The Institute for Advanced Architectures and Algorithms (IAA)

Sudip S. Dosanjh
Director
Center of Excellence
Institute for Advanced Architectures and Algorithms
Sandia National Laboratories
Albuquerque, NM

Jeff Nichols
Director
Center of Excellence
Institute for Advanced Architectures and Algorithms
Oak Ridge National Laboratory
Oak Ridge, TN

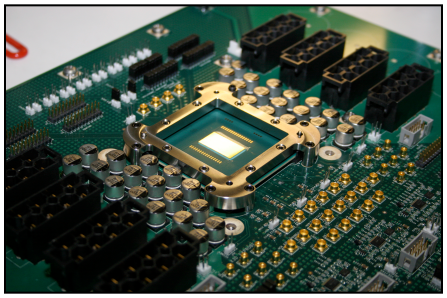
Sandia is a Multiprogram Laboratory Operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy Under Contract DE-ACO4-94AL85000.

Nationally, the DOE Institute for advanced Architectures and Algorithms (IAA) is moving forward

- Language signed by President in FY08 E&W Appropriations bill creates DOE Institute for Advanced Architectures
  - Centers of Excellence at SNL and ORNL
- Memory Workshop (1/08)
- Presentation to DOE HQ (1/08)
- IAA Proposal Approved (3/08)
- Interconnect workshop (7/08)
- Architecture-Aware Algorithms Project Approved (9/08)
- IAA Advisory Committee Formed (1/09)
  - Detailed set of recommendations
- HQ Review (2/09)
- CS/Math Institute Proposal Funded by ASCR (7/09)
- System Simulation Workshop (9/09)
- IAA Advisory Committee Meetings (9/09 and 1/10)

1997 – 1 Teraflop in a room
- 2,500 ft^2 & 500,000 Watts

Grand Challenge: Getting the science done by the system above done by the chip below. IAA is designed to take on this challenge.



2007 – 1 Teraflop on a chip
- 275 mm^2 (size of a dime) & 62 Watts

Oak Ridge National Laboratory

Sandia National Laboratories

Advisory Committee

- Responsibility
  - Help Directors and Steering Committee set research direction
- Meetings
  - 1/09 in Santa Fe
  - 9/09 in Boulder
  - 1/10 in Knoxville
- Members
  - Almadena Chtchelkanova (NSF)
  - Candy Culhane (DoD)
  - Jack Dongarra (UT)
  - Bill Gropp (UIUC)
  - Charlie Holland (DARPA)
  - Dean Klein (Micron)
  - Peter Kogge (Notre Dame)
  - John Morrison (LANL)
  - Kunle Olukotun (Stanford)
  - Steve Scott (Cray)
  - Marc Snir (UIUC)
  - Thomas Sterling (LSU)
  - Craig Stunkel (IBM)
  - Jim Sundet (DoD)

IAA Mission and Strategy

IAA is the medium through which architectures and applications can be co-designed in order to create synergy in their respective evolutions.

- Focused R&D on key impediments to high performance in partnership with industry and academia
- Partner with other agencies (e.g., DARPA, DoD…) 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

