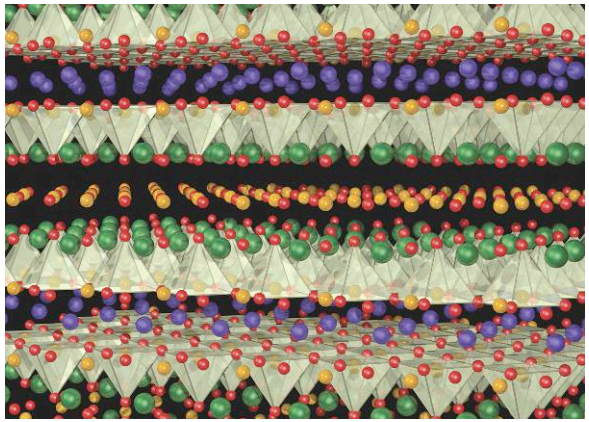
The Institute for Advanced Architectures and Algorithms (IAA)

David H. Rogers
Sudip Dosanjh

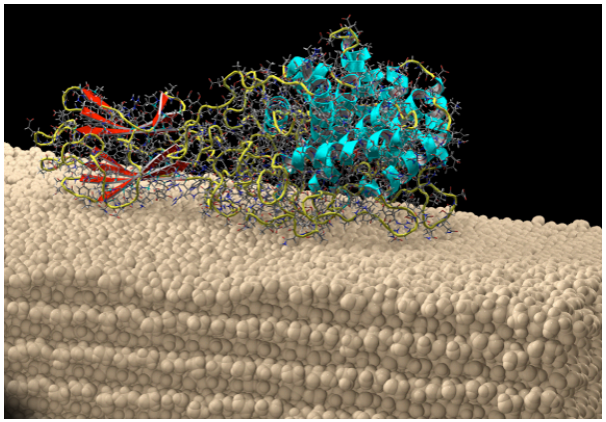
Sandia National Laboratories

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.

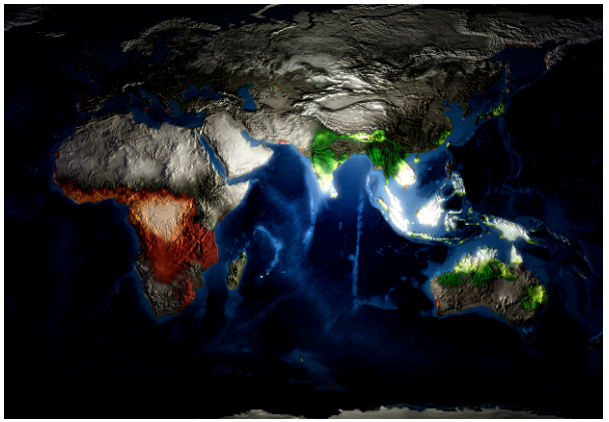
Leadership computing is advancing scientific discovery



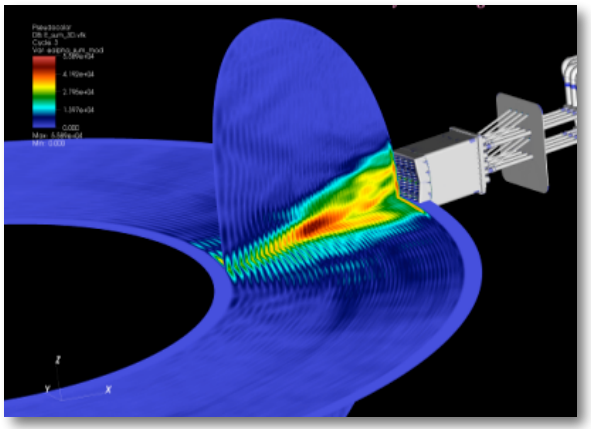
Resolved decades-long controversy about modeling physics of high temperature superconducting cuprates



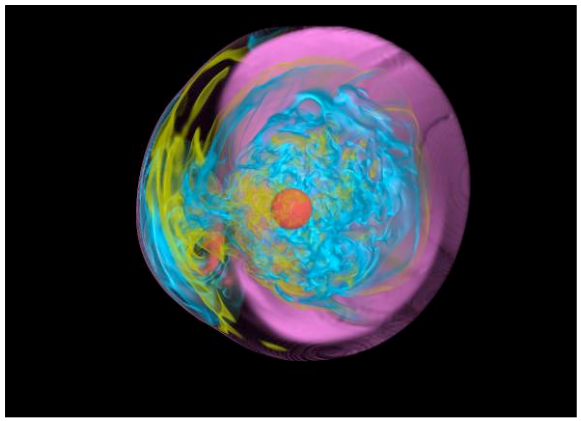
New insights into protein structure and function leading to better understanding of cellulose-to-ethanol conversion



