ASCR Workshop on Extreme Heterogeneity
January 23-25, 2018

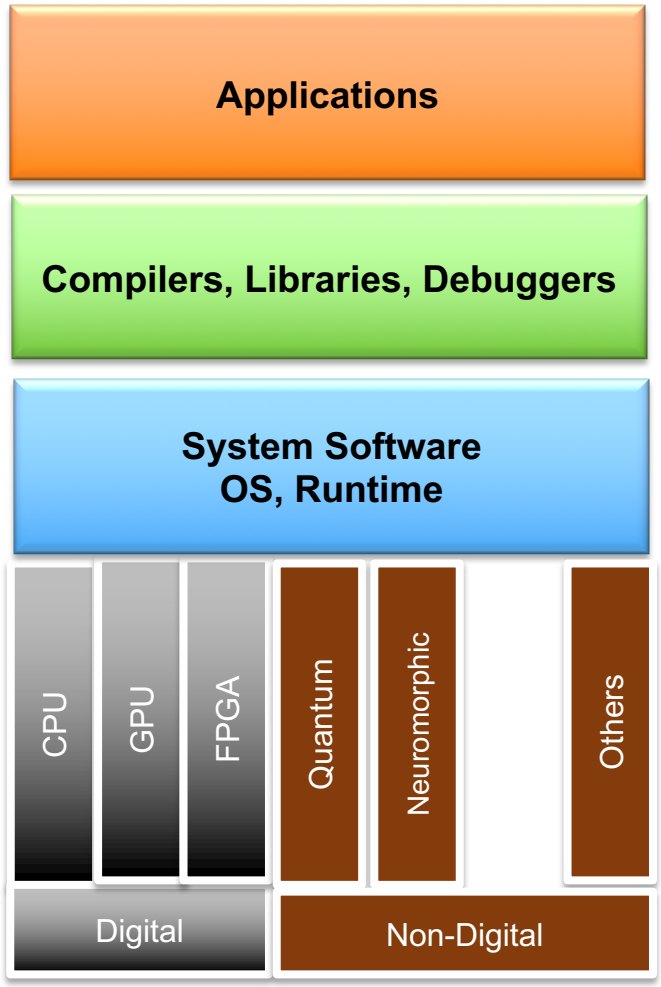
Storage Systems and I/O Workshop
Sept. 20, 2018

Lucy Nowell, PhD
Computer Science Program Manager
Advanced Scientific Computing Research

ASCR Extreme Heterogeneity Workshop

January 23-25, 2018 Virtual Meeting

- **Goal: Identify Priority Research Directions for Computer Science needed to make future supercomputers usable, useful and secure for science applications in the 2025-2040 timeframe**
- Primary focus on the software stack and programming models/environments/tools.
- Factual Status Document (FSD) [Pre-Workshop Report]



Workshop Organizers & Program Committee

- Jeff Vetter (ORNL), Lead Organizer and Program Committee Chair
- Organizing Committee and Program Committee Members: Ron Brightwell (Sandia-NM), Pat McCormick (LANL), Rob Ross (ANL), John Shalf (LBNL)
- Program Committee Members: Katy Antypas (LBNL, NERSC), David Donofrio (LBNL), Maya Gokhale (LLNL), Travis Humble (ORNL), Catherine Schuman (ORNL), Bryan VanEssen (LLNL), Shinjae Yoo (BNL)

Workshop Charge Excerpts

The purpose of this workshop is to identify the priority research directions for ASCR in providing a smart software stack that includes techniques, such as deep learning to make future computers composed of a variety of complex processors, new interconnects and deep memory hierarchies easily used by a broad community of computational scientists…The primary aim for the workshop is to **identify the new algorithms and software tools needed from basic research in computer science to enable ASCR's supercomputing facilities to support future scientific and technological advances on SC program's grand challenge problems**. ASCR's grand challenges and the resulting priority basic research directions should be identified by spanning existing and next generation computer architectures, including novel technologies that may be developed in the "Post-Moore's Law era" and the promising tools and techniques that are essential to efficient and productive utilization of such architectures. The workshop and subsequent report should **define basic research needs and opportunities in computer science research to develop smart and trainable operating and runtime systems, execution models, and programming environments that will make future systems easier to tailor to scientists' computing needs and for facilities to securely deploy**.