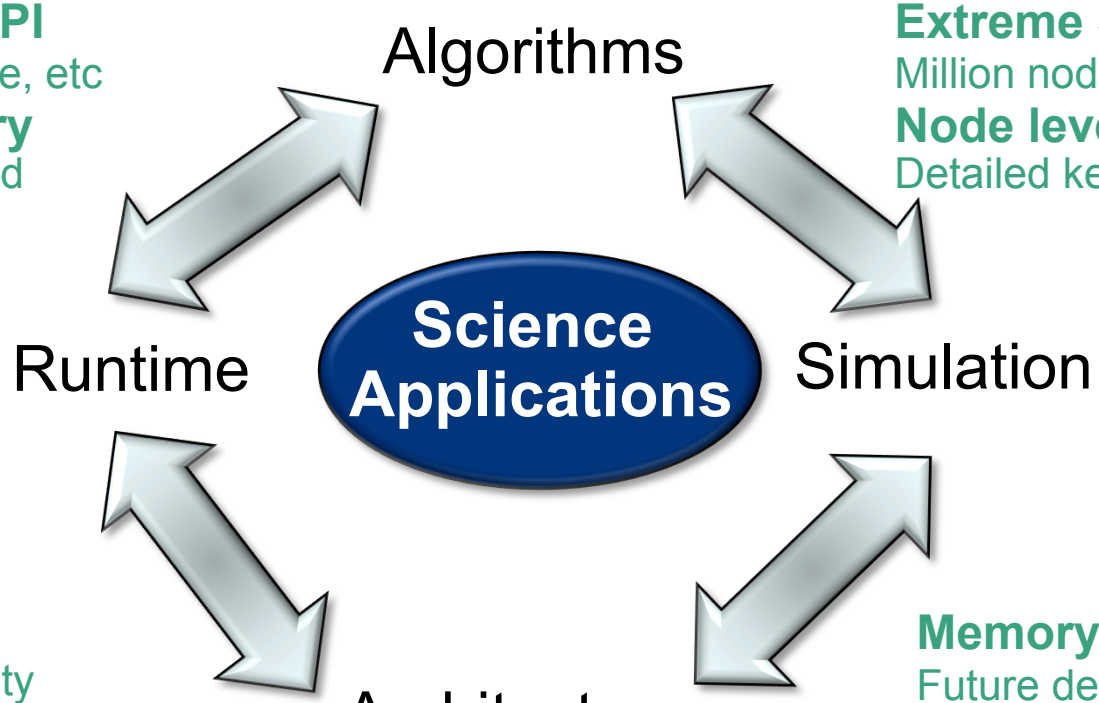
Co-design is key to the success of IAA

It all revolves around the science

Multi-core Aware
Hybrid, Parallel in time

Multi-precision
Krylov, Poisson, Helmholtz

Hierarchical MPI
MPI_Comm_Node, etc
Shared memory
MPI_Alloc_Shared



Extreme Scale
Million node systems
Node level
Detailed kernel studies

Multi-core
Processor affinity
Memory affinity
Scheduling

Memory hierarchy
Future designs
Interconnect
Latency/BW effects

Influence design

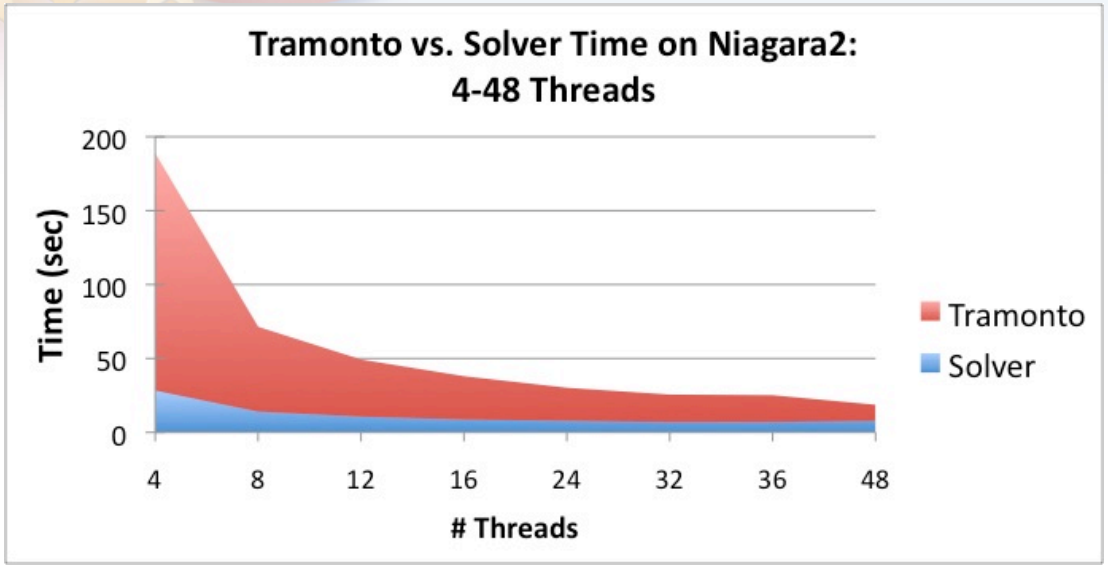
OAK RIDGE National Laboratory

