Sandia Initiatives in HPC Architectures

Overview for

Dr. Carlos Dengo
VP Geoscience,
ExxonMobil Upstream Research Company

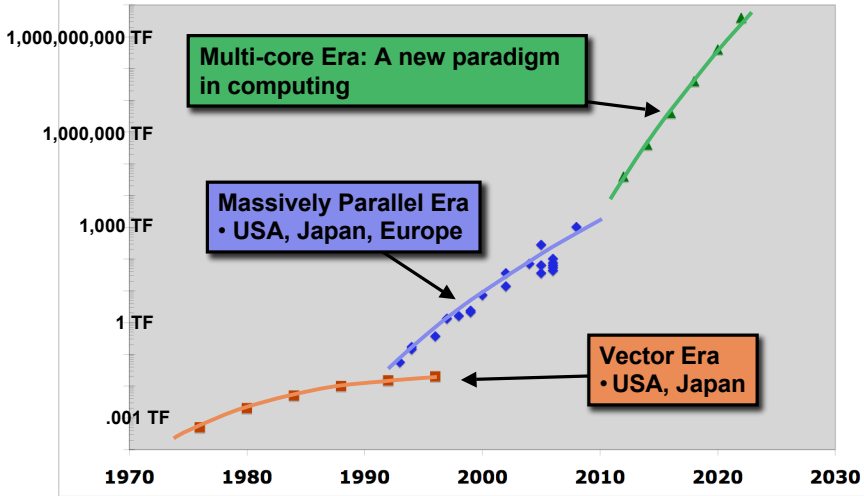
November 3, 2008

James A. Ang, Ph.D.
Manager, Scalable Computer Architectures
Computation, Computers, and Mathematics Center

Sandia is a Multiprogram Laboratory Operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy Under Contract DE-ACO4-94AL85000.

Sandia National Laboratories

Sandia's Computer Architecture R&D is focused on our transition to a new era of computing



Sandia National Laboratories

Industry Trends

Existing industry trends not going to meet HPC application needs

Semi-conductor industry trends

- Moore's Law still holds, but clock speed now constrained by power and cooling limits
- Processors are shifting to multi/many core with attendant parallelism
- Compute nodes with added hardware accelerators are introducing additional complexity of heterogeneous architectures
- Processor cost is increasingly driven by pins and packaging, which means the memory wall is growing in proportion to the number of cores on a processor socket

Development of large-scale Leadership-class supercomputers from commodity computer components requires collaboration

- Harder to integrate commodity components into a large scale massively parallel supercomputer architecture that performs well on full scale real applications
- Leadership-class supercomputers cannot be built from only commodity components
- Supercomputer architectures must be designed with an understanding of the applications they are intended to run

Sandia National Laboratories

Software Trends

Science is getting harder to solve on Leadership systems

Application trends

- Scaling limitations of present algorithms
- More complex multi-physics requires large memory per node
- Need for automated fault tolerance, performance analysis, and verification
- Software strategies to mitigate high memory latencies
- Hierarchical algorithms to deal with bandwidth across the memory hierarchy
- Innovative algorithms for multi-core, heterogeneous nodes
- Model coupling for more realistic physical processes

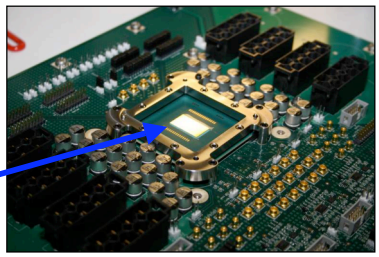
Emerging Applications

- Growing importance of data intensive applications
- Mining of experimental and simulation data

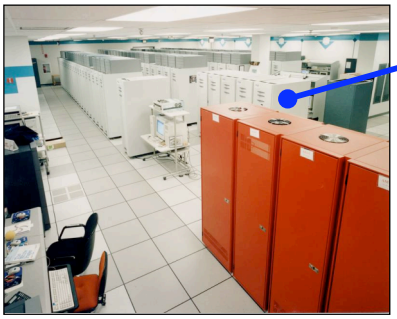
Sandia National Laboratories

Moore's Law + Multicore → Rapid Growth in Peak Computing Power

2007 - Intel Polaris R&D Processor
1 TeraFLOPs on a chip
• 275 mm² (size of a dime) & 62 W



