchical organization of hardware systems to enable productive performance analysis. With CHAMPVis, users can make definitive performance comparisons across applications or hardware platforms. In addition, CHAMPVis provides methods to rank and cluster based on performance metrics to identify common optimization opportunities. Our thorough task analysis reveals three types of datacenter-scale performance analysis tasks: summarization, detailed comparative analysis, and interactive performance bottleneck identification. We propose techniques for each class of tasks including (1) 1-D feature space projection for similarity analysis; (2) Hierarchical parallel co-ordinates for comparative analysis; and (3) User interactions for rapid diagnostic queries to identify optimization targets. We evaluate CHAMPVis by analyzing standard datacenter applications and machine learning benchmarks in two different case studies.

In Situ Visualization of Performance Metrics in Multiple Domains

As application scientists develop and deploy simulation codes on to leadership-class computing resources, there is a need to instrument these codes to better understand performance to efficiently utilize these resources. This instrumentation may come from independent third-party tools that generate and store performance metrics or from custom instrumentation tools built directly into the
application. The metrics collected are then available for visual analysis, typically in the domain in which they were collected. In this paper, we introduce an approach to visualize and analyze the performance metrics in situ in the context of the machine, application, and communication domains (MAC model) using a single visualization tool. This visualization model provides a holistic view of the application performance in the context of the resources where it is executing.

**Toward a Programmable Analysis and Visualization Framework for Interactive Performance Analytics**

Understanding the performance characteristics of applications in modern HPC environments is becoming more challenging due to the increase in the architectural and programming complexities. HPC software developers rely on sources such as hardware counters and event traces to infer performance problems while focusing on designing mitigation strategies. A large number of in-house tools exist in the community, which indicate replicated effort. This paper presents a customizable framework for analyzing performance measurements and visualizing through a web-based interactive dashboard for interactively exploring a large volume of hierarchical information. In this paper, we analyze three ECP applications as use cases and identify as well as optimize problematic resource utilization behaviors exposed by our visualizations. This framework is a step towards a unified platform for visual identification of performance scaling bottlenecks to ease the collaboration between application developers, performance analysts, and hardware vendors.

9:00 am - 5:30 pm

**The 14th Workshop on Workflows in Support of Large-Scale Science (WORKS19)**

**Session Description:** Data-intensive workflows (aka 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
Data-intensive workflows (aka 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.

WORKS19 Welcome
Sandra Gesing (University of Notre Dame), Rafael Ferreira da Silva (University of Southern California)

WORKS19 Keynote: Priority Research Directions for In Situ Data Management: Enabling Scientific Discovery from Diverse Data Sources
Thomas Peterka (Argonne National Laboratory)
Scientific computing will increasingly incorporate a number of different tasks that need to be managed along with the main simulation or experimental tasks—ensemble analysis, data-driven science, artificial intelligence, machine learning, surrogate modeling, and graph analytics—all nontraditional applications unheard of in HPC just a few years ago. Many of these tasks will need to execute concurrently, that is, in situ, with simulations and experiments sharing the same computing resources.

There are two primary, interdependent motivations for processing and managing data in situ. The first motivation is the need to decrease data volume. The in situ methodology can make critical
contributions to managing large data from computations and experiments to minimize data movement, save storage space, and boost resource efficiency—often while simultaneously increasing scientific precision. The second motivation is that the in situ methodology can enable scientific discovery from a broad range of data sources—HPC simulations, experiments, scientific instruments, and sensor networks—over a wide scale of computing platforms: leadership-class HPC, clusters, clouds, workstations, and embedded devices at the edge.

The successful development of in situ data management capabilities can potentially benefit real-time decision making, design optimization, and data-driven scientific discovery. This talk will feature six priority research directions that highlight the components and capabilities needed for in situ data management to be successful for a wide variety of applications: making in situ data management more pervasive, controllable, composable, and transparent, with a focus on greater coordination with the software stack, and a diversity of fundamentally new data algorithms.

WORKS19 Morning Break

Provenance Data in the Machine Learning Lifecycle in Computational Science and Engineering
Renan Souza (Federal University of Rio de Janeiro, IBM Research), Leonardo Azevedo (IBM Corporation), Vítor Lourenço (IBM Corporation), Elton Soares (IBM Corporation), Raphael Thiago (IBM Corporation), Rafael Brandão (IBM Corporation), Daniel Civitarese (IBM Corporation), Emilio Brazil (IBM Corporation), Marcio Moreno (IBM Corporation), Patrick Valduriez (French Institute for Research in Computer Science and Automation (INRIA)), Marta Mattoso (Federal University of Rio de Janeiro, IBM Research), Renato Cerqueira (IBM Corporation), Marco Netto (IBM Corporation)

Machine Learning (ML) has become essential in several industries. In Computational Science and Engineering (CSE), the complexity of the ML lifecycle comes from the large variety of data, scientists’ expertise, tools, and workflows. If data are not tracked properly during the lifecycle, it becomes unfeasible to recreate a ML model from scratch or to explain to stackholders how it was created. The main limitation of provenance tracking solutions is that they cannot cope with provenance capture and integration of domain and ML data processed in the multiple workflows in the lifecycle, while keeping the provenance capture overhead low. To handle this problem, in this paper we contribute with a detailed characterization of provenance data in the ML lifecycle in CSE; a new provenance data representation, called PROV-ML, built on top of W3C PROV and ML Schema; and extensions to a system that tracks provenance from multiple workflows to address the characteristics of ML and CSE, and to allow for provenance queries with a standard vocabulary. We show a practical use in a real case in the O&G industry, along with its evaluation using 239,616
CUDA cores in parallel.

A Codesign Framework for Online Data Analysis and Reduction
Kshitij Mehta (Oak Ridge National Laboratory), Bryce Allen (University of Chicago, Argonne National Laboratory), Matthew Wolf (Oak Ridge National Laboratory), Jeremy Logan (Oak Ridge National Laboratory), Eric Suchyta (Oak Ridge National Laboratory), Jong Choi (Oak Ridge National Laboratory), Keichi Takahashi (Nara Institute of Science and Technology), Igor Yakushin (Argonne National Laboratory), Todd Munson (Argonne National Laboratory), Ian Foster (Argonne National Laboratory), Scott Klasky (Oak Ridge National Laboratory)

In this paper we discuss our design of a toolset for automating performance studies of composed HPC applications that perform online data reduction and analysis. We describe Cheetah, a new framework for performing parametric studies on coupled applications. Cheetah facilitates understanding the impact of various factors such as process placement, synchronicity of algorithms, and storage vs. compute requirements for online analysis of large data. Ultimately, we aim to create a catalog of performance results that can help scientists understand tradeoffs when designing next-generation simulations that make use of online processing techniques. We illustrate the design choices of Cheetah by using a reaction-diffusion simulation (Gray-Scott) paired with an analysis application to demonstrate initial results of fine-grained process placement on Summit, a pre-exascale supercomputer at Oak Ridge National Laboratory.

Top-Down Performance Analysis Methodology for Workflows: Tracking Performance Issues from Overview to Individual Operations
Ronny Tschueter (Technical University Dresden), Christian Herold (Technical University Dresden), William Williams (Technical University Dresden), Maximilian Knespel (Technical University Dresden), Matthias Weber (Technical University Dresden)

Scientific workflows are well established in parallel computing. A workflow represents a conceptual description of work items and their dependencies. Researchers can use workflows to abstract away implementation details or resources to focus on the high-level behavior of their work items. Due to these abstractions and the complexity of scientific workflows, finding performance bottlenecks along with their root causes can quickly become involving. This work presents a top-down methodology for performance analysis of workflows to support users in this challenging task. Our work provides summarized performance metrics covering different workflow perspectives, from general overview to individual jobs and their job steps. These summaries allow to identify inefficiencies and determine the responsible job steps. In addition, we record detailed performance data about job steps, enabling a fine-grained analysis of the associated execution to exactly pinpoint performance issues. The introduced methodology provides a powerful tool for comprehensive performance analysis of complex workflows.
Data-Aware and Simulation-Driven Planning of Scientific Workflows on IaaS Clouds
Tchimou N’Takpé (Nangui Abrogoua University, Ivory Coast), Jean Edgar Gnimassoun (Nangui Abrogoua University, Ivory Coast), Souleymane Oumtanaga (Nangui Abrogoua University, Ivory Coast), Frédéric Suter (CC IN2P3 / CNRS)

The promise of an easy access to a virtually unlimited number of resources makes Infrastructure as a Service Clouds a good candidate for the execution of data-intensive workflow applications composed of hundreds of computational tasks. Thanks to a careful execution planning, Workflow Management Systems can build a tailored compute infrastructure by combining a set of virtual machine instances. However, these applications usually rely on files to handle dependencies between tasks. A storage space shared by all virtual machines may become a bottleneck and badly impact the application execution time.

In this paper, we propose an original data-aware planning algorithm that leverages two characteristics of a family of virtual machines instances, i.e., a large number of cores and a dedicated storage space on fast SSD drives, to improve data locality, hence reducing the amount of data transfers over the network during the execution of a workflow. We also propose a simulation-driven approach to solve a cost-performance optimization problem and correctly dimension the virtual infrastructure onto which execute a given workflow. Experiments conducted with real application workflows show the benefits of the presented algorithms. The data-aware planning leads to a clear reduction of both execution time and volume of data transferred over the network while the simulation-driven approach allows us to dimension the infrastructure in a reasonable time.

Exploration of Workflow Management Systems Emerging Features from Users Perspectives
Ryan Mitchell (University of Southern California), Loïc Pottier (University of Southern California), Steve Jacobs (National Ecological Observatory Network), Rafael Ferreira da Silva (University of Southern California), Mats Rynge (University of Southern California), Karan Vahi (University of Southern California), Ewa Deelman (University of Southern California)

There has been a recent emergence of new workflow applications focused on data analytics and machine learning. This emergence has precipitated a change in the workflow management landscape, causing the development of new data-oriented workflow management systems (WMSs) as opposed to the earlier standard of task-oriented WMSs. In this paper, we summarize three general workflow use-cases and explore the unique requirements of each use-case in order to understand how WMSs from both workflow management models (task-driven workflow management models and data-driven workflow management models) meet the requirements of each workflow use-case from the user’s perspective. We analyze the applicability of the two main workflow models by carefully describing each model and by providing an examination of the
different variations of WMSs that fall under the task-driven model. To illustrate the strengths and weaknesses of each workflow management model, we summarize the key features of four production-ready WMSs: Pegasus, Makeflow, Apache Airflow, and Pachyderm. Of these production-ready WMSs, three belong to the task-driven workflow management model (i.e., Pegasus, Makeflow, Apache Airflow) and one belongs to the data-driven workflow management model (i.e., Pachyderm). To deepen our analysis of the four WMSs examined in this paper, we implement three real-world use-cases to highlight the specifications and features of each WMS. The application of these real-world use-cases demonstrates how each workflow management model operates with the different applications. We present our final assessment of each WMS after considering the following factors: usability, performance, ease of deployment, and relevance. The purpose of this work is to offer insights from the user's perspective into the research challenges that WMSs currently face due to the evolving workflow landscape.

WORKS19 Lunch Break

Incorporating Scientific Workflows in Computing Research Processes
Shantenu Jha (Rutgers University), Scott Lathrop (University of Illinois), Jarek Nabrzyski (University of Notre Dame), Lavanya Ramakrishnan (Lawrence Berkeley National Laboratory)
The articles in this special section explore scientific workflows in computer research processes. The goal is to increase awareness of the benefits of workflows to enhance computational and data-enabled research and to foster the exchange of lessons learned and good practices that can benefit the community. The issue highlights some of the activities and approaches that are underway in the scientific workflow community. Scientific workflows evolved as a way to manage computation on High Performance Computing (HPC) and distributed systems. Early workflow efforts started as domain-specific efforts that managed directed acyclic graphs on high-performance and distributed systems. They considered the systems and the applications as black boxes and focused on distributed resource management, workload, and execution management.

Comparing GPU Power and Frequency Capping: A Case Study with the MuMMI Workflow
Tapasya Patki (Lawrence Livermore National Laboratory), Zachary Frye (Lawrence Livermore National Laboratory), Harsh Bhatia (Lawrence Livermore National Laboratory), Francesco Di Natale (Lawrence Livermore National Laboratory), James Glosli (Lawrence Livermore National Laboratory), Helgi Ingolfsson (Lawrence Livermore National Laboratory), Barry Rountree (Lawrence Livermore National Laboratory)
Accomplishing the goal of exascale computing under a potential power limit requires HPC clusters to maximize both parallel efficiency and power efficiency. As modern HPC systems embark on a trend toward extreme heterogeneity leveraging multiple GPUs per node, power management becomes even more challenging, especially when catering to scientific workflows with co-scheduled components. The impact of managing GPU power on workflow performance and run-to-run reproducibility has not been adequately studied.

In this paper, we present a first-of-its-kind research to study the impact of the two power management knobs that are available on NVIDIA Volta GPUs: frequency capping and power capping. We analyzed performance and power metrics of GPU's on a top-10 supercomputer by tuning these knobs for more than 5,300 runs in a scientific workflow. Our data found that GPU power capping in a scientific workflow is an effective way of improving power efficiency while preserving performance, while GPU frequency capping is a demonstrably unpredictable way of reducing power consumption. Additionally, we identified that frequency capping results in higher variation and anomalous behavior on GPUs, which is counterintuitive to what has been observed in the research conducted on CPUs.

WORKS19 Afternoon Break

Inter-Job Scheduling for High-Throughput Material Screening Applications
Zhihui Du (Tsinghua University, China), Xinning Hui (Tsinghua University, China), Yurui Wang (Tsinghua University, China), Jun Jiang (Beijing Computing Center), Jason Liu (Florida International University), Baokun Lu (Tsinghua University, China), Chongyu Wang (Tsinghua University, China)

Material screening entails a large number of electronic structure simulations that solve the Kohn-Sham (KS) equations based on the high-throughput density functional theory. The computational efficiency for solving the KS equations, however, is largely dependent on the initial displacement of the particles. Traditionally, simulation runs are treated separately as solving independent KS equations with their own initial conditions. As such, one cannot take advantage of the potential correlation between different simulations. In this paper, we propose to explore the job inter-dependencies and formulate material screening as an inter-job scheduling problem. More specifically, We schedule the material screening jobs according to the potential correlation between them and run the simulations by reusing the particle distribution from previous runs as the initial condition. We propose a heuristic inter-job scheduling algorithm and compare that with an optimal case where the pair-wise correlation between different simulations is known {\em a priori}. Experiments on the Sunway TaihuLight supercomputer show that the total time to run the large
number of jobs in one practical material screening example is significantly reduced, by as much as 89%.

Empowering Agroecosystem Modeling with HTC Scientific Workflows: The Cycles Model Use Case

Rafael Ferreira da Silva (University of Southern California), Rajiv Mayani (University of Southern California), Yuning Shi (Pennsylvania State University), Armen R. Kemanian (Pennsylvania State University), Mats Rynge (University of Southern California), Ewa Deelman (University of Southern California)

Scientific workflows have enabled large-scale scientific computations and data analysis, and lowered the entry barrier for performing computations in distributed heterogeneous platforms (e.g., HTC and HPC). In spite of impressive achievements to date, large-scale modeling, simulation, and data analytics in the long-tail still face several challenges such as efficient scheduling and execution of large-scale workflows ($O(10^6)$) with very short-running tasks (few seconds). While the current trend to support next-generation workflows on leadership class machines have gained much attention in the past years, at the other end of the spectrum scientific workflows from the long-tail science have become larger and require processing massive volumes of data. In this paper, we report on our experience in designing and implementing an HTC workflow for agroecosystem modeling. We leverage well-known (task clustering and co-scheduling) and emerging (hierarchical workflows and containers) workflow optimization techniques to make the workflow planning problem tractable, and maximize resource utilization and the degree of task parallelism.

Experimental results, via the implementation of a use case, show that by strategically combining the above strategies and defining an appropriate set of optimization parameters, the overall workflow makespan can be improved by 3.5 orders of magnitude when compared to a regular (non-optimized) execution of the workflow.

A performance comparison of Dask and Apache Spark for data-intensive neuroimaging pipelines

Mathieu Dugré (Concordia University), Valérie Hayot-Sasson (Concordia University), Tristan Glatard (Concordia University)

In the past few years, neuroimaging has entered the Big Data era due to the joint increase in image resolution, data sharing, and study sizes. However, no particular Big Data engines have emerged in this field, and several alternatives remain available. We compare two popular Big Data engines with Python APIs, Apache Spark and Dask, for their runtime performance in processing neuroimaging pipelines. Our evaluation uses two synthetic pipelines processing the 81GB BigBrain image, and a real pipeline processing anatomical data from more than 1,000 subjects. We benchmark these pipelines using various combinations of task durations, data sizes, and numbers of workers,
deployed on an 8-node (8 cores ea.) compute cluster in Compute Canada’s Arbutus cloud. We evaluate PySpark’s RDD API against Dask’s Bag, Delayed and Futures. Results show that despite slight differences between Spark and Dask, both engines perform comparably. However, Dask pipelines risk being limited by Python’s GIL depending on task type and cluster configuration. In all cases, the major limiting factor was data transfer. While either engine is suitable for neuroimaging pipelines, more effort needs to be placed in reducing data transfer time.

On a Parallel Spark Workflow for Frequent Itemset Mining Based on Array Prefix-Tree
Xinzheng Niu (University of Electronic Science and Technology of China), Mideng Qian (University of Electronic Science and Technology of China), Chase Wu (New Jersey Institute of Technology), Aiqin Hou (Northwest University, China)
Frequent Itemset Mining (FIM) is a fundamental procedure in various data mining techniques such as association rule mining. Among many existing algorithms, FP-Growth is considered as a milestone achievement that discovers frequent itemsets without generating candidates. However, due to the high complexity of its mining process and the high cost of its memory usage, FP-Growth still suffers from a performance bottleneck when dealing with large datasets. In this paper, we design a new Array Prefix-Tree structure, and based on that, propose an Array Prefix-Tree Growth (APT-Growth) algorithm, which explicitly obviates the need of recursively constructing conditional FP-Tree as required by FP-Growth. To support big data analytics, we further design and implement a parallel version of APTGrowth, referred to as PAPT-Growth, as a Spark workflow. We conduct FIM workflow experiments on both real-life and synthetic datasets for performance evaluation, and extensive results show that PAPT-Growth outperforms other representative parallel FIM algorithms in terms of execution time, which sheds light on its potential applications to big data mining.

WORKS19 Panel

9:00 am - 5:30 pm

The 2019 International Workshop on Software Engineering for HPC-Enabled Research (SE-HER 2019)

Session Description: Developers who build research software for High Performance Computing (HPC) or High Performance Data Analysis/Analytics (HPDA) face software engineering (SE) challenges at scales not often addressed by traditional SE approaches. For example, HPC and
HPDA 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 efficiently use compute resources. SE researchers have developed tools and practices to support development tasks, including: requirements, design, validation and verification, testing, continuous integration, and maintenance. Because of the scale of HPC and HPDA, there is a need to adapt these SE tools/methods that are standard elsewhere. This workshop brings together members of the SE and HPC/HPDA communities to present findings relative to these problems and to generate an agenda to advance software engineering tools and practices for HPC/HPDA software.

http://SE4Science.org/workshops/seher19/

Developers who build research software for High Performance Computing (HPC) or High Performance Data Analysis/Analytics (HPDA) face software engineering (SE) challenges at scales not often addressed by traditional SE approaches. For example, HPC and HPDA 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 efficiently use compute resources. SE researchers have developed tools and practices to support development tasks, including: requirements, design, validation and verification, testing, continuous integration, and maintenance. Because of the scale of HPC and HPDA, there is a need to adapt these SE tools/methods that are standard elsewhere. This workshop brings together members of the SE and HPC/HPDA communities to present findings relative to these problems and to generate an agenda to advance software engineering tools and practices for HPC/HPDA software.

Software Engineering Practices and Related Challenges in the LSST Dark Energy Science Collaboration
Katrin Heitmann (Argonne National Laboratory)

A Study of Hydrodynamics Based Community Codes in the Astrophysics

Advances in mathematical models and numerical algorithms combined with increasing reliance on simulations for understanding multi-physics and multi-scale phenomena has made the task of
software development for simulations a large and complex enterprise. Development and adoption of community codes is one way to address this challenge. The astrophysics community has been ahead of many other science communities in making research codes publicly available, and therefore in the development and adoption of community codes. ZEUS-2D was one of the earliest codes to become public, and it has been followed by several others. A study of the landscape of publicly available software and their penetration in the research conducted by the community can provide important insight for other communities that are struggling with similar issues. In this paper we analyze a subset of software available in ASCL to understand a part of the landscape. The category of software that we focused on involves modeling of hydrodynamics. We use the citation history of the software for our analysis.

Large Group Discussion

Large group discussion on presentations.

SE-HER 2019 Morning Break

Ice Breaking Exercise
Neil Chuo Hong (Software Sustainability Institute)

Lightweight Software Process Improvement Using Productivity and Sustainability Improvement Planning (PSIP)

Productivity and Sustainability Improvement Planning (PSIP) is a lightweight, iterative workflow that allows software development teams to identify development bottlenecks and track progress to overcome them. In this paper, we present an overview of PSIP and how it compares to other software process improvement (SPI) methodologies, and provide two case studies that describe how the use of PSIP lead to successful improvements in team effectiveness and efficiency.

Lightning Talk
Role Oriented Code Generation in an Engine for Solving Hyperbolic PDE Systems

The development of a high performance PDE solver requires the combined expertise of interdisciplinary teams w.r.t. application domain, numerical scheme and low-level optimization. In this paper, we present how the ExaHyPE engine facilitates the collaboration of such teams by isolating three roles -- application, algorithms, and optimization expert -- thus allowing team members to focus on their own area of expertise, while integrating their contributions into an HPC production code.

To do so, ExaHyPE takes inspiration from web application development practices and relies on two custom code generation modules, the Toolkit and the Kernel Generator, that follow a Model-View-Controller architectural pattern on top of the Jinja2 template engine library. By using Jinja2's templates to abstract the critical components of the engine and generated glue code, we isolate the application development from the engine. Furthermore the template language allows us to define and use custom macros that isolate low-level optimizations from the numerical scheme described in the templates.

We present three use cases, each focusing on one of our user roles, showcasing how the design of the code generation modules allows to easily expand the solver schemes to support novel demands from applications, to add optimized algorithmic schemes (with reduced memory footprint, e.g.), or provide improved low-level SIMD vectorization support.
FQL: An Extensible Feature Query Language and Toolkit on Searching Software Characteristics for HPC Applications

The amount of large-scale scientific computing software is dramatically increasing. In this work, we designed a new query language, named Feature Query Language (FQL), to collect and extract HPC-related software features or metadata from a quick static code analysis. We also designed and implemented an FQL-based toolkit to automatically detect and present software features using an extensible query repository. A number of large-scale, high performance computing (HPC) scientific applications have been studied in the paper with the FQL toolkit to demonstrate the HPC-related feature extraction and information/metadata collection. Different from the existing static software analysis and refactoring tools which focus on software debug, development and code transformation, the FQL toolkit is simpler, significantly lightweight and strives to collect various and diverse software metadata with ease and rapidly.

Small Group Discussion

SE-HER 2019 Afternoon Break

Finish Small Group Discussions

Large Group Report Out

Final Discussion
9:00 am - 5:30 pm

Fifth Computational Approaches for Cancer Workshop (CAFCW19)

Session Description: 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, with exceptional growth in deep learning, are expected to continue to reach unprecedented scales. Such opportunities to impact cancer computationally are underscored in the 2016 US Twenty-first Century Cures Act and international efforts such as the ITCC-P4 focusing on development treatments for pediatric cancer. The workshop focuses on bringing together interested individuals ranging from clinicians, mathematicians, data scientists, computational scientists, 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. 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. A special emphasis for the 2019 workshop is the role of computing in drug discovery for cancer, with a special emphasis on the role of AI, machine learning, and deep learning approaches applied to drug discovery and precision medicine. [https://ncihub.org/groups/cafcw19](https://ncihub.org/groups/cafcw19)
Welcome - Fifth Computational Approaches for Cancer

Keynote: Fifth Computational Approaches for Cancer
Warren Kibbe (Duke University)
Keynote presentation for Fifth Computational Approaches for Cancer

CAFCW19 Morning Break

Integrating High-Performance Simulations and Learning toward Improved Cancer Therapy
Shantenu Jha (Rutgers University, Molecular Sciences Software Institute (MolSSI))

We develop a novel deep learning workflow to effectively combine expensive but accurate molecular dynamics (MD) based BFE calculations with fast machine learning models to predict the affinity of compounds. In this approach, candidates are sampled from a large billion-compound synthetically accessible space (such as Enamine REAL) or de novo molecule generator. Sampling is designed to rapidly and parsimoniously train an inexpensive surrogate model from BFE calculations, creating a tool capable of accurately scoring vast libraries of molecules. This will be achieved by using an active learning approach to guide the choice of BFE calculations performed. Development of the workflow will be driven by applications and datasets provided by partners in both the pharmaceutical and healthcare sectors. Common to both is the need to react to the evolution of resistance to selective pharmaceutical compounds which undermines treatment of infections (both bacterial and viral) and human cancers. The dramatic reduction in the cost of genome sequencing has enabled genes to be identified in which variation potentially confers resistance, thereby creating the possibility of screening individual samples of infections or tumors to ascertain which drugs will be most effective. Learning from progress made in the INSPIRE project (which focused on tyrosine kinase resistance), we develop a platform combining molecular modeling and machine learning approaches to support representative workflow which brings together physical modeling and machine learning to identify novel potent and selective compounds.

Machine Learning Algorithms in Histology and Radiology for Cancer Drug Discovery and
Development
Partha Paul (LPixel Inc)

Background: Lung cancer is one of the most common cancers in the world. It is a leading cause of cancer death in men and women in the United States. Computational approaches such as deep learning could help accurate and efficient analysis of biomarkers, both histopathology and radiology, to assist cancer drug discovery and development. PD-L1 expression in the tumor micro-environment is one of the efficacy predictors of cancer treatment using immune checkpoint inhibitors. Tumor proportion score (the ratio of PD-L1 expressed area within the total tumor area) is often used as a measure to distinguish low and high PD-L1 expression. However, due to difficulty in quantitative measurement of microscopic images, the proportion score is prone to be evaluated subjectively. A more objective way by using computational techniques which can quantify PD-L1 expression only in the tumor cells is highly desirable. In addition, nodules detected on a chest radiograph (CR) is a good clinical marker for detection and progression of the disease. Interpretation of CR has problems of overlooking nodules and the reported error rates for missed lung cancer on CR are 20%-50%. Deep learning techniques can overcome the accuracy challenges of conventional interpretation and can improve the quality of the clinical biomarker.

Objective: One of our objectives is to create an image analysis system to automatically recognize the “total tumor area” and “tumor area with PD-L1 expression” in order to derive a quantitatively measured “tumor proportion score”. The second objective is to develop an algorithm with deep learning techniques for detecting lung cancer on CR and assess its performance.

Materials and Methods: For the PD-L1 expression study, whole slide images (WSI) of lung cancer pathology were used. Small patches were randomly selected from WSI to train our segmentation model. Deep convolutional neural networks were built and the segmentation model was trained using the extracted and annotated patches. Training has been done with 34 patches of 1000 x 1000 pixel images. The algorithm for lung nodule detection was developed by segmentation method with encoder-decoder architecture. The encoder used Inception-ResNet-v2 model as the backbone algorithm to extract features. The decoder was designed to compute up convolution to detect lung nodule or mass with segmentation. Training and test data were obtained from University of Osaka Medical School.

Results and Discussions: We have demonstrated that application of deep learning to microscopic images has potential in the following tasks: Distinguishing cancerous regions from other tissue regions, Segmentation of regions with PD-L1 expression, Quantitative evaluation of tumor proportion score. A accurate and fast quantitative evaluation of PD-L1 expression on tumor cells (not confounded by PL-L1 expression in non-cancer area) would provide a more accurate indication of drug efficacy in high content screening. The algorithm for detection of nodules in CR achieved
A Scalable, Validated Platform for Generative Lead Optimization of De Novo Molecules: Case Study in Discovery of Potent, Selective Aurora Kinase Inhibitors with Favorable Secondary Pharmacology
Andrew Weber (Accelerating Therapeutics for Opportunities in Medicine (ATOM) Consortium, GlaxoSmithKline)

De Novo design of therapeutic agents is currently a slow, expensive process generally relying on a large high throughput screen and several follow up cycles of iterative design to enhance the potency, eliminate safety liabilities, and enable favorable pharmacokinetic behavior. Computer aided drug discovery (CADD) has significantly aided this though structure and ligand-based design techniques, however current application often focuses on optimizing a single molecular property at a time, leading to long design cycle times. Recent advances in generative models for chemistry, including variational autoencoders (VAEs) and generative adversarial networks (GANs) have enabled a continuous representation of chemical space, allowing for application of numerical optimization techniques to this multi-factorial problem. As a proof of concept exercise, optimization of generative network across a diverse range of pharmacological properties was performed utilizing a high-performance scalable framework. Specifically, the algorithm was challenged to design a potent inhibitor of Aurora Kinase B, an enzyme involved with cell division and drug target in at least 12 ongoing and completed oncology trials. For a more realistic lead optimization challenge, the algorithm was also tasked with selectivity against the related Aurora Kinase A while simultaneously maintaining favorable secondary pharmacology properties (i.e. safety and pharmacokinetics). Running on a multi-node cluster, the framework was able to generate, evaluate and rank 3 million compounds, including >200 de novo molecules with predicted properties meeting the acceptability criteria. Experimental validation was performed for approximately 100 of the de novo molecules through chemical synthesis and in vitro testing.

Deep Learning Enabled Unsupervised State Identification in KRAS Dimers and Interacting Lipids
Gautham Dharuman (Lawrence Livermore National Laboratory)

Mutations of the KRAS gene are a prevalent driver in nearly 30% of all human cancers. It is hypothesized that the KRAS dimerization may facilitate RAF clustering, which is known to be required for RAF activation [1]. A recent study [2] has demonstrated the requirement of KRAS dimerization to sustain the oncogenic function of mutant KRAS and reveals that the disruption of dimerization could be a potential therapeutic strategy that may be effective in KRAS mutant
cancers. Consequently, it is crucial to identify the stable states of KRAS dimers. Identifying conformational states in KRAS monomers can be achieved using hand engineered features, such as tilt and rotation angles [3]. However, a similar approach is challenging in KRAS dimers because the complexity of the structure makes it a nontrivial task to identify the right set of angles for adequate feature representation.

We present our ongoing efforts on the state identification in KRAS dimers and associated lipids using ML-based models. In particular, with the advent of deep learning based approaches, it has become possible to use the raw coordinates of molecules to identify their states and the corresponding stability [4,5]. Our initial experiments support the hypothesis that state identification can be reduced to an unsupervised clustering problem in a latent space, which represents the underlying manifold governing the data distribution. We employ variational autoencoders using deep neural networks [6] to encode the molecular coordinates of the KRAS dimer and the interacting lipid density fields into a meaningful latent space.

Our training data comes from a massive simulation campaign using our MuMMI framework [7] that was run on up to 4000 nodes of the Sierra supercomputer at LLNL. Using data from 120,000 coarse-grained MD simulations each over 1 microsecond long, we train the neural network using NVIDIA Tesla V100 GPUs to reduce the spatial coordinates into a low-dimensional latent space. We perform spectral clustering in the latent space to determine, in an unsupervised manner, the number of distinct clusters—each corresponding to a distinct state of the RAS dimer. Identified distinct clusters are then analyzed further for their geometrical information and lifetime to understand their stability. Moreover, this information is in the context of different lipid compositions, so the lipid-dependence of the KRAS dimerization can also be tested.