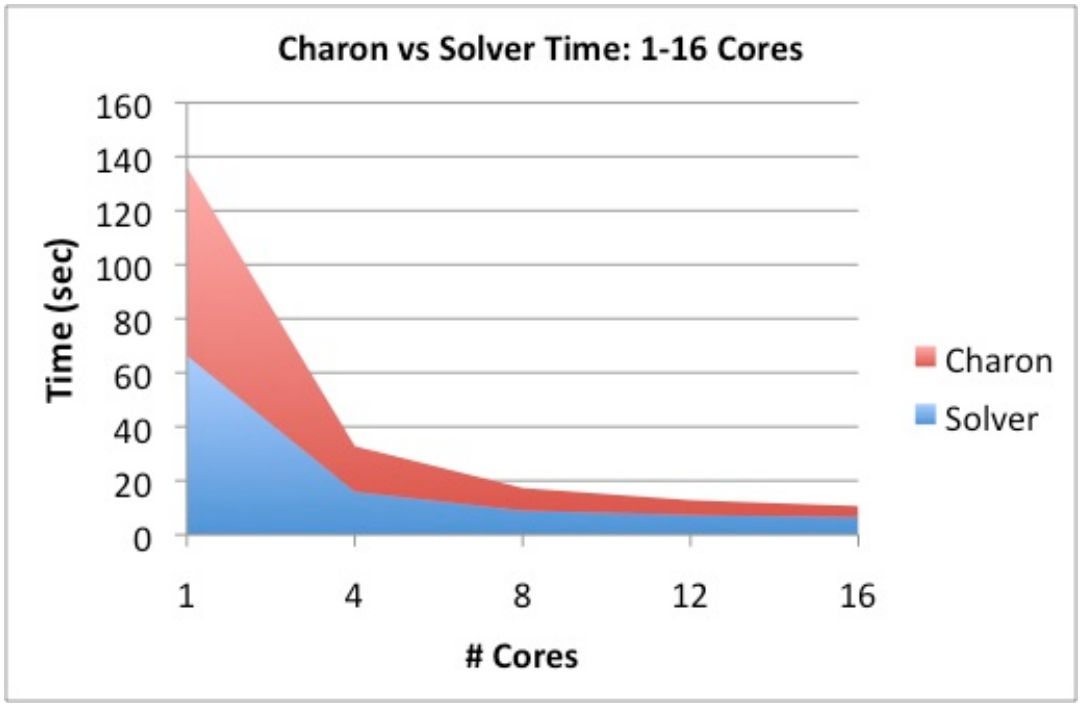
Sandia National Laboratories

IAA Areas

- Co-design of algorithms and runtime environments
  - CS/Math Institute
- Algorithms and architectures to tackle the memory wall
- System simulation as an enabling tool for co-design

Multicore Scaling: App vs. Solver





Application:

- Scales well (sometimes superlinear)
- MPI-only sufficient.

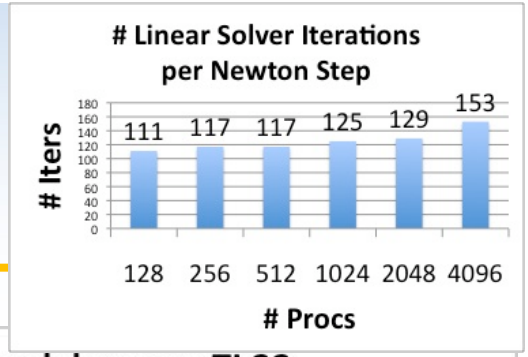
Solver:

- Scales more poorly.
- Memory system-limited.
- MPI+threads can help.

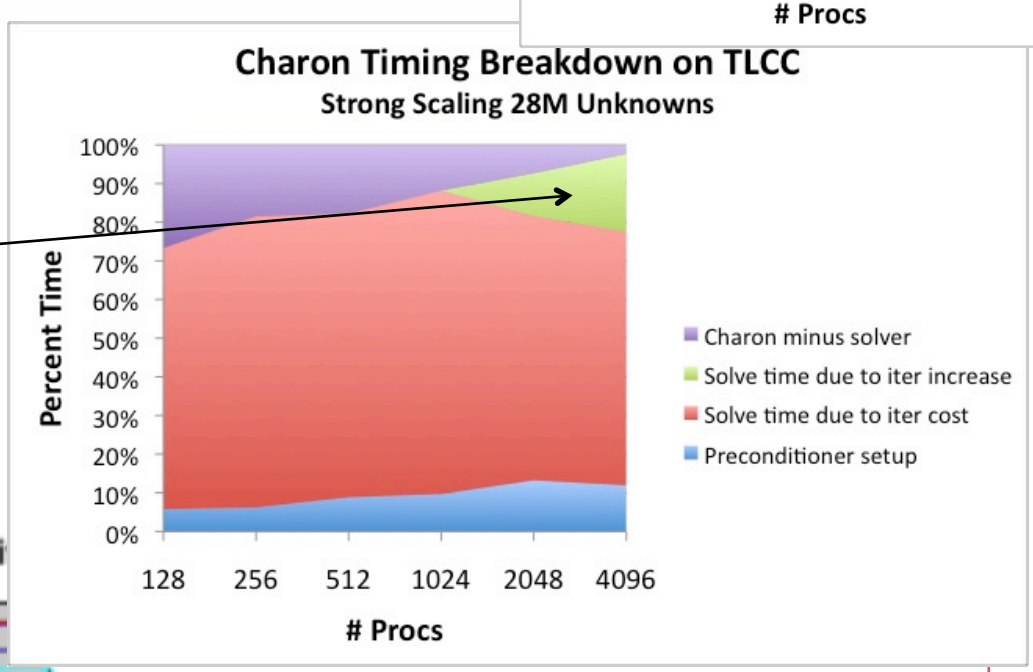
*** All Charon Results: Lin & Shadid TLCC Report**