1997 - Intel ASCI Red
1 TeraFLOPs in a house-size room
• 2,500 ft² & 500,000 W



Sandia National Laboratories

The Need for HPC Innovation and Investment is Well Documented

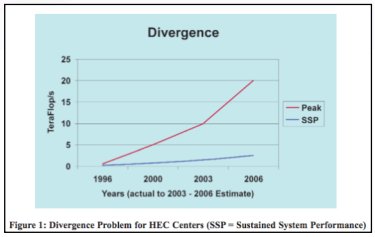
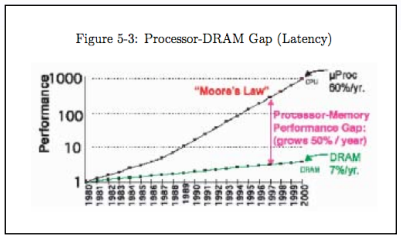
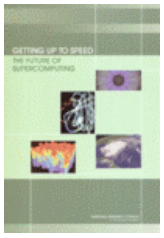


Figure 1: Divergence Problem for HEC Centers (SSP = Sustained System Performance)

Report of the High-End Computing Revitalization Task Force (HECRTF), May 2004



"Requirements for ASCI", JASONs Report, Sept 2002



National Research Council, "Getting Up To Speed The Future of Supercomputing", Committee on the Future of Supercomputing, 2004

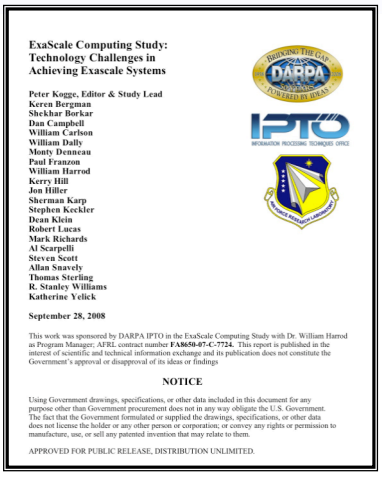
"Recommendation 1. To get the maximum leverage from the national effort, the government agencies that are the major users of supercomputing should be jointly responsible for the strength and continued evolution of the supercomputing infrastructure in the United States, from basic research to suppliers and deployed platforms. The Congress should provide adequate and sustained funding."

Sandia National Laboratories

Drivers for Sandia's Computer Architecture R&D

- The DARPA/IPTO *Exascale Computing Study*, is the latest report to describe four major technical challenges for HPC Computer Architectures:
  - Energy and Power
  - Memory and Storage
  - Concurrency and Locality
  - Resiliency

- In 2008 Sandia established two Strategic Collaborations in Computer Architectures
  - ACES (SNL & LANL)
  - IAA (SNL & ORNL)



ExaScale Computing Study:
Technology Challenges in
Achieving Exascale Systems

Peter Kogge, Editor & Study Lead
Keren Bergman
Shekhar Borkar
Dan Campbell
William Carlson
William Dally
Monty Denneau
Paul Franzon
William Harrod
Kerry Hill
Jon Hiller
Sherman Karp
Stephen Keckler
Dean Klein
Robert Lucas
Mark Richards
Al Scarpelli
Steven Scott
Allan Snavely
Thomas Sterling
R. Stanley Williams
Katherine Yelick

September 28, 2008

This work was sponsored by DARPA IPTO in the ExaScale Computing Study with Dr. William Harrod as Program Manager; AFRL contract number FA8650-07-C-7724. This report is published in the interest of scientific and technical information exchange and its publication does not constitute the Government's approval or disapproval of its ideas or findings

NOTICE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

APPROVED FOR PUBLIC RELEASE, DISTRIBUTION UNLIMITED.