We believe that meaningful state identification in KRAS dimers and associated lipids using ML-based unsupervised methods can provide key insights for experiments facilitating therapeutic strategies. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-ABS-790038-DRAFT


Fusion of Structure Based Deep Learning to Accelerate Molecular Docking Predictions
Derek Jones (Lawrence Livermore National Laboratory)
Modeling interactions with biological targets is a necessary step to begin reasoning about the therapeutic potential of a novel molecule in the drug discovery process. Molecular docking aids drug discovery researchers by searching over potential binding 'poses' of a drug molecule, ranking these poses according a scoring function, and using this information to identify the binding mode of a protein-small molecule complex, allowing for the quantification of the strength of the molecules binding interaction with their intended protein targets. While these simulations are able to scale on large computing clusters, the results that are produced are not generally accurate. Often the ‘best’ pose as identified by the docking algorithm is not actually the ‘correct’ pose. The score of the ‘best’ pose is also not highly correlated with an experimental binding affinity when it is known. The issues that arise with scoring functions in many molecular docking pipelines make this step of in silico prediction of binding strength even more difficult. A productive step forward would suggest a scoring function that is able to address these issues. Deep learning methods such as Convolutional Neural Networks (CNNs) and Graph Convolutional Neural Networks (GCNNs) have shown promise in providing researchers frameworks from which it is possible to model binding affinity with reasonable predictivity on experimental datasets. As part of an ongoing collaboration effort between Lawrence Livermore National Laboratory and the American Heart Association towards developing a human-protein drug atlas, we are investigating the effectiveness of convolutional neural networks and spatial graph methods for improving the accuracy and throughput of molecular docking pipelines to help address scaling challenges in the development of this resource.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Panel: AI and Cancer Drug Discovery

Panel session - AI and Cancer Drug Discovery will provide opportunity for interactive discussion on issues and opportunities related to drug discovery for cancer including use of AI techniques, data, active learning, model development and sharing, and other key areas leading the development of new treatments for cancer.

CAFCW19 Lunch Break

CAFCW19: Afternoon Welcome
Machine Learning Directed Multiscale Simulations To Explore RAS Biology

Peer-Timo Bremer (Lawrence Livermore National Laboratory)

Computational modeling has the potential to provide significant new insights into the detailed molecular mechanisms underlying many diseases such as cancer. However, existing approaches have difficulty bridging the experimentally observable effects at the macro scale with the molecular level of detail needed to provide new insight. Here we present a first step towards a new multi-resolution framework able to address these challenges. Using emerging deep learning approaches, we couple a continuum macro-scale model at experimentally relevant time- and length-scales with a large ensemble of molecular dynamic simulations of the micro scale. Given sufficient resources, the system will converge to macro-scale results at micro-scale resolution and provide fundamentally new insights. We discuss the general ideas and initial implementation and report on our experience with the first large-scale simulation campaign that utilized all of Sierra, the second fastest supercomputer in the world.

Semi-Supervised Method for Countering Batch Effect

Stewart He (Lawrence Livermore National Laboratory)

Predictive modeling of patient tumor growth response to drug treatment is severely limited by a lack of experimental data for training. Combining experimental data from several studies is an attractive approach, but presents two problems: batch effects and limited data for individual drugs and cell lines. Batch effects are caused by systematic procedural differences among studies, which causes systematic experimental outcome differences. Directly using these experimental results as features for machine learning commonly causes problems when training on one study and testing on another. This severely limits a model’s ability to perform well on new experiments. Even after combining studies, predicting outcomes on new patient tumors remains an open challenge.

We propose a semi-supervised, autoencoder-based, machine learning procedure, which learns a smaller set of gene expression features that are robust to batch effects using background information on a cell line or tissue’s tumor type. We implemented this reduced feature representation and show that the new feature space clusters strongly according to tumor type. This experiment is carried out across multiple studies: CCLE, CTRP, gCSI, GDSC, NCI60, and patient derived tumors. We hypothesize that using a batch effect resistant feature set across studies will improve prediction performance.

Genome Data Commons (GDC) gene expression profiles for publicly available human tissues and cell lines from NCI60 and CCLE were processed using the semi-supervised learning procedure. Our
autoencoder repurposes the ‘center loss’ (CL) cost function of Wen et. al. to learn a more generalized set of features using the cell line or tissue’s tumor type. Classification is performed by branching network the ‘pinch’ layer of the autoencoder. The ‘pinch’ layer now gets fed into a classification layer as well as the decoder portion of the autoencoder.

The new cost function balances the reconstruction performance, with the classification and ‘center loss’ performance. Reconstruction performance ensures that the ‘pinch’ layer retains information about original gene expression while classification performance shapes the space so tumors of the same type of close together regardless of the source study. Using the ‘pinch’ layer as new features reduces the number of features from 17,000 genes to approximately 1000 features or as few as 20 features. The performance of this method is compared with traditional batch correction methods (e.g. ComBat). Before applying these methods, individual samples clustered more strongly along study, a property that is not useful in many machine learning applications. We compare the new features from our ‘center loss’ autoencoder and ComBat using Silhouette score, the Calinski – Harabasz index, and the Davies – Bouldin index. All metrics show that using the prosed ‘center loss’ autoencoder features provide a latent space with better clusters than applying ComBat.

**Acceleration of Hyperparameter Optimization via Task Parallelism for Information Extraction from Cancer Pathology Reports**

Hong-Jun Yoon (Oak Ridge National Laboratory), John Gounley (Oak Ridge National Laboratory)

Recent advances in high-performance computing systems for artificial intelligence enable large-scale training of information extraction models from free-form natural language texts. The development of these models is essential to the cancer surveillance research and automation. In this study, we propose an approach to accelerate training of machine learning models by introducing task parallelism. For the information extraction from cancer pathology reports, we implement task parallelism by splitting the task of identifying multiple cancer topography and morphology into several sub-problems. This allows for the hyperparameters for each sub-problem to be optimized in parallel. Further, we introduce the model parameter inheritance to improve the convergence rate of the hyperparameter optimization runs themselves. We evaluate the feasibility of the proposed method on the Summit supercomputer and demonstrate that it improves time-to-solution by a factor of 10, when compared to the traditional model-based optimization algorithm, while maintaining the same level of clinical task performance scores.

**CAFCW19 Afternoon Break**
Digital Twins for Predictive Cancer Care: an HPC-Enabled Community Initiative
Paul Macklin (Indiana University)

Cancer is a complex multiscale dynamical systems problem with interactions between the tumor and host at the molecular, cellular, tissue, and organism levels. Moreover, treatment occurs within a larger dynamical system that couples clinical care teams, hospital systems, industry, and government policies. For individuals, effective patient-tailored therapeutic guidance and planning will require bridging spatiotemporal scales with modeling from atomistic to organismal levels. At the societal level, improving the patient population’s overall health and quality of life requires understanding how actions by patients, clinicians, researchers, and government agencies impact one another.

Predictive oncology aims to predict and steer each patient’s future disease dynamics, but it cannot be fully realized by any single technology, laboratory, or discipline. It will require a coordinated multidisciplinary, multi-institutional effort with access to sufficient data and computational resources. The past decade has seen tremendous advances in data-rich medical measurements; biological theories of cancer progression, therapeutic response and resistance; artificial intelligence (AI) and deep learning techniques; next-generation high-performance computing; multiscale modeling and simulation; and secure high-speed data infrastructures. The time is ripe to leverage these advances to create digital twins for predictive cancer care: patient-tailored models that can evaluate thousands of potential therapeutic plans, help clinicians understand and choose the plan that best meets the patient’s objectives, benchmark clinical performance, and continuously integrate new data and knowledge to refine treatment plans. This talk will discuss a community-driven initiative to design HPC-enabled digital twins for predictive cancer care, available technologies, barriers that must be overcome, and ongoing opportunities to help translate digital twins to clinical practice.

In 2016, the National Cancer Institute (NCI) and the Department of Energy (DOE) partnered to create Joint Design of Advanced Computing Solutions for Cancer (JDACS4C): a collaboration of multidisciplinary experts in the biological, computational, data, and physical sciences at the NCI, the Frederick National Laboratory for Cancer Research, and four DOE national laboratories. JDACS4C aims to develop, demonstrate, and disseminate advanced computational capabilities that help answer driving scientific questions across molecular, cellular and population scales. These efforts frame forward-looking approaches for integrating and analyzing large, heterogeneous data collections with advanced computational modeling that will accelerate predictive oncology.

In March 2019, the agencies built on JDACS4C’s team science approach to launch the Envisioning Computational Innovations for Cancer Challenges (ECICC) community. A series of ECICC events identified the digital twin initiative as a key opportunity to push the limits of computational
approaches to cancer care, compel the development of innovative computational technologies, and bring about new paradigms for care.

Digital twins have great potential to help patients reach their treatment goals. Aggregated data from populations of digital twins could help improve national strategies for cancer screening, prevention, research investment, and structuring healthcare systems. The ECICC community has initiated work to create this national resource, but it can only succeed with expanded contributions by the experimental, computational, and clinical communities. After outlining next steps, we will present opportunities to contribute to and drive the effort.

**Deep Kernel Learning for Information Extraction from Cancer Pathology Reports**

Devanshu Agrawal (University of Tennessee)

Cancer pathology reports comprise a rich source of data for surveilling cancer incidents and tracking cancer trends across the United States. Cancer registries manually extract key pieces of information from these reports including tumor site, histology, laterality, behavior, grade, and metastatic status. Automating this task is critical for an efficient and scalable processing pipeline of these reports. Deep neural networks have recently been shown to perform well on this information extraction task by casting it as a document classification problem. However, neural networks are prone to overfitting in low-sample regimes and are unable to quantify their own uncertainty. Deep kernel learning (DKL) has recently emerged as a simple and scalable paradigm to hybridize deep neural networks and Bayesian models, which may help to remedy some of these shortcomings of neural networks. A DKL model is obtained by feeding a neural network feature extractor into a Gaussian process (GP) classifier and training the resulting model with gradient descent in a variational inference framework. In this project, we build a DKL model with a shallow-wide convolutional neural network (CNN) feature extractor and use it to extract primary tumor site information from a dataset of de-identified cancer pathology reports. As far as we are aware, this marks the first application of DKL to document classification. Our DKL model outperforms the state-of-the-art CNN on this dataset. We also show that pretraining a CNN with the weights of a DKL model boosts performance, suggesting that DKL is beneficial not just because of GP inference at test time but also because DKL is able to extract better feature representations from the pathology reports through Bayesian training. We conclude that DKL has the potential to boost the performance of neural networks for information extraction on pathology reports while requiring little modification of the original network architecture, and that DKL can offer a path forward to develop scalable deep Bayesian models for such tasks.

**Massively Parallel Large-Scale Multi-Model Simulation of Tumor Development including Treatments**

Marco Berghoff (Karlsruhe Institute of Technology)
The temporal and spatial resolution in the microscopy of tissues has increased significantly within the last years, yielding new insights into the dynamics of tissue development and the role of the single-cell within it. A thorough theoretical description of the connection of single-cell processes to macroscopic tissue reorganizations is still lacking. Especially in tumor development, single cells play a crucial role in advance of tumor properties. We developed a simulation framework that can model tissue development up to the centimeter scale with micrometer resolution of single cells. Through a full parallelization, it enables the efficient use of high-performance computing systems, therefore enabling detailed simulations on a large scale. We developed a generalized tumor model that respects adhesion driven cell migration, cell-to-cell signaling, and mutation-driven tumor heterogeneity. We scan the response of the tumor development depending on division inhibiting substances such as cytostatic agents. Furthermore, we are investigating the interaction with radiation therapy to find a suitable therapy plan.

Panel Session: Collaborations, Community, and Computational Approaches for Cancer

CAFCW19 Wrap-Up

2:00 pm - 5:30 pm

HPC for Urgent Decision Making (UrgentHPC)

Session Description: Responding to disasters such as wildfires, hurricanes, flooding, earthquakes, tsunamis, winter weather conditions, and accidents; technological advances are creating exciting new opportunities that have the potential to move HPC well beyond traditional computational workloads. While HPC has a long history of simulating disasters, what’s missing to support emergency, urgent, decision making is fast, real-time acquisition of data and the ability to guarantee time constraints. Our ability to capture data continues to grow very significantly, and combining high velocity data and live analytics with HPC models can aid in urgently responding to real-world problems, ultimately saving lives and reducing economic loss. Not just responding to disasters, but also using HPC to make urgent decisions addressing more general issues, such as human health emergencies and global diseases, requires expertise in a wide range of areas. From dealing with real-time data, to experience in generating results within specific time frames (real-time
constraints), and generating visualizations enabling front-line decision makers to make correct choices first time, every time. The challenges here are significant, but if HPC can be proven as a tool in responding to these real-world issues, the impact for our community is huge. This workshop will bring together stakeholders, researchers and practitioners from across the HPC community to identify and tackle issues involved in using HPC for urgent decision making. Success stories, case-studies and challenges will be shared, with the goal of further building up a community around leveraging HPC as an important tool in urgently responding to disasters and societal challenges.

[https://www.urgenthpc.com](https://www.urgenthpc.com)

Responding to disasters such as wildfires, hurricanes, flooding, earthquakes, tsunamis, winter weather conditions, and accidents; technological advances are creating exciting new opportunities that have the potential to move HPC well beyond traditional computational workloads. While HPC has a long history of simulating disasters, what’s missing to support emergency, urgent, decision making is fast, real-time acquisition of data and the ability to guarantee time constraints.

Our ability to capture data continues to grow very significantly, and combining high velocity data and live analytics with HPC models can aid in urgently responding to real-world problems, ultimately saving lives and reducing economic loss. Not just responding to disasters, but also using HPC to make urgent decisions addressing more general issues, such as human health emergencies and global diseases, requires expertise in a wide range of areas. From dealing with real-time data, to experience in generating results within specific time frames (real-time constraints), and generating visualizations enabling front-line decision makers to make correct choices first time, every time. The challenges here are significant, but if HPC can be proven as a tool in responding to these real-world issues, the impact for our community is huge.

This workshop will bring together stakeholders, researchers and practitioners from across the HPC community to identify and tackle issues involved in using HPC for urgent decision making. Success stories, case-studies and challenges will be shared, with the goal of further building up a community around leveraging HPC as an important tool in urgently responding to disasters and societal challenges.

**Keynote: Wildfire Risk Modeling, simulating at the gazillion scale**

Joaquin Ramirez (Technosylva Inc)

Evaluating wildfire risk is needed at high spatial resolution and an hourly scale. This a technological
challenge as on top of high-resolution weather models, fire behavior needs other complex inputs. Being a hazard with global coverage, the problem to provide quantitative evaluations can only be addressed by using HPC infrastructures that integration of a vast amount of data sources. Technosylva applies HPC methods to evaluate hourly wildland fire risk for the three biggest utilities of California and has been recently selected among 131 companies in the request for Innovation Ideas California’s Governor to provide these capabilities to CAL FIRE.

**Quantifying Uncertainty in Source Term Estimation with Tensorflow Probability**

Fast and accurate location and quantification of a dangerous chemical, biological or radiological release plays a significant role in evaluating emergency situations and their consequences. Thanks to the advent of Deep Learning frameworks (e.g. Tensorflow) and new specialized hardware (e.g. Tensor Cores), the excellent fitting ability of Artificial Neural Networks (ANN) has been used by several researchers to model atmospheric dispersion. Despite the high accuracy and fast prediction, regular ANNs do not provide any information about the uncertainty of the prediction. Such uncertainty can be the result of a combination of measurement noise and model architecture. In an urgent decision making situation, the ability to provide fast prediction along with a quantification of the uncertainty is of paramount importance. In this work, a Probabilistic Deep Learning model for source term estimation is presented, using the Tensorflow Probability framework.

**Statistical Parameter Selection for Clustering Persistence Diagrams**

In urgent decision making applications, ensemble simulations are an important way to determine different outcome scenarios based on currently available data. In this paper, we will analyze the output of ensemble simulations by considering so-called persistence diagrams, which are reduced representations of the original data, motivated by the extraction of topological features. Based on a recently published progressive algorithm for the clustering of persistence diagrams, we determine the optimal number of clusters, and therefore the number of significantly different outcome scenarios, by the minimization of established statistical score functions. Furthermore, we present a proof-of-concept prototype implementation of the statistical selection of the number of clusters and provide the results of an experimental study, where this implementation has been applied to real-world ensemble data sets.

**UrgentHPC Afternoon Break**
On-Demand Urgent High Performance Computing Utilizing the Google Cloud Platform

In this paper we describe how high performance computing in the Google Cloud Platform can be utilized in an urgent and emergency situation to process large amounts of traffic data efficiently and on demand. Our approach provides a solution to an urgent need for disaster management using massive data processing and high performance computing. The traffic data used in this demonstration is collected from the public camera systems on Interstate highways in the Southeast United States. Our solution launches a parallel processing system that is the size of a Top 5 supercomputer using the Google Cloud Platform. Results show that the parallel processing system can be launched in a few hours, that it is effective at fast processing of high volume data, and can be de-provisioned in a few hours. We processed 211TB of video utilizing 6,227,593 core hours over the span of about eight hours with an average cost of around $0.008 per vCPU hour, which is less than the cost of many on-premise HPC systems.

The Technologies Required for Fusing HPC and Real-Time Data to Support Urgent Computing

The use of High Performance Computing (HPC) to compliment urgent decision making in the event of disasters is an important future potential use of supercomputers. However, the usage modes involved are rather different from how HPC has been used traditionally. As such, there are many obstacles that need to be overcome, not least the unbounded wait times in the batch system queues, to make the use of HPC in disaster response practical. In this paper, we present how the VESTEC project plans to overcome these issues and develop a working prototype of an urgent computing control system. We describe the requirements for such a system and analyse the different technologies available that can be leveraged to successfully build such a system. We finally explore the design of the VESTEC system and discuss ongoing challenges that need to be addressed to realise a production level system.

An Interactive Data-Driven HPC System for Forecasting Weather, Wildland Fire, and Smoke

We present an interactive HPC framework for coupled fire and weather simulations. The system is suitable for urgent simulations and forecast of wildfire propagation and smoke. It does not require expert knowledge to set up and run the forecasts.

The core of the system is a coupled weather, wildland fire, fuel moisture, and smoke model, running in an interactive workflow and data management system. The system automates job setup, data
acquisition, preprocessing, and simulation on an HPC cluster. It provides animated visualization of the results on a dedicated mapping portal in the cloud as well as delivery as GIS files and Google Earth KML files. The system also serves as an extensible framework for further research, including data assimilation and applications of machine learning to initialize the simulations from satellite data.

**Urgent Tsunami Computing**

Tsunamis pose a hazard that may strike a coastal population within a short amount of time. To effectively forecast and warn for tsunamis, extremely fast simulations are needed. However, until recently such urgent tsunami simulations have been infeasible in the context of early warning and even for high-resolution rapid post-event assessment. The implementation of efficient tsunami numerical codes using Graphical Processing Units (GPUs) has now allowed much faster simulations, which have opened a new avenue for carrying out simulations Faster Than Real Time (FTRT). This paper discusses the need for urgent computing in computational tsunami science, and presents workflows for two applications, namely FTRT itself and Probabilistic Tsunami Forecasting (PTF). PTF relies on a very high number of FTRT simulations addressing forecasting uncertainty, whose full quantification will be made more and more at reach with the advent of exascale computing resources.

**Panel: What Role Can HPC Play in Urgent Decision Making?**

Andreas Gerndt (German Aerospace Center (DLR)), Mike Ringenburg (Cray Inc), Jason Knievel (National Center for Atmospheric Research (NCAR)), Nick Brown (Edinburgh Parallel Computing Centre)

The use of HPC for urgent decision making is still in its infancy, but this workshop believes that it is a really important future use-case for our community. Based on the discussions and papers presented thus far, we will be exploring the challenges and opportunities that panelists believe face our community in realising the fusion of HPC with real-time data for addressing urgent workloads.

We will explore whether the panelists agrees or disagree with common sentiments discussed thus far, and their own take on some of the themes that the workshop has uncovered.

**Monday, November 18**

9:00 am - 12:30 pm
Third Workshop on Interactive High Performance Computing

Session Description: Interactive exploration and analysis of large data sets, intelligent simulation ("cog-sim") workflows that combine interactive analysis and AI techniques with modeling and simulation, interactive preparation and debugging of large-scale scientific simulations, in-situ visualization, and application steering are all compelling scenarios for exploratory science, design optimizations, and signal processing. However, a range of technical, organizational and sociological challenges must be overcome to make these interactive workflows mainstream in HPC centers: What simulation scenarios or problem domains can benefit most from interactivity? How can we simplify the toolchain? What center policies are needed to support highly interactive workflows? The goal of this workshop 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.

Interactive exploration and analysis of large data sets, intelligent simulation ("cog-sim") workflows that combine interactive analysis and AI techniques with modeling and simulation, interactive preparation and debugging of large-scale scientific simulations, in-situ visualization, and application steering are all compelling scenarios for exploratory science, design optimizations, and signal processing. However, a range of technical, organizational and sociological challenges must be overcome to make these interactive workflows mainstream in HPC centers: What simulation scenarios or problem domains can benefit most from interactivity? How can we simplify the toolchain? What center policies are needed to support highly interactive workflows? The goal of this workshop 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.

Accelerating Experimental Science Using Jupyter and NERSC HPC

Interactive Supercomputing for Experimental Data-Driven Workflows

Interactive High Performance Computing Morning Break
Portals for Interactive Steering of HPC Workflows

Pangeo Ecosystem, Interactive Computing Tools for Geosciences: Benchmark on HPC

Interactivity Use Case Panel Discussion

Interactive High Performance Computing Closing Remarks

9:00 am - 12:30 pm

6th International Workshop on HPC User Support Tools (HUST-19)

Session Description: 6th International Workshop on HPC User Support Tools (HUST19) – Supercomputing centers exist to drive scientific discovery by supporting researchers in computational science fields. To make users more productive in complex HPC environment, HPC centers employ specialized support teams. These teams have many functions covering basic system administration, managing 100's of Petabytes with complex hierarchal data-storage systems, supporting high performance networks, consulting on math libraries, code optimization, and 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 100's or even 1000's of HPC users can be extremely demanding. The HUST workshop, has been the ideal forum where new and innovative tools such as XALT, SPACK, and Easybuild have been widely announced to the broader HPC community. This has created communities and special interest
groups about these tools, many of which now hold their own BoFs, workshops, and tutorials at SC, ISC and other HPC conferences. HUST 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. [https://hust-workshop.github.io](https://hust-workshop.github.io)

6th International Workshop on HPC User Support Tools (HUST19) – Supercomputing centers exist to drive scientific discovery by supporting researchers in computational science fields. To make users more productive in complex HPC environment, HPC centers employ specialized support teams. These teams have many functions covering basic system administration, managing 100’s of Petabytes with complex hierarchal data-storage systems, supporting high performance networks, consulting on math libraries, code optimization, and 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 100’s or even 1000’s of HPC users can be extremely demanding.

The HUST workshop, has been the ideal forum where new and innovative tools such as XALT, SPACK, and Easybuild have been widely announced to the broader HPC community. This has created communities and special interest groups about these tools, many of which now hold their own BoFs, workshops, and tutorials at SC, ISC and other HPC conferences. HUST 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.

**Buildtest: A Software Testing Framework with Module Operations for HPC systems**

*Shahzeb Siddiqui (Pfizer Inc)*

HPC support teams are often tasked with installing scientific software for their user community and the complexity of managing a large software stack gets very challenging. Software installation brings forth many challenges that requires a team of domain expertise and countless hours troubleshooting to build an optimal software state that is tuned to the architecture. In the past decade, two software build tools (Easybuild, Spack) have emerged that are widely accepted in HPC community to accelerate building a complete software stack for HPC systems. The support team
are constantly involved in fulfilling software request for end-users which leads to an ever-growing software ecosystem. Once a software is installed, the support team hands it off to the user without any testing because scientific software requires domain expertise in order to test software. Some software packages are shipped with a test suite that can be run at post build while many software have no mechanism for testing. This poses a knowledge gap between HPC support team and end-users on the type of testing to do. Some HPC centers may have developed in-house test scripts that are suitable for testing their software, but these tests are not portable due to hardcoded paths and are often site dependent. In addition, there is no collaboration between HPC sites in building a test repository that will benefit the community. This paper presents buildtest, a framework to automate testing for a software stack along with several module operations that would be of interest to HPC support teams.

**Using Malleable Task Scheduling to Accelerate Package Manager Installations**

Samuel Knight (Sandia National Laboratories)

Package managers, containers, automated testing, and Continuous Integration (CI), are becoming an essential part of HPC development workflows. These automated tools often require software recompilation. However, large stacks such as those deployed on HPC clusters can have combinatorial dependencies, and may take a system several days to compile. Despite the use of simple parallelization (such as 'make -j'), build execution time often do not scale with system resources. For such cases, it is possible to improve overall installation time by compiling parts of software stack independently, each scheduled on a subset of available cores. We apply malleable-task scheduling algorithms to better exploit available parallelism in build system workflows and improve stack build time overall. Using a prototype implementation in the Spack package manager, malleable-task scheduling can improve build times by more than 2x.

**HUST-19 Morning Break**

**Enabling Continuous Testing of HPC Systems Using ReFrame**

Vasileios Karakasis (Swiss National Supercomputing Centre (CSCS))

Regression testing of HPC systems is of crucial importance when it comes to ensure the quality of service offered to the end users. At the same time, it poses a great challenge to the systems and application engineers to continuously maintain regression tests that cover as many aspects as possible of the user experience. In this paper, we briefly present ReFrame, a framework for writing regression tests for HPC systems and how this is used by CSCS, NERSC and OSC to continuously
test their systems. ReFrame is designed to abstract away the complexity of the interactions with the system and to separate the logic of a regression test from the low-level details, which pertain to the system configuration and setup. Regression tests in ReFrame are simple Python classes that specify the basic parameters of the test plus any additional logic. The framework will load the test and send it down a well-defined pipeline which will take care of its execution. ReFrame can be easily set up on any cluster and its straightforward invocation allows it to be easily integrated with common continuous integration/deployment (CI/CD) tools, in order to perform continuous testing of an HPC system. Finally, its ability to feed the collected performance data to well known log channels, such as Syslog, Graylog or, simply, parsable log files, make it also a powerful tool for continuously monitoring the health of the system from user’s perspective.

**Tools for Monitoring CPU Usage and Affinity in Multicore Supercomputers**

Kent Milfeld (Texas Advanced Computing Center (TACC)), Si Liu (Texas Advanced Computing Center (TACC))

Without other naturally efficient methods to further boost CPU performance, supercomputer designers and vendors have packed more and more physical processors(cores) into HPC processors. When working with multi-processor compute nodes, process and thread affinity/pinning can influence performance in a significant way. However, existing tools for monitoring an application’s CPU usage and affinity information are sometimes inconvenient and unsatisfying. Three innovative HPC user-support tools, core_usage, show_affinity, and amask, have been designed and implemented to change this status. These tools are helping HPC users and administrators easily monitor CPU usage and affinity. They provide a convenient mechanism for evaluating an application's process scheduling and process occupation on processors, especially on systems with a large core counts.

**HUST19 Panel: HPC User Tickets Discussion**

Robert McLay (Texas Advanced Computing Center (TACC)), Lissa Moore (Los Alamos National Laboratory)

9:00 am - 12:30 pm

**SuperCompCloud: Workshop on Interoperability of Supercomputing and Cloud Technologies**

Session Description: Exascale computing initiatives are expected to enable breakthroughs for
multiple scientific disciplines. Increasingly these systems may utilize cloud technologies, enabling complex and distributed workflows that can improve not only scientific productivity, but accessibility of resources to a wide range of communities. Such an integrated and seamlessly orchestrated system for supercomputing and cloud technologies is indispensable for experimental facilities that have been experiencing unprecedented data growth rates. While a subset of high performance computing (HPC) services have been available within a public cloud environments, petascale and beyond data and computing capabilities are largely provisioned within HPC data centres using traditional, bare-metal provisioning services to ensure performance, scaling and cost efficiencies. At the same time, on-demand and interactive provisioning of services that are commonplace in cloud environments, remain elusive for leading supercomputing ecosystems. This workshop aims at bringing together a group of experts and practitioners from academia, national laboratories, and industry to discuss technologies, use cases and best practices in order to set a vision and direction for leveraging high performance, extreme-scale computing and on-demand cloud ecosystems. Topics of interest include tools and technologies enabling scientists for adopting scientific applications to cloud interfaces, interoperability of HPC and cloud resource management and scheduling systems, cloud and HPC storage convergence to allow a high degree of flexibility for users and community platform developers, continuous integration/deployment approaches, reproducibility of scientific workflows in distributed environment, and best practices for enabling X-as-a-Service model at scale while maintaining a range of security constraints.

https://sites.google.com/view/supercompcloud

Exascale computing initiatives are expected to enable breakthroughs for multiple scientific disciplines. Increasingly these systems may utilize cloud technologies, enabling complex and distributed workflows that can improve not only scientific productivity, but accessibility of resources to a wide range of communities. Such an integrated and seamlessly orchestrated system for supercomputing and cloud technologies is indispensable for experimental facilities that have been experiencing unprecedented data growth rates. While a subset of high performance computing (HPC) services have been available within a public cloud environments, petascale and beyond data and computing capabilities are largely provisioned within HPC data centres using traditional, bare-metal provisioning services to ensure performance, scaling and cost efficiencies. At the same time, on-demand and interactive provisioning of services that are commonplace in cloud environments, remain elusive for leading supercomputing ecosystems.

This workshop aims at bringing together a group of experts and practitioners from academia, national laboratories, and industry to discuss technologies, use cases and best practices in order to
set a vision and direction for leveraging high performance, extreme-scale computing and on-demand cloud ecosystems. Topics of interest include tools and technologies enabling scientists for adopting scientific applications to cloud interfaces, interoperability of HPC and cloud resource management and scheduling systems, cloud and HPC storage convergence to allow a high degree of flexibility for users and community platform developers, continuous integration/deployment approaches, reproducibility of scientific workflows in distributed environment, and best practices for enabling X-as-a-Service model at scale while maintaining a range of security constraints.

**Scalability and Data Security: Deep Learning with Health Data on Future HPC Platforms**

Performing health data analytics at scale presents several challenges to classic HPC environments. Datasets contain personal health information (PHI) and are updated regularly, complicating data access on publicly accessible HPC systems. Moreover, the diverse group of tasks and models – ranging from neural networks for information extraction to knowledge bases for predictive modeling – have widely varying scales, hardware preferences, and software requirements. Both exascale systems and cloud-based environments have the opportunity to play important roles by addressing data security and performance portability. Cloud platforms provide out-of-the-box solutions for maintaining data security, while recent work has extended secure computing environments to systems like OLCF Summit. In this talk I will discuss how we are handling the need for scalable HPC resources with the data security requirements inherent in working with personal health information, in the context of the interagency partnership between the Department of Energy and the National Cancer Institute. As part of this partnership, we are developing state-of-the-art deep learning models to perform information extraction from cancer pathology reports for near real-time cancer incidence reporting. Our approach to addressing the patient privacy complexities involves integral roles for both traditional HPC resources and cloud-like platforms, playing to the relative strengths of both modalities.