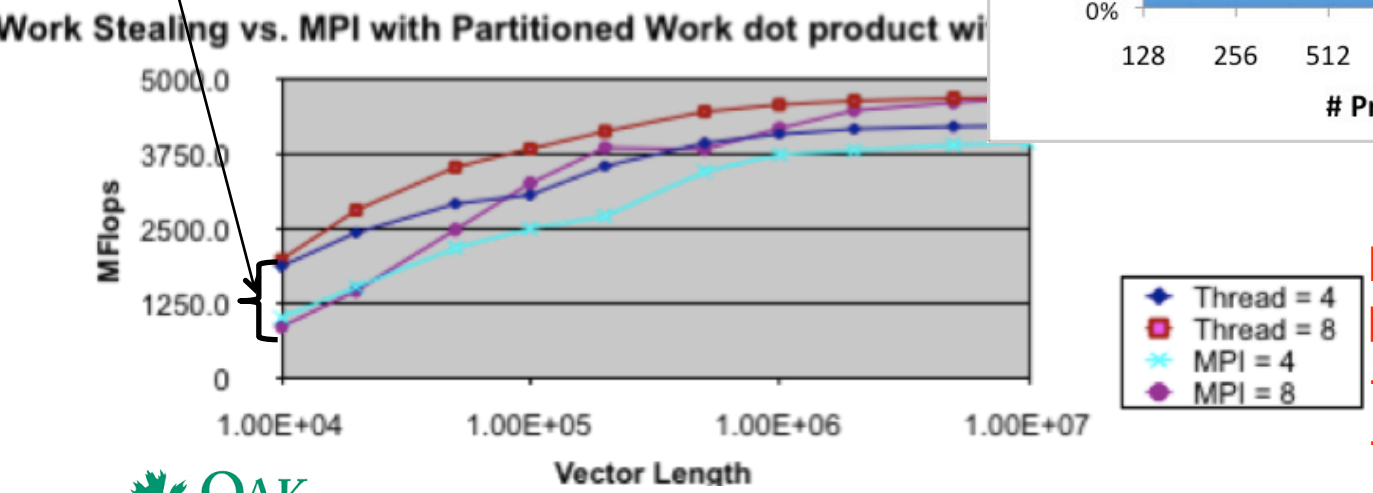
Sandia National Laboratories

Co-design of algorithms and runtime libraries (IAA)



- **Selective Shared Memory Use**:
  - App: 4096 MPI tasks
  - Solver: 256 MPI tasks, 16-way threading
- **Robustness:**
  - 117 iterations (not 153)
  - Eliminates green region
- **Speed: Threading (carefully used):**
  - Same asymptotic speed as MPI
  - Faster ramp up: 2X or more
  - Better load imbalance tolerance





Bottom line: Hybrid parallelism promises better:

- Robustness,
- Strong scaling and
- Load balancing.

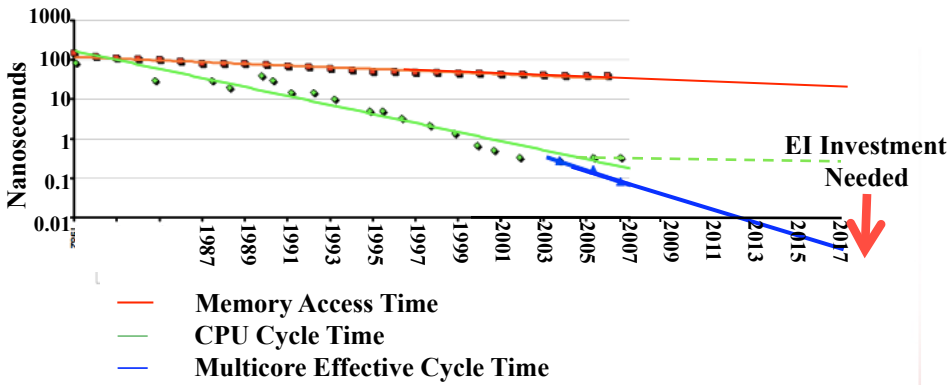
Oak Ridge National Laboratory

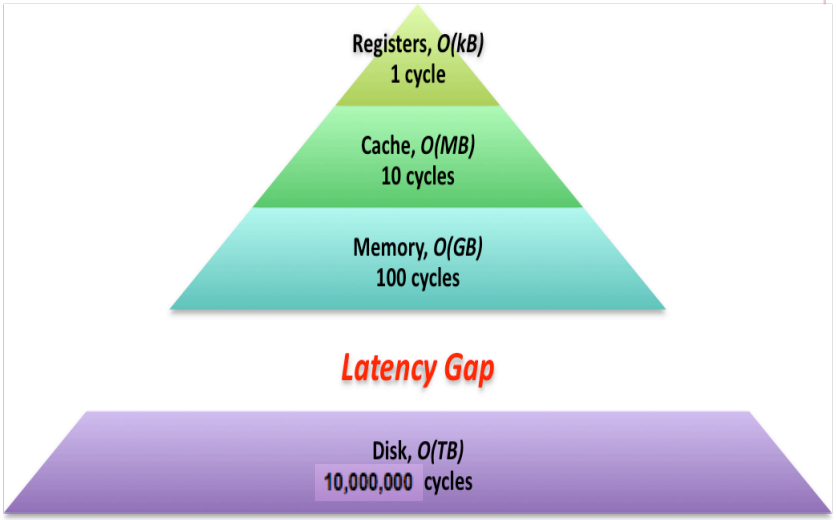
IAA

Sandia National Laboratories

Memory and Storage

- **Barriers**
  - *Per-disk performance, failure rates, and energy efficiency no longer improving*
  - *Linear extrapolation of DRAM vs. Multi-core performance means the height of the memory wall is accelerating*
  - *Off-chip bandwidth, latency, combined with poor concurrency are throttling delivered performance*
- **Technical Focus Areas**
  - *Efficient Data Movement*
    - Photonic DRAM interfaces
    - Optical interconnects / routers
    - Communications optimal algorithms
  - *New Storage Approaches*
    - Non-volatile memory gap fillers
    - Advanced packaging (chip stacking)
    - Storage efficient programming models (Global Address Space)
- **Technical Gap**
  - *Need 10X improvement in memory access speeds to keep current balance with computation.*