Sandia National Laboratories

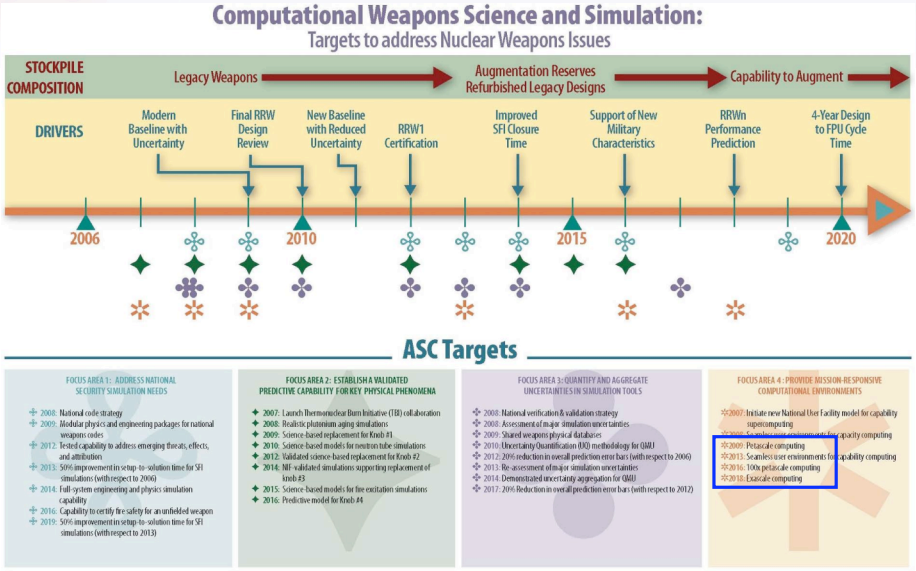
Motivation for ACES Partnership

- 3/7/2008: SNL & LANL Lab Directors sign Memorandum of Understanding to establish ACES: The New Mexico Alliance for Computing at Extreme Scale
  - Joint design, architecture, development, deployment and operation of production capability systems for NNSA/ASC
- Driven by mission needs
- Commitment to the development and use of world class computing
- Continued leadership in high performance computing
- Sharing intellectual capabilities of both laboratories

Sandia National Laboratories

www.nnsa.doe.gov/ASC

ASC Roadmap



Institute for Advanced Architectures and Algorithms

IAA was established by Congressional Legislation in FY08 with Centers of Excellence at SNL and ORNL to co-develop architectures and algorithms and create synergy in their respective evolutions.

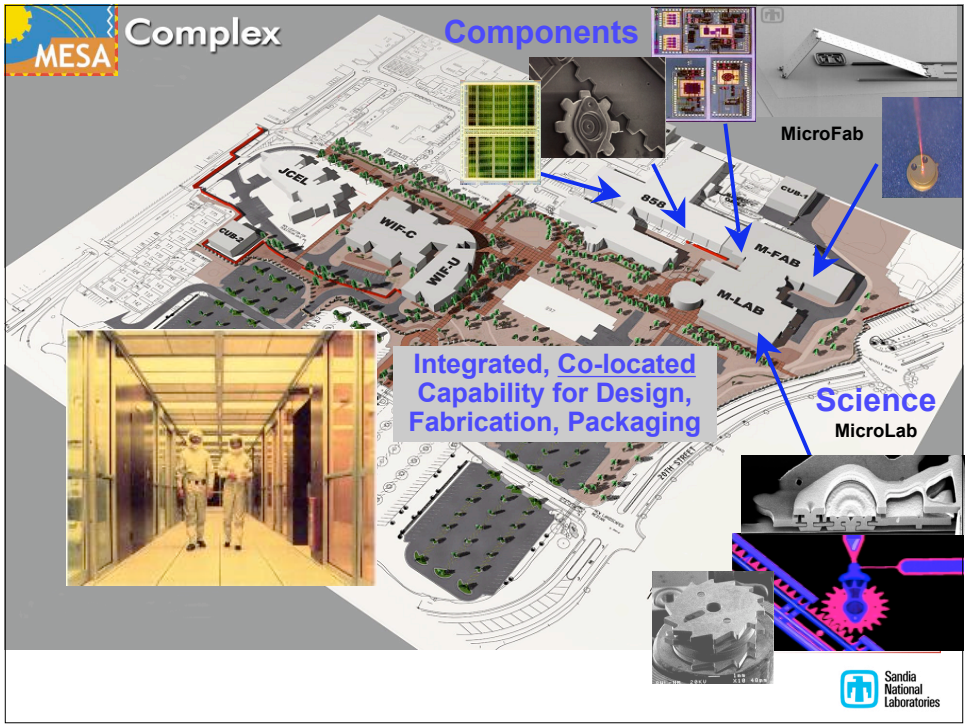
- **Focused R&D on key impediments to high performance in partnership with industry and academia**
- **Foster the integrated co-design of architectures and algorithms to enable more efficient and timely solutions to mission critical problems**
- **Partner with other agencies (e.g., DARPA, NSA …) to leverage our R&D and broaden our impact**
- **Impact vendor roadmaps by committing National Lab staff and funding the Non-Recurring Engineering (NRE) costs of promising technology development and thus lower risks associated with its adoption**
- **Train future generations of computer engineers, computer scientists, and computational scientists, thus enhancing American competitiveness**
- **Deploy prototypes to prove the technologies that allow application developers to explore these architectures and to foster greater algorithmic richness**