**Cloud and Supercomputing Platforms at NCI Australia: the Why, the How, and the Future.**

Australia’s National Computational Infrastructure (NCI Australia) is a tier 1 provider of high performance computing and data services for Australian researchers spanning the higher education, government agency and industry sectors. NCI’s HPC facility lies in the range 20-150 on the top500 list, depending on stage within its lifecycle. The facility also manages on the order of 70PB of high-performance data storage capacity, comprising both projectized data spaces as well as high curated, functionalized FAIR data collections of national significance. Alongside its HPC capability provisioning, NCI has run a cloud architecture for internal and selected external purposes and over the past 5 years has progressively evaluated the most effective functional role that a cloud
infrastructure might play in the context of a national facility. Strategically, its current focus with cloud is to build the infrastructure for major research communities that have the demand; the national strategic priority and the resourcing capabilities to partner NCI in development of services and functionalities beyond the provision of “bare metal” hardware as an infrastructure. One of the significant technical challenges associated with this is the need for data analytics that access the petabyte-scale datasets residing on NCI’s high performance storage file systems, necessitating a level of data transfer bandwidth and compute resourcing that are not typical of “conventional” cloud. In this presentation I will give an overview of the above issues as NCI Australia encounters them presently, providing examples of current activities and sketching the future as we see it at this point.

SuperCompCloud Morning Break

HPC and Cloud Operations at CERN

CERN was established in 1954, with the mission of advancing science for peace and exploring fundamental physics questions — primarily through elementary particle research. The Large Hadron Collider (LHC) at CERN is the world’s most powerful particle accelerator colliding bunches of protons 40 million times every second. This extremely high rate of collisions makes it possible to identify rare phenomenon and to declare new discoveries such as the Higgs boson in 2012. The high-energy physics (HEP) community has long been a driver in processing enormous scientific datasets and in managing the largest scale high-throughput computing centres. Today, the Worldwide LHC Computing Grid is a collaboration of more than 170 computing centres in 42 countries, spread across five continents. Recently demonstrations at scale of both commercial cloud providers and HPC centers have been performed.

In 2026 we will launch the High Luminosity LHC (HL-LHC), which will represent a true exa-scale computing challenge. The processing capacity required by the experiments is expected to be 50 to 100 times greater than today, with storage needs expected to be on the order of exabytes. Neither the rate of technology improvement nor the computing budget will increase fast enough to satisfy these needs and new sources of computing and new ways of working will be needed to fully exploit the physics potential of this challenging accelerator. The growth of commercial clouds and HPC centres into the exa-scale represents a huge opportunity to increase the potential total resource pool, but even together this ecosystem may not be sufficient to satisfy the needs of our scientific workflows. The total computing required is pushing us to investigate alternative architectures and
alternative methods of processing and analysis. In this presentation we will discuss the R&D activities to utilize HPC and cloud providers. We will summarize our progress and challenges in operating on dedicated resources and on shared and purchased allocations on HPC and cloud. We will outline the biggest impedance issues to interoperating these facilities, which often have similar challenges for data handing and scale but very different challenges in flexibility and operations. We will close by addressing forward looking projects together with industry partners to utilize techniques like Machine Learning and optimized hardware to fundamentally change how many resources are needed to extract science from the datasets.

**Computing Without Borders: Combining Cloud and HPC to Advance Experimental Science.**

Computing has always been part of experimental science, and is increasingly playing a central role in enabling scientific discovery. Technological advances are providing more powerful supercomputers and super-efficient specialized chip architectures. It is also advancing instrument technology, resulting in higher resolution detectors that produce orders of magnitude more data and allow access to exciting new scientific domains.

The scale of the compute needs of such instruments is typically mixed, sometimes requiring cloud or small cluster computing and sometimes requiring dedicated access to a supercomputer. There is a vital dependence on how quickly the compute power can be available and how quickly the data can be transferred to the compute site, as many experiments require real-time data analysis. Using real-life examples from high energy physics, microscopy and genomics, I will discuss how experimental science is taking advantage of both cloud and near-exascale HPC resources. I will outline some of the challenges we see in operating such workflows across multiple sites, with a focus on system responsiveness and data management issues.

**OpenStack and the Software-Defined Supercomputer**

Two long-held aspirations, "agile supercomputing" and "performant cloud", are converging. Supercomputing performance can now be achieved in a cloud-native context. Cloud flexibility ensures different classes of workload may be targeted at the most effective resources.

For private cloud, OpenStack has become the de facto standard for IaaS. This presentation will introduce an open project to create private and hybrid cloud infrastructure specifically aimed at addressing the requirements of HPC. Through pooled development effort and a continuous process of validation, many complexities and overheads commonly associated with OpenStack operation are eliminated.
By combining extensive experience of both cloud and HPC infrastructure the functional gaps and performance bottlenecks of conventional cloud infrastructure have been identified and addressed, while maintaining the advantages in agility, flexibility and interoperability heralded by the cloud-native era.

This presentation will provide technical insights, an update on the capabilities of modern OpenStack, and an opportunity to become involved in a project aimed to deliver on the promise of cloud without sacrifice to performance.

**Perform Like a Supercomputer, Run Like a Cloud**

Without a doubt cloud has reshaped the enterprise datacenter and there is an ongoing debate as to what extent this disruption will also spill into Supercomputing. Many of the concepts and capabilities that we see today in cloud were spearheaded by the Grid computing initiatives that this community introduced in the early 90s. However, the technology and economics have dramatically involved since then. Cloud is an overloaded term used to capture a wide range of characteristics including business models, technologies, operational models and user access paradigms. Although some of these considerations will be assessed based on the individual requirements of an organization, there are cloud technologies that are becoming industry standards and an integral part of the Supercomputing industry.

This presentation will focus on how leveraging cloud technologies adapted to meet the needs of Supercomputing workloads is a dominant trend that is empowering system administrators to be more productive and end users to have an experience familiar and analogous to cloud environments.

In a cloud based system management solution all capabilities are exposed to the system administrators and DevOps personnel as an open, programmable fabric that can be either integrated to a broader management ecosystem or operated separately with open source, commercial, or home-grown tools. This is the foundation for providing the necessary automation and programmability to easily expose capabilities such as policy-based self-service, multi-tenancy, and elasticity.

The cloudification of the supercomputer system management infrastructure is also a requirement for the interoperation and integration of supercomputers with public cloud environments. Although today most organizations make a binary decision to deploy their solution to a public or private cloud, it is possible that in the future hybrid cloud might become more prevalent in our industry, especially
with the proliferation of processor architectures and silicon specialization.

Finally, our community has been in the forefront of researching technologies and solutions that meet the extreme requirements of our industry. The availability of an open and programable computer fabric removes any barriers to collect data and experiment.

**SuperCompCloud Workshop Q&A Panel and Closing**
Sadaf Alam (Swiss National Supercomputing Centre (CSCS))
Q&A Panel with workshop presenters (Maria Girone, Debbie Bard, Stig Telfer and Stathis Papaefstathiou) followed by closing and next steps by Sadaf Alam.

9:00 am - 5:30 pm

**MCHPC'19: Workshop on Memory Centric High Performance Computing**

**Session Description:** The growing disparity between CPU speed and memory speed, known as the memory wall problem, has been one of the most critical and long-standing challenges in the computing industry. The situation is further complicated by the recent expansion of the memory hierarchy, which is becoming deeper and more diversified with the adoption of new memory technologies and architectures including 3D-stacked memory, non-volatile random-access memory (NVRAM), memristor, hybrid software and hardware caches, etc. Computer architecture and hardware system, operating systems, storage and file systems, programming stack, performance model and tools are being enhanced, augmented, or even redesigned to address the performance, programmability and energy efficiency challenges of the increasingly complex and heterogeneous memory systems for HPC and data-intensive applications. The MCHPC workshop aims to bring together computer and computational science researchers, from industry, government labs and academia, concerned with the challenges of efficiently using existing and emerging memory systems. [https://passlab.github.io/mchpc/mchpc2019/](https://passlab.github.io/mchpc/mchpc2019/)

The growing disparity between CPU speed and memory speed, known as the memory wall problem, has been one of the most critical and long-standing challenges in the computing industry. The situation is further complicated by the recent expansion of the memory hierarchy, which is becoming deeper and more diversified with the adoption of new memory technologies and
architectures including 3D-stacked memory, non-volatile random-access memory (NVRAM), memristor, hybrid software and hardware caches, etc. Computer architecture and hardware system, operating systems, storage and file systems, programming stack, performance model and tools are being enhanced, augmented, or even redesigned to address the performance, programmability and energy efficiency challenges of the increasingly complex and heterogeneous memory systems for HPC and data-intensive applications.

The MCHPC workshop aims to bring together computer and computational science researchers, from industry, government labs and academia, concerned with the challenges of efficiently using existing and emerging memory systems.

**MCHPC’19 Keynote Talk:**
*J. Thomas Pawlowski (Independent)*

**MCHPC’19 Morning Break**

**Performance Evaluation of the Intel Optane DC Memory With Scientific Benchmarks**

Intel Optane technology is a cost-effective solution to create large non-volatile memory pools, which is attractive to many scientific applications. Last year Intel Optane DC Persistent Memory in DIMM (PMM) form-factor was introduced, which is capable of being used as main memory device. This technology promises faster memory access comparing to Intel Optane DC P4800X SSD with Intel Memory Drive Technology devices available previously.

In this paper, we present new benchmark data for the Intel Optane DC Persistent Memory. We studied the performance of scientific application from domains of quantum chemistry and computational astrophysics. To put performance of the Intel Optane DC PMM in comparison, we used two memory configurations: DDR4 only system and Optane DC SSD with Intel Memory Drive Technology (IMDT) that is another option for memory extension. We see that PMM is superior to IMDT in almost all our benchmarks, to no surprise. However, PMM was 20-30% more performant in quantum chemistry and only marginally better for astrophysics simulations. We also found, that for practical calculations hybrid setup demonstrates on the same order of magnitude performance as DDR4 system.
Optimizing Data Layouts for Irregular Applications on a Migratory Thread Architecture

Applications that operate on sparse data induce irregular data access patterns and cannot take full advantage of caches and prefetching. Novel hardware architectures have been proposed to address the disparity between processor and memory speeds by moving computation closer to memory. One such architecture is the Emu system, which employs light-weight threads that migrate to the location of the data being accessed. While smart heuristics and profile-guided techniques have been developed to derive good data layouts for traditional machines, these methods are largely ineffective when applied to a migratory thread architecture. In this work, we present an application-independent framework for data layout optimizations that targets the Emu architecture. We discuss the necessary tools and concepts to facilitate such optimizations, including a data-centric profiler, data distribution library, and cost model. To demonstrate the framework, we have designed a block placement optimization that distributes blocks of data across the system such that access latency is reduced. The optimization was applied towards sparse matrix-vector multiplication on an Emu FPGA implementation, achieving a geometric mean speed up of 12.5% across 57 matrices. Only one matrix experienced a loss of performance of 6%, while the maximum runtime speedup was 50%.

Optimizing Post-Copy Live Migration with System-Level Checkpoint Using Fabric-Attached Memory

Emerging Non-Volatile Memories have byteaddressability and low latency, close to the latency of main memory, together with the non-volatility of storage devices. Similarly, recently emerging interconnect fabrics, such as Gen-Z, provide high bandwidth, together with exceptionally low latency. These concurrently emerging technologies are making possible new system architectures in the data centers including systems with Fabric-Attached Memories (FAMs). FAMs can serve to create scalable, high-bandwidth, distributed, shared, byteaddressable, and non-volatile memory pools at a rack scale, opening up new usage models and opportunities.

Based on these attractive properties, in this paper we propose FAM-aware, checkpoint-based, post-copy live migration mechanism to improve the performance of migration. We have implemented our prototype with a Linux open source checkpoint tool, CRIU (Checkpoint/Restore In Userspace). According to our evaluation results, compared to the existing solution, our FAM-aware post-copy can improve at least 15% the total migration time, at least 33% the busy time, and can let the migrated application perform at least 12% better during migration.
Optimizing Memory Layout of Hyperplane Ordering for Vector Supercomputer SX-Aurora TSUBASA

This paper describes the performance optimization of hyperplane ordering methods applied to the high cost routine of the turbine simulation code called "Numerical Turbine" for the newest vector supercomputer. The Numerical Turbine code is a computational fluid dynamics code developed at Tohoku University, which can execute large-scale parallel calculation of the entire thermal flow through multistage cascades of gas and steam turbines. The Numerical Turbine code is a memory-intensive application that requires a high memory bandwidth to achieve a high sustained performance. For this reason, it is implemented in a vector supercomputer equipped with a high-performance memory subsystem. The main performance bottleneck of the Numerical Turbine code is the time-integration routine. To vectorize the lower-upper symmetric Gauss-Seidel method used in this time integration routine, a hyperplane ordering method is used. We clarify the problems of the current hyperplane ordering methods for the newest vector supercomputer NEC SX-Aurora TSUBASA and propose an optimized hyperplane ordering method that changes the data layout in the memory to resolve this bottleneck. Through the performance evaluation, it is clarified that the proposed hyperplane ordering can achieve further improvement of the performance by up to 2.77x, and 1.27x on average.

Generalized Sparse Matrix-Matrix Multiplication for Vector Engines and Graph Applications

Generalized sparse matrix-matrix multiplication (SpGEMM) is a key primitive kernel for many high-performance graph algorithms as well as for machine learning, and data analysis algorithms. Although many SpGEMM algorithms have been proposed, such as ESC and SPA, there is currently no SpGEMM kernel optimized for vector engines (VEs). NEC SX-Aurora is the new vector computing system that can achieve high performance by leveraging high bandwidth memory of 1.2TB/s and long vector of VEs, where the execution of scientific applications is limited by memory bandwidth. In this paper, we demonstrate significant initial work of SpGEMM kernel for a vector engine and implement it to vectorize several essential graph analysis algorithms: Butterfly counting and Triangle counting. We propose a SpGEMM algorithm with a novel hybrid method based on sparse vectors and loop raking to maximize the length of vectorizable code for vector machine architectures. The experimental results show that the vector engine has advantages on more massive data sets. This work contributes to high performance and portability of the SpGEMM kernel to a new family of heterogeneous computing systems, which is Vector Host (VH) equipped with different accelerators or VEs.

A Distributed Deep Memory Hierarchy System for Content-based Image Retrieval of Big
Whole Slide Image Datasets

Whole slide images (WSIs) are very large (30-50GB each in uncompressed format), multiple resolution tissue images produced by digital slide scanners, and are widely used by pathology departments for diagnostic, educational and research purposes. Content-based Image Retrieval (CBIR) applications allow pathologists to perform a sub-region search on WSIs to automatically identify image patterns that are consistent with a given query patch containing cancerous tissue patterns. The results can then be used to draw comparisons among patient samples in order to make informed decisions regarding likely prognoses and most appropriate treatment regimens, leading to new discoveries in precision and preventive medicine. CBIR applications often require repeated, random or sequential access to WSIs, and most of the time the images are preprocessed into smaller tiles, as it is infeasible to bring the entire WSI into the memory of a computer node. In this study, we have designed and implemented a distributed deep memory hierarchy data staging system that leverages Solid-State Drives (SSDs) and provides an illusion of a very large memory space that can accommodate big WSI datasets and prevent subsequent accesses to the file system. An I/O intensive sequential CBIR workflow for searching cancerous patterns in prostate carcinoma datasets was parallelized and the I/O paths were altered to include the proposed memory system. Our results indicate that the parallel performance of the CBIR workflow improves and our deep memory hierarchy, staging framework produces negligible overheads for the application performance even when the number of staging servers and their memory sizes are limited.

MCHPC’19 Lunch Break

Performance Evaluation of Advanced Features in CUDA Unified Memory

CUDA Unified Memory improves the GPU programmability and also enables GPU memory oversubscription. Recently, two advanced memory features, memory advises and asynchronous prefetch, have been introduced. In this work, we evaluate the new features on two platforms that feature different CPUs, GPUs, and interconnects. We derive a benchmark suite for the experiments and stress the memory system to evaluate both in-memory and oversubscription performance.

The results show that memory advises on the Intel-Volta/Pascal-Pcie platform bring negligible improvement for in-memory executions. However, when GPU memory is oversubscribed by about
50%, using memory advises results in up to 25% performance improvement compared to the basic CUDA Unified Memory. In contrast, the Power9-Volta-NVLink platform can substantially benefit from memory advises, achieving up to 34% performance gain for in-memory executions. However, when GPU memory is oversubscribed on this platform, using memory advises increases GPU page faults and results in considerable performance loss. The CUDA prefetch also shows different performance impact on the two platforms. It improves performance by up to 50% on the Intel-Volta/Pascal-PCI-E platform but brings little benefit to the Power9-Volta-NVLink platform.

Explicit Data Layout Management for Autotuning Exploration on Complex Memory Topologies

The memory topology of high-performance computing platforms is becoming more complex. Future exascale platforms in particular are expected to feature multiple types of memory technologies, and multiple accelerator devices per compute node.

In this paper, we discuss the use of explicit management of the layout of data in memory across memory nodes and devices for performance exploration purposes. Indeed, many classic optimization techniques rely on reshaping or tiling input data in specific ways to achieve peak efficiency on a given architecture.

With autotuning of a linear algebra code as the end goal, we present AML: a framework to treat three memory management abstractions as first-class citizens: data layout in memory, tiling of data for parallelism, and data movement across memory types. By providing access to these abstractions as part of the performance exploration design space, our framework eases the design and validation of complex, efficient algorithms for heterogeneous platforms.

Using the Intel Knights Landing architecture in one of its most NUMA configurations as a proxy platform, we showcase our framework by exploring tiling and prefetching schemes for a DGEMM algorithm.

Machine Learning Guided Optimal Use of GPU Unified Memory

NVIDIA’s unified memory (UM) creates a pool of managed memory on top of physically separated CPU and GPU memories. UM automatically migrates page-level data on-demand so programmers can quickly write CUDA codes on heterogeneous machines without tedious and error-prone manual memory management. To improve performance, NVIDIA allows advanced programmers to pass additional memory use hints to its UM driver. However, it is extremely difficult for programmers to decide when and how to efficiently use unified memory, given the complex interactions between
applications and hardware. In this paper, we present a machine learning-based approach to choosing between discrete memory and unified memory, with additional consideration of different memory hints. Our approach utilizes profiler-generated metrics of CUDA programs to train a model offline, which is later used to guide optimal use of UM for multiple applications at runtime. We evaluate our approach on NVIDIA Volta GPU with a set of benchmarks. Results show that the proposed model achieves 96% prediction accuracy in correctly identifying the optimal memory advice choice.

MCHPC’19 Afternoon Break

UMap: Enabling Application-driven Optimizations for Page Management

Leadership supercomputers feature a diversity of storage, from node-local persistent memory and NVMe SSDs to network-interconnected flash memory and HDD. Memory mapping files on different tiers of storage provides a uniform interface in applications. However, system-wide services like mmap are optimized for generality and lack flexibility for enabling application-specific optimizations. In this work, we present UMap to enable user-space page management that can be easily adapted to access patterns in applications and storage characteristics. UMap uses the userfaultfd mechanism to handle page faults in multi-threaded applications efficiently. By providing a data object abstraction layer, UMap is extensible to support various backing stores. The design of UMap supports dynamic load balancing and I/O decoupling for scalable performance. UMap also uses application hints to improve the selection of caching, prefetching, and eviction policies. We evaluate UMap in five benchmarks and real applications on two systems. Our results show that leveraging application knowledge for page management could substantially improve performance. On average, UMap achieved 1.25 to 2.5 times improvement using the adapted configurations compared to the system service.

Extending OpenMP map Clause to Bridge Storage and Device Memory

Heterogeneous architectures for high performance computing, particularly those systems that have GPU devices attached to the host CPU system, offer accelerated performance for a variety of workloads. To use those systems, applications are commonly developed to offload most computation and data onto an accelerator while utilizing host processors for helper tasks such as I/O and data movement. The approach requires users to program I/O operations for reading and
writing data from and to storage, and to and from host memory. Then users are required to program
operations for moving data between host memory and device memory. In this paper, we present
our extension to the OpenMP map clause for programming directly reading and writing data
between storage and device memory. The extension includes mechanism for handling metadata
such that metadata can be manipulated independently from data itself. This work demonstrates a
prototype runtime, and the support for binary and image data format, including jpeg and png, with
OpenCV. Experiments on matrix and image processing kernels show that the designed extension
can significantly reduce programming efforts for manipulating data and metadata among storage,
host memory, and device memory.

MCHPC'19 Panel
Maya B. Gokhale (Lawrence Livermore National Laboratory)

9:00 am - 5:30 pm
Machine Learning in HPC Environments

Session Description: 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

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.
Opening Remarks: Machine Learning in HPC Environments
Janis Keuper (Fraunhofer Institute for Industrial Mathematics), Seung-Hwan Lim (Oak Ridge National Laboratory), Michael Houston (Nvidia Corporation), Xipeng Shen (North Carolina State University)

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.

Morning Keynote – Torsten Hoefler
Torsten Hoefler (ETH Zurich)

Machine Learning in HPC Environments Morning Break

Understanding Scalability and Fine-Grain Parallelism of Synchronous Data Parallel Training
Jiali Li (University of Tennessee)

In the age of big data, deep learning has emerged as a powerful tool to extract insight and exploit its value, both in industry and scientific applications. With increasing complexity of learning models and amounts of training data, data-parallel approaches based on frequent all-reduce synchronization steps are increasingly popular. Despite the fact that high-performance computing (HPC) technologies have been designed to address such patterns efficiently, the behavior of data-parallel approaches on HPC platforms is not well understood. To address this issue, in this paper we study the behavior of Horovod, a popular data-parallel approach that relies on MPI, on Theta, a pre-Exascale machine at Argonne National Laboratory. Using two representative applications, we explore two aspects: (1) how performance and scalability is affected by important parameters such as number of nodes, number of workers, threads per node, batch size; (2) how computational phases are interleaved with all-reduce communication phases at fine granularity and what consequences this interleaving has in terms of potential bottlenecks. Our findings show that pipelining of back-propagation, gradient reduction and weight updates mitigate the effects of
stragglers during all-reduce only partially. Furthermore, there can be significant delays between weights update, which can be leveraged to mask the overhead of additional background operations that are coupled with the training.

**Fine-Grained Exploitation of Mixed Precision for Faster CNN Training**  
*J. Travis Johnston (Oak Ridge National Laboratory)*

As deep convolutional neural networks (CNNs) have become increasingly popular and successful at an ever-widening number of machine learning tasks specialized hardware has become increasingly available for training and deploying them. NVIDIA's recent Volta architecture includes tensor cores which perform a fused operation reduced and mixed precision (16-bit multiply, 32-bit accumulate). Recent research indicates that, typically, very little is lost (in terms of training accuracy) when half precision is used in place of single precision, and performance gains can be made by doing arithmetic in reduced precision. In this work we demonstrate that making layer-by-layer choices as to the arithmetic/data precision can lead to further performance improvement. In our study of 25,200 CNNs we demonstrate an average speedup (over purely half precision) of 1.27x and speedups as high as 3.64x by appropriately combining single and half precision arithmetic and data types on a layer-by-layer basis.

**Metaoptimization on a Distributed System for Deep Reinforcement Learning**  
*Iuri Frosio (Nvidia Corporation)*

Training intelligent agents through reinforcement learning (RL) is a notoriously unstable procedure. Massive parallelization on GPUs and distributed systems has been exploited to generate a large amount of training experiences and consequently reduce instabilities, but the success of training remains strongly influenced by the choice of the hyperparameters. To overcome this issue, we introduce HyperTrick, a new metaoptimization algorithm, and show its effective application to tune hyperparameters in the case of deep RL, while learning to play different Atari games on a distributed system. Our analysis provides evidence of the interaction between the identification of the optimal hyperparameters and the learned policy, that is peculiar of the case of metaoptimization for deep RL. When compared with state-of-the-art metaoptimization algorithms, HyperTrick is characterized by a simpler implementation and it allows learning similar policies, while making a more effective use of the computational resources in a distributed system.

**Scheduling Optimization of Parallel Linear Algebra Algorithms Using Supervised Learning**  
*Gabriel Laberge (Polytechnique Montréal)*

Linear algebra algorithms are used widely in a variety of domains, e.g. machine learning, numerical physics and video games graphics. For all these applications, loop-level parallelism is required to
achieve high performance. However, finding the optimal way to schedule the workload between threads is a non-trivial problem because it depends on the structure of the algorithm being parallelized and the hardware the executable is run on. In the realm of Asynchronous Many Task runtime systems, a key aspect of the scheduling problem is predicting the proper chunk-size, where the chunk-size is defined as the number of iterations of a for-loop are assigned to a thread as one task. In this paper, we study the applications of supervised learning models to predict the chunk-size which yields maximum performance on multiple parallel linear algebra operations using the HPX backend of Blaze's linear algebra library. More precisely, we generate our training and test sets by measuring performance of the application with different chunk-sizes for multiple linear algebra operations; vector-addition, matrix-vector-multiplication, matrix-matrix addition and matrix-matrix-multiplication. We compare the use of logistic regression, neural networks and decision trees with a newly developed decision tree based model in order to predict the optimal value for chunk-size. Our results show that classical decision trees and our custom decision tree model are able to forecast a chunk-size which results in good performance for the linear algebra operations.

Machine Learning in HPC Environments Lunch Break

Parallel Data-Local Training for Optimizing Word2Vec Embeddings for Word and Graph Embeddings
Gordon E. Moon (Ohio State University)

The Word2Vec model is a neural network-based unsupervised word embedding technique widely used in applications such as natural language processing, bioinformatics and graph mining. As Word2Vec repeatedly performs Stochastic Gradient Descent (SGD) to minimize the objective function, it is very compute-intensive. However, existing methods for parallelizing Word2Vec are not optimized enough for data locality to achieve high performance. In this paper, we develop a parallel data-locality-enhanced Word2Vec algorithm based on Skip-gram with a novel negative sampling method that decouples loss calculation with positive and negative samples; this allows us to efficiently reformulate matrix-matrix operations for the negative samples over the sentence. Experimental results demonstrate our parallel implementations on multi-core CPUs and GPUs achieve significant performance improvement over the existing state-of-the-art parallel Word2Vec implementations while maintaining evaluation quality. We also show the utility of our Word2Vec implementation within the Node2Vec algorithm which accelerates embedding learning for large graphs.

Scalable Hyperparameter Optimization with Lazy Gaussian Processes
Most machine learning methods require careful selection of hyper-parameters in order to train a high performing model with good generalization abilities. Hence, several automatic selection algorithms have been introduced to overcome tedious manual (try and error) tuning of these parameters. Due to its very high sample efficiency, Bayesian Optimization over a Gaussian Processes modeling of the parameter space has become the method of choice. Unfortunately, this approach suffers from a cubic compute complexity due to underlying Cholesky factorization, which makes it very hard to be scaled beyond a small number of sampling steps.

In this paper, we present a novel, highly accurate approximation of the underlying Gaussian Process. Reducing its computational complexity from cubic to quadratic allows an efficient strong scaling of Bayesian Optimization while outperforming the previous approach regarding optimization accuracy. First experiments show speedups of a factor of 162 in single node and further speed up by a factor of 5 in a parallel environment.

Machine Learning in HPC Environments Afternoon Break

Afternoon Keynote - TBD

GradVis: Visualization and Second Order Analysis of Optimization Surfaces During the Training of Deep Neural Networks
Avraam Chatzimichailidis (Fraunhofer Institute for Industrial Mathematics)

Current training methods for deep neural networks boil down to very high dimensional and non-convex optimization problems which are usually solved by a wide range of stochastic gradient descent methods. While these approaches tend to work in practice, there are still many gaps in the theoretical understanding of key aspects like convergence and generalization guarantees, which are induced by the properties of the optimization surface (loss landscape). In order to gain deeper insights, a number of recent publications proposed methods to visualize and analyze the optimization surfaces. However, the computational cost of these methods are very high, making it hardly possible to use them on larger networks.
In this paper, we present the GradVis Toolbox, an open source library for efficient and scalable visualization and analysis of deep neural network loss landscapes in TensorFlow and PyTorch. Introducing more efficient mathematical formulations and a novel parallelization scheme, GradVis allows to plot 2d and 3d projections of optimization surfaces and trajectories, as well as high resolution second order gradient information for large networks.

DisCo: Physics-Based Unsupervised Discovery of Coherent Structures in Spatiotemporal Systems
Adam Rupe (University of California, Davis)

Extracting actionable insight from complex unlabeled scientific data is an open challenge and key to unlocking data-driven discovery in science. Complementary and alternative to supervised machine learning approaches, unsupervised physics-based methods based on behavior-driven theories hold great promise. Due to computational limitations, practical application on real-world domain science problems has lagged far behind theoretical development. However, powerful modern supercomputers provide the opportunity to narrow the gap between theory and practical application. We present our first step towards bridging this divide - DisCo - a high-performance distributed workflow for the behavior-driven local causal state theory. DisCo provides a scalable unsupervised physics-based representation learning method that decomposes spatiotemporal systems into their structurally relevant components, which are captured by the latent local causal state variables. In several firsts we demonstrate the efficacy of DisCo in capturing physically meaningful coherent structures from observational and simulated scientific data. To the best of our knowledge, DisCo is also the first application software developed entirely in Python to scale to over 1000 machine nodes, providing good performance along with ensuring domain scientists' productivity. Our capstone experiment, using newly developed and optimized DisCo workflow and libraries, performs unsupervised spacetime segmentation analysis of CAM5.1 climate simulation data, processing an unprecedented 89.5 TB in 6.6 minutes end-to-end using 1024 Intel Haswell nodes on the Cori supercomputer obtaining 91% weak-scaling and 64% strong-scaling efficiency. This enables us to achieve state-of-the-art unsupervised segmentation of coherent spatiotemporal structures in complex fluid flows.

Concluding Remarks - Machine Learning in HPC Environment

9:00 am - 5:30 pm
IA^3 2019: 9th Workshop on Irregular Applications: Architectures and Algorithms

Session Description: Due to the heterogeneous data sets they process, data intensive applications employ a diverse set of methods and data structures, exhibiting 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 programming effort, and often leads to poor performance. Holistic solutions to these challenges emerge only by considering the problem from multiple perspectives: from micro- to system-architectures, from compilers to languages, from libraries to runtimes, and from algorithm design to data characteristics. Only strong collaborative efforts among researchers with different expertise, including domain experts and end users, can lead to significant breakthroughs. This workshop brings together scientists with different backgrounds to discuss methods and technologies for efficiently supporting irregular applications on current and future architectures. [https://hpc.pnl.gov/IA3/](https://hpc.pnl.gov/IA3/)

Due to the heterogeneous data sets they process, data intensive applications employ a diverse set of methods and data structures, exhibiting 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 programming effort, and often leads to poor performance. Holistic solutions to these challenges emerge only by considering the problem from multiple perspectives: from micro- to system-architectures, from compilers to languages, from libraries to runtimes, and from algorithm design to data characteristics. Only strong collaborative efforts among researchers with different expertise, including domain experts and end users, can lead to significant breakthroughs. This workshop brings together scientists with different backgrounds to discuss methods and technologies for efficiently supporting irregular applications on current and future architectures.

Keynote 1
Michael Wong (Codeplay Software Ltd)
Conveyors for Streaming Many-to-Many Communication
F. Miller Maley (Institute for Defense Analyses, Center for Communications Research)

We report on a software package that offers high-bandwidth and memory-efficient ways for a parallel application to transmit numerous small data items among its processes. The package provides a standalone library that can integrated into any SHMEM, UPC, or MPI application. It defines a simple interface to parallel objects called conveyors, and it provides a variety of conveyor implementations. Often the most efficient type of conveyor is an asynchronous three-hop conveyor, which makes heavy use of fast intranode communication. This type also uses the least memory internally. Conveyors of this type scale well to 100,000 processes and beyond.

