



Hardware Architecture Group

Scott Hemmert (SNL), Patrick Chiang (OSU), Ken Koch (LANL), Parks Fields (LANL), Cheryl Wampler (LANL), Jeff Vetter (ORNL), Doug Doerfler (SNL), Brian Carnes (LLNL), Matt Leininger (LLNL)



Working Group Scope



- Inform other working groups on hardware challenges that will impact various levels of stack
- Owner of abstract machine description
 - Follow vendor roadmaps, feed back ideas to vendors
- Facilitate co-design with vendors
 - Architectural simulators, emulators, test beds, prototypes
 - Understand direction of applications and help to ensure future architectures will enable useful science
- Identify other markets that have similar requirements in order to help influence industry



Working Group Dependencies



- System Software
- Solvers, Algorithms & Libraries
 - Tie to applications, centroid of vendor to app co-design
- Programming Models
- I/O & Networking
 - Mainly resiliency issues
- Tools
 - What performance, power and resilience counters would be useful?
- Vendors
 - Business strategy issues
 - Roadmaps



Exascale Challenges



- Technical Challenges
 - Power/Energy
 - Memory
 - Increased Parallelism
 - Interconnect/Data Movement
 - Resilience

- Non-technical Challenge
 - Exascale requirements are diverging from commodity requirements
 - Games, multimedia, streaming and internet drive vendors
 - SIMD widths
 - Resilience



Power and Energy



- Drives many of the other challenges
- Power constraints (technical)
 - What can be cooled
 - Chip
 - Rack
 - Facilities issues
 - Data Center cooling
 - Data Center electrical power limits
- Energy constraints (non-technical)
 - Primary constraint from DOE
 - Operation Costs
 - Facilities upgrades
 - Floor space, power, cooling
 - Siting issues (utilities companies)



Increased Parallelism



- Move to simpler cores
- Levels of parallelism
 - Inter-node
 - Intra-node
 - Latency hiding mechanisms
 - Vector/SIMD
 - Wider units will require scatter/gather to be useful to many apps
- Applications will identify more parallelism, but it needs to be less bulk synchronous.
 - Non-uniformity
 - Adaptivity
 - Hardware support for scheduling
 - Load balancing
 - Work queue models



Memory



- Reduced Memory Capacity Ratios
 - Technological
 - Area issues involved with stacking sufficient fast memory
 - Cost issues
 - Total memory cost
- More memory hierarchy
 - Unknown market for multi-level memory (stacked + DIMMs + NVRAM)
 - Chip connectivity constraints
- Memory Power
 - Memory predicted to consume a large portion of system energy
- Pre-exascale machine may have different characteristics because of unavailability of stacked memories in 2015



Interconnect/Data Movement



- Data movement will dominate energy consumption
 - Need to reduce transmission and switching energy
- System level interconnect needs better power management capabilities
 - Ability to shut down links and quickly restart
- Apps need to more efficiently utilize network bandwidth
 - More smaller messages, more asynchronous, more computation communication overlap
- Scalability of collective operations
 - Machine size will increase to over 100,000 nodes



Resilience



- Key cross cutting topic that hasn't seemed to get enough attention at this series of workshops
- Increased part counts drive down MTBF
 - Need to maintain reasonable MTTI
 - Increased research needed
- How does resilience map across the many layers of the system?
 - How much do hardware, system software, libraries, apps worry about
- Extreme case of commodity drivers not matching exascale drivers
- Power constraints conflict to some extent with resilience constraint
 - Near threshold voltage operation, etc



Path Forward





Memory Technology Development



- Project: Lower power, higher bandwidth, more efficient, 3D stacked memory
- Timeline: Hit in Phase II, revolutionary approach
- Required Funding: \$\$\$\$
- Need to partner with memory vendor
 - Requires processor vendor to support resulting technology
- Risks
 - Funding profile would require other agency participation



Silicon Photonics



- Proposed technology: lower power, lower cost optical interconnects
 - Questions remain: to package, to die, other...
- Timeline: Phases II and III
- Required Funding
 - Probably small for one-off demonstrators
 - Probably large for production capabilities
- Risks
 - Will it be commercially viable in early exascale timeframe?
- Does DOE need independent project, or can we join other agencies??



Recommended Co-Design Strategy





Using Mini-Apps to Drive Co-design



- Primarily work between HWA & SAL and SAL & Apps
- SAL roles
 - Distill mini-apps and skeleton apps from applications working with Apps group
 - Along with context
 - Provide feedback to apps about architectural directions and ways to exploit them
 - Be available to vendors to help ensure they understand the operating environment
- HWA roles
 - Provide architecture information to SAL
 - Facilitate interaction with vendors using mini-apps
 - Early access to new platforms
 - Extract important architectural features of interest from mini-apps



Defining APIs to Enable Innovation



- Goal is to distill application requirements in various areas to functional requirements for the architecture, system software, etc.
 - Do we know what 10 functions we would put in hardware?
 - What would an ideal system look like?
- Groups involved depend on API
- Examples: Communication API, power management API, Resiliency API, Storage Class Memory API



Understanding Memory Hierarchy Issues



- Primarily work between HWA & SAL
- SAL role
 - Explore algorithms that can take advantage (or tolerate depending on your point of view) multiple levels of memory
- HWA role
 - Provide relative capacity and performance characteristics of different memory levels



Big Picture Issues



- Coordination
 - Abstract machine model
 - Test beds
 - Simulators
- Remaining gaps
 - Resilience co-design strategy
- Others
 - IP issues in vendor interactions