Sandia National Laboratories

Sandia Brings Multi-disciplinary Capabilities to our ACES and IAA Collaborations

- Partnerships with industry, as opposed to contract management
- Cuts across DOE and other government agencies and laboratories
- A focus on impacting commercial product lines
  - National competitiveness
  - Impact on a broad spectrum of platform acquisitions
- A focus on problems of interest to DOE
  - National Security
  - Science
- Sandia, LANL and ORNL have capabilities across a broad and deep range of disciplines
  - Applications & Algorithms
  - System performance modeling and simulation
  - Application performance modeling
  - System software
  - Computer architectures
- Sandia also has a Unique Capability - Microelectronics Fab …

Sandia National Laboratories



Sandia National Laboratories

Official Use Only

Simulator Project

Long Term Vision: Become **the** HPC community standard simulator

Near Term-Goals

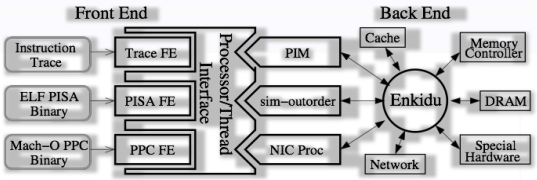
- Prototype parallel simulator
- x86 Front-/Back-end models
- Integrate MPI Models
- Tracing for Interconnect Sim.

Long Term Goals

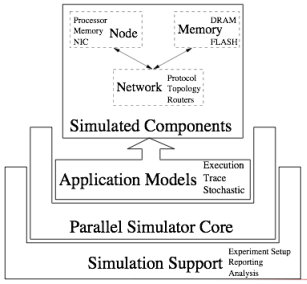
- Highly scalable parallel simulator
- Multi-scale simulation
- Technology model interface

Partners

- B. Jacob (U. Maryland): Improve DRAM model
- S. Yalamanchili (Georgia Tech): Parallel SST
- D. Chiou (Texas): FPGA Acceleration of Simulation



The modular simulation structure allows flexible simulation

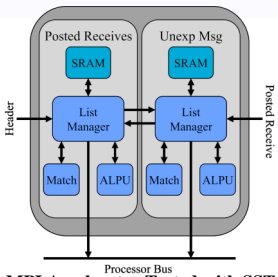


Sandia National Laboratories

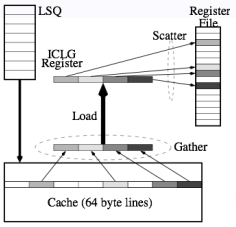
Official Use Only

Projects Supported

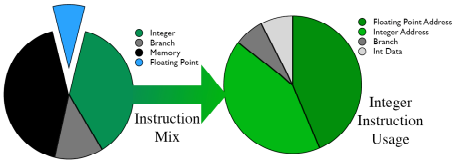
- **Microarchitecture**
  - **Inter CacheLine Gather (ICGL)**
  - **Reconfigurable FU (Wisc./SNL)**
  - **FP Aggregates**
  - **In-Memory Ops**
- **Application analysis**
  - **Memory Footprint**
  - **Instruction Usage**
- **Network/MPI**
  - **NIC Tradeoffs**
  - **MPI Acceleration**
- **Programming Models**
  - **PIM Compiler work (SNL/Rice)**
  - **ParallX (LSU/SNL)(FastOS)**
  - **Transactional Memory (ORNL)**
  - **QThreads**
- **Processor-In-Memory (LDRD)**



MPI Accelerator Tested with SST



Inter CacheLine Scatter/Gather Microarchitecture



Instruction Mix & Usage for Sandia Applications

Sandia National Laboratories

CSRI Facilities for Collaboration

"Where Research meets Innovation"