Our experience with conveyors applied to irregular algorithms at scale has convinced us of the necessity and profitability of message aggregation. The conveyor interface is a low-level C API that is intended to guide future hardware and runtime improvements and to be a foundation for future parallel programming models.

Extending a Work-Stealing Framework with Priorities and Weights
Ryusuke Nakashima (Kyushu Institute of Technology)

This paper proposes priority- and weight-based steal strategies for an idle worker (thief) to select a victim worker in work-stealing frameworks. Typical work-stealing frameworks employ uniformly random victim selection. We implemented the proposed strategies on a work-stealing framework called Tascell; Tascell programmers can let each worker estimate and declare, as a real number, the amount of remaining work required to complete its current task so that declared values are used as priorities or weights in the enhanced Tascell framework. To reduce the total task-division cost, the proposed strategies avoid stealing small tasks. With a priority-based strategy, a thief selects the victim that has the highest known priority at that point in time. With a weight-based non-uniformly random strategy, a thief uses the relative weights of victim candidates as their selection probabilities. The proposed selection strategies outperformed uniformly random victim selection. Our evaluation uses a parallel implementation of the "highly serial" version of the Barnes-Hut force-calculation algorithm in a shared memory environment and five benchmark programs in a distributed memory environment.

RDMA vs. RPC for Implementing Distributed Data Structures
Benjamin A. Brock (University of California, Berkeley; Lawrence Berkeley National Laboratory)
Distributed data structures are key to implementing scalable applications for scientific simulations and data analysis. In this paper we look at two implementation styles for distributed data structures: remote direct memory access (RDMA) and remote procedure call (RPC). We focus on operations that require individual accesses to remote portions of a distributed data structure, e.g., accessing a hash table bucket or distributed queue, rather than global operations in which all processors collectively exchange information. We look at the trade-offs between the two styles through microbenchmarks and a performance model that approximates the cost of each. The RDMA operations have direct hardware support in the network and therefore lower latency and overhead, while the RPC operations are more expressive but higher cost and can suffer from lack of attentiveness from the remote side. We also run experiments to compare the real-world performance of RDMA- and RPC-based data structure operations with the predicted performance to evaluate the accuracy of our model, and show that while the model does not always precisely predict running time, it allows us to choose the best implementation in the examples shown. We believe this analysis will assist developers in designing data structures that will perform well on current network architectures, as well as network architects in providing better support for this class of distributed data structures.

A Mixed Precision Multicolor Point-Implicit Solver for Unstructured Grids on GPUs
Aaron Walden (NASA)

This paper presents a new mixed-precision implementation of a linear-solver kernel used in practical large-scale CFD simulations to improve GPU performance. The new implementation reduces memory traffic by using the half-precision format for some critical computations while maintaining double-precision solution accuracy. As the linear-solver kernel is memory bound on GPUs, a reduction in memory traffic directly translates to improved performance. The performance of the new implementation is assessed for a benchmark steady flow simulation and a large-scale unsteady turbulent flow application. Both studies were conducted using NVIDIA® Tesla V100 GPUs on the Summit system at the Oak Ridge Leadership Computing Facility.

Mixed-Precision Tomographic Reconstructor Computations on Hardware Accelerators

The computation of tomographic reconstructions (ToR) is at the core of a simulation framework to design the next generation of adaptive optics (AO) systems to be installed on future Extremely Large Telescopes (ELT). In fact, it is also a critical component for their operation on sky. The goals of these instruments range from the detection of the light from the most distant galaxies to the analysis of the composition of exoplanets terrestrial atmospheres. Based on advanced AO techniques, the instrument MOSAIC relies on a computational framework to filter out the Earth
atmospheric turbulence and eventually enhance the images quality out of ground-based telescopes. The ToR calculation is a compute-bound operation based on the Cholesky factorization. Due to its cubical algorithmic complexity, the ToR may represent a major bottleneck for the E-ELT when scaling up the large number of wavefront sensors used in the baseline MOSAIC design. To mitigate this increasing dimensionality overhead, this paper presents the implementation of a novel mixed-precision Cholesky-based dense matrix solver on hardware accelerators. The new algorithm takes into account the data-sparse structure of the covariance matrix operator and uses the tensor cores of NVIDIA V100 GPUs to leverage performance at an unprecedented scale. To our knowledge, this is the first computational astronomy application that exploits V100's tensor cores outside of the traditional arena of artificial intelligence. Experimental results demonstrate the accuracy robustness and the high performance of the mixed-precision ToR on synthetic datasets, which paves the way for future instrument deployments on the E-ELT.

**Metall: A Persistent Memory Allocator Enabling Graph Processing**

Keita Iwabuchi (Lawrence Livermore National Laboratory)

We present Metall, a persistent memory allocator designed to provide developers with an API to allocate custom C++ data structures in both block-storage and byte-addressable persistent memories (e.g., NVMe and Intel Optane DC Persistent Memory). Metall incorporates state-of-the-art allocation algorithms in Supermalloc with the rich C++ interface developed by Boost.Interprocess, and provides persistent memory snapshotting (versioning) capabilities. We demonstrate Metall processing large graphs in a variety of conditions and data-structure configurations, indicating a bright future for data-analytics leveraging emerging persistent memory technologies.

**IA^3 2019 Lunch Break**

**Keynote 2: Sparse Linear Algebra in Facebook's Deep Learning Models**

Jongsoo Park (Facebook)

Sparse linear algebra plays surprisingly important roles in deep learning models used in social network. The use cases share similar challenges as scientific computing applications using sparse linear algebra. I believe both communities can benefit from each other by working together on basic SW building blocks, processor architecture innovations, and interconnect network.

**iPregel: Strategies to Deal with an Extreme Form of Irregularity in Vertex-Centric Graph Processing**
Over the last decade, the vertex-centric programming model has attracted significant attention in the world of graph processing, resulting in the emergence of a number of vertex-centric frameworks. Its simple programming interface, where computation is expressed from a vertex point of view, offers both ease of programming to the user and inherent parallelism for the underlying framework to leverage.

However, vertex-centric programs represent an extreme form of irregularity, both inter and intra core. This is because they exhibit a variety of challenges from a workload that may greatly vary across supersteps, through fine-grain synchronisations, to memory accesses that are unpredictable both in terms of quantity and location.

In this paper, we explore three optimisations which address these irregular challenges; a hybrid combiner carefully coupling lock-free and lock-based combinations, the partial externalisation of vertex structures to improve locality and the shift to an edge-centric representation of the workload. We also assess the suitability of more traditional optimisations such as dynamic load-balancing and software prefetching.

The optimisations were integrated into the iPregel vertex-centric framework, enabling the evaluation of each optimisation in the context of graph processing across three general purpose benchmarks common in the vertex-centric community, each run on four publicly available graphs covering all orders of magnitude from a million to a billion edges.

The result of this work is a set of techniques which we believe not only provide a significant performance improvement in vertex-centric graph processing, but are also applicable more generally to irregular applications.
diagonal dominant matrices, the iterative Jacobi-method presents itself as a scalable alternative on massively-parallel systems. In this work, we present modifications to the Jacobi algorithm, extending its applicability to a wider class of matrices. Our approach cuts the matrix into blocks to allow register-based pivoting in batched CUDA factorization kernels and, for the first time on GPUs, also flexible permutations on the block level in an a posteriori threshold full pivoting scheme. Experiments show our batched kernels to be on par with partially pivoted approaches and the resulting sparse factorizations compare favorable in quality and speed with their traditionally scheduled counterparts.

A Hardware Prefetching Mechanism for Vector Gather Instructions
Hikaru Takayashiki (Tohoku University)

Vector gather instructions are responsible for handling indirect memory accesses in vector processing. Since the indirect memory accesses usually express irregular access patterns, they have relatively low spatial and temporal locality compared with regular access patterns. As a result, an application with many vector gather instructions suffers from long latencies of the indirect memory accesses. Thus, the long latencies cause a significant performance degradation in vector processing.

This paper proposes a hardware prefetching mechanism to hide memory access latencies of indirect memory accesses. The mechanism prefetches cacheable index data before executing a vector gather instruction, and predicts the addresses of the memory requests issued by the vector gather instruction. The mechanism then tries to prefetch the data based on the predicted addresses. As a result, the mechanism can reduce the memory access latencies of vector gather instructions. Moreover, this paper discusses how many cache blocks should be loaded per prediction regarding a single vector gather instruction by varying the prefetching parameters of distance and degree. In the evaluation, the performance of a simple kernel is examined with two types of index data: sequential and random. The evaluation results show that the prefetching mechanism improves the performance of the sequential-indexed and random-indexed kernels by 2.2x and 1.2x, respectively.

Performance Impact of Memory Channels on Sparse and Irregular Algorithms
Oded Green (Nvidia Corporation, Georgia Institute of Technology)

Graph processing is typically considered to be a memory-bound rather than compute-bound problem. One common line of thought is that more available memory bandwidth corresponds to better graph processing performance. However, in this work we demonstrate that the key factor in the utilization of the memory system for graph algorithms is not necessarily the raw bandwidth or even the latency of memory requests. Instead, we show that performance is proportional to the number of memory channels available to handle small data transfers with limited spatial locality.
Using several widely used graph frameworks, including Gunrock (on the GPU) and GAPBS and Ligra (for CPUs), we evaluate key graph analytics kernels using two unique memory hierarchies, DDR-based and HBM/MCDRAM. Our results show that the differences in the peak bandwidths of several Pascal-generation GPU memory subsystems aren't reflected in the performance of various analytics. Furthermore, our experiments on CPU and Xeon Phi systems (see arXiv extended version) demonstrate that the number of memory channels utilized can be a decisive factor in performance across several different applications. For CPU systems with smaller thread counts, the memory channels can be underutilized while systems with high thread counts can oversaturate the memory subsystem, which leads to limited performance. Finally, we model the potential performance improvements of adding more memory channels with narrower access widths than are found in current platforms. We analyze performance trade-offs for the two most prominent types of memory accesses found in graph algorithms, streaming and random accesses.

**Cascaded DMA Controller for Speedup of Indirect Memory Access in Irregular Applications**

Tomoya Kashimata (Waseda University)

Indirect memory accesses caused by sparse linear algebra calculations are widely used in important real applications. However, they also cause serious inefficient memory accesses and pipeline stalls resulting in low execution efficiency even with high memory bandwidth and much computational resource. One of the important issues of indirect memory accesses, such as accessing $A[B[i]]$, is that it requires two successive memory accesses: the index loads ($B[i]$) and the following data element accesses ($A[B[i]]$). To overcome this situation, we propose the Cascaded-DMAC (CDMAC). This CDMAC is intended to be attached in each core of a multicore chip in addition to a CPU core, a vector accelerator, and a local data memory. It performs data transfers between an off-chip main memory and an in-core local data memory, which provides data to the accelerator. The key idea of the CDMAC is cascading two DMACs so that the first one loads indices, then the second one accesses data elements by using these indices. Thus, this organization realizes the autonomous indirect memory accesses by giving an index array and an element array, and obtains the efficient SIMD computations by lining up the sparse data into the local data memory. We implemented a multicore processor having the proposed CDMAC on an FPGA board. The evaluation result of sparse matrix-vector multiplications on the FPGA shows that the CDMAC achieves a maximum speedup of 17x compared with the CPU data transfer.

**Debate**

Tyler Sorensen (Princeton University), Nesreen Ahmed (Intel Corporation), Jonathan Beard (ARM Ltd), Ana Verbanescu (University of Amsterdam), Mark Raugas (Pacific Northwest National Laboratory (PNNL))
International Parallel Data Systems Workshop (PDSW)

Session Description: We are pleased to announce that the 4th International Parallel Data Systems Workshop (PDSW’19) will be hosted at SC19: The International Conference for High Performance Computing, Networking, Storage and Analysis. The objectives of this one day workshop are to promote and stimulate researchers’ interactions to address some of the most critical challenges for scientific data storage, management, devices, and processing infrastructure for both traditional compute intensive simulations and data-intensive high performance computing solutions. Special attention will be given to issues in which community collaboration can be crucial for problem identification, workload capture, solution interoperability, standards with community buy-in, and shared tools. Many scientific problem domains continue to be extremely data intensive. Traditional HPC systems and the programming models for using them such as MPI were designed from a compute-centric perspective with an emphasis on achieving high floating point computation rates. But processing, memory, and storage technologies have not kept pace and there is a widening performance gap between computation and the data management infrastructure. Hence data management has become the performance bottleneck for a significant number of applications targeting HPC systems. Concurrently, there are increasing challenges in meeting the growing demand for analyzing experimental and observational data. In many cases, this is leading new communities to look toward HPC platforms. In addition, the broader computing space has seen a revolution in new tools and frameworks to support Big Data analytics and machine learning.
widening performance gap between computation and the data management infrastructure. Hence data management has become the performance bottleneck for a significant number of applications targeting HPC systems. Concurrently, there are increasing challenges in meeting the growing demand for analyzing experimental and observational data. In many cases, this is leading new communities to look toward HPC platforms. In addition, the broader computing space has seen a revolution in new tools and frameworks to support Big Data analytics and machine learning.

Alluxio - Data Orchestration for Analytics and AI in the Cloud
Haoyuan Li (Alluxio Inc)
The data eco-system has heavily evolved over the past two decades. There is an explosion of data-driven frameworks including Presto, Hive, Spark, and MapReduce to run data analytics and ETL queries, as well as TensorFlow, PyTorch to train and serve models. On the data side, the approach to manage and store data has evolved from HDFS to cheaper, more scalable and separated services typified by cloud object stores like AWS S3. Data engineering has become increasingly complex, inefficient, and hard, particularly in the hybrid and cloud environments.

Alluxio Open Source Software is to address these challenges. Alluxio, born from UC Berkeley AMPLab, is a data orchestration system that provides a unified data access and caching layer for single cloud, hybrid and multi-cloud deployments. Alluxio enables distributed compute engines like Presto, Hive, or TensorFlow to transparently access data from various storage systems (including S3, HDFS, Azure etc.) while actively leveraging in-memory cache to accelerate data access. Alluxio community has 1000+ open source contributors and the software is used by 100+ companies worldwide with the large production deployment over 1000 nodes.

In this talk, we will present - New trends and challenges in the data ecosystem in cloud era - Key innovation of Alluxio Project - Production use cases of using popular stacks like {Presto, Spark, Flink, Tensorflow}/Alluxio/{S3, HDFS}

PDSW Morning Break

In Search of a Fast and Efficient Serverless DAG Engine
Benjamin Carver (George Mason University)

Python-written data analytics applications can be modeled as and compiled into a directed
acyclic graph (DAG) based workflow, where the nodes are fine-grained tasks and the edges are task dependencies. Such analytics workflow jobs are increasingly characterized by short, fine-grained tasks with large fan-outs. These characteristics make them well-suited for a new cloud computing model called serverless computing or Function-as-a-Service (FaaS), which has become prevalent in recent years. The auto-scaling property of serverless computing platforms accommodates short tasks and bursty workloads, while the pay-per-use billing model of serverless computing providers keeps the cost of short tasks low.

In this paper, we thoroughly investigate the problem space of DAG scheduling in serverless computing. We identify and evaluate a set of techniques to make DAG schedulers serverless-aware. These techniques have been implemented in WUKONG, a serverless, DAG scheduler attuned to AWS Lambda. WUKONG provides decentralized scheduling through a combination of static and dynamic scheduling. We present the results of an empirical study in which WUKONG is applied to a range of microbenchmark and real-world DAG applications. Results demonstrate the efficacy of WUKONG in minimizing the performance overhead introduced by AWS Lambda — WUKONG achieves competitive performance compared to a serverful DAG scheduler, while improving the performance of real-world DAG jobs by as much as 3.1× at larger scale.

**Enabling Transparent Asynchronous I/O Using Background Threads**

**Houjun Tang (Lawrence Berkeley National Laboratory)**

With scientific applications moving toward exascale levels, an increasing amount of data is being produced and analyzed. Providing efficient data access is crucial to the productivity of the scientific discovery process. Compared to improvements in CPU and network speeds, I/O performance lags far behind, such that moving data across the storage hierarchy can take longer than data generation or analysis. To alleviate this I/O bottleneck, asynchronous read and write operations have been provided by the POSIX and MPI-I/O interfaces and can overlap I/O operations with computation, and thus hide I/O latency. However, these standards lack support for non-data operations such as file open, stat, and close, and their read and write operations require users to both manually manage data dependencies and use low-level byte offsets. This requires significant effort and expertise for applications to utilize.

To overcome these issues, we present an asynchronous I/O framework that provides support for all I/O operations and manages data dependencies transparently and automatically. Our prototype asynchronous I/O implementation as an HDF5 VOL connector demonstrates the effectiveness of hiding the I/O cost from the application with a low overhead and easy-to-use programming interface.
PDSW Works in Progress I

Four presenters will provide brief (5-minute) talks on their on-going work, with fresh problems/solutions.

Active Learning-Based Automatic Tuning and Prediction of Parallel I/O Performance

Parallel I/O is an indispensable part of scientific applications. The current stack of parallel I/O contains many tunable parameters. While changing these parameters can increase I/O performance many-fold, the application developers usually resort to default values because tuning is a cumbersome process and requires expertise. We propose two auto-tuning models, based on active learning that recommend a good set of parameter values (currently tested with Lustre parameters and MPI-IO hints) for an application on a given system. These models use Bayesian optimization to find the values of parameters by minimizing an objective function. The first model runs the application to determine these values, whereas, the second model uses an I/O prediction model for the same. Thus the training time is significantly reduced in comparison to the first model (e.g., from 800 seconds to 18 seconds). Also both the models provide flexibility to focus on improvement of either read or write performance. To keep the tuning process generic, we have focused on both read and write performance. We have validated our models using an I/O benchmark (IOR) and 3 scientific application I/O kernels (S3D-IO, BT-IO and GenericIO) on two supercomputers (HPC2010 and Cori). Using the two models, we achieve an increase in I/O bandwidth of up to 11x over the default parameters. We got up to 3x improvements for 37 TB writes, corresponding to 1 billion particles in GenericIO. We also achieved up to 3.2x higher bandwidth for 4.8 TB of non-contiguous I/O in BT-IO benchmark.

Applying Machine Learning to Understand the Write Performance of Large-Scale Parallel Filesystems

Bing Xie (Oak Ridge National Laboratory)

In high-performance computing (HPC), I/O performance prediction offers the potential to improve the efficiency of scientific computing. In particular, accurate prediction can make runtime estimates more precise, guide users toward optimal checkpoint strategies, and better inform facility provisioning and scheduling policies. HPC I/O performance is notoriously difficult to predict and model, however, in large part because of inherent variability and a lack of transparency in the behaviors of constituent storage system components. In this work we seek to advance the state of the art in HPC I/O performance prediction by (1) modeling the mean performance to address high variability, (2) deriving model features from write patterns, system architecture and system configurations, and (3) employing Lasso regression model to improve
model accuracy. We demonstrate the efficacy of our approach by applying it to a crucial subset of common HPC I/O motifs, namely, file-per-process checkpoint write workloads. We conduct experiments on two distinct production HPC platforms—Titan at the Oak Ridge Leadership Computing Facility and Cetus at the Argonne Leadership Computing Facility—to train and evaluate our models. We find that we can attain ≤ 30% relative error for 92.79% and 99.64% of the samples in our test set on these platforms, respectively.

**PDSW Lunch Break**

**A House Divided: Why Don't Cloud Storage and HPC Storage Share More Technology?**
Brent Welch (Google LLC), Raghu Raja (Amazon Web Services), Evan Burness (Microsoft Corporation), Philip Carns (Argonne National Laboratory), Glenn K. Lockwood (Lawrence Berkeley National Laboratory)
This panel of experts who've straddled the worlds of both high-performance computing and cloud storage and I/O will discuss what technologies the cloud has to offer that could solve some outstanding challenges in HPC, how the data-driven challenges of HPC workloads are shaping cloud technologies, and what gaps remain that must be bridged between the world of HPC and cloud computing.

**PDSW Works in Progress II**
Jay Lofstead (Sandia National Laboratories)
Four presenters will provide brief (5-minute) talks on their on-going work, with fresh problems/solutions.

**PDSW Afternoon Break**

**Towards Physical Design Management in Storage Systems**

In the post-Moore era, systems and devices with new architectures will arrive at a rapid rate with significant impacts on the software stack. Applications will not be able to fully benefit from
new architectures unless they can delegate adapting to new devices to lower layers in the stack.

In this paper we introduce physical design management which deals with the problem of identifying and executing transformations on physical designs of stored data -- i.e. how data is mapped to storage abstractions like files, objects, or blocks -- in order to improve performance. Physical design is traditionally placed with applications, access libraries, and databases, using hard-wired assumptions about underlying storage systems. Yet, storage systems increasingly not only contain multiple kinds of storage devices with vastly different performance profiles but also move data among those storage devices, thereby changing the benefit of a particular physical design. We advocate placing physical design management in storage, identify interesting research challenges, provide a brief description of a prototype implementation in Ceph, and discuss the results of initial experiments at scale that are digitally replicable using CloudLab. These experiments show performance and resource utilization trade-offs associated with choosing different physical designs and choosing to transform between physical designs.

A Foundation for Automated Placement of Data

With the increasing complexity of memory and storage, it is important to automate the decision of how to assign data structures to memory and storage devices. On one hand, this requires developing models to reconcile application access patterns against the limited capacity of higher-performance devices. On the other, such a modeling task demands a set of primitives to build from, and a toolkit that implements those primitives in a robust, dynamic fashion. We focus on the latter problem, and to that end we present an interface that abstracts the physical layout of data from the application developer. This will allow the developers of memory cost models to use our abstracta as the basis for their implementation, while application developers will see a unified, scalable, and resilient memory environment.

Profiling Platform Storage Using IO500 and Mistral

Jay Lofstead (Sandia National Laboratories)

This paper explores how we used IO500 and the Mistral tool from Ellexus to observe detailed performance characteristics to inform tuning IO performance on Astra, a Sandia machine which uses an all flash, Lustre-based storage array. Through this case study, we demonstrate that IO500 serves as a meaningful, widely applicable storage benchmark, even for all flash storage. We also demonstrate that using fine-grained profiling tools, such as Mistral, is essential for revealing tuning requirement details. Overall, this paper provides a methodology and set of tools that can be used to analyze storage system performance to inform tuning for other large-scale supercomputers.
Understanding Data Motion in the Modern HPC Data Center

The utilization and performance of storage, compute, and network resources within HPC data centers have been studied extensively, but much less work has gone towards characterizing how these resources are used in conjunction to solve larger scientific challenges. To address this gap, we present our work towards characterizing workloads and workflows at a data center-wide level by examining all data transfers that occurred between storage, compute, and the external network at the National Energy Research Scientific Computing Center (NERSC) over a three-month period. Using a simple abstract representation of data transfers, we analyze over 100 million transfer logs from Darshan, HPSS user interfaces, and Globus to quantify the load on data paths between compute, storage, and the wide-area network based on transfer direction, user, transfer tool, source, destination, and time. We show that parallel I/O from user jobs, while undeniably important, is only one of several major I/O workloads that occurs throughout the execution of scientific workflows. We also show that this approach can be used to connect anomalous data traffic to specific users and file access patterns, and we construct time-resolved user transfer traces to demonstrate that it is possible to systematically identify coupled data motion for individual workflows.

PDSW Discussion and Closing Remarks
Suzanne McIntosh (New York University), Philip Carns (Argonne National Laboratory), Glenn K. Lockwood (Lawrence Berkeley National Laboratory)

9:00 am - 5:30 pm

The 10th International Workshop on Performance Modeling, Benchmarking, and Simulation of High-Performance Computer Systems (PMBS19)

Session Description: The PMBS19 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 exascale 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 towards the exascale era. This workshop continues to attract high quality input from industry, government, and academia. http://www.pmbsworkshop.org

The PMBS19 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 exascale 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 towards the exascale era. This workshop continues to attract high quality input from industry, government, and academia.

Automatic Throughput and Critical Path Analysis of x86 and ARM Assembly Kernels
Jan Laukemann (University of Erlangen-Nuremberg, Department of Computer Science)

Useful models of loop kernel runtimes on out-of-order architectures require an analysis of the in-core performance behavior of instructions and their dependencies. While an instruction throughput prediction sets a lower bound to the kernel runtime, the critical path defines an upper bound. Such predictions are an essential part of analytic (i.e., white-box) performance models like the Roofline
and Execution-Cache-Memory (ECM) models. They enable a better understanding of the performance-relevant interactions between hardware architecture and loop code.

The Open Source Architecture Code Analyzer (OSACA) is a static analysis tool for predicting the execution time of sequential loops. It previously supported only x86 (Intel and AMD) architectures and simple, optimistic full-throughput execution. We have heavily extended OSACA to support ARM instructions and critical path prediction including the detection of loop-carried dependencies, which turns it into a versatile cross-architecture modeling tool. We show runtime predictions for code on Intel Cascade Lake, AMD Zen, and Marvell ThunderX2 micro-architectures based on machine models from available documentation and semi-automatic benchmarking. The predictions are compared with actual measurements.

An Instruction Roofline Model for GPUs
Nan Ding (Lawrence Berkeley National Laboratory)

The Roofline performance model provides an intuitive approach to identify performance bottlenecks and guide performance optimization. However, the classic FLOP-centric approach is inappropriate for emerging applications that perform more integer operations than floating-point operations. In this paper, we propose an Instruction Roofline Model on NVIDIA GPUs. The Instruction Roofline incorporates instructions and memory transactions across all memory hierarchies together and provides more performance insights than the FLOP-oriented Roofline Model, i.e., instruction throughput, stride memory access patterns, bank conflicts, and thread predication. We use our Instruction Roofline methodology to analyze five proxy applications: HPGMG from AMReX, BatchSW from merAligner, Matrix Transpose benchmarks, cudaTensorCoreGemm, and cuBLAS. We demonstrate the ability of our methodology to understand various aspects of performance and performance bottlenecks on NVIDIA GPUs and motivate code optimizations.

PMBS19 Morning Break

Exploiting Hardware-Accelerated Ray Tracing for Monte Carlo Particle Transport with OpenMC
Simon McIntosh-Smith (University of Bristol)

OpenMC is a CPU-based Monte Carlo particle transport simulation code recently developed in the Computational Reactor Physics Group at MIT, and which is currently being evaluated by the UK Atomic Energy Authority for use on the ITER fusion reactor project. In this paper we present a novel
port of OpenMC to run on the new ray tracing (RT) cores in NVIDIA’s latest Turing GPUs. We show here that the OpenMC GPU port yields up to 9.8x speedup on a single node over a 16-core CPU using the native constructive solid geometry, and up to 13x speedup using approximate triangle mesh geometry. Furthermore, since the expensive 3D geometric operations required during particle transport simulation can be formulated as a ray tracing problem, there is an opportunity to gain even higher performance on triangle meshes by exploiting the RT cores in Turing GPUs to enable hardware-accelerated ray tracing. Extending the GPU port to support RT core acceleration yields between 2x and 20x additional speedup. We note that geometric model complexity has a significant impact on performance, with RT core acceleration yielding comparatively greater speedups as complexity increases. To the best of our knowledge, this is the first work showing that exploitation of RT cores for scientific workloads is possible. We finish by drawing conclusions about RT cores in terms of wider applicability, limitations and performance portability.

Enhancing Monte Carlo Proxy Applications on GPUs
Forrest Shriver (University of Florida)

In Monte Carlo neutron transport simulations, a computational routine commonly known as the "cross-section lookup" has been identified as being the most computationally expensive part of these applications. A tool which is commonly used as a proxy application for these routines, named "XSBench", was created to simulate popular algorithms used in these routines on CPUs. Currently, however, as GPU-based HPC resources have become more widely available, there has been significant interest and efforts invested in moving these traditionally CPU-based simulations to GPUs. Unfortunately, the algorithms commonly used in the cross-section lookup routine were originally devised and developed for CPU-based platforms, and have seen limited study on GPUs to date. Additionally, platforms such as XSBench implement approximations which may have a negligible effect on CPUs, but may be quite impactful to performance on GPUs given the more resource-limited nature of the latter. As a result, we have created VEXS, a new tool for modeling the cross-section lookup routine which removes or at least reduces the approximations made by XSBench in order to provide a more realistic prediction of algorithm performance on GPUs. In this paper, we detail our efforts to remove and reduce these approximations, show the resulting improvement in performance prediction in comparison to a reference production code, Shift, and provide some basic profiling analysis of the resulting application.

Comparing Managed Memory and UVM with and without Prefetching on NVIDIA Volta GPUs
Rahulkumar Gayatri (National Energy Research Scientific Computing Center (NERSC)), Kevin Gott (National Energy Research Scientific Computing Center (NERSC)), Jack Deslippe (National Energy Research Scientific Computing Center (NERSC))

One of the major differences in many-core versus multicore architectures is the presence of two
different memory spaces: a host space and a device space. In the case of NVIDIA GPUs, the device is supplied with data from the host via one of the multiple memory management API calls provided by the CUDA framework, such as CudaMallocManaged and CudaMemCpy. Modern systems, such as the Summit supercomputer, have the capability to avoid the use of CUDA calls for memory management and access the same data on GPU and CPU. This is done via the Address Translation Services (ATS) technology that gives a unified virtual address space for data allocated with malloc and new if there is an NVLink connection between the two memory spaces. In this paper, we perform a deep analysis of the performance achieved when using two types of unified virtual memory addressing: UVM and managed memory.

**Testing the Limits of Tapered Fat Tree Networks**

HPC system procurement with a fixed budget is an optimization problem with many trade-offs. In particular, the choice of an interconnection network for a system is a major choice, since communication performance is important to overall application performance and the network makes up a substantial fraction of a supercomputer’s overall price. It is necessary to understand how sensitive representative jobs are to various aspects of network performance to procure the right network. Unlike previous studies, which used mostly communication-only motifs or simulation, this work employs a real system and measures the performance of representative applications under controlled environments. We vary background congestion, job mapping, and job placement with different levels of network tapering on a fat tree. Overall, we find that a 2:1 tapered fat tree provides sufficiently robust communication performance for a representative mix of applications while generating meaningful cost savings relative to a full bisection bandwidth fat tree. Furthermore, our results advise against further tapering, as the resulting performance degradation would exceed cost savings. However, application-specific mappings and topology-aware schedulers may reduce global bandwidth needs, providing room for additional network tapering.

**Validation of the gem5 Simulator for x86 Architectures**

Ayaz Akram (University of California, Davis)

gem5 has been extensively used in computer architecture simulations and in the evaluation of new architectures for HPC (high performance computing) systems. Previous work has validated gem5 against ARM platforms. However, gem5 still shows high inaccuracy when modeling Intel’s x86 based processors. In this work, we focus on the simulation of a single node high performance system and study the sources of inaccuracies of gem5. Then we validate gem5 simulator against an Intel processor, Core-i7 (Haswell microarchitecture). We configured gem5 as close as possible to match Core-i7 Haswell microarchitecture configurations and made changes to the simulator to add some features, modified existing code, and tuned built-in configurations. As a result, we validated
the simulator by fixing many sources of errors to match real hardware results with less than 6%
mean error rate for different control, memory, dependency and execution microbenchmarks.

PMBS19 Lunch Break

A Generalized Statistics-Based Model for Predicting Network-Induced Variability

Shared network topologies, such as dragonfly, subject applications to unavoidable inter-job
interference arising from congestion on shared network links. Quantifying the impact of congestion
is essential for effectively assessing and comparing the application runtimes. We use network
performance counter-based metrics for this quantification. We claim and demonstrate that by using
a local view of congestion captured through the counters monitored during a given application run,
we can accurately determine the run conditions and thereby estimate the impact on the
application’s performance. We construct a predictive model that is trained using several
applications with distinctive communication characteristics run under production system conditions
with a 91% accuracy for predicting congestion effects.