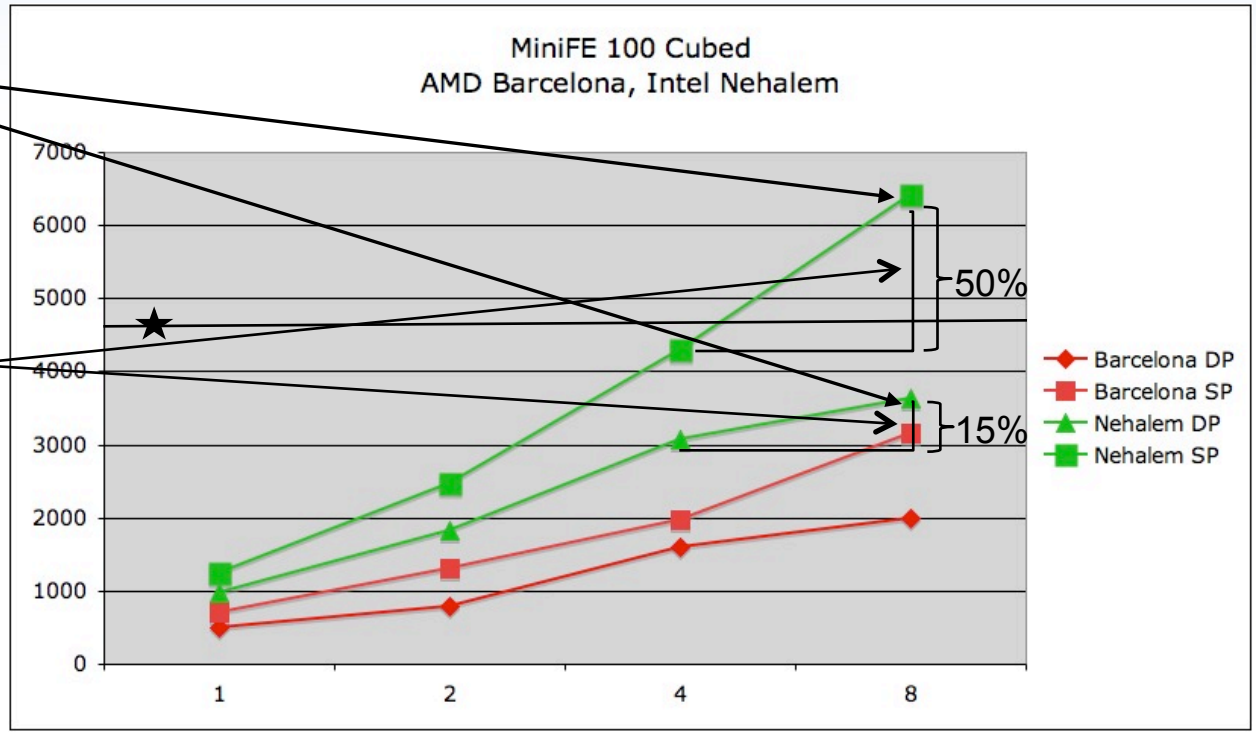


Oak Ridge National Laboratory

Sandia National Laboratories

The Case for Mixed Precision

- **Float useful:**
  - Always true.
  - More important now.
  - Mixed precision algorithms.
- **Bandwidth even more important:**
  - Saturation means loss of effective core use.
  - Loss of scaling opportunity for modern systems.



- Mixed precision & GPUs:
  - GEForce GTX280
    - SP: 624 GFLOPS/s
    - DP: 78 GFLOPS/s
- First MiniFE result on GPUs: 4.71 GFLOP/s (SP)
- Expected results: 12 GFLOP/s (SP), 6 GFLOP/s (DP)

Oak Ridge National Laboratory

Sandia National Laboratories

Tpetra Linear Algebra Library

- Tpetra is a templated version of the Petra distributed linear algebra model in Trilinos.
  - Objects are templated on the underlying data types:

```
MultiVector<scalar=double, local_ordinal=int,
                    global_ordinal=local_ordinal> …
CrsMatrix<scalar=double, local_ordinal=int,
                    global_ordinal=local_ordinal> …
```

  - Examples:

```
MultiVector<double, int, long int> V;
CrsMatrix<float> A;
```

Speedup of float over double in Belos linear solver.