- **The facility**
  - **In Sandia Science and Technology Park (off KAFB)**
  - **34,500 square feet**
  - **140 staff**
  - **50 collaborators**
  - **Large conference room**
  - **Additional collaborative work areas**

Sandia National Laboratories

Backup Slides

Sandia National Laboratories

Industry Support for IAA

"The US-DOE has long invested in R&D for High Performance Computing and in systems that build on that R&D. With the creation of the IAA, the US-DOE has the opportunity to explore, foster, and sponsor needed advances in hardware and software architectures that will serve companies like Intel well as we strive to maintain US intellectual and business leadership in computing and information." Pat Gelsinger, Senior Vice President, Intel

"We fully support the IAA as we believe it is truly critical to do joint research and develop key leading-edge technologies in order to optimize ultra-scale systems to advance the mission of the Department of Energy and U.S. leadership in supercomputing. Having the IAA jointly run by Sandia and ORNL will ensure that future peta and exascale systems will be viable across the wide breadth of applications in NNSA and the Office of Science. We are looking forward to working with the Institute leveraging this R&D into our future products, just as we have successfully done with the NNSA's Red Storm program which enabled a new line of Cray Supercomputers (the XT) and enabling it for the Leadership Computing program at the Office of Science." Pete Ungaro, President and CEO, Cray Inc.

Sandia National Laboratories

Industry Support for IAA

- *"Micron Technology supports the efforts of the IAA. Micron recognizes that advanced memory and system architecture will be required to meet upcoming high-performance computing needs and further believes that such architectural solutions will find broad commercial application. Micron looks forward to continuing advanced memory and system architecture work with Sandia and Oak Ridge."* Dean Klein, VP Memory System Development, Micron Technology
- *"SGI and the entire industry recognize the innovation and leadership Sandia National Laboratories and Oak Ridge National Laboratory have provided to the HPC industry in computation, storage and visualization. We are very pleased to support the proposal for an Institute for Advanced Architectures and Algorithms and believe that it will be a benefit to both the industry and the nation."* Bo Ewald, CEO, SGI
- *"Given the massive changes underway in the development of computer systems, IBM is pleased that DOE has identified several key areas of concern and launched the IAA. The initial focus areas agree with the challenges we see and we believe there is value in having participation from industry to complement the expertise present in the DOE labs and academia."* David Turek, Vice-President, IBM

Sandia National Laboratories

ACES Alliance Organization Chart

Senior Executives:
Charlie McMillian, LANL / ADWP
Rick Stulen, SNL / CTO & VP-STE

NNSA Defense Programs

Governance Board:
James Peery, SNL / CCIM & ASC
Rob Leland, SNL / 9300
Andy White, LANL / ADTSC
John Hopson, LANL / ASC

ASC Headquarters:
Bob Meisner, ASC Program Director

Los Alamos Sandia

Co-Directors:
John Morrison, LANL / HPC
Sudip Dosanjh, SNL / CCIM

Advisory Board
(TBD)

Computer Architecture Office
Sited at Sandia
SNL Office Leader, Jim Ang
LANL Deputy Leader, Karl-Heinz Winkler
Joint SNL & LANL team members

Deployment and Operations Office
Sited at Los Alamos
LANL Office Leader, Nick Nagy
SNL Deputy Leader, John Zepper
Joint LANL & SNL team members

User Requirements Office
Mike Haertling, LANL
Ken Alvin, SNL

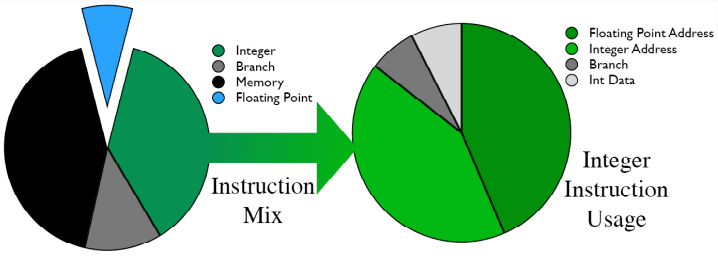
Sandia National Laboratories

ACES Architecture Office

- **Led by Jim Ang (SNL) and Karl-Heinz Winkler (LANL)**
- **Approximately 10 participants from each laboratory**
- **Developed a LANL/SNL platform strategy document**
- **Numerous meetings and discussions on**
  - **RFP for a 2010 petascale production capability platform**
  - **Areas for technology development**
  - **2014 production capability platform for predictivity**
- **Half-day to full day meetings with:**
  - **AMD, Appro, Cisco, Open MPI, Corning, Cray, HP, IBM, Intel, Luxtera, Micron, Myricom, Panasas, Qlogic, Seagate, SGI, Spansion, Sun**