CUDA Flux: A Lightweight Instruction Profiler for CUDA Applications
Lorenz Braun (Heidelberg University)

GPUs are powerful, massively parallel processors, which require a vast amount of thread parallelism
to keep their thousands of execution units busy, and to tolerate latency when accessing its high-
throughput memory system. Understanding the behavior of massively threaded GPU programs can
be difficult, even though recent GPUs provide an abundance of hardware performance counters,
which collect statistics about certain events. Profiling tools that assist the user in such analysis for
their GPUs, like NVIDIA’s nvprof and cupti, are state-of-the-art. However, instrumentation based on
reading hardware performance counters can be slow, in particular when the number of metrics is
large. Furthermore, the results can be inaccurate as instructions are grouped to match the available
set of hardware counters.

PMBS19 Afternoon Break
OMB-UM: Design, Implementation, and Evaluation of CUDA Unified Memory Aware MPI Benchmarks

Unified Memory (UM) has significantly simplified the task of programming CUDA applications. With UM, the CUDA driver is responsible for managing the data movement between CPU and GPU and the programmer can focus on the actual designs. However, the performance of Unified Memory codes has not been on par with explicit device buffer based code. To this end, the latest NVIDIA Pascal and Volta GPUs with hardware support such as fine-grained page faults offer the best of both worlds, i.e., high-productivity and high-performance. However, these enhancements in the newer generation GPU architectures need to be evaluated in a different manner, especially in the context of MPI+CUDA applications.

In this paper, we extend the widely used MPI benchmark — OSU Micro-benchmarks (OMB) to support Unified Memory or Managed Memory based MPI benchmarks. The current version of OMB cannot effectively characterize UM-Aware MPI design because CUDA driver movements are not captured appropriately with standardized Host and Device buffer based benchmarks. To address this key challenge, we propose new designs for the OMB suite and extend point to point and collective benchmarks that exploit sender and receiver side CUDA kernels to emulate the effective location of the UM buffer on Host and Device. The new benchmarks allow the users to better understand the performance of codes with UM buffers through user-selectable knobs that enable or disable sender and receiver side CUDA kernels. In addition to the design and implementation, we provide a comprehensive performance evaluation of the new UM benchmarks in the OMB-UM suite on a wide variety of systems and MPI libraries. From these evaluations we also provide valuable insights on the performance of various MPI libraries on UM buffers which can lead to further improvement in the performance of UM in CUDA-Aware MPI libraries.

Fine-Grained Analysis of Communication Similarity between Real and Proxy Applications
Omar Aaziz (Sandia National Laboratories)

In this work we investigate the dynamic communication behavior of parent and proxy applications, and investigate whether or not the dynamic communication behavior of the proxy matches that of its respective parent application. The idea of proxy applications is that they should match their parent well, and should exercise the hardware and perform similarly, so that from them lessons can be learned about how the HPC system and the application can best be utilized. We show here that some proxy/parent pairs do not need the extra detail of dynamic behavior analysis, while others can benefit from it, and through this we also identified a parent/proxy mismatch and improved the proxy application.
Performance Analysis of Deep Learning Workloads on Leading-Edge Systems
Yihui Ren (Brookhaven National Laboratory)

This work examines the performance of leading-edge systems designed for machine learning computing, including the NVIDIA DGX-2, Amazon Web Services (AWS) P3, IBM Power System Accelerated Compute Server AC922, and a consumer-grade Exxact TensorEX TS4 GPU server. Representative deep learning workloads from the fields of computer vision and natural language processing are the focus of the analysis. Performance analysis is performed along with a number of important dimensions. Performance of the communication interconnects and large and high-throughput deep learning models are considered. Different potential use models for the systems as standalone and in the cloud also are examined. The effect of various optimization of the deep learning models and system configurations is included in the analysis.

9:00 am - 5:30 pm

Tenth Annual Workshop for the Energy Efficient HPC Working Group

Session Description: Despite significant pressure on both Moore’s Law and Dennard scaling, the appetite for ever larger systems continues to grow. Computing trends, in terms of highly optimized hardware platforms that may leverage accelerators or other non-traditional components, scalable and highly performing applications, and the requirements to manage exponentially larger data sets are driving facility demands not envisioned just a few years ago. The facility demands for supercomputing centers (SCs) are characterized by electrical power demands for computing systems that scale to tens of megawatts (MW) and millisecond voltage fluctuations approaching 10MW for the largest systems. The demand for primary electrical distribution capabilities to current large-scale facilities can exceed 60MW. The operational costs of these facilities must be balanced versus the demand from the systems owners and users for high availability, high utilization, and low-impact facility maintenance and service demands. To achieve this balance, many SCs continue to innovate their operational design practices and technologies. The workshop for the Energy Efficient HPC Working Group brings together those concerned with HPC energy efficiency and energy conservation of both HPC systems and facilities. The annual workshop has a broad scope of energy efficiency in HPC ranging from silicon and components, through platform integration, software, and firmware and up to the data center, facility, and connection to the utility provider. Because of this broad scope, we engage a wide spectrum of participants - both facilities and operationally focused people as well as computer engineers and architects - which enhances the discussions and sharing that occurs. https://eehpcwg.llnl.gov/conf_sc19.html
Despite significant pressure on both Moore’s Law and Dennard scaling, the appetite for ever larger systems continues to grow. Computing trends, in terms of highly optimized hardware platforms that may leverage accelerators or other non-traditional components, scalable and highly performing applications, and the requirements to manage exponentially larger data sets are driving facility demands not envisioned just a few years ago. The facility demands for supercomputing centers (SCs) are characterized by electrical power demands for computing systems that scale to tens of megawatts (MW) and millisecond voltage fluctuations approaching 10MW for the largest systems. The demand for primary electrical distribution capabilities to current large-scale facilities can exceed 60MW. The operational costs of these facilities must be balanced versus the demand from the systems owners and users for high availability, high utilization, and low-impact facility maintenance and service demands. To achieve this balance, many SCs continue to innovate their operational design practices and technologies.

The workshop for the Energy Efficient HPC Working Group brings together those concerned with HPC energy efficiency and energy conservation of both HPC systems and facilities. The annual workshop has a broad scope of energy efficiency in HPC ranging from silicon and components, through platform integration, software, and firmware and up to the data center, facility, and connection to the utility provider. Because of this broad scope, we engage a wide spectrum of participants—both facilities and operationally focused people as well as computer engineers and architects—which enhances the discussions and sharing that occurs.

Keynote: 30 Year Perspective on HPC and Energy Efficiency
Jeff Broughton (National Energy Research Scientific Computing Center (NERSC), Lawrence Berkeley National Laboratory)

EE HPC WG Morning Break

State of the Working Group
The Power Grid

Novel Cooling Technologies and Experiences

EE HPC WG Lunch Break

Machine Installations - Pre-Exascale and Beyond

EE HPC WG Afternoon Break

Silicon Manufacturing Variability

Operational Data Analytics - Global Survey Results

Panel: Challenges with Manufacturing Variation and Other Challenges to Platform Level Monitoring and Management
Closing Remarks

9:00 am - 5:30 pm

10th Workshop on Latest Advances in Scalable Algorithms for Large-Scale Systems

Session Description: 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 and 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 and AI 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. [https://www.csm.ornl.gov/srt/conferences/Scala/2019](https://www.csm.ornl.gov/srt/conferences/Scala/2019)
overlap, have minimal communication, and have no synchronization points. With the advent of Big Data and AI 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 1: The Extreme-Scale Scientific Software Stack and Its Promise for the Exascale Computing Era**

*Michael Heroux (Sandia National Laboratories)*

Open source, community-developed reusable scientific software represents a large and growing body of capabilities. Linux distributions, vendor software stacks and individual disciplined software product teams provide the scientific computing community with usable holistic software environments containing core open source software components. At the same time, new software capabilities make it into these distributions in a largely ad hoc fashion.

The Extreme-scale Scientific Software Stack (E4S), first announced in November 2018, along with its community-organized scientific software development kits (SSDKs), is a new community effort to create lightweight cross-team coordination of scientific software development, delivery and deployment and a set of support tools and processes targeted at improving scientific software quality via improved practices, policy, testing and coordination.

E4S (https://e4s.io) is an open architecture effort, welcoming teams that are developing technically compatible and high-quality products to participate in the community. E4S and the SSDKs are sponsored by the US Department of Energy Exascale Computing Project (ECP), driven by our need to effectively develop, test, deliver and deploy our open source software products on next generation platform to the scientific community.

In this presentation, we introduce E4S, discuss its design and implementation goals and show examples of success and challenges so far. We will also discuss our connection with other key community efforts we rely upon for our success.
Keynote 2: Toward Scaling Deep Learning to 100,000 Processors - The Fugaku Challenge
Satoshi Matsuoka (RIKEN Center for Computational Science (R-CCS), Tokyo Institute of Technology)
Modern AI with deep learning poses significant overhead in training over very large data sets, whereby the use of HPC techniques to compute in parallel on a large machine is becoming increasingly popular. However, most of the efforts have been on GPUs at relatively low scale, in the order of a few hundreds, up to a thousand except on fairly limited sets of cases, due to inherent difficulties. On Fugaku we plan on extending the capabilities of deep learning by allowing training to be done on the full machine, or more than 100,000 nodes. This requires various technological underpinnings as well as new algorithms for scalable training, the ongoing effort whose current state will be described.

GPU Acceleration of Communication Avoiding Chebyshev Basis Conjugate Gradient Solver for Multiphase CFD Simulations
Yasuhiro Idomura (Japan Atomic Energy Agency)
Iterative methods for solving large linear systems are common parts of computational fluid dynamics (CFD) codes. The Preconditioned Conjugate Gradient (P-CG) method is one of the most widely used iterative methods. However, in the P-CG method, global collective communication is a crucial bottleneck especially on accelerated computing platforms. To resolve this issue, communication avoiding (CA) variants of the P-CG method are becoming increasingly important. In this paper, the P-CG and Preconditioned Chebyshev Basis CA CG (P-CBCG) solvers in the multiphase CFD code JUPITER are ported on the latest V100 GPUs. All GPU kernels are highly optimized to achieve about 90% of the roofline performance, the block Jacobi preconditioner is re-designed to extract high computing power of GPUs, and the remaining bottleneck of halo data communication is avoided by overlapping communication and computation. The overall performance of the P-CG and P-CBCG solvers is determined by the competition between the CA properties of the global collective communication and the halo data communication, indicating an importance of the inter-node interconnect bandwidth per GPU. The developed GPU solvers are accelerated up to 2x compared with the former CPU solvers on KNLs, and excellent strong scaling is achieved up to 7,680 GPUs on the Summit.
Volta and AMD Instinct GPUs
Mohammad Zubair (Old Dominion University), James Warner (NASA)

The Scalable Implementation of Finite Elements by NASA (ScIFEN) is a software package developed to solve complex computational materials and structures problems using the finite element method (FEM). In this paper, we describe optimization techniques to speed up the linear solver computation that occurs within the ScIFEN application. We consider GPUs from two different vendors, NVIDIA and AMD as our target platforms for optimization and highlight differences in performance and optimization techniques. The NVIDIA GPU Volta V100 is used in the Summit system deployed at Oak Ridge National Laboratory, and the new exascale system, Frontier, will be using AMD Radeon Instinct GPU. We evaluated the performance of various optimization techniques on test matrices, ranging in size from 100K to 4M, that are representative of ScIFEN applications. The linear solver computation is memory-bound on both GPUs. Our experiments show that on the NVIDIA GPU we obtained up to 79% of the theoretical peak bandwidth, while the AMD GPU achieved 59%. Overall, the NVIDIA V100 GPU outperforms the AMD MI 25 GPU. We observed an overall speedup of up to 37X on an NVIDIA V100 compared to an Intel Skylake 12-core machine. The solver for a 4M degree of freedom system took under 2.5 seconds.

Toward Half-Precision Computation for Complex Matrices: A Case Study for Mixed Precision Solvers on GPUs
Ahmad Abdelfattah (University of Tennessee, Innovative Computing Laboratory)

Low-precision computations are popular in machine learning and artificial intelligence (AI) applications. Hardware architectures, such as high-end GPUs, now support native 16-bit floating point arithmetic (i.e. half-precision). While half-precision provides a natural 2x/4x speedups against the performance of single/double precisions, modern GPUs are equipped with hardware accelerators for even more FP16 performance. These accelerators, which are called tensor cores, have a theoretical peak performance that is 8x/16x faster than FP32/FP64 performance, respectively. Such a high level of performance has encouraged researchers to harness the compute power of the tensor cores outside AI applications.

This paper presents a mixed-precision dense linear solver (Ax = b) for complex matrices using the tensor core units of the GPU. Unlike similar efforts that have discussed accelerating Ax=b using real FP16 arithmetic, this paper focuses on complex precisions. The developed solution uses a “half-complex” precision to accelerate the solution of Ax=b while maintaining single-complex precision accuracy. The proposed solver requires a matrix multiplication kernel that can accept half-complex inputs. We discuss two possible designs for such a kernel, and integrate both of them into a mixed-precision LU factorization. The other component of our solution is an iterative refinement solver, which recovers the single-complex accuracy using a preconditioned GMRES solver. Our
experiments, which are conducted on a V100 GPU, show that the mixed-precision solver can be up to 2.5x faster than a full single-complex precision solver.

Lunch Break

Keynote 3: Exascale Application Progress and Challenges
Douglas B. (Doug) Kothe (Oak Ridge National Laboratory)
For exascale applications under development in the U.S Department of Energy (DOE) Exascale Computing Project (ECP), they are charged with delivering comprehensive science-based computational applications that effectively exploit exascale HPC technologies to provide breakthrough modeling and simulation and data science solutions. These solutions must yield high-confidence insights and answers to our Nation’s most critical problems and challenges in scientific discovery, energy assurance, economic competitiveness, health enhancement, and national security.

Exascale applications (and their companion co-designed computational motifs) are a foundational element of the ECP and are the vehicle for delivery of consequential solutions and insight from exascale systems. The breadth of these applications runs the gamut: chemistry and materials; energy production and transmission; earth and space science; data analytics and optimization; and national security. Each ECP application is focused on targeted development to address a unique mission challenge problem, i.e., one that possesses solution amenable to simulation insight, represents a strategic problem important to a DOE mission program, and is currently intractable without the computational power of exascale. Any tangible progress requires close coordination between exascale application and software ecosystem development to adequately address the application development challenges currently encountered in the ECP: flat performance profiles; strong scaling; understanding accelerator performance; choice of programming model; selecting math models to fit architectures; and software dependencies.

Each ECP application possesses a unique development plan base on its requirements-based combination of physical model enhancements and additions, algorithm innovations and improvements, and software architecture design and implementation. Illustrative examples of these development activities will be given along with results achieved to date on existing DOE supercomputers such as the Summit system at Oak Ridge National Laboratory.

Afternoon Break
Extreme Scale Phase-Field Simulation of Sintering Processes
Johannes Hötzer (University of Applied Science Karlsruhe)

The sintering process, which turns loose powders into dense materials, is naturally found in the formation of glaciers, but is also the indispensable process to manufacture ceramic materials. The dynamically evolving microstructure, which is established during this process, largely influences the resulting material properties. To investigate this complex three-dimensional, scale-bridging evolution in realistic domain sizes, a highly optimized and parallelized multiphysics phase-field solver is developed. The solver is optimized in a holistic way, from the application level over the time integration and parallelization, down to the hardware. Optimizations include communication hiding, explicit vectorization, implicit schemes and local reduction of degrees of freedom. With this, we are able to investigate large-scale, three-dimensional domains and long integration times. We have achieved a single-core peak performance of 32.5%, scaled up to 172032 cores on Hazel Hen, and simulated a multi-million particle system.

Generic Matrix Multiplication for Multi-GPU Accelerated Distributed-Memory Platforms Over PaRSEC
Thomas Herault (University of Tennessee, Innovative Computing Laboratory)

This paper introduces a generic and flexible matrix-matrix multiplication algorithm $C = A \times B$ for state-of-the-art computing platforms. Typically, these platforms are distributed-memory machines whose nodes are equipped with several accelerators (e.g., 6 GPUs per node for Summit. To the best of our knowledge, SLATE is the only library that provides a publicly available implementation on such platforms, and it is currently limited to problem instances where the $C$ matrix can entirely fit in the memory of the GPU accelerators. Our algorithm relies on the classical tile-based outer-product algorithm, but enhances it with several control dependences to increase data re-use and to optimize communication flow from/to the accelerators within each node. The algorithm is written within the Parsec runtime system, which allows for a fast and generic implementation, while achieving close-to-peak performance for a large variety of situations.

Toward Accelerated Unstructured Mesh Particle-in-Cell
Gerrett Diamond (Rensselaer Polytechnic Institute (RPI))

The effective execution of unstructured mesh based particle-in-cell, PIC, simulations on GPUs requires careful design and implementation choices to ensure performance while maintaining productive programmability. This paper overviews the developing PUMIPic library that employs a
set of mesh centric data structures and algorithms upon which unstructured mesh PIC simulation codes can be developed. Current performance results for component tests on up to 96 GPUs of the Summit system at Oak Ridge National Laboratory are presented.

**Parallel Multigrid Methods on Manycore Clusters with IHK/McKernel**

Kengo Nakajima (University of Tokyo, RIKEN Center for Computational Science (R-CCS))

The parallel multigrid method is expected to play an important role in large-scale scientific computing on exa-scale supercomputer systems. Previously we proposed Hierarchical Coarse Grid Aggregation (hCGA), which dramatically improved the performance of the parallel multigrid solver when the number of MPI processes was $O(10^4)$ or more. Because hCGA can handle only two layers of parallel hierarchical levels, the computation overhead due to coarse grid solver may become significant when the number of MPI processes reaches $O(10^5)$- $O(10^6)$ or more. In the present work, we propose AM-hCGA (Adaptive Multilevel hCGA) that can take into account multiple layers of three or more levels, and show preliminary results using the Oakforest-PACS (OFP) system by JCAHPC. Additionally, we also examine the impact of a lightweight multi-kernel operating system, called IHK/McKernel, for parallel multigrid solvers running on OFP.

**Making Speculative Scheduling Robust to Incomplete Data**

Ana Gainaru (Vanderbilt University)

We study in this work the robustness of Speculative Scheduling to the incompleteness of data. Speculative scheduling has been introduce as a solution to incorporate future types of applications into the design of HPC schedulers, specifically applications whose runtime is not perfectly known but can be modeled with probability distributions. Preliminary studies show the importance of speculative scheduling when dealing with stochastic applications when the application runtime model is completely known. In this work we show how one can extract even from incomplete data on the behavior of HPC applications enough information so that speculative scheduling performs well.

Specifically, we show that for synthetic runtimes who follow usual probability distributions such as truncated normal distribution, we can extract enough data from as little as 10 previous runs, to be within 5% of the solution which has all the exact information. For real traces of applications, the performance with 10 data points varies with the applications (within 20% of the full-knowledge solution), but converges fast (5% with 100 previous samples).

Finally a side effect of this study is to show the importance of the theoretical results obtained on continuous probability distributions for speculative scheduling. Indeed, we observe that the solutions for such distributions are more robust to incomplete data than the solutions for discrete
Parallel SFC-Based Mesh Partitioning and Load Balancing
Ricard Borrell (Barcelona Supercomputing Center)

Modern supercomputers allow the simulation of complex phenomena with increased accuracy. Eventually, this requires finer geometric discretizations with larger numbers of mesh elements. In this context, and extrapolating to the Exascale paradigm, meshing operations such as generation, adaptation or partition, become a critical issue within the simulation workflow. In this paper, we focus on mesh partitioning. In particular, we present some improvements carried out on an in-house parallel mesh partitioner based on the Hilbert Space-Filling Curve.

Additionally, taking advantage of its performance, we present the application of the SFC-based partitioning for dynamic load balancing. This method is based on the direct monitoring of the imbalance at runtime and the subsequent re-partitioning of the mesh. The target weights for the optimized partitions are evaluated using a least-squares approximation considering all measurements from previous iterations. In this way, the final partition corresponds to the average performance of the computing devices engaged. Results of this strategy are presented for both homogeneous and heterogeneous executions, including CPU and GPU accelerators.

9:00 am - 5:30 pm

Sixth Workshop on Accelerator Programming Using Directives (WACCPD)

Session Description: The ever-increasing heterogeneity in supercomputing applications has given rise to complex compute node architectures offering multiple, heterogeneous levels of massive parallelism. Exploiting the maximum available parallelism out of such systems necessitates sophisticated programming approaches that can provide scalable as well as portable solutions without compromising on performance. Software abstraction-based programming models, such as OpenMP and OpenACC, have been raising the abstraction of code to reduce the burden on the programmer while improving productivity. Recent architectural trends indicate a heavy reliance of future exascale machines on accelerators for performance. Toward this end, the workshop will highlight the improvements over state-of-art through the accepted papers and prompt discussion through keynotes and panel. The workshop aims to showcase all aspects of heterogeneous systems discussing innovative high-level language features, lessons learned while using directives to migrate scientific legacy code to parallel processors, compilation and runtime scheduling techniques among others.  
https://waccpd.org/
The ever-increasing heterogeneity in supercomputing applications has given rise to complex compute node architectures offering multiple, heterogeneous levels of massive parallelism. Exploiting the maximum available parallelism out of such systems necessitates sophisticated programming approaches that can provide scalable as well as portable solutions without compromising on performance. Software abstraction-based programming models, such as OpenMP and OpenACC, have been raising the abstraction of code to reduce the burden on the programmer while improving productivity.

Recent architectural trends indicate a heavy reliance of future exascale machines on accelerators for performance. Toward this end, the workshop will highlight the improvements over state-of-art through the accepted papers and prompt discussion through keynotes and panel. The workshop aims to showcase all aspects of heterogeneous systems discussing innovative high-level language features, lessons learned while using directives to migrate scientific legacy code to parallel processors, compilation and runtime scheduling techniques among others.

**WACCPD Opening Remarks**
Sandra Wienke (RWTH Aachen University), Sridutt Bhalachandra (Lawrence Berkeley National Laboratory)
Welcome, agenda and statistics

**Keynote: Perlmutter - A 2020 Pre-Exascale GPU-Accelerated System for NERSC: Architecture and Application Performance Optimization**
Nicholas James Wright (National Energy Research Scientific Computing Center (NERSC))
In 2020, NERSC will take delivery of its next-generation supercomputer, Perlmutter. In this talk, we will describe the architecture of the machine and how it was optimized to meet the performance and usability goals of NERSC’s more than 7000 users. We will discuss the current usage of different programming models at NERSC and our plans for supporting them on Perlmutter, and on future machines.

**WACCPD Morning Break**
GPU Implementation of a Sophisticated Implicit Low-Order Finite Element Solver with FP21-32-64 Computation Using OpenACC

Takuma Yamaguchi (University of Tokyo), Kohei Fujita (University of Tokyo, RIKEN Center for Computational Science (R-CCS))

Accelerating applications with portability and maintainability is one of the big challenges in science and engineering. Previously, we have developed a fast implicit low-order three-dimensional finite element solver, which has a complicated algorithm including artificial intelligence and transprecision computing. In addition, all possible tunings for the target architecture were implemented; accordingly, the solver has inferior portability and maintainability.

In this paper, we apply OpenACC to the solver. The directive-based implementation of OpenACC enables GPU computation to be introduced with a smaller developmental cost even for complex codes. In performance measurements on AI Bridging Cloud Infrastructure (ABCI), we evaluated that a reasonable speedup was attained on GPUs, given that the elapsed time of the entire solver was reduced to 1/14 of that on CPUs based on the original CPU implementation. Our proposed template to use transprecision computing with our custom FP21 data type is available to the public; therefore, it can provide a successful example for other scientific computing applications.

Acceleration in Acoustic Wave Propagation Modeling Using OpenACC/OpenMP and Its Hybrid for the Global Monitoring System

Noriyuki Kushida (Comprehensive Nuclear-Test-Ban Treaty Organization)

CTBTO is operating and maintaining the international monitoring system of Seismic, Infrasound, Hydroacoustic and Airborne radionuclide to detect a nuclear explosion over the globe. The monitoring network of CTBTO, especially with regard to infrasound and hydroacoustic, is quite unique because the network covers the globe, and the data is opened to scientific use. CTBTO has been developing and improving the methodologies to analyze observed signals intensively. In this context, hydroacoustic modelling software, especially which that solves the partial differential equation directly, is of interest. As seen in the analysis of the Argentinian submarine accident, the horizontal reflection can play an important role in identifying the location of an underwater event, and as such, accurate modelling software may help analysts find relevant waves efficiently. Thus, CTBTO has been testing a parabolic equation based model (3D-SSFPE) and building a finite difference time domain (FDTD) model. At the same time, using such accurate models require larger computer resources than simplified methods such as ray-tracing. Thus we accelerated them using OpenMP and OpenACC, or the hybrid of those. As a result, in the best case scenarios, (1) 3D-SSFPE was accelerated by approximately 19 times to the original Octave code, employing the GPU-enabled Octfile technology, and (2) FDTD was accelerated by approximately 160 times to the original Fortran code using the OpenMP/OpenACC hybrid technology, on our DGX—Station with
V100 GPUs.

**Accelerating the Performance of Modal Aerosol Module of E3SM Using OpenACC**
Zhengji Zhao (Lawrence Berkeley National Laboratory)

Using GPUs to accelerate the performance of HPC applications has recently gained great momentum. Energy Exascale Earth System Model (E3SM) is a state-of-the-science earth system model development and simulation project and has gained national recognition. It has a large code base with over a million lines of code. How to make effective use of GPUs remains a challenge. In this paper, we use the modal aerosol module (MAM) of E3SM as a driving example to investigate how to effectively offload computational tasks to GPUs, using the OpenACC directives. In particular, we are interested in the performance advantage of using GPUs and understanding the limiting factors from both the application characteristics and the GPU or OpenACC sides.

**Evaluation of Directive-Based GPU Programming Models on a Block Eigensolver with Consideration of Large Sparse Matrices**
Fazlay Rabbi (Michigan State University)

Achieving high performance and performance portability for large-scale scientific applications is a major challenge on heterogeneous computing systems such as many-core CPUs and accelerators like GPUs. In this work, we implement a widely used block eigensolver, Locally Optimal Block Preconditioned Conjugate Gradient (LOBPCG), using two popular directive based programming models (OpenMP and OpenACC) for GPU-accelerated systems. Our work differs from existing work in that it adopts a holistic approach that optimizes the full solver performance rather than narrowing the problem into small kernels (e.g., SpMM, SpMV). Our LOPBCG GPU implementation achieves a 2.8x - 4.3x speedup over an optimized CPU implementation when tested with four different input matrices. The evaluated configuration compared one Skylake CPU to one Skylake CPU and one NVIDIA V100 GPU. Our OpenMP and OpenACC LOBPCG GPU implementations gave nearly identical performance. We also consider how to create an efficient LOBPCG solver that can solve problems larger than GPU memory capacity. To this end, we create microbenchmarks representing the two dominant kernels (inner product and SpMM kernel) in LOBPCG and then evaluate performance when using two different programming approaches: tiling the kernels, and using Unified Memory with the original kernels. Our tiled SpMM implementation achieves a 2.9x and 48.2x speedup over the Unified Memory implementation on supercomputers with PCIe Gen3 and NVLink 2.0 CPU to GPU interconnects, respectively.

**WACCPD Lunch Break**
Invited Talk: The SPEC ACCEL Benchmark – Results and Lessons Learned
Robert Henschel (Indiana University)
The High-Performance Group (HPG) of the Standard Performance Evaluation Corporation (SPEC) is a forum for discussing and developing benchmark methodologies for High-Performance Computing (HPC) systems. The group released the SPEC ACCEL benchmark in 2014, containing OpenCL and OpenACC components. In 2017, an OpenMP 4.5 target offload component was added by porting the OpenACC applications to OpenMP 4.5. This talk will introduce the benchmark, show results and talk about the lessons learned from developing and maintaining this directive based benchmark. In addition, current challenges of creating a follow on suite are discussed.

Performance of the RI-MP2 Fortran Kernel of GAMESS on GPUs via Directive-Based Offloading with Math Libraries
JaeHyuk Kwack (Argonne National Laboratory)
The US Department of Energy (DOE) started operating two GPU-based pre-exascale supercomputers in 2018 and plans to deploy another pre-exascale in 2020, and three exascale supercomputers in 2021/2022. All of the systems are GPU-enabled systems, and they plan to provide optimized vendor-promoted programming models for their GPUs such as CUDA, HIP and SYCL. However, due to their limited functional portability, it is challenging for HPC application developers to maintain their applications in an efficient and effective way with good productivity across all US DOE pre-exascale/exascale systems. Directive-based programming models for accelerators can be one of the solutions for HPC applications on the DOE supercomputers. In this study, we employ OpenMP and OpenACC offloading models to port and re-implement the RI-MP2 Fortran kernel of the GAMESS application on a pre-exascale GPU system, Summit. We compare and evaluate the performance of the re-structured offloading kernels with the original OpenMP threading kernel. We also evaluate the performance of multiple math libraries on the Nvidia V100 GPU in the RI-MP2 kernel. Using the optimized directive-based offloading implementations, the RI-MP2 kernel on a single V100 GPU becomes more than 7 times faster than on dual-socket Power9 processors, which is near the theoretical speed-up based on peak performance ratios. MPI + directive-based offloading implementations of the RI-MP2 kernel perform more than 40 times faster than a MPI + OpenMP threading implementation on the same number of Summit nodes. This study demonstrates how directive-based offloading implementations can perform near what we expect based on machine peak ratios.

WACCPD Afternoon Break
Performance Portable Implementation of a Kinetic Plasma Simulation Mini-App
Yuuichi Asahi (National Institutes for Quantum and Radiological Science and Technology, Japan)

Performance portability is considered to be an inevitable requirement in the exascale era. We explore a performance portable approach for fusion plasma turbulence simulation code employing kinetic model, namely GYSELA code. For this purpose, we extract the key features of GYSELA such as high dimensionality and Semi-Lagrangian scheme, and encapsulate them into a mini-application which solves the similar but simplified Vlasov-Poisson system. We implement the mini-app with a mixed OpenACC/OpenMP and Kokkos implementation, where we suppress unnecessary duplications of code lines. For a reference case with the problem size of 128 to the 4, the Skylake (Kokkos), Nvidia Tesla P100 (OpenACC), and P100 (Kokkos) versions achieve an acceleration of 1.45, 12.95, and 17.83, respectively, with respect to the baseline OpenMP version on Intel Skylake. In addition to the performance portability, we discuss the code readability and productivity of each implementation. Based on our experience, Kokkos can offer a readable and productive code at the cost of initial porting efforts, which would be enormous for a large scale simulation code like GYSELA.

A Portable SIMD Primitive Using Kokkos for Heterogeneous Architectures
Damodar Sahasrabudhe (University of Utah, Scientific Computing and Imaging Institute)

As computer architectures are rapidly evolving (e.g. those designed for exascale), multiple portability frameworks have been developed to avoid new architecture-specific development and tuning. However, portability frameworks depend on compilers for auto-vectorization and may lack support for explicit vectorization on heterogeneous platforms. Alternatively, programmers can use intrinsics-based primitives to achieve more efficient vectorization, but the lack of a gpu back-end for these primitives makes such code non-portable. A unified, portable, Single Instruction Multiple Data (SIMD) primitive proposed in this work, allows intrinsics-based vectorization on cpus and many-core architectures such as Intel Knights Landing (KNL), and also facilitates Single Instruction Multiple Threads (SIMT) based execution on gpus. This unified primitive, coupled with the Kokkos portability ecosystem, makes it possible to develop explicitly vectorized code, which is portable across heterogeneous platforms. The new SIMD primitive is used on different architectures to test the performance boost against hard-to-auto-vectorize baseline, to measure the overhead against efficiently vectroized baseline, and to evaluate the new feature called the "logical vector length" (LVL). The SIMD primitive provides portability across cpus and gpus without any performance degradation being observed experimentally.
WACCPD Best Paper Award
Sandra Wienke (RWTH Aachen University), Sridutt Bhalachandra (Lawrence Berkeley National Laboratory)
From all accepted submissions that included reproducibility information, the WACCPD chairs award the best paper.