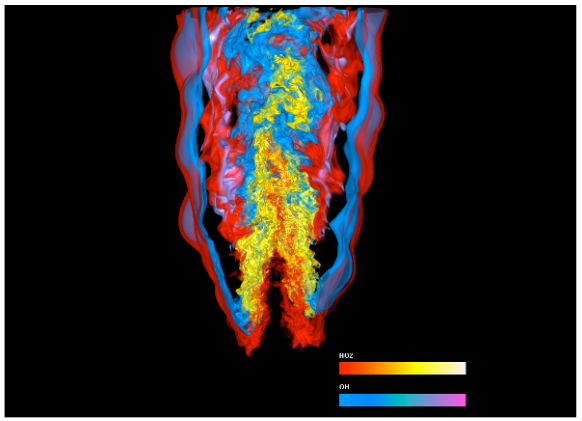
Addition of vegetation models in climate code for global, dynamic CO_2 exploration



First fully 3D plasma simulations shed new light on engineering superheated ionic gas in ITER

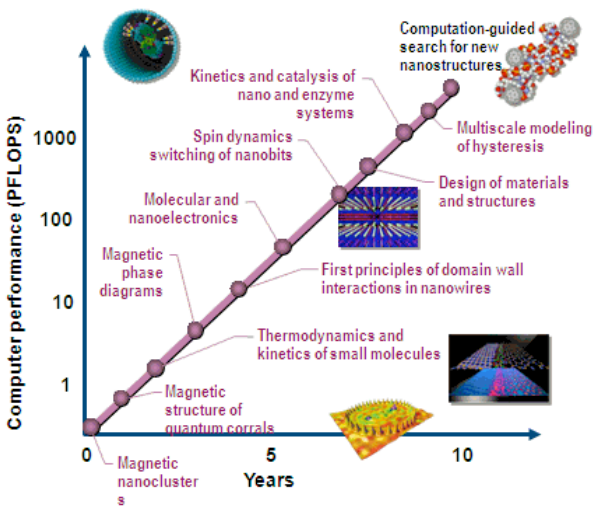


Fundamental instability of supernova shocks discovered directly through simulation



First 3-D simulation of flame that resolves chemical composition, temperature, and flow

DOE-SC Science Drivers



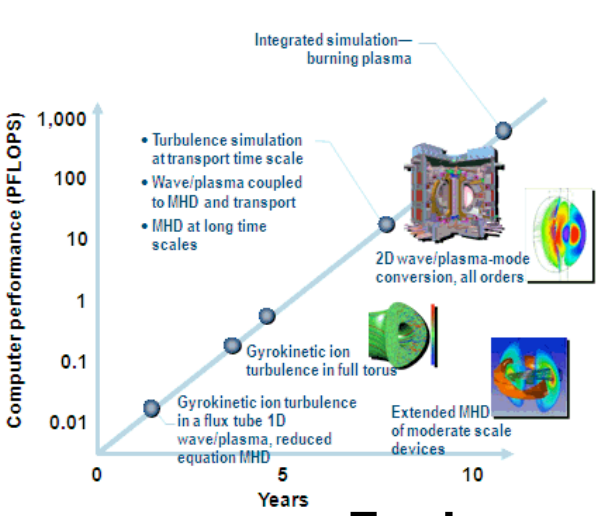
Expected outcomes

5 years

- Realistic simulation of self-assembly and single-molecule electron transport
- Finite-temperature properties of nanoparticles/quantum corrals

10 years

- Multiscale modeling of molecular electronic devices
- Computation-guided search for new materials and nanostructures



Expected outcomes

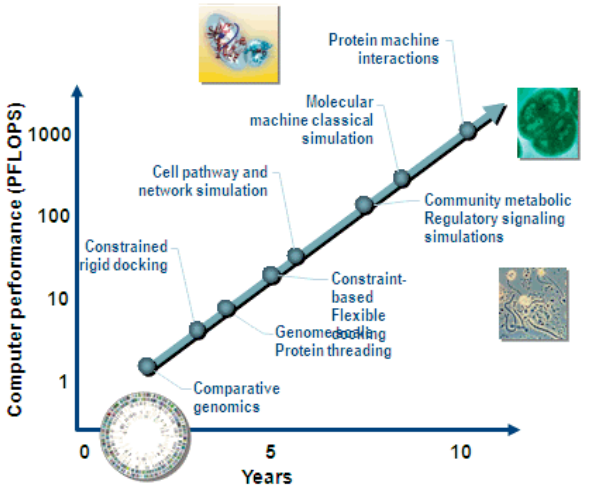
5 Years

- Full-torus, electromagnetic simulation of turbulent transport with kinetic electrons for simulation times approaching transport time-scale
- Develop understanding of internal reconnection events in extended MHD, with assessment of RF heating and current drive techniques for mitigation

