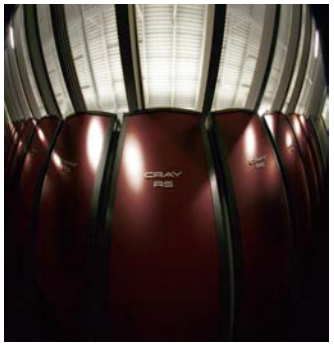
Exceptional service in the national interest

Sandia National Laboratories






Exascale Runtime System Architecture and Design

Ron Brightwell

R&D Manager, Scalable System Software Department

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Outline

- Vision for Exascale runtime systems
- Issues/Concerns

Outline

- Fantasy
- Reality

Existing Approach

- Generalized abstractions and machine models that allow algorithm designers and application developers to create code that works reasonably well on a broad spectrum of systems
- Compilers, libraries, RTS, and OS work within the constraints of these abstractions to map the application to the underlying hardware as efficiently as possible
- Performance tools identify shortcomings in the mapping
- Refine the mapping on a per-platform basis
- Adjust the abstractions and models in response to evolving hardware
- Leverage RTS adaptivity within bounded set of resources and relatively fixed cost models

Vision for Exascale Runtime Systems

- Responsible for mapping the machine to the application
- Require dynamic discovery
  - Determine the goals of the application
  - Develop knowledge on how well resources are being used
  - Make informed optimization decisions
  - Understand behavior in response to decisions
  - Constantly changing cost models
- Respond to elastic system and application resources
- Richer abstractions and models at the system level
- Improve the productivity of application and library developer as well as the scalability and efficiency of the system

Transition to Reality

- Realizing this vision, even for current systems, is a daunting task that should involve the exploration of fundamentally new and more holistic approaches that are informed, but not encumbered, by current methods. The expectation is that discoveries and the knowledge gained through more forward-looking research will continue to inform and accelerate the path of incremental advancements, at least until the gravity of the breakthroughs causes a fundamental shift to more promising alternatives. Realizing this vision cannot be done by research in runtime systems alone, as there are strong connections and dependencies on other critical aspects of application development, system software, and architecture, including parallel programming models, code generation and optimization, and operating systems.

Issues/Concerns

- Managing the memory hierarchy
  - Lots of evidence that the RTS/OS are not good at this for HPC
- Increasing complexity and responsibility of the RTS
  - Pushing complexity to the RTS with less info
- Resource requirements of the RTS
  - Potentially significant overhead
- Compelling application evaluation
  - Applications need to exercise the advanced RTS functionality
  - Implementation bias
- Application performance portability
  - From laptop to exascale
- Transparency is in the eye of the application developer
  - Need to support both experts and ambivalent

More Issues/Concerns

- Cost of modularity
  - Not all RTS services should be componentized
- Ability to constrain the problem
  - Too many hardware and application “knobs”
- Performance portability of the RTS
  - Not any easier to solve than application performance portability
- Dependence on hardware advancements
  - Inability to demonstrate compelling results on current systems
- Lack of standard low-level network API
  - Fundamental issue for RTS communication
- HPC market pressure
  - Influence of non-HPC “solutions”

Final Issues/Concerns

- Amount of asynchrony
  - Ability of algorithms to reduce global operations
- Jitter
  - Will lack of balance absorb noise effects
- Mechanisms to support event-driven capability
  - More efficient ways to enable adaptivity
- Walking before running
  - Make progress at small scale while working towards large scale