Panel: Convergence, Divergence, or New Approaches? - The Future of Software-Based Abstractions for Heterogeneous Supercomputing
Jeff R. Hammond (Intel Corporation), Jack Deslippe (Lawrence Berkeley National Laboratory), Christian Robert Trott (Sandia National Laboratories), Michael Wolfe (Nvidia Corporation), Johannes Doerfert (Argonne National Laboratory), Fernanda Foertter (Nvidia Corporation)
With an ongoing shift towards accelerator-based architectures given the promise of improved performance per watt, the complexity of parallelizing and tuning applications is also increasing. Today's high-performance computing (HPC) developers face a serious dilemma to decide on processor architecture and parallel programming paradigm.

Software-based abstractions for accelerators give hope to lift this burden but also leave the developers spoilt for choice - from open standards over Domain-specific languages (DSLs) to proprietary approaches, and from open-source to closed-source solutions. The uncertainty surrounding the usability, portability, and interoperability of these abstractions in the future makes it unsettling for the developers. It is but imperative that we try and resolve these shortcomings for their greater adoption.

This panel hopes to work towards this resolution by bringing together standards committee, supporters, and users of these abstractions, e.g., from OpenACC, OpenMP, Kokkos or SYCL. The panelists will share their insights on the future of these abstractions - Will they converge, diverge, or will there be new approaches that would be needed? What makes a good accelerator programming model? Is there a measure for this "goodness"? The audience is also encouraged to challenge the panelists with their questions or share their insights.

WACCPD Closing Remarks
Sandra Wienke (RWTH Aachen University), Sridutt Bhalachandra (Lawrence Berkeley National Laboratory)
ISAV 2019: In Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization

Session Description: 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.  

http://dav.lbl.gov/events/ISAV2019/
In-Situ Visualization for the Large Scale Computing Initiative Milestone
Jeffrey Mauldin (Sandia National Laboratories), Thomas Otahal (Sandia National Laboratories)

The Sandia National Laboratories (SNL) Large-Scale Computing Initiative (LSCI) milestone required running two parallel simulation codes at scale on the Trinity supercomputer at Los Alamos National Laboratory (LANL) to obtain presentation quality visualization results via in-situ methods. The two simulation codes used were Sandia Parallel Aerosciences Research Code (SPARC) and Nalu, both fluid dynamics codes developed at SNL. The codes were integrated with the ParaView Catalyst in-situ visualization library via the SNL developed Input Output SubSystem (IOSS). The LSCI milestone had a relatively short time-scale for completion of two months. During setup and execution of in-situ visualization for the milestone, there were several challenging issues in the areas of software builds, parallel startup-times, and in the a priori specification of visualizations. This paper will discuss the milestone activities and technical challenges encountered in its completion.

A New Approach for In Situ Analysis in Fully Coupled Earth System Models
Ufuk turuncoglu (National Center for Atmospheric Research (NCAR))

This work focuses on enabling the use of existing in situ analysis and visualization methods within the model coupling framework, which is specifically designed software layer to combine physical components of Earth system (i.e., atmosphere, ocean) through a standardized calling interface to simplify common tasks such as interpolation among different numerical grids and synchronization of the model components. The new approach enables simultaneous analysis of the vast amount of data produced by multi-component Earth system model in an integrated way and facilitate efficient, interoperable and optimized data analysis and visualization workflows for Earth system science applications. To demonstrate the flexibility and the capability of the new approach, a fully coupled regional atmosphere-ocean modeling application, that aims to investigate air-sea interaction in a very high spatial and temporal resolution have been implemented. In the use case application, both the simulation and in situ visualization application run in parallel on a large-scale HPC platform with both software and hardware rendering support through the use of accelerators.

In Situ Adaptive Timestep Control and Visualization Based on the Spatio-Temporal Variations of the Simulation Results
Naohisa Sakamoto (Kobe University, RIKEN Center for Computational Science (R-CCS))

Effective visualization of time-varying volumetric data is considered challenging, since in addition to the traditional visualization parameter selections, there is also the need to take care about the dynamic changings between the timesteps. In situ adaptive sampling control based on the amount of change between the simulation timesteps might be helpful for the I/O and visualization cost savings. In this paper, we propose the use of Kernel Density Estimation (KDE) and Kullback-Leibler
(KL) divergence, in order to estimate the amount of internal variations between the timesteps, and use this parameter for controlling the sampling rate of the timesteps. Needless to say that these selected timesteps should be sufficient to understand the underlying phenomena without missing important data features. We confirmed the effectiveness on reducing the number of timepsteps, by using a CFD simulation of irregular volume data by using OpenFOAM. However, we could also observe that some parameter selection highly influences the visualization smoothness, and this remains as a future work.

ISAV 2019 Morning Break

The Impact of Work Distribution on In Situ Visualization: A Case Study
Tobias Rau (University of Stuttgart)

Large-scale computer simulations generate data at rates that necessitate visual analysis tools to run in situ. The distribution of work on and across nodes of a supercomputer is crucial to utilize compute resources as efficiently as possible. In this paper, we study two work distribution problems in the context of in situ visualization and jointly assess the performance impact of different variants. First, especially for simulations involving heterogeneous loads across their domain, dynamic load balancing can significantly reduce simulation run times. However, the adjustment of the domain partitioning associated with this also has a direct impact on visualization performance. The exact impact of this is side effect is largely unclear a priori as generally different criteria are used for balancing simulation and visualization load. Second, on node level, the adequate allocation of threads to simulation or visualization tasks minimizes the performance drain of the simulation while also enabling timely visualization results. In our case study, we jointly study both work distribution aspects with the visualization framework MegaMol coupled in situ on node level to the molecular dynamics simulation ls1 Mardyn on Stampede2 at TACC.

The Challenges of Elastic In Situ Analysis and Visualization
Matthieu Dorier (Argonne National Laboratory)

In situ analysis and visualization have been proposed in high-performance computing (HPC) to enable executing analysis tasks while a simulation is running, bypassing the parallel file system and avoiding the need for storing massive amounts of data. One aspect of in situ analysis that has not been extensively researched to date, however, is elasticity. Current in situ analysis frameworks use a fixed amount of resources and can hardly be scaled up or down dynamically throughout the simulation’s run time as a response to changes in the requirements.
In this paper, we present the challenges posed by elastic in situ analysis and visualization. We emphasize that elasticity can take various forms. We show the difficulties of supporting each form of elasticity with the state-of-the-art HPC technologies, and we suggest solutions to overcome these difficulties. The resulting four-way classification can be seen as a taxonomy for future elastic in situ analysis and visualization systems.

**In Situ Particle Advection Via Parallelizing Over Particles**
Roba Binyahib (University of Oregon)

We extend the method for particle advection that parallelizes over particles to work in an in situ setting. We then compare our method with the typical method for in situ, parallelizing over data. Our experiments consist of parallelism at 512 cores, a data set with 67 million cells, and ten billion total advection steps. Our findings show that parallelizing over particles can be more than ten times faster for some workloads, for reasonable memory cost. Overall, the significance of these findings is to demonstrate that moving data can be worthwhile in some in situ settings.

**Spack Meets Singularity: Creating Movable In-Situ Analysis Stacks with Ease**
Sergei Shudler (Argonne National Laboratory), Silvio Rizzi (Argonne National Laboratory)

In-situ data analysis and visualization is a promising technique to handle the enormous amount of data an extreme-scale application produces. One challenge users often face in adopting in-situ techniques is setting the right environment on a target machine. Platforms such as SENSEI require complex software stacks that consist of various analysis packages and visualization applications. The user has to make sure all these prerequisites exist on the target machine, which often involves compiling and setting them up from scratch. In this paper, we leverage the containers technology (e.g., light-weight virtualization images) and provide users with Singularity containers that encapsulate ready-to-use, movable in-situ software stacks. Moreover, we make use of Spack to ease the process of creating these containers. Finally, we evaluate this solution by running in-situ analysis from within a container on an HPC system.

**HDF5 as a Vehicle for In Transit Data Movement**
Junmin Gu (Lawrence Berkeley National Laboratory), Kesheng Wu (Lawrence Berkeley National Laboratory), Burlen Loring (Lawrence Berkeley National Laboratory)

For in transit processing, one of the fundamental challenges is the efficient movement of data from producers to consumers. We investigate the performance characteristics of an in transit data transport mechanism that leverages the HDF5 parallel I/O library and is accessible as one of several different in transit transport mechanisms in the SENSEI generic in situ framework. For in transit use
cases at scale on HPC platforms, one might expect that an in transit data transport mechanism that uses faster layers of the storage hierarchy, such as DRAM memory, would always outperform a transport that uses slower layers of the storage hierarchy, such as an NVRAM-based persistent storage presented as a distributed file system. However, our results show that the performance of the transport using NVRAM is competitive with the transport that uses socket-based data movement across varying levels of producer and consumer concurrency.

In Situ and In Transit Visualization for Numerical Simulations in HPC
Seiji Tsutsumi (Japan Aerospace Exploration Agency (JAXA))

To overcome the difficulty in visualizing large-scale simulation results in HPC, we implemented in situ and in transit visualization using VisIt/Libsim and ADIOS2 into an in-house application for computational fluid dynamics, and evaluated both forms on a heterogeneous HPC system of the Japan Aerospace Exploration Agency. It is found that both frameworks can be used for batch and interactive visualization commonly conducted by researchers and engineers. The in situ approachcosts less to implement in applications and is easier to use. Conversely, the in transit approach is advantageous in terms of lower execution overhead and its flexibility for using separate resources in a heterogeneous HPC system. It is therefore desirable to use both approaches as needed. Based on the knowledge obtained herein, this paper summarizes suggestions for research and development of the in situ and in transit visualization frameworks, the HPC system, and queue policy for interactive jobs, in order to realize in situ and in transit visualization in a heterogeneous HPC system.

Metadata Enabled Optimized Partitioning for M to N In Transit Processing with SENSEI
Burlen Loring (Lawrence Berkeley National Laboratory)

In an M to N in transit setting, a simulation runs on M MPI ranks while an endpoint, such as analysis or visualization, runs on N MPI ranks in a separate application. Before data is moved or processed, a partitioning step occurs where decisions are made about how data is mapped from the simulation’s M MPI ranks onto the end-point’s N MPI ranks. We investigate metadata enabled, optimized partitioning for M to N in transit processing. We extended the SENSEI generic in situ infrastructure for M to N in transit processing. The design incorporates a partitioner object and API that is used to compute the mapping of simulation data onto end-point resources, and a metadata object that comprehensively yet compactly describes data and its mapping onto MPI ranks. In a scaling study using up to 16384 cores on NERSC’s Cori system, we compare optimized and unoptimized partitioning approaches for two common analysis operations, slicing and isosurfacing, and vary N, the level of end-point concurrency, and show optimal ratios of M to N for both types of end-point operations and partitioners. We find that metadata enabled partitioning optimizations significantly reduce time to solution, amount of data moved, memory footprint, and cost of solution in node
hours, especially when running in the regime where $N << M$.

**In Situ Visualization and Analysis Workflows for WarpX**  
Lixin Ge (SLAC National Accelerator Laboratory)

This paper summarizes ongoing work on in situ visualization and analysis workflows for the particle-in-cell code WarpX, using the in situ tools SENSEI and Ascent. We report on preliminary tests performed on the DOE NERSC Cori Haswell system. We also report on the progress to generate “compiled” visualizations represented as an image database (e.g. Cinema databases) to support interactive post-processing of large datasets. An example of using SENSEI with ParaView Catalyst backend and VisIt Libsim backend is presented. Other backends for SENSEI on NERSC’s Cori system are under evaluation. A detailed description of installation and execution processes with WarpX using in situ visualization (via SENSEI or Ascent) is given in the WarpX online documentation.

**In Transit Management and Creation of Data Extracts**  
Brad Whitlock (Intelligent Light)

Scoreboard brings together in transit, extract management and creation, and interactive computational steering into a single system. Scoreboard is flexible and supports both in situ and in transit where the method is selected at run time. Scoreboard can be built with minimal dependencies and produce data extracts specified by an input file. The set of data extracts can be interactively managed from a dynamically generated user interface that also supports simulation steering during in transit.

**Lightning Talks Q&A**  
Wes Bethel (Lawrence Berkeley National Laboratory)

**ISAV 2019 Lunch Break**

**Keynote Talk: In-Situ Analysis for Extreme-Scale Cosmological Simulations**  
Katrin Heitmann (Argonne National Laboratory)

In the talk, I will discuss our in-situ analysis approach within HACC, the Hardware/Hybrid
Accelerated Cosmology Code. We have carried out several extreme-scale simulations on Summit and Mira recently employing our "on the fly" analysis tool suite. We have identified certain bottlenecks and limitations of the approach and have started developing solutions to prepare for the arrival of the first exascale machines in 2021.

**ISAV 2019 Afternoon Break**

**Interactive In Situ Visualization and Analysis Using Ascent and Jupyter**

*Cyrus Harrison (Lawrence Livermore National Laboratory)*

There is a gap in interactive in situ solutions for HPC simulations. While we have access to fully featured visualization UIs through tools like ParaView Catalyst and VisIt LibSim, we lack in situ infrastructure to access more general interactive environments like Jupyter. The Jupyter ecosystem of tools provides a rich paradigm for interactive data analysis and is well suited to help expand interactive use of in situ. The complexity of the Jupyter ecosystem and HPC center security requirements pose challenges to develop software infrastructure that enables direct use of Jupyter in simulation codes.

With this work, we describe a system that enables simulations instrumented with Ascent to connect to Jupyter. This system allows simulation users to interact with their data in situ using Jupyter Notebooks. The system combines Ascent’s embedded Python filter infrastructure with a Client/Server Jupyter Bridge Kernel design that simplifies both deployment and security considerations on HPC systems. We describe the design of this system, demonstrate basic usage, and describe a prototype Ascent rendering UI built on top of this system.

**Invited Presentation: Enabling Scientific Discovery from Diverse Data Sources through In Situ Data Management**

*Tom Peterka (Argonne National Laboratory)*

The successful development of in situ data management capabilities can potentially benefit real-time decision making, design optimization, and data-driven scientific discovery. There are two primary motivations for processing and managing data in situ. The first is the need to decrease data volume. The second motivation is that the in situ methodology can enable scientific discovery from a broad range of data sources, over a wide scale of computing platforms. This talk will highlight the findings of a recent U.S. Department of Energy workshop on in situ data management, outlining six priority research directions. The research directions identify the components and capabilities
needed for in situ data management to be successful for a wide variety of applications: making in situ data management more pervasive, controllable, composable, and transparent, with a focus on greater coordination with the software stack, and a diversity of fundamentally new data algorithms.

The ISAV 2019 UnPanel
Matthew Wolf (Oak Ridge National Laboratory)
The UnPanel session will be returning again this year! As an unscripted, audience-driven conversation, we want to focus this year on capturing the state of the practice from all of the angles and will be producing a community-authored survey paper from the conversation.

Closing Remarks
Kenneth Moreland (Sandia National Laboratories)

9:00 am - 5:30 pm

PHOTONICS: Photonics-Optics Technology Oriented Networking, Information, and Computing Systems

Session Description: Computing systems, from HPC and data center to automobile, aircraft, and cellphone, are integrating growing numbers of processors, accelerators, memories, and peripherals to meet the burgeoning performance requirements of new applications under tight cost, energy, thermal, space, and weight constraints. Recent photonic advances promise great energy efficiency, ultra-high bandwidth, and low latency to alleviate inter-rack/board/chip communication bottlenecks and innovate post-Moore’s Law optical computing. <https://eexu.home.ece.ust.hk/PHOTONICS.html>
Silicon photonics technologies piggyback onto developed silicon fabrication processes to provide viable and cost-effective solutions. Many companies and institutes have been actively developing silicon photonics technologies for more than a decade. A large number of silicon photonics devices and circuits have been demonstrated in CMOS-compatible fabrication processes. Silicon photonics technologies open up new opportunities for applications, systems, architectures, design techniques, and design automation tools to fully explore new approaches and address the challenges of post-Moore’s Law computing systems. PHOTONICS (Photonics-Optics Technology Oriented Networking, Information, and Computing Systems) workshop presents the latest progress and provides insight into the challenges and future developments of this emerging area. Please visit PHOTONICS website (https://eexu.home.ece.ust.hk/PHOTONICS.html) for more information.
[KEYNOTE] Silicon Photonics: Technology Advancement and Applications  
*Jin Hong* (Intel Corporation)  
Keynote by Intel VP of Data Center Group and GM of Silicon Photonics R&D

Coffee Break

[Morning Session] Shrinking Photonic Interconnection Networks from Kilometers to Millimeters

The morning session of PHOTONICS 2019.

**HPC Interconnects at the End of Moore’s Law**  
*John Shalf* (Lawrence Berkeley National Laboratory)  
The tapering of lithography advances that have been associated with Moore’s Law will substantially change requirements for future interconnect architectures for large-scale datacenters and HPC systems. Architectural specialization is creating new datacenter requirements such as emerging accelerator technologies for machine learning workloads and rack disaggregation strategies will push the limits of current interconnect technologies. Whereas photonic technologies are often sold on the basis of higher bandwidth and energy efficiency (e.g. lower picojoules per bit), these emerging workloads and technology trends will shift the emphasis to other metrics such as bandwidth density (as opposed to bandwidth alone) and reduced latency, and performance consistency. Such metrics cannot be accomplished with device improvements alone, but require a systems view of photonics in datacenters.

**Past, Current and Future Technologies for Optical Submarine Cables**  
*Yuichi Nakamura* (NEC Corporation)  
Optical submarine cables are a crucial infrastructure as they convey 99% of internet traffic between countries and continents. Although, their history is rooted in the middle in the 19th Century with the first transatlantic telegraph cable, the uninterrupted strong growth of current internet traffic has
been a prime motivation for the introduction of several disruptive technologies. In this talk, we review technologies used in past and current optical submarine cables, as well as future technological trends to increase their capacity.

**Simulation-Driven Design of Photonic Quantum Communication Networks**  
*Martin Suchara (Argonne National Laboratory)*

This work models metropolitan-scale photonic quantum networks that use time bin encoding for quantum key distribution and quantum state teleportation. We develop and validate theoretical models by comparing them with prior experimental results. We use our newly developed simulator of quantum network communication, called SeQUeNCe, to perform simulations at the individual photon level with picosecond resolution. The simulator integrates accurate models of optical components including light sources, interferometers, detectors, beam splitters, and telecommunication fiber, allowing studies of their complex interactions. Optical quantum networks have been generating significant interest because of their ability to provide secure communication, enable new functionality such as clock synchronization with unprecedented accuracy, and reduce the communication complexity of certain distributed computing problems. In the past few years experimental demonstrations moved from table-top experiments to metropolitan-scale deployments and long-distance repeater network prototypes. As the number of optical components in these experiments increases, simulation tools such as SeQUeNCe will simplify experiment planning and accelerate designs of new network protocols. The modular design of our tool will also allow modeling future technologies such as network nodes with quantum memories and quantum transducers as they become available.

**Overview of Silicon Photonics Components for Commercial DWDM Applications**  
*Ashkan Seyedi (Hewlett Packard Enterprise)*

In this talk, I will present an overview of the work done by Hewlett Packard Labs on high bandwidth, scalable and cost-effective interconnect solution. I will outline recent progress on our opticalBio: Ashkan received a dual Bachelor’s in Electrical and Computer Engineering from the University of Missouri-Columbia and a Ph.D. from University of Southern California working on photonic crystal devices, high-speed nanowire photodetectors, efficient white LEDs and solar cells. While at Hewlett Packard Enterprise as a research scientist, he has been working on developing high-bandwidth, efficient optical interconnects for exascale and high performance computing applications.

**An Approximate Thermal-Aware Q-Routing for Optical NoCs**  
*Yaoyao Ye (Shanghai Jiao Tong University)*
Optical networks-on-chips (NoCs) based on silicon photonics have been proposed as an emerging communication architecture for many-core chip multiprocessors. However, one of the key challenges is the thermal sensitivity of silicon photonic devices under on-chip temperature variations, which would result in significant thermal-induced optical power loss in optical NoCs. In this work, we propose an approximate Q-routing to find optimal low-loss paths in the presence of on-chip temperature variations. With the method of linear function approximation, the proposed approximate Q-routing does not require Q-tables which are necessary in traditional table-based Q-routing. Simulation results of an 8x8 mesh-based optical NoC under a set of synthetic traffic patterns and real applications show that the proposed approximate Q-routing can converge faster and its optimization effect is very close to the best optimization effect of the traditional table-based Q-routing.

On the Feasibility of Optical Circuit Switching for Distributed Deep Learning

Truong Thao Nguyen (National Institute of Advanced Industrial Science and Technology (AIST))

Data parallelism is the dominant method used to training deep learning (DL) model on High-Performance Computing systems such as large-scale GPU clusters. In which, collective communication large message, e.g., up to hundreds of MB, between GPUs becomes one of the major bottlenecks. Especially when training a deep learning model on a large number of node, inter-node communication becomes bottle-neck due to its relatively higher latency and lower link bandwidth (than intra-node communication). To cope with this problem, some techniques have been proposed to (a) optimize the collective communication algorithms that take into account the network topology, (b) reduce the message size, and (c) overlap the communication and computation. All of these approaches target to deal with the large message size issue while diminishing the effect of the limitation of the inter-node network. In this study, we investigate the benefit of increasing inter-node link bandwidth by using the hybrid switching systems, i.e., Electrical Packet Switching and Optical Circuit Switching. We find that the typical data-transfer of synchronous data-parallelism training are long-live and rarely changed that can be speed-up with optical switching. Simulation results on Simgrid simulator show that our approach speed-up the training time of deep learning application.

Lunch Break

[Afternoon Session] Photonic Switching and Computing are Coming
The afternoon session of PHOTONICS 2019.

**High-Radix Sub-Microsecond Photonic Switch on a Chip**  
*Ming C. Wu (University of California, Berkeley)*

The advent of silicon photonic micro-electro-mechanical-system (MEMS) has made it possible to integrate large-scale switches (radix > 100) on single chips of silicon. In this talk, we will review the current state of the art of silicon photonic switches and describe in detail the performance of 240x240 switches.

**Sub-Microsecond Optical Circuit Switched Networks for Data Centers and HPCs**  
*Georgios Zervas (University College London)*

There have been over two decades of research on optical switched technologies for Data Centers and HPCs. However, there is still no commercial uptake and adoption by Cloud and HPC providers. This presentation will delve into why that is and elaborate on the roadblocks, opportunities and choices to address the Moore’s Law slow down and replace electronic switching. It will explain the reasons why optical circuit switching rather than optical packet switching is the best way forward for Data Centers and HPCs and what are the main challenges. PULSE, a scalable network architecture that can reconfigure optical time-slotted circuits in nanoseconds will be introduced to support heterogeneous high performance systems. An insight on technologies we are developing for sub-microsecond network scheduling and sub-nanosecond switching that aim to maximise throughput and minimise latency across thousands of end points (Servers, CPUs, accelerators, memory, etc.) will be provided. Crucially, I’ll present how the network can offer zero data loss, without in-network a) buffering, b) active switching and routing, and c) network header addressing and processing to minimize complexity, and crucially to consume very low power. Furthermore, the system also aims to inherently support 1-to-1, 1-to-N, N-to-N and N-to-1 connectivity in a synchronous manner without the need for data replication for multi/broad-casting, currently not possible. This is key to support diverse workloads such as storage caching, large-scale database lookups, training distributed DNNs, parallel computing that use communication primitives such as allreduce, broadcast and reduce, gather and scatter, all-to-all among others.

**Low-Cost High-Radix Photonic Switch for Rack-Scale Computing Systems**  
*Jiang Xu (Hong Kong University of Science and Technology)*

The demand for more computational power by scientific computing, big data processing, and machine learning is pushing the development of HPC (high-performance computing) systems. As the basic HPC building blocks, modularized server racks with a large number of multicore nodes are facing performance and energy efficiency challenges. This talk introduces RSON, an optical network
for rack-scale computing systems. RSON connects processor cores, caches, local memories, and remote memories through a novel inter/intra-chip silicon photonic network architecture. We develop a low-latency scalable channel partition and low-power dynamic path priority control scheme for RSON. APEX benchmarks show that rack-scale computing systems based on RSON can achieve about one order of magnitude higher energy efficiency comparing to the state-of-the-art systems.

Coffee Break

Using the Optical Processing Unit for Large-Scale Machine Learning
Laurent Daudet (LightOn)

An Optical Neural Network Architecture Based on Highly Parallelized WDM-Multiplier-Accumulator
Tohru Ishihara (Nagoya University)
Future applications such as anomaly detection in a network and autonomous driving require extremely low, submicrosecond latency processing in pattern classification. Towards the realization of such an ultra-fast inference processing, this paper proposes an optical neural network architecture which can classify anomaly patterns at sub-nanosecond latency. The architecture fully exploits optical parallelism of lights using wavelength division multiplexing (WDM) in vector-matrix multiplication. It also exploits a linear optics with passive nanophotonic devices such as microring resonators, optical combiners, and passive couplers, which make it possible to construct low power and ultra-low latency optical neural networks. Optoelectronic circuit simulation using optical circuit implementation of multi-layer perceptron (MLP) demonstrates sub-nanosecond processing of optical neural network.

Ising Machine: an Approach to Solve Combinational Optimization Problems
Fumiyo Takano (NEC Corporation)
A combinational optimization is one of difficult problems, since almost all optimal solutions belong to NP-hard, and near optimal solution is also taken long time to calculate. Recently, an Ising machine approach is proposed and it could obtain good solution in a short time for various combinational problems instead of conventional heuristic methods. In this talk, hardware and software architectures of Ising machines are presented.
Session Description: Ensuring correctness in high-performance computing (HPC) applications is one of the fundamental challenges that the HPC community faces today. While significant advances in verification, testing, and debugging have been made to isolate software errors (or defects) in the context of non-HPC software, several factors make achieving correctness in HPC applications and systems much more challenging than in general systems software—growing heterogeneity (architectures with CPUs, GPUs, and special purpose accelerators), massive scale computations (very high degree of concurrency), use of combined parallel programing models (e.g., MPI+X), new scalable numerical algorithms (e.g., to leverage reduced precision in floating-point arithmetic), and aggressive compiler optimizations/transformations are some of the challenges that make correctness harder in HPC. As the complexity of future architectures, algorithms, and applications in HPC increases, the ability to fully exploit exascale systems will be limited without correctness. With the continuous use of HPC software to advance scientific and technological capabilities, novel techniques and practical tools for software correctness in HPC are invaluable. The goal of the Correctness Workshop is to bring together researchers and developers to present and discuss novel ideas to address the problem of correctness in HPC. The workshop will feature contributed papers and invited talks in this area. [https://correctness-workshop.github.io/2019/](https://correctness-workshop.github.io/2019/)
algorithms (e.g., to leverage reduced precision in floating-point arithmetic), and aggressive compiler optimizations/transformations are some of the challenges that make correctness harder in HPC. As the complexity of future architectures, algorithms, and applications in HPC increases, the ability to fully exploit exascale systems will be limited without correctness. With the continuous use of HPC software to advance scientific and technological capabilities, novel techniques and practical tools for software correctness in HPC are invaluable. The goal of the Correctness Workshop is to bring together researchers and developers to present and discuss novel ideas to address the problem of correctness in HPC. The workshop will feature contributed papers and invited talks in this area.

**Keynote 1 - Correctness 2019**  
Alex Aiken (Stanford University)

**Correctness 2019 Morning Break**

**Debugging and optimization of HPC programs with the Verrou tool**  
Bruno Lathuilière (EDF Research and Development)

The analysis of Floating-Point-related issues in HPC codes is becoming a topic of major interest: parallel computing and code optimization often break the reproducibility of numerical results across machines, compilers and even executions of the same program. This paper presents how the Verrou tool can help during all stages of the Floating-Point analysis of HPC codes: diagnostic, debugging and optimization. Recent developments of Verrou are presented, along with examples illustrating the interest of these new features for industrial codes such as code aster. More specifically, the Verrou arithmetic back-ends now allow analyzing or emulating mixed-precision programs. Interlibm, an interposition layer for the mathematical library, is introduced to mitigate long-standing issues with algorithms from the libm. Finally, debugging algorithms are extended in order to produce useful information as soon as it is available. All these features are available in released version 2.1.0 and upcoming version 2.2.0.

**Exploring Regression of Data Race Detection Tools Using DataRaceBench**  
Markus Schordan (Lawrence Livermore National Laboratory)

DataRaceBench is an OpenMP benchmark suite designed to systematically and quantitatively evaluate data race detection tools. It has been used by several research and development groups to
measure the quality of their tools. In this paper we explore how to evaluate the regression of data race detection tools in the presence of observed tool errors. We define major factors related to generating consistent, reproducible, and comparable evaluation results and a detailed evaluation process with a set of configuration and execution rules. We also outline differences in the evaluation of dynamic and static data race detection tools. On top of the evaluation results, we further explore and suggest different ways to process and present the data, with focus on how to consider tool errors. Using DataRaceBench we can show a regression for several popular data race detection tools in recent release cycles.

**Toward Multi-Precision, Multi-Format Numerics**
David Thien (University of California, San Diego)

New, domain-specific number systems that accelerate modern workloads. Using these number systems effectively requires analyzing subtle multi-format, multi-precision (MPMF) code. Ideally, recent programming tools that automate numerical analysis tasks could help make MPMF programs both accurate and fast. However, three key challenges must be addressed: existing automated tools are difficult to compose due to subtle incompatibilities; there is no “gold standard” for correct MPMF execution; and no methodology exists to guide generalizing existing, IEEE-754-specialized tools to support MPMF. In this paper we report on recent work towards mitigating these related challenges. First, we extend the FPBench standard to support multi-precision, multi-format (MPMF) applications. Second, we present Titanic, a tool which provides reference results for arbitrary MPMF computations. Third, we describe our experience adapting an existing numerical tool to support MPMF programs.

**Tool Integration for Source-Level Mixed Precision**
Michael O. Lam (James Madison University, Lawrence Livermore National Laboratory)

Mixed-precision computing offers potential data size reduction and performance improvement at the cost of accuracy, a tradeoff that many practitioners in high-performance computing and related fields are becoming more interested in as workloads become increasingly communication-bound. However, it can be difficult to build valid mixed-precision configurations and navigate the performance/accuracy space without the help of automated tools. We present FloatSmith, an open-source, end-to-end source-level mixed-precision tuner that incorporates several software tools (CRAFT, TypeForge, and ADAPT) into an integrated tool chain.

Correctness 2019 Lunch Break
Bounded Verification of Sparse Matrix Computations
Tristan Dyer (North Carolina State University)

We show how to model and reason about the structure and behavior of sparse matrices, which are central to many applications in scientific computation. Our approach is state-based, relying on a formalism called Alloy to show that one model is a refinement of another. We present examples of sparse matrix-vector multiplication, transpose, and translation between formats using ELLPACK and compressed sparse row formats to demonstrate the approach. To model matrix computations in a declarative language like Alloy, a new idiom is presented for bounded iteration with incremental updates. Mechanical verification is performed using SAT solvers built into the tool.