What Do We Mean by Extreme Heterogeneity?

- **Exponentially Increasing Parallelism** (central challenge for ECP, but will be even worse)
  - **Trend**: *End of exponential clock frequency scaling (end of Dennard scaling)*
  - **Consequence:** *Exponentially increasing parallelism*
- **End of Lithography as Primary Driver for Technology Improvements**
  - **Trend:** *Tapering of lithography Scaling*
  - **Consequence:** *Many forms of heterogeneous acceleration (not just GPGPUs anymore)*
- **Data Movement Heterogeneity and Increasingly Hierarchical Machine Model**
  - **Trend**: *Moving data operands costs more than computation performed on them*
  - **Consequence**: *More heterogeneity in data movement performance and energy cost*
- **Performance Heterogeneity**
  - **Trend**: *Heterogeneous execution rates from contention and aggressive power management*
  - **Consequence**: *Extreme variability and heterogeneity in execution rates*
- **Diversity of Emerging Memory and Storage Technologies**
  - **Trend**: *Emerging memory technologies and stall in disk performance improvements*
  - **Consequence**: *Disruptive changes to our storage environment*
- **Increasingly Diverse User Requirements**
  - **Trend**: *Diverse and Complex and heterogeneous scientific workflows*
  - **Consequence**: *Complex mapping of heterogeneous workflows on heterogeneous systems.*

Productive Computational Science in the Era of Extreme Heterogeneity: Revolutionizing How We Utilize Leadership Class Computers for Scientific Discovery

Drivers

- **Scientific**
  - Increasingly diverse science workflows
  - Increasingly diverse science users
- **Technological**
  - Market-driven innovation
    - Data driven workloads
    - AI/ML
  - Market-based constraints
    - Conserve power & heat
    - International competition

Key Questions

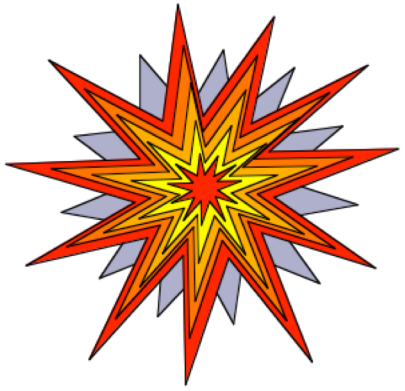
- What will improve productivity of scientific software developers?
- Can AI /ML help coordinate & control diverse computing resources?
- Can advanced mod-sim predict performance for application designers & improve system ROI?
- What will improve verifiable scientific findings in the future?
- What software infrastructure will we need for productive scientific workflow across multiple complex computing environments?

Priority Research Directions

- **Maintaining and improving programmer productivity**
- **Managing resources intelligently**
- **Modeling & predicting performance**
- **Reproducible science**
- **Facilitating data management, analytics, & workflows**

Facilitating Data Management, Analysis, & Workflows

- **Workflow execution: EH systems and applications bring significant challenges to usability through an unprecedented variety and number of data, resources, and services.**
  - **Discovery and mapping of science workflow requirements to appropriate data, hardware, and software services**
  - **New methods of composing scientific workflows (e.g., via workflow motifs)**
  - **Interfaces that facilitate HW and service composition – programming environments for data services**
  - **Ties to Programmer Productivity**
- **Autonomous workflows: Extracting the highest value from EH systems requires rapid, complex balancing across a wide variety of storage and networking technologies in response to a widely varying workload and significant rate of faults.**
  - **Profiling workflow telemetric data and learn what and how to instrument these systems**
  - **Online learning to adapt computation and data organization to available resources and emerging results**
  - **Transfer learning of the knowledge of workflow optimization to other types of workflows and EH systems**
  - **Ties to Managing System Resources**
- **Rapid adaptive analytics: Data analytic workflows on EH systems must employ new, specialized algorithms and data organizations and facilitate use by domain scientists, not just data scientists.**
  - **Data and algorithm abstractions for domain scientists (e.g., probabilistic execution vs deterministic execution)**
  - **New data organizations and specialized robust algorithm developments for specific EH components**



Thank you!

Lucy Nowell

lucy.nowell@science.doe.gov

https://science.energy.gov/ascr/