float	double	speedup
18 s	26 s	1.42x

Scalar	float	double	double - double	quad-double
Solve time (s)	2.6	5.3	29.9	76.5
Accuracy	10^{-6}	10^{-12}	10^{-24}	10^{-48}

Arbitrary precision solves using Tpetra and Belos linear solver package

Sandia National Laboratories

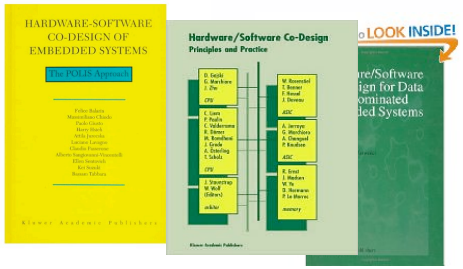
Summary of Micron-Sandia Interactions

Activity	Outputs
Kickoff Meeting	Micron-Sandia Collaboration Begins (7/18/2006)
Foundation for Discussions	2-way NDA (11/12/2007)
Advanced Memory for DOE Architectures	Simulations, Papers, PIM LDRD Effort, ASC/CSSE L2 Milestone: Evaluate Advanced Memory Subsystems - Due Sept'10
University Collaborations	UMaryland, LSU, Notre Dame, USC/ISI, IndianaU, UIUC
IAA Activities	Dean Klein is a member of the IAA Advisory Board IAA Workshop on Memory Opportunities for HPC (Jan'08) IAA Workshop on HPC Architectural Simulation (Sept'09)
Collaborations with other agencies	Alignment of ASC/CSSE, DoD/ACS, & IAA support to Integrate U-MD's Memory Simulator (DRAMsim) with Sandia's SST
Proposal Partnerships	IAA Memory Proposal, DARPA/UHPC - pending
CRADA	Micron-Sandia collaboration to analyze advanced concepts for error correction in advanced memory designs - pending, Feb'10

- **Technical Exchanges from July 2006 - Present**
  - Approximately two dozen face-to-face technical meetings
  - Catalyst for Collaboration with other DOE/NNSA labs, DoD, and Universities

Hardware/Software co-design is a mature field

- Design of an integrated system that contains hardware and software
- Focus on embedded systems (cell phones, appliances, engines, controllers, etc.)
- Concurrent development of hardware and software
  - Interactions and tradeoffs
  - Partitioning is a focus
  - Must satisfy real-time and/or other performance/energy metrics/constraints