Investigating the Impact of Mixed Precision on Correctness for a Large Climate Code
Daniel J. Milroy (Lawrence Livermore National Laboratory)

Earth system models (ESMs) are computationally expensive and represent many complex processes on a wide range of scales from molecular to global. Certain ESM computations require high-precision while others, such as atmospheric microphysics (e.g., precipitation) which are approximated by bulk properties, should not. As such, atmospheric microphysics models are prime candidates for conversion to single-precision, which afford distinct computational and memory advantages over typical double-precision numbers. However, care must be taken as indiscriminate type casting to single precision can result in numerical instability and divergent output when applied naively. In this work we relate our experiences attempting to improve the performance of the Morrison-Gettelman microphysics package (MG2) in a popular ESM by modifying it to compute in
single precision without sacrificing correctness. We find that modification of the entire MG2 package to compute with single-precision floats achieves a respectable performance increase but does not appear to be correct in terms of maintaining consistency with double-precision MG2. On the other hand, narrowing the scope of our conversion to a couple expensive subprograms yields more satisfying results in terms of correctness but with negligible overall performance improvement. We evaluate correctness with both an objective statistical tool and traditional approaches more familiar to climate scientists. While we are still working toward our ultimate goal of improving the performance of MG2 without negatively affecting model output, we believe that our experiences may be helpful to other groups pursuing similar goals.

2:00 pm - 5:30 pm

1st Annual Workshop on Large-Scale Experiment-in-the-Loop-Computing (XLOOP)

Session Description: Continued advancement in computational power and high-speed networking is enabling a new model of scientific experiment, experiment-in-the-loop computing (EILC). In this model, one or more high-fidelity simulations are run as data is collected from observational and experimental sources. At the same time, the amount and complexity of data generated by simulations and by observational and experimental sources, such as distributed sensor networks and large-scale scientific user facilities, continues to increase. Several research and development challenges are posed by this paradigm, many of which are independent of the scientific application domain. New algorithms, including artificial intelligence and machine learning algorithms, to merge simulation ensembles and experimental data sets must be developed. High performance data management and transfer techniques must be developed to control the ensembles and move simulated and observed data sets. Workflows must be constructed with high-quality provenance metadata to enable post-analysis and improvement of the computing solution. The Workshop on Large-scale Experiment-in-the-Loop Computing (XLOOP 2019) will be a unique opportunity to promote this cross-cutting, interdisciplinary topic area. We invite papers, presentations, and participants from both the physical and computer sciences. https://press3.mcs.anl.gov/xloop2019/
At the same time, the amount and complexity of data generated by simulations and by observational and experimental sources, such as distributed sensor networks and large-scale scientific user facilities, continues to increase. Several research and development challenges are posed by this paradigm, many of which are independent of the scientific application domain. New algorithms, including artificial intelligence and machine learning algorithms, to merge simulation ensembles and experimental data sets must be developed. High performance data management and transfer techniques must be developed to control the ensembles and move simulated and observed data sets. Workflows must be constructed with high-quality provenance metadata to enable post-analysis and improvement of the computing solution. The Workshop on Large-scale Experiment-in-the-Loop Computing (XLOOP 2019) will be a unique opportunity to promote this cross-cutting, interdisciplinary topic area. We invite papers, presentations, and participants from both the physical and computer sciences.

**Computational Strategies to Increase Efficiency of Gaussian-Process-Driven Autonomous Experiments**

Natural sciences are increasingly reliant on large volumes of experimental data obtained from highly-automated scientific equipment to drive scientific discovery. While the instrumentation available to the scientific community enables individual researchers to acquire experimental data at spectacular rates, the decision making and experimental design process of the measurements is still largely driven by the individual researcher operating the instrument. In recent years, some scientific areas, like X-Ray scattering and synchrotron infra-red microscopy, have started to shift towards autonomously executed experimentation, using tools from Artificial Intelligence and Machine Learning. Gaussian Process Regression (GPR) is a popular technique that enables the construction of a surrogate model required to drive an experiment in an autonomous fashion. GPRs have been successfully implemented at synchrotron radiation beam lines at the Advanced Light Source (ALS) and the National Synchrotron Light Source II (NSLS-II). Given that a single measurement on state-of-the-art synchrotron equipment can be acquired in a fraction of a second, being able to obtain feedback at a similar time scale is essential. Traditionally, specific tasks in GPR, such as hyper-parameter tuning and covariance estimation, are compute-intensive, limiting the utility of GPR as an appropriate tool when near real-time feedback is an absolute necessity. In this paper we discuss, present and review computational strategies that allow the numerical acceleration of Gaussian Process analyses within a framework of autonomous sequential experiments. The results show that significant time savings can be achieved by taking advantage of a number of some rather well-established mathematical and computational approaches.

**Scientific Image Restoration Anywhere**
The use of deep learning models within scientific experimental facilities frequently requires low-latency inference, so that, for example, quality control operations can be performed while data are being collected. Edge computing devices can be useful in this context, as their low cost and compact form factor permit them to be co-located with the experimental apparatus. Can such devices, with their limited resources, can perform neural network feed-forward computations efficiently and effectively? We explore this question by evaluating the performance and accuracy of a scientific image restoration model, for which both model input and output are images, on edge computing devices. Specifically, we evaluate deployments of TomoGAN, an image-denoising model based on generative adversarial networks developed for low-dose x-ray imaging, on the Google Edge TPU and NVIDIA Jetson. We adapt TomoGAN for edge execution, evaluate model inference performance, and propose methods to address the accuracy drop caused by model quantization. We show that these edge computing devices can deliver accuracy comparable to that of a full-fledged CPU or GPU model, at speeds that are more than adequate for use in the intended deployments, denoising a 1024x1024 image in less than a second. Our experiments also show that the Edge TPU models can provide 3x faster inference response than a CPU-based model and 1.5x faster than an edge GPU-based model. This combination of high speed and low cost permits image restoration anywhere.

Distributed Global Digital Volume Correlation by Optimal Transport

Because of the speed and data rates of timeresolved experiments at facilities such as synchrotron beamlines, automation is critical during time-resolved experiments. In 3D imaging experiments like microCT (CT), this includes recognizing features of interest and “zooming in” spatially and temporally to those features; ideally without requiring advanced information about which features are being imaged. Digital Volume Correlation (DVC) can achieve this by measuring the deformation field between images, but has not been used during autonomous experiments because of the scalability of the codes. In this work, we propose a model for global DVC and a parallel algorithm for solving it for large-scale images, suitable for giving feedback for autonomous experiments at synchrotron-based microCT beamlines. In particular, we leverage recent advancements in entropy-regularized optimal transport to develop efficient, simple-to-implement, parallel algorithms which scale linearly (O(N)) in space and time, where N is the number of voxels, and well with an increasing number of processors. As a demonstration, we compute the deformation field for every voxel from a CT volume with dimensions 2560x2560x2160. We discuss implementation details, drawbacks and future directions.

XLOOP Afternoon Break
Sparse Data Management in HDF5

The importance of sparse data management is growing with data produced by large-scale experimental and observational facilities that contain small amounts of non-zero values. In this document, we explore different design options to support sparse data in HDF5, one of the most popular high-performance I/O libraries and file formats used for scientific data. We discuss several use cases and requirements. Our main hard design constraint is that any changes to the HDF5 dataset API would be a burden on users and not acceptable. The remaining options are discussed below. We provide a rationale for what we believe is the strongest candidate and describe how its potential benefits can be simulated with the existing HDF5 library. We have conducted a set of computational experiments the results of which are reported here. They show that our candidate meets all relevant requirements and gives us a certain degree of confidence for an HDF5 library-native implementation.

Data Processing at the Linac Coherent Light Source

The increase in velocity, volume, and complexity of the data generated by the upcoming LCLS-II upgrade presents a considerable challenge for data acquisition, data processing, and data management. These systems face formidable challenges due to the extremely high data throughput, hundreds of GB/s to multi-TB/s, generated by the detectors at the experimental facilities and to the intensive computational demand for data processing and scientific interpretation. The LCLS-II Data System offers a fast, powerful, and flexible architecture that includes a feature extraction layer designed to reduce the data volumes by at least one order of magnitude while preserving the science content of the data. Innovative architectures are required to implement this reduction with a configurable approach that can adapt to the multiple science areas served by LCLS. In order to increase the likelihood of experiment success and improve the quality of recorded data, a real-time analysis framework provides visualization and graphically-configurable analysis of a selectable subset of the data on the timescale of seconds. A fast feedback layer offers dedicated processing resources to the running experiment in order to provide experimenters feedback about the quality of acquired data within minutes. We will present an overview of the LCLS-II Data System architecture with an emphasis on the Data Reduction Pipeline (DRP) and
online monitoring framework.

**Balsam: Near Real-time Experimental Data Analysis on Supercomputers**

The growing sophistication of experimental data analysis at DOE scientific user facilities is poised to outstrip local resources for storage and computing. At the same time, experimental facilities generally underuse DOE supercomputers for analysis because 1) logging into a supercomputer to schedule computations complicates the experimental workflow, and 2) scientists must often interpret data on-the-fly and cannot abide by batch queueing policies. In this work, we describe how the Balsam edge service has enabled near-real-time analysis of data collected at the Advanced Photon Source (APS) with an X-ray Photon Correlation Spectroscopy (XPCS) application running on Argonne Leadership Computing Facility (ALCF) resources. Balsam provides a uniform API for remote job submission and can leverage transfer protocols like Globus to orchestrate data movement. The Balsam service then interfaces with the system scheduler to perform automated job submission and process tasks in a high-throughput, fault-tolerant fashion. We highlight the effectiveness of backfill queues when used with Balsam to process XPCS data on idle nodes with relatively low latency.

**XLOOP Panel: Big Science, Big Computing**

**XLOOP Awards and Conclusion**

**2:00 pm - 5:30 pm**

**Containers and New Orchestration Paradigms for Isolated Environments in HPC (CANOPIE HPC)**

**Session Description:** We are pleased to introduce the 1st International Workshop on Containers and New Orchestration Paradigms for Isolated Environments in HPC (CANOPIE HPC), to be hosted at SC19: The International Conference for High Performance Computing, Networking, Storage and Analysis. The objective of this workshop is to serve as the principal venue for leaders in the field to
stimulate research and interactions in relation to cutting-edge container technologies, virtualization, and OS system software as it relates to supporting High Performance Computing (HPC). The first iteration of this workshop will give special attention to work which can provide real-world experiences to containerization in HPC, as well as promote the use of containers and engage in a collaborative manner with the larger HPC community. Container computing has revolutionized the way groups are developing, sharing, and running software and services. This has initially been led by growth with Docker, which has provided an ecosystem of tools to enable container based computing. This paradigm shift has since made in-roads in the HPC community enabled by container runtimes like Singularity, Shifter, and Charliecloud, which allow end-users to run containers in environments where standard Docker tools would not be feasible. Orchestration systems such as Kubernetes allow for quick, scalable service deployments in conjunction with HPC. While this adoption is growing, the larger HPC community still has many questions around this new model. There still exist several undefined research questions related to containerization in HPC that the community must address. http://canopie-hpc.org/

We are pleased to introduce the 1st International Workshop on Containers and New Orchestration Paradigms for Isolated Environments in HPC (CANOPIE HPC), to be hosted at SC19: The International Conference for High Performance Computing, Networking, Storage and Analysis. The objective of this workshop is to serve as the principal venue for leaders in the field to stimulate research and interactions in relation to cutting-edge container technologies, virtualization, and OS system software as it relates to supporting High Performance Computing (HPC). The first iteration of this workshop will give special attention to work which can provide real-world experiences to containerization in HPC, as well as promote the use of containers and engage in a collaborative manner with the larger HPC community.

Container computing has revolutionized the way groups are developing, sharing, and running software and services. This has initially been led by growth with Docker, which has provided an ecosystem of tools to enable container based computing. This paradigm shift has since made in-roads in the HPC community enabled by container runtimes like Singularity, Shifter, and Charliecloud, which allow end-users to run containers in environments where standard Docker tools would not be feasible. Orchestration systems such as Kubernetes allow for quick, scalable service deployments in conjunction with HPC. While this adoption is growing, the larger HPC community still has many questions around this new model. There still exist several undefined research questions related to containerization in HPC that the community must address.
Evaluation and Benchmarking of Singularity MPI containers on EU Research e-Infrastructure

Linux Containers with the build-once run-anywhere principle have gained huge attention in the research community where portability and reproducibility are key concerns. Unlike virtual machines (VMs), containers run the underlying host OS kernel. The container filesystem can include all necessary non-default prerequisites to run the container application at unaltered performance. For this reason, containers are popular in HPC for use with parallel/MPI applications. Some use cases include also abstraction layers, e.g. MPI applications require matching of MPI version between the host and the container, and/or GPU applications require the underlying GPU drivers to be installed within the container filesystem. In short, containers can only abstract what is above the OS kernel, not below. Consequently, portability is not completely granted. This paper presents the experience of PRACE (Partnership for Advanced Computer in Europe) in supporting Singularity containers on HPC clusters and provides notes about possible approaches for deploying MPI applications using different use cases. Performance comparison between bare metal and container executions are also provided, showing a negligible overhead in the container execution.

Enabling HPC workloads on Cloud Infrastructure using Kubernetes Container Orchestration Mechanisms

Containers offer a broad array of benefits, including a consistent lightweight runtime environment through OS-level virtualization, as well as low overhead to maintain and scale applications with high efficiency. Moreover, containers are known to package and deploy applications consistently across varying infrastructures. Container orchestrators manage a large number of containers for microservices based cloud applications. However, the use of such service orchestration frameworks towards HPC workloads remains relatively unexplored.

In this paper we study the potential use of Kubernetes on HPC infrastructure for use by the scientific community. We directly compare both its features and performance against Docker Swarm and bare metal execution of HPC applications. Herein, we detail the configurations required for Kubernetes to operate with containerized MPI applications, specifically accounting for operations such as (1) underlying device access, (2) inter-container communication across different hosts, and (3) configuration limitations. This evaluation quantifies the performance difference between representative MPI workloads running both on bare metal and containerized orchestration frameworks with Kubernetes, operating over both Ethernet and InfiniBand interconnects. Our results show that Kubernetes and Docker Swarm can achieve near bare metal performance over RDMA communication when high performance transports are enabled. Our results also show that Kubernetes presents overheads for several HPC applications over TCP/IP protocol. However,
Docker Swarm's throughput is near bare metal performance for the same applications.

**CANOPIE HPC Afternoon Break**

**Long-term Preservation of Repeatable Builds in Occam**

In order to provide transparency, wide availability, and easier reuse of scholarly software, there is a need for greater emphasis on code preservation. Yet, not just mirroring the source code, but preserving the ability to build it. Occam is a tool that offers preservation and distribution using containerization to provide repeatable execution in both building and running software. This paper gives detail about the design of Occam and its potential use within the scholarly community and beyond.

**KBase: A Platform for Reproducible Bioinformatics Research**

Reproducibility is a core tenet of the scientific process, yet it remains elusive for much of the sophisticated analysis required in modern science. In this paper we describe how reproducibility is addressed in the KBase platform, a web-based platform for performing sophisticated analysis of biological data with the goal of enabling reproducible, predictive biology. We give an overview of the architecture and some of the key design considerations. Containers play a key role in the KBase design and how it achieves a measure of strong reproducibility. We explain how containers are utilized in the platform and some of the additional considerations that aid in the goal for reproducibility. Finally, we compare KBase with other similar platforms and systems and discuss future plans.

**HPC container runtimes have minimal or no performance impact**

HPC centers are facing increasing demand for greater software flexibility to support faster and more diverse innovation in computational scientific work. Containers, which use Linux kernel features to allow a user to substitute their own software stack for that installed on the host, are an increasingly popular method to provide this flexibility. Because standard container technologies such as Docker are unsuitable for HPC, three HPC-specific technologies have emerged: Charliecloud, Shifter, and Singularity.
A common concern is that containers may introduce performance overhead. To our knowledge, no comprehensive, rigorous, HPC-focused assessment of container performance has previously been performed. Our present experiment compares the performance of all three HPC container implementations and bare metal on multiple dimensions using industry-standard benchmarks (SysBench, STREAM, and HPCG).

We found no meaningful performance differences between the four environments, with the possible exception of modest variation in memory usage.

These results suggest that HPC users should feel free to containerize their applications without concern about performance degradation, regardless of the container technology used. It is an encouraging development towards greater adoption of user-defined software stacks to increase the flexibility of HPC systems.

**On-node resource manager for containerized HPC workloads**

This document presents a new containerized architecture to enable fine-grain control over the management of on-node resources for complex scientific high-performance workloads. Our approach is introducing a node-local, application-specific resource manager by extending a container runtime, which can coordinate with the global resource manager, i.e., the system-wide manager that assigns resources to jobs. The proposed work is based on the extension of a container runtime to interface running containers with global resource managers, as well as the implementation of advanced resource management capabilities to address all the running application’s needs.

Based on this design, the various runtimes that are required for the execution of scientific applications can interact with the container runtime under which it is running. This interaction enables the scalable and dynamic allocation of resources based on runtime requirements, in opposition to job-level requirements that are traditionally handled by the global resource manager. It also enables fine-grain control over the placement of all processes and threads running in a container on specific hardware components, which is critical to achieve performance. Our approach therefore enables an efficient, scalable, dynamic and trackable management of resources on behalf of scientific applications; bridging a gap observed with current solutions.

**A Case for Portability and Reproducibility of HPC Containers**
Containerized computing is quickly changing the landscape for the development and deployment of many HPC applications. Containers are able to lower the barrier of entry for emerging workloads to leverage supercomputing resources. However, containers are no silver bullet for deploying HPC software and there are several challenges ahead in which the community must address to ensure container workloads can be reproducible and inter-operable.

In this paper, we discuss several challenges in utilizing containers for HPC applications and the current approaches used in many HPC container runtimes. These approaches have been proven to enable high-performance execution of containers at scale with the appropriate runtimes. However, the use of these techniques are still ad hoc, test the limits of container workload portability, and several gaps likely remain. We discuss those remaining gaps and propose several potential solutions, including custom container label tagging and runtime hooks as a first step in managing HPC system library complexity.

**Lightening Session - CANOPIE-HPC**
Andrew Younge (Sandia National Laboratories)
Lightening round of talks featuring various container technologies and R&D activities, moderated by Andrew Younge.

**Friday, November 22**
8:30 am - 12:00 pm

**2nd International Workshop on Performance, Portability, and Productivity in HPC (P3HPC)**

**Session Description:** The ability for applications to achieve both portability and high performance across computer architectures remains an open challenge. It is often unrealistic or undesirable for developers to maintain separate implementations for each target architecture, yet in many cases, achieving high performance and fully utilizing an architecture’s underlying features requires the use of specialized language constructs and libraries. Likewise, abstractions and standards that promise portability cannot necessarily deliver high performance without additional algorithmic considerations, and performance compromises are often made to remain portable. Application developers, who strive to work productively while balancing these concerns, often find the goal to be elusive. There is a clear need to develop ways of managing the complexity that arises from system diversity that balance the need for performant specializations with the economy of appropriate and efficient abstractions. Despite growth in the number of available architectures,
there are similarities that represent general trends in current and emerging HPC hardware: increased thread parallelism; wider vector units; specialized hardware units; and deep, complex, memory hierarchies. This in turn offers some hope for common programming techniques and language support as community experience matures. The purpose of this workshop is to provide an opportunity for attendees to share ideas, practical experiences, and methodologies for tackling the challenge of achieving performance portability and developer productivity across current and future homogeneous and heterogeneous computer architectures.  

https://p3hpc2019.lbl.gov/

The ability for applications to achieve both portability and high performance across computer architectures remains an open challenge. It is often unrealistic or undesirable for developers to maintain separate implementations for each target architecture, yet in many cases, achieving high performance and fully utilizing an architecture’s underlying features requires the use of specialized language constructs and libraries. Likewise, abstractions and standards that promise portability cannot necessarily deliver high performance without additional algorithmic considerations, and performance compromises are often made to remain portable. Application developers, who strive to work productively while balancing these concerns, often find the goal to be elusive.

There is a clear need to develop ways of managing the complexity that arises from system diversity that balance the need for performant specializations with the economy of appropriate and efficient abstractions. Despite growth in the number of available architectures, there are similarities that represent general trends in current and emerging HPC hardware: increased thread parallelism; wider vector units; specialized hardware units; and deep, complex, memory hierarchies. This in turn offers some hope for common programming techniques and language support as community experience matures.

The purpose of this workshop is to provide an opportunity for attendees to share ideas, practical experiences, and methodologies for tackling the challenge of achieving performance portability and developer productivity across current and future homogeneous and heterogeneous computer architectures.

Opening Remarks
Performance Portability across Diverse Computer Architectures
Tom Deakin (University of Bristol), Simon McIntosh-Smith (University of Bristol)

Previous studies into performance portability have typically analysed a single application (and its various implementations) in isolation. In this study we explore the wider landscape of performance portability by considering a number of applications from across the space of dwarfs, written in multiple parallel programming models, and across a diverse set of architectures. We apply rigorous performance portability metrics, as defined by Pennycook et al [Pennycook2016]. We believe this is the broadest and most rigorous performance portability study to date, representing a far reaching exploration of the state of performance portability that is achievable today. We will present a summary of the performance portability of each application and programming model across our diverse range of twelve computer architectures, including six different server CPUs from five different vendors, five different GPUs from two different vendors, and one vector architecture. We will conclude with an analysis of the performance portability of key programming models in general, across different application spaces as well across differing architectures, allowing us to comment on more general performance portability principles.

Performance Portability of a Wilson Dslash Stencil Operator Mini-App using Kokkos and SYCL
Bálint Joó (Thomas Jefferson National Accelerator Facility)

We describe our experiences in creating mini-apps for the Wilson-Dslash stencil operator for Lattice Quantum Chromodynamics using the Kokkos and SYCL programming models. In particular we comment on the performance achieved on a variety of hardware architectures, limitations we have reached in both programming models and how these have been resolved by us, or may be resolved by the developers of these models.

Performance Portability of Multi-Material Kernels
Istvan Z. Reguly (Pázmány Péter Catholic University, Hungary)

Trying to improve performance, portability, and productivity of an application presents non-trivial trade-offs, which are often difficult to quantify. Recent work has developed metrics for performance portability, as well some aspects of productivity - in this case study, we present a set of challenging computational kernels and their implementations from the domain of multi-material simulations, and evaluate them using these metrics. Three key kernels are implemented using OpenMP, OpenMP offload, OpenACC, CUDA, SYCL, and KOKKOS, and tested on ARM ThunderX2, IBM Power 9, Intel KNL, Broadwell, and Skylake CPUs, as well as NVIDIA P100 and V100 GPUs. We also consider the choice of compilers, evaluating LLVM/Clang, GCC, PGI, Intel, IBM XL, and Cray compilers, where available. We present a detailed performance analysis, calculate performance portability and code divergence metrics, contrasting performance, portability, and productivity.
An Approach for Indirectly Adopting a Performance Portability Layer in Large Legacy Codes
John K. Holmen (University of Utah, Scientific Computing and Imaging Institute)

Diversity among supported architectures in current and emerging high performance computing systems, including those for exascale, makes portable codebases desirable. Portability of a codebase can be improved using a performance portability layer to provide access to multiple underlying programming models through a single interface. Direct adoption of a performance portability layer, however, poses challenges for large pre-existing software frameworks that may need to preserve legacy code and/or adopt other programming models in the future. This paper describes an approach for indirect adoption that introduces a framework-specific portability layer between the application developer and the adopted performance portability layer to help improve legacy code support and long-term portability for future architectures and programming models. This intermediate layer uses loop-level, application-level, and build-level components to ease adoption of a performance portability layer in large legacy codebases. Results are shown for two challenging case studies using this approach to make portable use of OpenMP and CUDA via Kokkos in an asynchronous many-task runtime system, Uintah. These results show performance improvements up to 2.7x when refactoring for portability and 2.6x when more efficiently using a node. Good strong-scaling to 442,368 threads across 1,728 Knights Landing processors are also shown using MPI+Kokkos at scale.

Panel Discussion - Application Perspectives

P3HPC Morning Break

On Applying Performance Portability Metrics
Daniela Ferreira Daniel (Aeronautics Institute of Technology, Brazil)

As we prepare for further technological advancement in supercomputing, the diversity of hardware architectures and parallel programming languages has increased to new levels. At the same time, extracting performance from so many architectures is even more difficult. In this context, the appearance of portable languages capable of generating executable code for multiple architectures has become a recurrent research target. We port a set of seven parallel benchmarks from SPEC
ACCEL suite and a wave propagation code to one such portable language: the Kokkos C++ programming library. Using the original OpenACC versions of the eight codes, we apply a known performance portability metric on the OpenACC and Kokkos versions of those codes across a variety of hardware platforms and problem sizes. We observe that the portability metric is sensitive to the problem size. To remedy this deficiency, we propose a novel metric for performance portability, apply the proposed metric to the eight codes and discuss the results.

**mdspan in C++: A Case Study in the Integration of Performance Portable Features into International Language Standards**  
David S. Hollman (Sandia National Laboratories)

Multi-dimensional arrays are ubiquitous in high-performance computing (HPC), but their absence from the C++ language standard is a long-standing and well-known limitation of their use for HPC. This paper describes the design and implementation of `mdspan`, a proposed C++ standard multidimensional array view (planned for inclusion in C++23). The proposal is largely inspired by work done in the Kokkos project—a C++ performance-portable programming model deployed by numerous HPC institutions to prepare their code base for exascale-class supercomputing systems. This paper describes the final design of mdspan after a five-year process to achieve consensus in the C++ community. In particular, we will lay out how the design addresses some of the core challenges of performance-portable programming, and how its customization points allow a seamless extension into areas not currently addressed by the C++ Standard but which are of critical importance in the heterogeneous computing world of today’s systems. Finally, we have provided a production-quality implementation of the proposal in its current form. This work includes several benchmarks of this implementation aimed at demonstrating the zero-overhead nature of the modern design.

**RAJA: Portable Performance for Large-Scale Scientific Applications**  
Thomas RW Scogland (Lawrence Livermore National Laboratory)

Modern high-performance computing systems are diverse, with hardware designs ranging from homogeneous multi-core CPUs to GPU or FPGA accelerated systems. Achieving desirable application performance often requires choosing a programming model best suited to a particular platform. For large codes used daily in production that are under continual development, architecture-specific ports are untenable. Maintainability requires single-source application code that is performance portable across a range of architectures and programming models.

In this paper we describe RAJA, a portability layer that enables C++ applications to leverage various programming models, and thus architectures, with a single-source codebase. We describe preliminary results using RAJA in three large production codes at Lawrence Livermore National
Laboratory, observing 17x, 13x and 12x speedup on GPU-only over CPU-only nodes with single-source application code in each case.

**ClangJIT: Enhancing C++ with Just-in-Time Compilation**  
Hal Finkel (Argonne National Laboratory)

The C++ programming language is not only a keystone of the high-performance-computing ecosystem but has proven to be a successful base for portable parallel-programming frameworks. As is well known, C++ programmers use templates to specialize algorithms, thus allowing the compiler to generate highly-efficient code for specific parameters, data structures, and so on. This capability has been limited to those specializations that can be identified when the application is compiled, and in many critical cases, compiling all potentially-relevant specializations is not practical. ClangJIT provides a well-integrated C++ language extension allowing template-based specialization to occur during program execution. This capability has been implemented for use in large-scale applications, and we demonstrate that just-in-time-compilation-based dynamic specialization can be integrated into applications, often requiring minimal changes (or no changes) to the applications themselves, providing significant performance improvements, programmer-productivity improvements, and decreased compilation time.

**Panel Session - P3 Enabling Technology**

**Closing Remarks**

8:30 am - 12:00 pm

**HPC Systems Professionals Workshop (HPCSYSPROS19)**

**Session Description:** The complexity of High Performance Computing (HPC) systems necessitates advanced techniques in system administration, configuration, and engineering and (by proxy) staff who are well-versed on the best practices in this field. 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
separate conferences exist for the HPC field and for the systems administration field, none exist that focus specifically on the needs of HPC systems professionals. As such, it can be difficult to find support resources who are able to help with the issues encountered in this specialized field. The ACM SIGHPC SYSPROS Virtual Chapter, the sponsor for this workshop, has been established to provide opportunities to develop and grow relationships among HPC systems administration practitioners and to act as a support resource for them. This workshop is designed to share best practices for common HPC system deployment and maintenance, to provide a platform to discuss upcoming technologies, and to present state of the practice techniques that increase performance and reliability of systems, and in turn increase researcher and analyst productivity.  
http://sighpc-syspros.org/workshops/2019/

The complexity of High Performance Computing (HPC) systems necessitates advanced techniques in system administration, configuration, and engineering and (by proxy) staff who are well-versed on the best practices in this field. 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 separate conferences exist for the HPC field and for the systems administration field, none exist that focus specifically on the needs of HPC systems professionals. As such, it can be difficult to find support resources who are able to help with the issues encountered in this specialized field. The ACM SIGHPC SYSPROS Virtual Chapter, the sponsor for this workshop, has been established to provide opportunities to develop and grow relationships among HPC systems administration practitioners and to act as a support resource for them.

This workshop is designed to share best practices for common HPC system deployment and maintenance, to provide a platform to discuss upcoming technologies, and to present state of the practice techniques that increase performance and reliability of systems, and in turn increase researcher and analyst productivity.

**Chameleon: How to Build a Cloud++**

Kate Keahey (Argonne National Laboratory)

Chameleon is a large-scale, deeply reconfigurable experimental platform built to support Computer Sciences systems research. Community projects range from systems research developing exascale operating systems, virtualization methods, performance variability studies, and power management research to projects in software defined networking, machine learning, and resource management.
What makes Chameleon unique is that it provides these sophisticated capabilities based on a mainstream infrastructure cloud system (OpenStack). In this talk, I will explain the challenges we faced in building Chameleon, lessons learned, operations experiences, and describe our packaging of the system that integrates both the developed capabilities and the operational experience and facilitates managing platforms of this kind.

Decoupling OpenHPC Critical Services
Jacob Chappell (University of Kentucky), Bhushan Chitre (University of Kentucky)

High-Performance Computing (HPC) cluster-management software often consolidates cluster-management functionality into a centralized management node, using it to provision the compute nodes, manage users, and schedule jobs. A consequence of this design is that the management node must typically be up and operating correctly for the cluster to schedule and continue executing jobs. This dependency de-incentivizes administrators from upgrading the management node because the entire cluster may need to be taken down during the upgrade. Administrators may even avoid performing minor updates to the management node for fear that an update error could bring the cluster down.

To address this problem, we redesigned the structure of management nodes, specifically OpenHPC’s System Management Server (SMS), breaking it into components that allow portions of the SMS to be taken down and upgraded without interrupting the rest of the cluster. Our approach separates the time-critical SMS tasks from tasks that can be delayed, allowing us to keep a relatively small number of time-critical tasks running while bringing down critical portions of the SMS for long periods of time to apply OpenHPC upgrades, update applications, and perform acceptance tests on the new system.

We implemented and deployed our solution on the University of Kentucky’s HPC cluster, and it has already helped avoid downtime from an SMS failure. It also allows us to reduce, or completely eliminate our regularly scheduled maintenance windows.