Sandia National Laboratories

The Memory Wall significantly impacts the performance of our applications



- **Most of DOE's Applications (e.g., climate, fusion, shock physics, …) spend most of their instructions accessing memory or doing integer computations, not floating point**
- **Additionally, most integer computations are computing memory Addresses**
- **Advanced development efforts are focused on accelerating memory subsystem performance for both scientific and informatics applications**

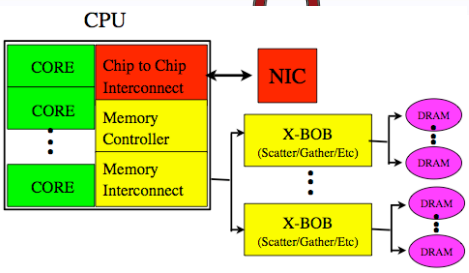
Sandia National Laboratories

Memory Project

Vision:
- Create a **commodity** memory part with support for HPC data movement operations

Approach:
- New high-speed memory signaling technology inserts an ASIC (the Buffer-on-Board, or BOB) between the CPU and memory
- Add data movement support in the ASIC

X-BOB



Near Term Goals:
- Define in-memory operations (scatter/gather, atomic memory operations, etc.)
- Define CPU/X-BOB coherency

Long Term Goals:
- Create a commodity memory part that increases **effective** bandwidth utilization

Potential Partners:
- Industry: Micron (since June 05), Intel, IBM, Cray, SUN, AMD
- Academia: USC/ISI (Draper/Hall), LSU (Sterling)

Sandia National Laboratories

Interconnect Project

Vision: Ensure next generation interconnects satisfy HPC needs

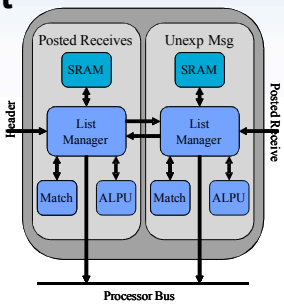
Approach: Provide understanding of application needs, explore designs with simulation, prototype features with vendors

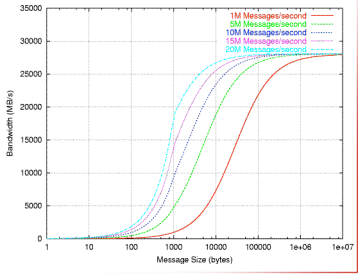
Near Term Goals:

- Identify interconnect simulation strategy
- Characterize interconnect requirements on mission apps
- Develop MPI models & tracing methods
- Pursue small collaboration project with industry partner

Long Term Goals:

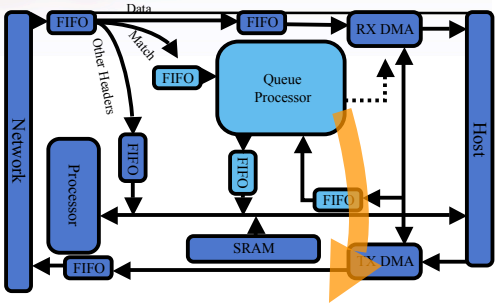
- *Scalability*: >100,000 ports (including power, cabling, cost, failures, etc.)
- *High Bandwidth*: 1TF sockets will require >100GBps
- *High Message Throughput*: >100M for MPI; >1000M for load/store
- *Low Latency*: Maintain ~1us latency across system
- *High Reliability*: <10^{-23} unrecovered bit error rate