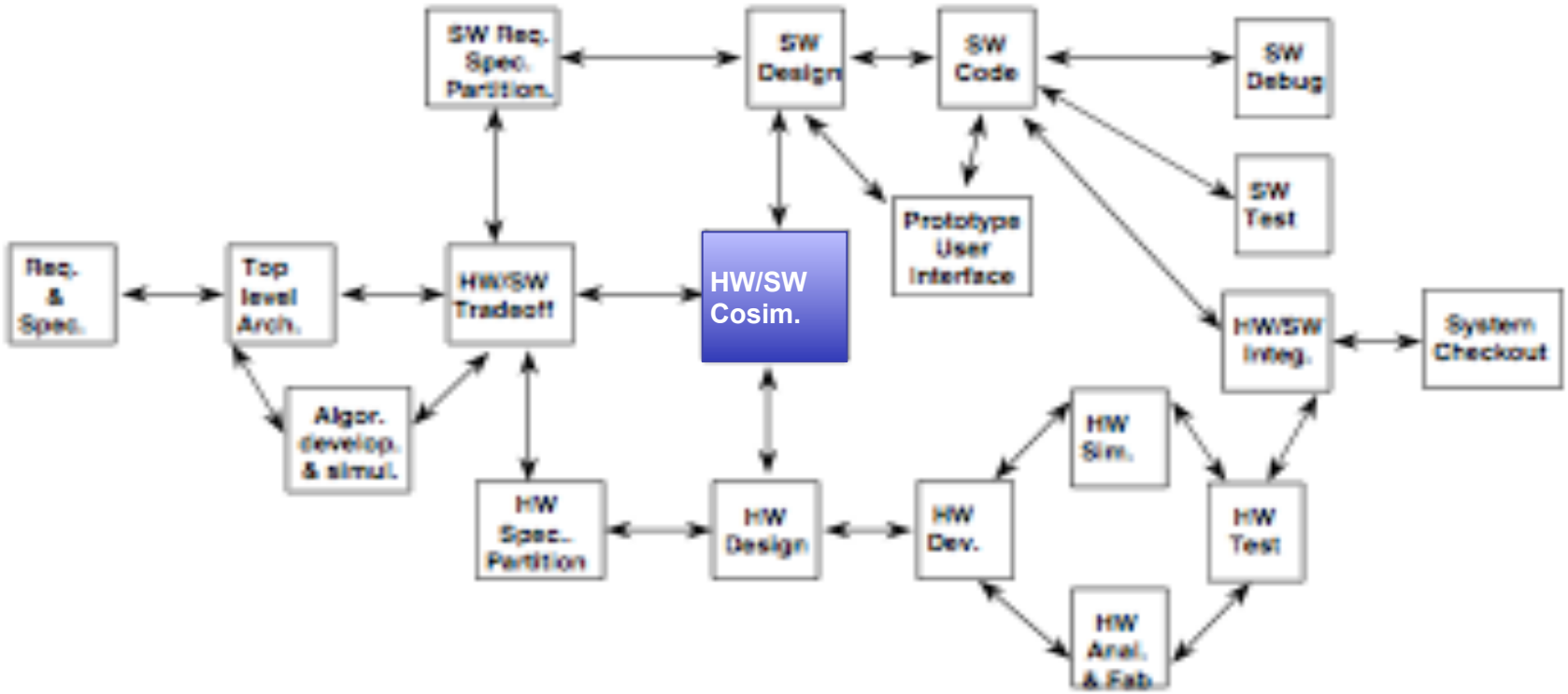


Oak Ridge National Laboratory

iAA

Sandia National Laboratories

Lockheed Martin Co-design Methodology



Oak Ridge National Laboratory

Sandia National Laboratories

System Simulation Workshop was co-sponsored by IAA and LBNL

September 2009 in Boulder, CO

- Goals
  - Define requirements for a common community product for HPC architectural simulation.
  - Addresses the needs of multiple HPC communities
  - Form the foundation of an ecosystem for HPC simulation
- Audience: ~50 participants from labs, universities, industry
  - Brought together simulation "providers" and "customers"
- Findings
  - HPC is faced by fundamentally new challenges (Hardware, Software, Scale, Power) and needs new simulation capabilities to confront them
  - Many simulation models share common implementation themes
  - Most participants believe a common simulation platform is desirable, beneficial and technically feasible
  - Verification and Validation are major concerns

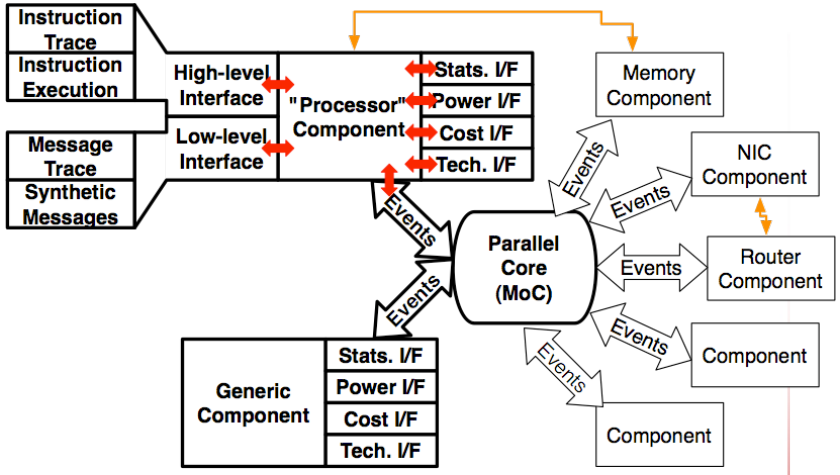
Oak Ridge National Laboratory

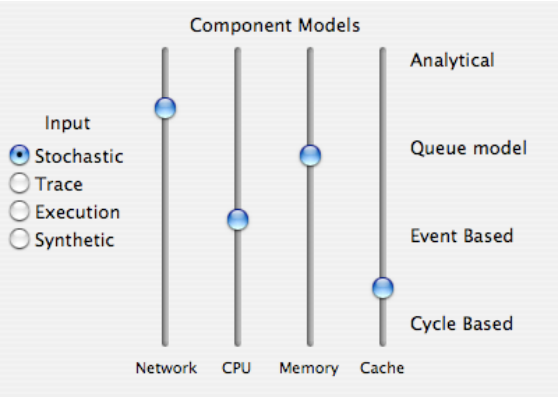
IAA

Sandia National Laboratories

System simulation should be a key enabling technology

- Co-simulation of hardware and software
  - Assess architectural choices and their impact on applications
  - Identify bottlenecks and enable the development of algorithms for future architectures
- Key features
  - Open source with the ability to interface to proprietary software
  - Holistic: performance, power, area, cost, reliability analysis
  - Modular and multiscale (cycle accurate to analytical)
  - Input traces as well as joint execution
  - Parallel
  - FPGA acceleration