Implementing a Common HPC Environment in a Multi-User Spack Instance
Carson Woods (University of Tennessee, Chattanooga)

High performance computing is highly dependent on validated and tested environments that are
tuned for specific hardware to yield good performance. Spack enables users to construct their own software stacks, but does not provide the benefits of a centrally curated environment. We implemented new features in Spack to support co-existing system-wide deployments and user Spack environments, and a Spack-enabled global environment with them. This had the result of improving the global environment's portability and the efficiency of Spack use on a production system.

HPCSYSPROS19 Morning Break

Arbiter: Dynamically Limiting Resource Consumption on Login Nodes
Dylan Gardner (University of Utah, Center for High Performance Computing), Brian Haymore (University of Utah, Center for High Performance Computing)

Login nodes are shared resources that are meant for small, lightweight tasks such as submitting to the batch system, scripting, compiling, and staging data. Because they are shared resources, the responsiveness of login nodes depends on users being good citizens with respect to CPU time and memory usage. Many HPC centers have policies that define behavior that is acceptable. However, the unfortunate reality is that users frequently violate these policies and negatively impact the work of others. This impact is exacerbated by the fact that administrators often have to police users’ behavior.

Arbiter is a service that addresses the misuse of login nodes by automatically enforcing policies using cgroups. When users log in, Arbiter sets a default hard memory and CPU limit on the user to prevent them from dominating the whole machine’s memory and CPU resources. To enforce policies, Arbiter tracks the usage of individual users over a set interval and looks for policy violations. When a violation occurs, the violating user is emailed about what behavior constituted the violation and the acceptable usage policy for login nodes. In addition, Arbiter also temporarily lowers the hard memory and CPU limit to discourage excessive usage. The length of time and severity of the lower hard limit depends on whether a user has repeatedly violated policies to penalize users for continued excessive usage. The result of the Arbiter service is that login nodes stay responsive, with users informed of policies and discouraged from running computationally heavy jobs on login nodes.

Using GUFI in Data Management
Christopher Hoffman (National Center for Atmospheric Research (NCAR))
Storage systems at the peta-scale and beyond are now a reality. This unprecedented system scale can generate tens of billions of files within its lifetime. Due to this amount of metadata generated, managing it can be a significant challenge. Traditional tools such as ls, find, and du are quickly becoming insufficient for peta-scale storage and beyond.

In this talk, we will share information about the Grand Unified File Index (GUFI) tool and how it can help sites manage large numbers of files and their metadata. As this tool is still under development at the time of investigation, we will go over our methods that are used to index the storage systems at NCAR. Then, the on-going effort of how NCAR is making use of this tool to query metadata will be discussed. Finally, we will share initial testing results that show that find and du commands on some directory trees run over 100 times faster using GUFI.

**Monitoring HPC Services with CheckMK**

Kieran Leach (Edinburgh Parallel Computing Centre), Craig Manzi (Edinburgh Parallel Computing Centre)

Administrative monitoring of a range of HPC systems can be time consuming and inefficient with many HPC systems being provided with their own integrated monitoring solutions and an expectation that system managers will monitor each system separately. In order to save staff time, effort and in order to improve the potential for rapid and effective response to emerging problems where systems interact, a "single pane of glass" approach is considered optimal. HPC systems typically utilise relatively boutique technology however which is commonly not monitored by existing out-of-the-box monitoring solutions. In this presentation we detail the application of CheckMK, a general use monitoring system, to HPC systems using non-commodity hardware and software. We focus on the development and use of "check" scripts, which at EPCC have enabled the System Administrators team to simply and reliably monitor all relevant and service-critical aspects of a variety of HPC systems through a single "pane of glass" approach.

**The Road to Devops HPC Cluster Management**

Ken P. Schmidt (Pacific Northwest National Laboratory (PNNL))

EMSL has been using xCAT for cluster administration since 2011. As supercomputing strategies have changed, we've developed a configuration approach which addresses many of xCAT’s shortcomings and enables us to manage heterogeneous cluster configurations in a manner concordant with software engineering best practices. Some changes have been submitted upstream to the xCAT project, and others exist as a suite of automation tools and architectural practices. This presentation is an overview of some of the challenges we've overcome and best practices derived from 8 years of real-world xCAT deployment across multiple clusters.
What Deploying MFA Taught Us About Changing Infrastructure
Rebecca Hartman-Baker (National Energy Research Scientific Computing Center (NERSC))

NERSC is not the first organization to implement multi-factor authentication (MFA) for its users. We had seen multiple talks by other supercomputing facilities who had deployed MFA, but as we planned and deployed our MFA implementation, we found that nobody had talked about the more interesting and difficult challenges, which were largely social rather than technical. Our MFA deployment was a success, but, more importantly, much of what we learned could apply to any infrastructure change. Additionally, we developed the sshproxy service, a key piece of infrastructure technology that lessens user and staff burden and has made our MFA implementation more amenable to scientific workflows. We found great value in using robust open-source components where we could and developing tailored solutions where necessary.

A Better Way of Scheduling Jobs on HPC Systems: Simultaneous Fair-Share
Craig P. Steffen (University of Illinois, National Center for Supercomputing Applications (NCSA))

Typical HPC job scheduler software determines scheduling order by a linear sum of weighted priority terms. When a system has a rich mix of job types, this makes it difficult to maintain good productivity across diverse user groups. Currently-implemented fair-share algorithms tweak priority calculations based on past job handling by modifying priority, but don’t fully solve problems of queue-stuffing and various classes of under-served job types, because of coupling between different terms of the linear calculated priority.

This paper proposes a new scheme of scheduling jobs on an HPC system called “Simultaneous Fair-share” (or “SFS”) that works by considering the jobs already committed to run in a given time slice and adjusting which jobs are selected to run accordingly. This allows high-throughput collaborations to get lots of jobs run, but avoids the problems of some groups starving out others due to job characteristics, all while keeping system administrators from having to directly manage job schedules. This paper presents Simultaneous Fair-share in detail, with examples, and shows testing results using a job throughput and scheduler simulation.

Closing Remarks and Open Discussion

We will finish up with some closing remarks and leave the remaining 15-20 minutes for free discussion.

8:30 am - 12:00 pm
The 3rd Industry/University Joint International Workshop on Data-Center Automation, Analytics, and Control (DAAC)

**Session Description:** This proposal describes a plan to continue the Industry/University Joint International Workshop on Data-Center Automation, Analytics, and Control (DAAC) to be hosted at the 2019 ACM/IEEE Supercomputing Conference (SC19). This workshop proposal is an outcome of intensive discussions from academia, industry, and national laboratory researchers that led to successful previous instances hosted at SC’18 and UCC’17. Looking at the last year’s attendance at SC’18, DAAC has attracted up to ~80 attendees and had 50.6 attendees on average. DAAC’18 featured an industry panel, 3 invited talks from both academia and industry, and 10 paper talks from academia, national labs, and industry, selected through a rigorous review process (at least three reviews for each paper). DAAC’17 featured three invited speakers from industry and a panel of five experts and different stakeholders in addition to presentations from peer-reviewed papers. This proposed new instance of the DAAC workshop at SC remains a unique workshop that promotes collaboration among academia, industry, and national labs and remains jointly organized by academic and industry researchers. The objective is to promote and stimulate community’s interactions to address some of most critical challenges in automation, analytics, and control specifically aimed for the needs of large-scale data centers in high-performance/high-end computing. DAAC’19 will provide a valuable addition to main conference programs. Taking advantage of the opportune match to the SC19 audience, DAAC’19 expects to attract a larger number of attendees from academia, industry, and government labs who are interested in data center automated management, operation, and maintenance. [http://daac-workshop.org](http://daac-workshop.org)
promotes collaboration among academia, industry, and national labs and remains jointly organized by academic and industry researchers. The objective is to promote and stimulate community’s interactions to address some of most critical challenges in automation, analytics, and control specifically aimed for the needs of large-scale data centers in high-performance/high-end computing. DAAC’19 will provide a valuable addition to main conference programs. Taking advantage of the opportune match to the SC19 audience, DAAC’19 expects to attract a larger number of attendees from academia, industry, and government labs who are interested in data center automated management, operation, and maintenance.


**CloudTraceViz: A Visualization Tool for Tracing Dynamic Usage of Cloud Computing Resources**

This paper introduces CloudTraceViz, a visual analytic tool for analyzing the characteristics of modern cloud data centers. The goals of this tool are: 1) to fulfill a set of visual tasks on cloud computing retrieved from in-depth interviews with domain experts, 2) to visualize and monitor large real-world data in terms of both the number of profiles and number of time steps, and 3) to aid system administrator to trace and understand the causal relationship of multivariate data. To reach these goals, our system composes several interconnected visual components. The customized heatmap is used to capture the pattern of a single machine as well as a group of machines, the progressive rendering of parallel coordinate allows users to see the dynamic behavior of running task/job over time, the scatterplot matrices are used in conjunction with the parallel graph for anomaly extraction. The results on the Alibaba Cloud Trace dataset show that the visualization tools offer great support for users to have a high-level overview of a large dataset as well as understand the causal relations within multivariate data.

**MetricQ: A Scalable Infrastructure for Processing High-Resolution Time Series Data**

In this paper we present MetricQ, a novel infrastructure for collecting, archiving, and analyzing sensor data. Core components of MetricQ are a scalable message broker based on the Advanced Message Queuing Protocol, and a newly developed Hierarchical Timeline Aggregation (HTA) storage concept that is specifically designed for timeseries data. HTA requires moderate data processing during data collection and a storage space overhead of about 10 %, and in turn reduces the complexity of typical timeline request from O(N) to O(1). This enables access to very large metric timelines spanning years and billions of data points at a performance level that is sufficient for interactive use cases. In contrast to existing solutions in this domain, no relevant information
such as very short peaks in the data is discarded. We demonstrate how we use MetricQ with few metrics at very high update rates, e.g., for energy efficiency research, and for a very large number of metrics at moderate update rates, e.g., monitoring data from the electrical and cooling infrastructure of our data center.

**MELA: A Visual Analytics Tool for Studying Multifidelity HPC System Logs**

To maintain a robust and reliable supercomputing hardware system there is a critical need to understand various system events, including failures occurring in the system. Toward this goal, we analyze various system logs such as error logs, job logs and environment logs from Argonne Leadership Computing Facility’s (ALCF) Theta Cray XC40 supercomputer. This log data incorporates multiple subsystem and component measurements at various fidelity levels and temporal resolutions - a very diverse and massive dataset. To effectively identify various patterns that characterize system behavior and faults over time, we have developed a visual analytics tool, MELA, to better identify patterns and glean insights from these log data.

**(Late-Breaking Extended Abstract) Job Outcome Prediction via Text Log Analysis**

The massive scale of high performance computing machines necessitates the use of automatic statistical methods to assist human operators in monitoring day-to-day behavior. The largest high performance computing machines today produce on the order of terabytes of monitoring information each day. We specifically address the problem of identifying problematic compute jobs running on these massive machines by modeling the computer-generated text logs known as system logs, which record all activities on the machine in near-natural language form. We apply techniques from relational learning and human language technology, as well as making use of systems domain knowledge, in order to extract features from system logs produced by approximately 10,000 high performance computing jobs. We then evaluate the usefulness of these features by training a random forest model to predict job outcome (completion, failure, timeout, or node failure) in real time. We compare our models to a baseline which mimics state-of-the-art human operator behavior, and find that the best-performing feature set is one which combines domain knowledge with simple aggregate numerical content and temporal metrics. We find that in the average case, our method can predict job outcomes with an F1 score approaching 0.9 after a job has been running for only 30 minutes, giving an average lead time of 3 hours before failure during which a human operator could take mitigating action. The work is a proof-of-concept which suggests the ability to develop a production tool that would raise early alerts based on job outcome predictions.
HiperJobViz: Visualizing Resource Allocations in High-Performance Computing Center via Multivariate Health-Status Data

Scheduling, visualizing, and balancing resource allocations in High-Performance Computing Centers are complicated tasks due to a large amount of data and the dynamic natures of the resource allocation problem. This paper introduces HiperJobViz, a visual analytic tool for visualizing the resource allocations of data centers for jobs, users, and usage statistics. The goals of this tool are: 1) to provide an overview of the current resource usages, 2) to track changes of resource usages by users, jobs, and hosts and 3) to provide a detailed view of the resource usage via multi-dimensional representation of health metrics, such as CPU temperatures, memory usage, and power consumption. To support these goals, our visual analytics tool provides a full range of interactive features, including details on demands, brushing and links, and filtering.

DAAC Morning Break

Multi Source Cooling Control Algorithm

This paper describes how the automation team at ENI Green Data Center (GDC) solved the problem of controlling the temperature inside a data room that can be air cooled both with direct free-cooling and with water-based chillers and heat exchangers using a slide mode control design. They created an algorithm which provided: 1. a directional, incremental “walking” function to seek out and test alternative solutions, biased toward a preferred solution, 2. stabilization on an optimized solution, 3. a destabilization function to retest the solution periodically without allowing jitter, and 4. a method of changing preference to autonomously re-optimize for another configuration.

HyperOXN: A Novel Optical Cross-Connect Architecture for Data Centers

With the advent of new applications such as Cloud Computing, Big Data, and Machine Learning, modern data center network (DCN) architecture has been evolving to meet numerous challenging requirements such as scalability, agility, energy efficiency and high performance. The above new applications are expediting the convergence of High-Performance Computing and Data Centers. Inspired by hypermeshes, this paper presents HyperOXN, a novel cost-efficient topology for exascale Data Center Networks (DCNs). HyperOXN takes advantage of high-radix switch components
leveraging state-of-the-art colorless dense wavelength division multiplexing (DWDM) technologies, effectively supports broadcast traffic and at the same time meets the demands on high throughput and low latency. It’s shown that HyperOXN outperforms a comparable Fat-Tree topology in cost, throughput and power consumption under a variety of workload conditions.

DAAC Industrial/University Joint Panel Discussion

8:30 am - 12:00 pm

3rd International Workshop on Emerging Parallel and Distributed Runtime Systems and Middleware (IPDRM'2019)

Session Description: Node architectures of extreme-scale systems are rapidly increasing in complexity. Emerging homogeneous and heterogeneous designs provide massive multi-level parallelism, but developing efficient runtime systems and middleware that allow applications to efficiently and productively exploit these architectures is extremely challenging. Moreover, current state-of-the-art approaches may become unworkable once energy consumption, resilience, and data movement constraints are added. The goal of this workshop is to attract the international research community to share new and bold ideas that will address the challenges of design, implementation, deployment, and evaluation of future runtime systems and middleware. The previous two iterations of the workshop has been met with great success with the International Parallel and Distributed Processing Symposium with great attendance and high quality research papers. Due to the rise of beyond Moore Computing and Exascale systems, the organizing committee deems appropriate to co-locate this workshop with the premier research and development showcase: SC19. https://ipdrm.github.io/

Node architectures of extreme-scale systems are rapidly increasing in complexity. Emerging homogeneous and heterogeneous designs provide massive multi-level parallelism, but developing efficient runtime systems and middleware that allow applications to efficiently and productively exploit these architectures is extremely challenging. Moreover, current state-of-the-art approaches may become unworkable once energy consumption, resilience, and data movement constraints are
The goal of this workshop is to attract the international research community to share new and bold ideas that will address the challenges of design, implementation, deployment, and evaluation of future runtime systems and middleware.

The previous two iterations of the workshop has been met with great success with the International Parallel and Distributed Processing Symposium with great attendance and high quality research papers. Due to the rise of beyond Moore Computing and Exascale systems, the organizing committee deems appropriate to co-locate this workshop with the premier research and development showcase: SC19.

**Sequential Codelet Model of Program Execution. A Super-Codelet model based on the Hierarchical Turing Machine.**

The Sequential Codelet Model is a definition of a program execution model that aims to achieve parallel execution of programs that are expressed sequentially and in a hierarchical manner. The Sequential Codelet Model heavily borrows from the successful experience acquired through decades of sequential program execution, in particular, the use of Instruction Level Parallelism optimizations for implicit parallel execution of code. We revisit and re-define the Universal Turing Machine and the Von Neumann Architecture to account for the hierarchical organization of the whole computation system and its resources (i.e. memory, computational capabilities, and interconnection networks), as well as consider program complexity and structure in relation to its execution. This work defines the Sequential Codelet Model (SCM), the Hierarchical Turing Machine (HTM), and the Hierarchal Von Neumann Architecture, as well as explains how implicit parallel execution of programs could be achieved by using these definitions.

**Advert: An Asynchronous Runtime for Fine-Grained Network Systems**

The Data Vortex Network is a novel high-radix, congestion free interconnect able to cope with the fine-grained, unpredictable communication patterns of irregular applications. This paper presents ADVERT, an asynchronous runtime system that provides performance and productivity for the Data Vortex Network. ADVERT integrates a lightweight memory manager (DVMem) for the user accessible SRAM integrated in the network interface, and a communication library (DVComm) that implements active messaging primitives (get, put, and remote execution). ADVERT hides the complexity of controlling all the network hardware features through the low-level Data Vortex programming interface, while providing comparable performance. We discuss ADVERT’s design and present microbenchmarks to examine different runtime features. ADVERT provides the functionalities for building higher level asynchronous many tasking runtimes and partitioned global
address space (PGAS) libraries on top of the Data Vortex Network.

**Characterizing the Performance of Executing Many-tasks on Summit**

Many scientific workloads are comprised of many tasks, where each task is an independent simulation or analysis of data. The execution of millions of tasks on heterogeneous HPC platforms requires scalable dynamic resource management and multi-level scheduling. RADICAL-Pilot (RP) -- an implementation of the Pilot abstraction, addresses these challenges and serves as an effective runtime system to execute workloads comprised of many tasks. In this paper, we characterize the performance of executing many tasks using RP when interfaced with JSM and PRRTE on Summit: RP is responsible for resource management and task scheduling on acquired resource; JSM or PRRTE enact the placement of launching of scheduled tasks. Our experiments provide lower bounds on the performance of RP when integrated with JSM and PRRTE. Specifically, for workloads comprised of homogeneous single-core, 15 minutes-long tasks we find that: PRRTE scales better than JSM for > O(1000) tasks; PRRTE overheads are negligible; and PRRTE supports optimizations that lower the impact of overheads and enable resource utilization of 63% when executing O(16K), 1-core tasks over 404 compute nodes.

**Assessing the Performance Impact of using an Active Global Address Space in HPX: A Case for AGAS**

In this research, we describe the functionality of AGAS (Active Global Address Space), a subsystem of the HPX runtime system that is designed to handle data locality at runtime, independent of the hardware and architecture configuration. AGAS enables transparent runtime global data access and data migration, but incurs a an overhead cost at runtime. We present a method to assess the performance of AGAS and the amount of impact it has on the execution time of the Octo-Tiger application. With our assessment method we identify the four most expensive AGAS operations in HPX and demonstrate that the overhead caused by AGAS is negligible.

**Leveraging Network-level parallelism with Multiple Process-Endpoints for MPI Broadcast**

The Message Passing Interface has been the dominating programming model for developing scalable and high-performance parallel applications. Collective operations empower group communication operations in a portable, and efficient manner and are used by a large number of applications across different domains. Optimization of collective operations is the key to achieve good performance speed-ups and portability. Broadcast or One-to-all communication is one of the
most commonly used collectives in MPI applications. However, the existing algorithms for broadcast do not effectively utilize the high degree of parallelism and increased message rate capabilities offered by modern architectures. In this paper, we address these challenges and propose a Scalable Multi-Endpoint broadcast algorithm that combines hierarchical communication with multiple endpoints per node for high performance and scalability. We evaluate the proposed algorithm against state-of-the-art designs in other MPI libraries, including MVAPICH2, Intel MPI, and Spectrum MPI. We demonstrate the benefits of the proposed algorithm at benchmark and application level at scale on four different hardware architectures, including Intel Cascade Lake, Intel Skylake, AMD EPYC, and IBM POWER9, and with InfiniBand and Omni-Path interconnects. Compared to other state-of-the-art designs, our proposed design shows up to 2.5 times performance improvements at a microbenchmark level with 128 Nodes. We also observe up to 37% improvement in broadcast communication latency for the SPECMPI scientific applications.

IPDRM‘2019 Morning Break

Design and Evaluation of Shared Memory Communication Benchmarks on Emerging Architectures using MVAPICH2

Recent advances in processor technologies have led to highly multi-threaded and dense multi- and many-core HPC systems. The adoption of such dense multi-core processors is widespread in the Top500 systems. Message Passing Interface (MPI) has been widely used to scale out scientific applications. The communication designs for intra-node communication in MPI are mainly based on shared memory communication. The increased core-density of modern processors warrants the use of efficient shared memory communication designs to achieve optimal performance. While there have been various algorithms and data-structures proposed for the producer-consumer like scenarios in the literature, there is a need to revisit them in the context of MPI communication on modern architectures to find the optimal solutions that work best for modern architectures. In this paper, we first propose a set of low-level benchmarks to evaluate various data-structures such as Lamport queues, Fast-Forward queues, and Fastboxes (FB) for shared memory communication. Then, we bring these designs into the MVAPICH2 MPI library and measure their impact on the MPI intra-node communication for a wide variety of communication patterns. The benchmarking results are carried out on modern multi-/many-core architectures including Intel Xeon CascadeLake and Intel Knights Landing.
8:40 am - 12:00 pm

Workshop on Fault-Tolerance for HPC at Extreme Scale (FTXS)

Session Description: Addressing failures in extreme-scale systems remains a significant challenge to reaching exascale. Current projections suggest that at the scale necessary to sustain exaflops of computation, systems could experience failures as frequently as once per hour. As a result, robust and efficient fault tolerance techniques are critical to obtaining reasonable application performance. Additionally, it is also imperative that we develop an understanding of trends in hardware devices may affect the reliability of future systems. The emergence of high-bandwidth memory devices, the continued deployment of burst buffers, and the development of near-threshold devices to address power concerns will all impact fault tolerance on new systems. These design trends coupled with increases in the number, variety, and complexity of components required to compose an extreme-scale system means that systems will experience significant increases in aggregate fault rates, fault diversity, and the complexity of isolating root causes. Due to the continued need for research on fault tolerance in extreme-scale systems, the 9th Workshop on Fault-Tolerance for HPC at Extreme Scale (FTXS 2019) will present an opportunity for innovative research ideas to be shared, discussed, and evaluated by researchers in fault-tolerance, resilience, dependability, and reliability from academic, government, and industrial institutions. Building on the success of the previous editions of the FTXS workshop, the organizers will assemble quality publications, invited talks, and panels to facilitate a lively and thought-provoking group discussion.

https://sites.google.com/site/ftxsworkshop/home/ftxs-2019

Addressing failures in extreme-scale systems remains a significant challenge to reaching exascale. Current projections suggest that at the scale necessary to sustain exaflops of computation, systems could experience failures as frequently as once per hour. As a result, robust and efficient fault tolerance techniques are critical to obtaining reasonable application performance. Additionally, it is also imperative that we develop an understanding of trends in hardware devices may affect the reliability of future systems. The emergence of high-bandwidth memory devices, the continued deployment of burst buffers, and the development of near-threshold devices to address power concerns will all impact fault tolerance on new systems. These design trends coupled with increases in the number, variety, and complexity of components required to compose an extreme-scale system means that systems will experience significant increases in aggregate fault rates, fault diversity, and the complexity of isolating root causes. Due to the continued need for research on fault tolerance in extreme-scale systems, the 9th Workshop on Fault-Tolerance for HPC at Extreme Scale (FTXS 2019) will present an opportunity for innovative research ideas to be shared, discussed, and evaluated by researchers in fault-tolerance, resilience, dependability, and reliability from academic, government, and industrial institutions. Building on the success of the previous editions of the FTXS workshop, the organizers will assemble quality publications, invited talks, and panels to facilitate a lively and thought-provoking group discussion.

https://sites.google.com/site/ftxsworkshop/home/ftxs-2019
concerns will all impact fault tolerance on new systems. These design trends coupled with increases in the number, variety, and complexity of components required to compose an extreme-scale system means that systems will experience significant increases in aggregate fault rates, fault diversity, and the complexity of isolating root causes.

Due to the continued need for research on fault tolerance in extreme-scale systems, the 9th Workshop on Fault-Tolerance for HPC at Extreme Scale (FTXS 2019) will present an opportunity for innovative research ideas to be shared, discussed, and evaluated by researchers in fault-tolerance, resilience, dependability, and reliability from academic, government, and industrial institutions. Building on the success of the previous editions of the FTXS workshop, the organizers will assemble quality publications, invited talks, and panels to facilitate a lively and thought-provoking group discussion.

Asynchronous Receiver-Driven Replay for Local Rollback of MPI Applications
Aurelien Bouteiller (University of Tennessee, Innovative Computing Laboratory)

With the increase in scale and architectural complexity of supercomputers, the management of failures has become integral to successfully executing a long-running high-performance computing application. In many instances, failures have a localized scope, usually impacting a subset of the resources being used, yet widely used failure recovery strategies (like checkpoint/restart) fail to take advantage and rely on global, synchronous recovery actions. Even with local rollback recovery, in which only the fault impacted processes are restarted from a checkpoint, the consistency of further progress in the execution is achieved through the replay of communication from a message log. This theoretically sound approach encounters some practical limitations: the presence of collective operations forces a synchronous recovery that prevents survivor processes from continuing their execution, removing any possibility for overlapping further computation with the recovery; and the amount of resources required at recovering peers can be untenable. In this work, we solved both problems by implementing an asynchronous, receiver-driven replay of point-to-point and collective communications, and by exploiting remote-memory access capabilities to access the message logs. This new protocol is evaluated in an implementation of local rollback over the User Level Failure Mitigation fault tolerant Message Passing Interface (MPI). It reduces the recovery times of the failed processes by an average of 59%, while the time spent in the recovery by the survivor processes is reduced by 95% when compared to an equivalent global rollback protocol, thus living to the promise of a truly localized impact of recovery actions.

Enforcing Crash Consistency of Scientific Applications in Non-Volatile Main Memory Systems
Tyler Coy (Washington State University, Vancouver)

To fully leverage the emerging non-volatile main memory (NVMM) for high-performance computing,
Programmers need efficient data structures that are aware of NVMM memory models and provide crash consistency. Manual creation of NVMM-aware persistent data structures requires a deep understanding of how and when to create persistent snapshots of memory objects corresponding to the data structures and substantial code modification, which makes it very difficult to use in its manual form even for experienced programmers. To simplify the process, we design a compiler-assistant technique, NVPath. With the aid of compilers, it automatically generates NVMM-aware persistent data structures which provide the same level of guarantee of crash consistency compared to the baseline code. Compiler-assistant code annotation and transformation is general and can be applied to applications using various data structures. Furthermore, it is a gray-box technique which requires minimum users' input. Finally, it keeps the baseline code structure for good readability and maintenance. Our experimental results with real-world scientific applications (e.g., matrix multiplication, LU decomposition, adaptive-mesh refinement, and page ranking) show that the performance of annotated programs is commensurate with the version using the manual code transformation on the Titan supercomputer.

**FaultSight: A Fault Analysis Tool for HPC Researchers**

*Jon Calhoun (Clemson University, Holcombe Department of Electrical and Computer Engineering)*

System reliability is expected to be a significant challenge for future extreme-scale systems. Poor reliability results in a higher frequency of interruptions in high-performance computer (HPC) applications due to system/application crashes or data corruption due to soft errors. In response, application level error detection and recovery schemes are devised to mitigate the impact of these interruptions. Evaluating these schemes and the reliability of an application requires the analysis of thousands of fault injection trials, resulting in tedious and time-consuming process. Furthermore, there is no one data analysis tool that can work with all of the fault injection frameworks currently in use. In this paper, we present FaultSight, a fault injection analysis tool capable of efficiently assisting in the analysis of HPC application reliability as well as the effectiveness of resiliency schemes. FaultSight is designed to be flexible and work with data coming from a variety of fault injection frameworks. The effectiveness of FaultSight is demonstrated by exploring the reliability of different versions of the Matrix-Matrix Multiplication kernel using two different fault injection tools. In addition, the detection and recovery schemes are highlighted for the HPCCG mini-app.

**FTXS Morning Break**

**Self-stabilizing Connected Components**

*Piyush Sao (Oak Ridge National Laboratory, Georgia Institute of Technology)*
For the problem of computing the connected components of a graph, this paper considers the design of algorithms that are resilient to transient hardware faults, like bit flips. More specifically, it applies the technique of self-stabilization. A system is self-stabilizing if, when starting from a valid or invalid state, it is guaranteed to reach a valid state after a finite number of steps. Therefore on a machine subject to a transient fault, a self-stabilizing algorithm could recover if that fault caused the system to enter an invalid state.

We give a comprehensive analysis of the valid and invalid states during label propagation and derive algorithms to verify and correct the invalid state. The self-stabilizing label-propagation algorithm performs $O(V \log V)$ additional computation and requires $O(V)$ additional storage over its conventional counterpart (and, as such, does not increase asymptotic complexity over conventional labelprop). When run against a battery of simulated fault injection tests, the self-stabilizing label propagation algorithm exhibits more resilient behavior than a triple modular redundancy (TMR) based fault-tolerant algorithm in 80% of cases. From a performance perspective, it also outperforms TMR as it requires fewer iterations in total. Beyond the fault-tolerance properties of self-stabilizing label-propagation, we believe, they are useful from the theoretical perspective; and may have other use-cases.

**Evaluating Compiler IR-Level Selective Instruction Duplication with Realistic Hardware Errors**

Chun-Kai Chang (University of Texas)

Hardware faults (i.e., soft errors) are projected to increase in modern HPC systems. The faults often lead to error propagation in programs and result in silent data corruptions (SDCs), seriously compromising system reliability. Selective instruction duplication, a widely used software-based error detector, has been shown to be effective in detecting SDCs with low performance overhead. In the past, researchers have relied on compiler intermediate representation (IR) for program reliability analysis and code transformation in selective instruction duplication. However, they assumed that the IR-based analysis and protection are representative under realistic fault models (i.e., faults originated at lower hardware layers). Unfortunately, the assumptions have not been fully validated, leading to questions about the accuracy and efficiency of the protection since IR is a higher level of abstraction and far away from hardware layers. In this paper, we verify the assumption by injecting realistic hardware faults to programs that are guided and protected by IR-based selective instruction duplication. We find that the protection yields high SDC coverage with low performance overhead even under realistic fault models, albeit a small amount of such faults escaping the detector. Our observations confirm that IR-based selective instruction duplication is a cost-effective method to protect programs from soft errors.

**Node-failure-resistant preconditioned conjugate gradient method without replacement nodes**
Carlos Pachajoa (University of Vienna)

As HPC systems grow in scale to meet increased computational demands, the incidence of faults in a given window of time is expected to grow. This issue is addressed by the scientific community with research on solutions in every computational layer.

In this paper, we explore strategies for fault tolerance at the algorithmic level. We propose a node-failure-tolerant preconditioned conjugate gradient method, which is able to efficiently recover from node failures without the use of extra spare nodes, i. e., without any overhead in terms of available hardware. For purposes of load balancing, we redistribute the surviving and reconstructed solver data. The objective is to reconstruct the system either as it was before the node failure, or an equivalent, permuted version, and then continue the execution of the solver only on the surviving nodes.

In our experimental evaluations, the recovery stage of the solver typically takes around 10% or less of the solver runtime, including the time to retrieve the problem-defining static data from the hard disk, and, when using a suitable preconditioner, an average solver runtime overhead of 3.5% over that of a resilient solver that uses a replacement node. We investigate the influence of the preconditioner on a trade-off between load-balancing and communication cost in the recovery phase. The obtained solutions are correct, and our method is thus a feasible way to recover from a node failure and continue the execution of the solver only on the surviving nodes.

Closing remarks
Scott Levy (Sandia National Laboratories)
Closing remarks on FTXS 2019.