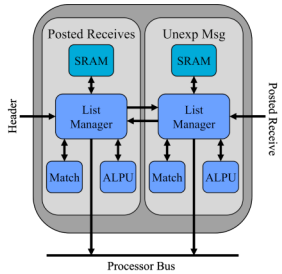




Sandia National Laboratories

MPI Acceleration

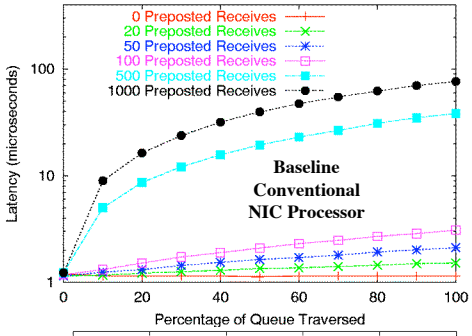


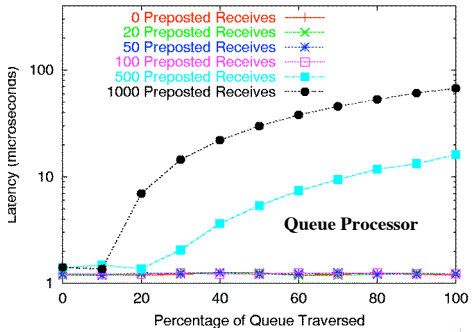


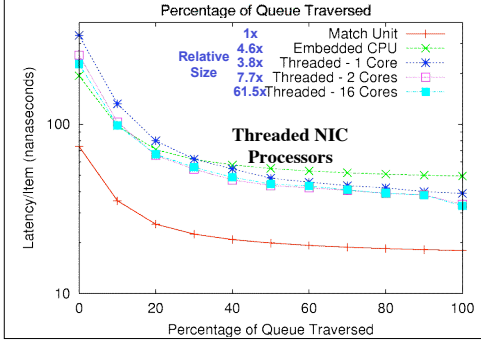
- **Problem: Message rate determines <u>effective</u> bandwidth**
- **SST Analysis: MPI matching limits performance**
- **Solution: Accelerate list management with hardware**
- **SST Simulation**
  - **Hardware**
    - **Implement NIC/Router based on RedStorm SeaStar**
    - **Implement List Manager, ALPU, and Match Unit, integrate with NIC**
      - **"Translated" from FPGA RTL**
    - **Validate against FPGA & Red Storm**
  - **Software**
    - **Create baseline offload MPI**
    - **Modify MPI to use acceleration HQ**
- **Impact: Collaboration w/ Intel**
  - **Intel licensing**
  - **Keith Underwood "on loan" to Intel**
  - **Foundation for IAA network project**

Sandia National Laboratories

MPI Acceleration Results





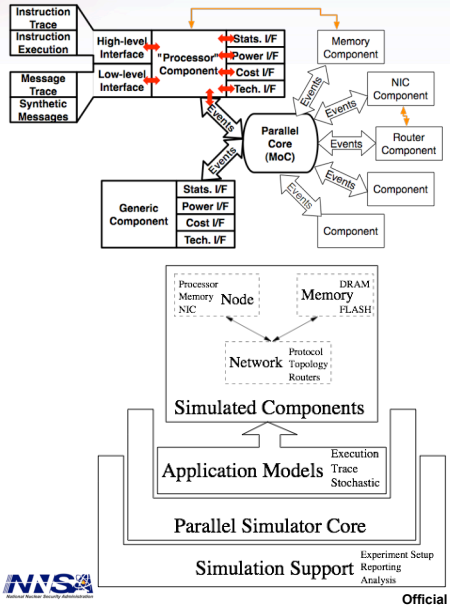


- Long MPI message queues increase effective latency dramatically
- Queue processor processes messages more quickly
- Queue processor more area effective than conventional or threaded NIC processor

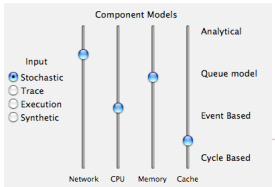
Sandia National Laboratories

Official Use Only

Planned Simulator Activities



- Project: Create a parallel, multi-scale architecture simulator
- Purpose
  - Assist in the design and procurement of large-scale HPC systems
  - Enable application/algorithm analysis and design
- Approach
  - Open-Source, vendor-neutral
  - Scalable parallel simulator
  - Multi-scale
  - Modular
- Long Term Vision: Become the HPC community standard simulator



NNSA

Sandia National Laboratories

Official Use Only