Sixth International Workshop on Emerging Parallel Distributed Runtime Systems and Middleware

Welcome on behalf of the IPDRM organizing committee

Topics

Runtime System/Middleware Techniques. Design, and Evaluation

- Runtime/Middleware for exascale/large scale computing
- Runtime/Middleware for accelerators or appliances
- Network and I/O middleware technology
- Modeling and Performance Analysis of Runtime Systems
- Interactions between runtime and middleware
- Runtime-architecture co-design
- Tuning and optimization studies
- Workflow/application-centric challenges and solutions for runtime systems

Constraints and Issues for Runtime Systems and Middleware

- Energy- and Power-aware schemes
- Fault Tolerance and Reliability
- Heterogenous resource management
- Data movement
- Memory models
- Scalability

Design Principles and Programming Support

- High-level programming models (e.g., thread and task-based models, data parallel models, and stream programming) and domain-specific languages
- Programming frameworks, parallel programming, and design methodologies
- Methodologies and tools for runtime and middleware design, implementation , verification, and evaluation
- Wild and crazy ideas on future Runtime System and Middleware

Program

Time	Talk	Speaker
2:00-2:02 PM	Welcome	Barbara Chapman, HPE
2:00-3:00 PM	DeepSpeed4Science: Enabling Future Large-Scale Scientific Discovery through Sophisticated AI System Technologies	Shuaiwen Leon Song, Microsoft
3:00-3:30 PM	Break	
3:30-4:00 PM	Invited Talk: HPC Software Scaling for ML Using CXL 3.0 GFAM	Patrick Estep, Micron
4:00-4:20 PM	Dask-Extended External Tasks for HPC/ML In Transit Workflows	Amal Gueroudji, Julien Bigot, Bruno Raffin, Robert Ross
4:20-4:50 PM	Invited Talk: Enabling Large Dynamic Neural Network Training with Learning-Based Runtime Memory Management	Dong Li, University of California, Merced
4:50-5:10 PM	MPI-xCCL: A Portable MPI Library over Collective Communication Libraries for Various Accelerators	Chen-Chun Chen, Kawthar Shafie Khorassani, Pouya Kousha, Qinghua Zhou, Jinghan Yao, Hari Subramoni, Dhabaleswar K. Panda
5:10-5:28 PM	A gem5 Implementation of the Sequential Codelet Model: Reducing Overhead and Expanding the Software Memory Interface	Dawson Fox, Jose Monsalve Diaz, Xiaoming Li
5:28-5:30 PM	Closing	Oceane Bel, PNNL

Distinguished Speaker:
Shuaiwen Leon Song, Microsoft

Title: DeepSpeed4Science: Enabling Future Large-Scale Scientific Discovery through Sophisticated AI System Technologies

Bio: Shuaiwen Leon Song is a senior principal scientist and manager at Microsoft. He leads the effort of Deepspeed4Science initiative to create a broad engagement between Microsoft, Microsoft research, DoE labs, academia and industry partners to enable sophisticated system technology research and development for supporting aspects of training and inference for large-scale AI-driven scientific models. At DeepSpeed, he also drives or co-drives several pathfinding projects and releases (e.g., ZeRO inference, scalable dialogue system design and DeepSpeed Chat) and co-manages the Brainwave team. Prior to Microsoft, he was the SOAR associate professor at University of Sydney and an adjunct professor at University of Washington. His past works in HPC have received several best paper nominations and were featured in U.S. DoE research highlights and other media outlets. He was the recipient of several awards including IEEE early-career award for HPC, IEEE mid-career award for scalable computing, Facebook faculty award, Google brain faculty award, Australian most innovative engineer award, AIR global faculty award. He is also an ACM distinguished speaker.

Thanks and Hope to see you next Year!!!

- General Chairs
  - Barbara Chapman, HPE, USA
  - Shirley Moore, University of Texas at El Paso, USA
  - Eun Jun Park, Qualcomm, USA
  - Joseph Manzano, Pacific Northwest National Laboratory, USA

- Program Chair
  - Joshua Suetterlein Pacific Northwest National Laboratory, USA

- Publicity Chair:
  - Jose Monsalve Diaz, University of Delaware, Argonne National Laboratory, USA

- Publication Chair:
  - Oceane Bel, Pacific Northwest National Laboratory, USA

- Diversity Chair:
  - Cimone Wright-Hamor, Pacific Northwest National Laboratory, USA

- TPC

  - Kevin J. Barker, Pacific Northwest National Laboratory, USA
  - Mehmet E Belviranli, Colorado School of Mines, USA
  - Nicolas Bohm Agostini, Northeastern University, USA
  - Vincent Cave, Intel Corporation, USA
  - Serena Curzel, Polytechnic University of Milan, Italy
  - Ivy Peng, KTH Royal Institute of Technology, Sweden
  - Bin Ren, College of William & Mary, USA
  - Omer Subasi, Pacific Northwest National Laboratory, USA
  - Shubbhi Taneja, Worcester Polytechnic Institute, USA
  - Li Tang, Los Alamos National Laboratory, USA
  - Zhijia Zhao, University of California, Riverside, USA
  - Christopher Zimmer, Oak Ridge National Laboratory, USA
  - Stephane Zuckerman, CY Cergy Paris University, France

Advanced Memory to Support Artificial Intelligence for Science

Special thanks to AIAMS from Pacific Northwest National Laboratory for their support

- Memory has always been a second-class citizens

  - Concepts in memory research and tools (PIM / dataflow designs, memory profilers, memory centric runtimes, etc.) have been "in the works" for decades

- Heterogeneous architectures have exposed the need to orchestrate memory more carefully

  - We are reaching the limits of current memory hierarchies
  - Challenges include variable latencies, different coherence and consistency requirements, different computation and capacity capabilities, etc.

- Support by DOE ASCR office of science, *PNNL is collaborating with major industry partners* to explore end-to-end solutions to memory challenges via a comprehensive approach to accelerator-based software / hardware codesign

We are looking for collaborators in the form of interns, postdocs or general discussions in the field!
Andres.Marquez@pnnl.gov / Joshua.Suetterlein@pnnl.gov