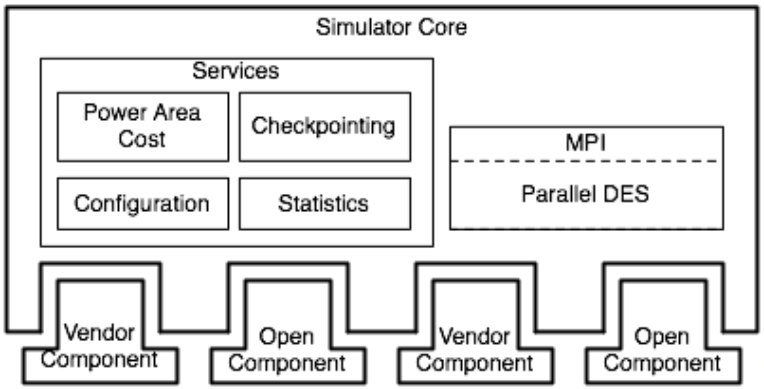


Oak Ridge National Laboratory

Sandia National Laboratories

SST Simulation Project

- Parallel
- Parallel Discrete Event core with conservative optimization over MPI
- Holistic
- Integrated Tech. Models for power
- McPAT, Sim-Panalyzer
- Multiscale
- Detailed and simple models for processor, network, and memory



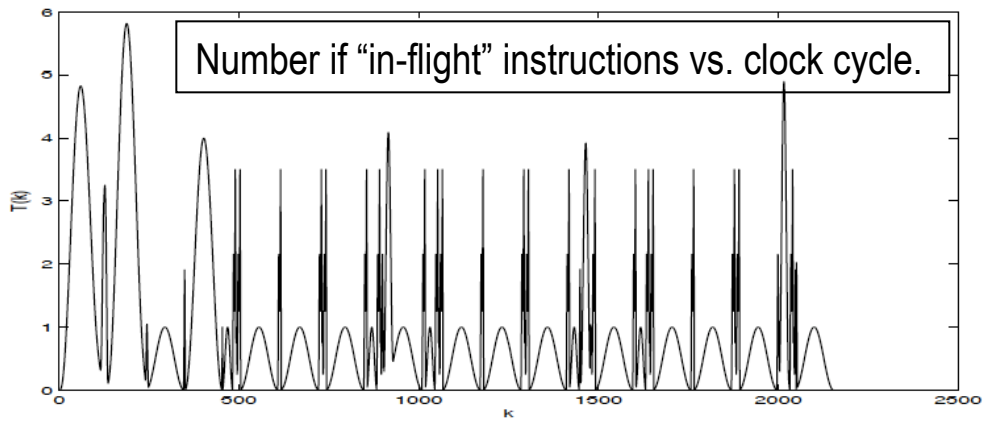
- Current Release (2.0) at

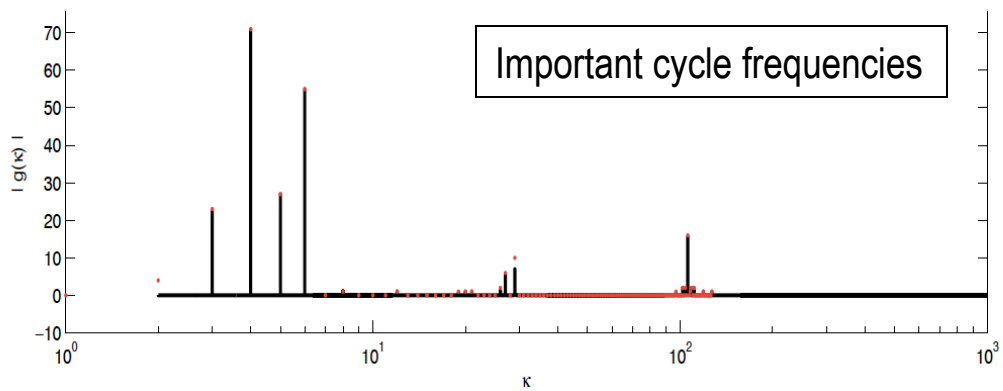
http://www.cs.sandia.gov/sst/

- Includes parallel simulation core, configuration, power models, basic network and processor models, and interface to detailed memory model

Analyzing Simulator Results

- Input: SST Trace for SpMV.
- Lots of instruction stream data.
- Model: Use restricted *sin*^{2} function to mark start/finish of each instruction.
- Use FFTs to analyze behavior.





Trace fragment from SpMV inner loop

j	I_j	issue	complete	κ
59	bc	737	741	4
60	lwz	738	744	6
61	lfd	740	746	6
62	addi	742	746	4
63	addi	742	746	4
64	rlwinm	743	746	3
65	lfdx	744	850	106
66	fmadd	849	854	5
67	bc	850	854	4
68	lwz	851	857	6
69	lfd	853	859	6
70	addi	855	859	4
71	addi	855	859	4
72	rlwinm	856	859	3
73	lfdx	857	886	29
74	fmadd	885	890	5
75	bc	886	890	4
76	lwz	887	893	6
77	lfd	889	895	6
78	addi	891	895	4
79	addi	891	895	4
80	rlwinm	892	895	3
81	lfdx	893	899	6
82	fmadd	898	903	5
83	bc	899	903	4

Need to define HPC co-design methodology

- Could range from discussions between architecture, software and application groups to tight collaboration centered on the co-simulation of hardware and applications
- Opportunity to influence future architectures
  - Cores/node, threads/core, scheduling width/thread
  - Logic in memory subsystem
  - Interconnect performance