10 years

- Develop quantitative, predictive understanding of disruption events in large tokamaks
- Begin integrated simulation of burning plasma devices –

Fusion



Expected outcomes

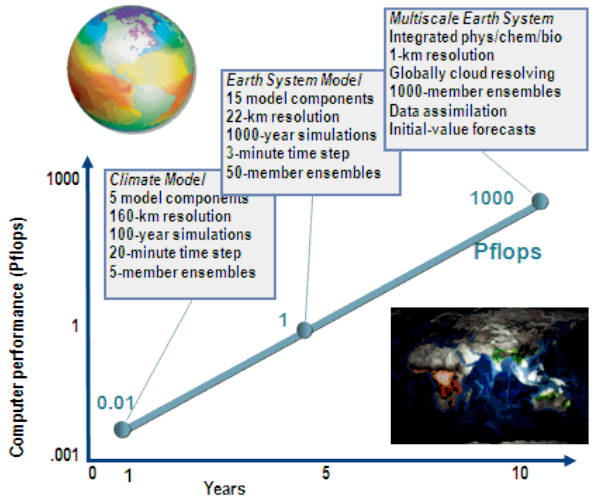
5 years

- Metabolic flux modeling for hydrogen and carbon fixation pathways
- Constrained flexible docking simulations of interacting proteins

10 years

- Multiscale stochastic simulations of microbial metabolic, regulatory, and protein interaction networks
- Dynamic simulations of complex molecular machines

Biology



Expected outcomes

5 years

- Fully coupled carbon-climate simulation
- Fully coupled sulfur-atmospheric chemistry simulation

10 years

- Cloud-resolving 1-km spatial resolution atmosphere
- Fully coupled, physics, chemistry, biology earth system model

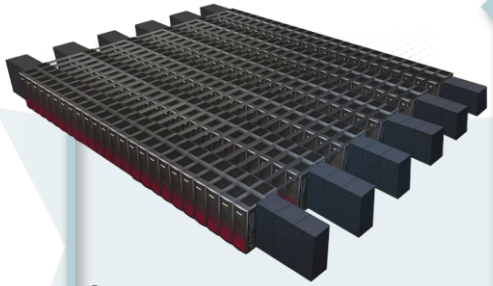
Climate

OAK RIDGE National Laboratory

Sandia National Laboratories

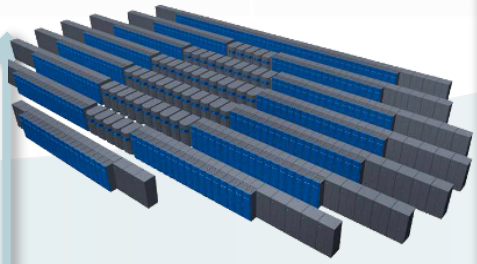
DOE Leadership Computing Roadmap

Mission: Deploy and operate the computational resources required to tackle global challenges	Vision: Maximize scientific productivity and progress on the largest scale computational problems
• Deliver transforming discoveries in materials, biology, climate, energy technologies, etc. • Ability to investigate otherwise inaccessible systems, from supernovae to energy grid dynamics	• Providing world-class computational resources and specialized services for the most computationally intensive problems • Providing stable hardware/software path of increasing scale to maximize productive applications development



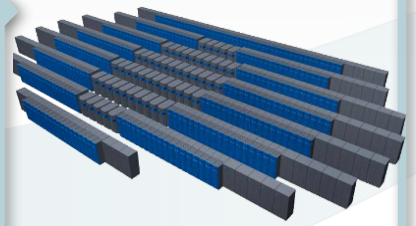
Cray XT5: 1 PF
10 PB Disk
40 PB Archive

FY2009



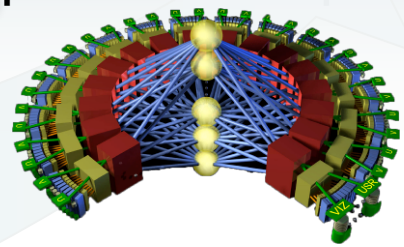
DARPA HPCS: 20 PF
50 PB Disk
200 PB Archive

FY2011



Follow on to DARPA
HPCS: 100 PF
150 PB Disk
1 EB Archive

FY2015



Future system: 1 EF
500 PB Disk
10 EB Archive

FY2018

OAK RIDGE National Laboratory

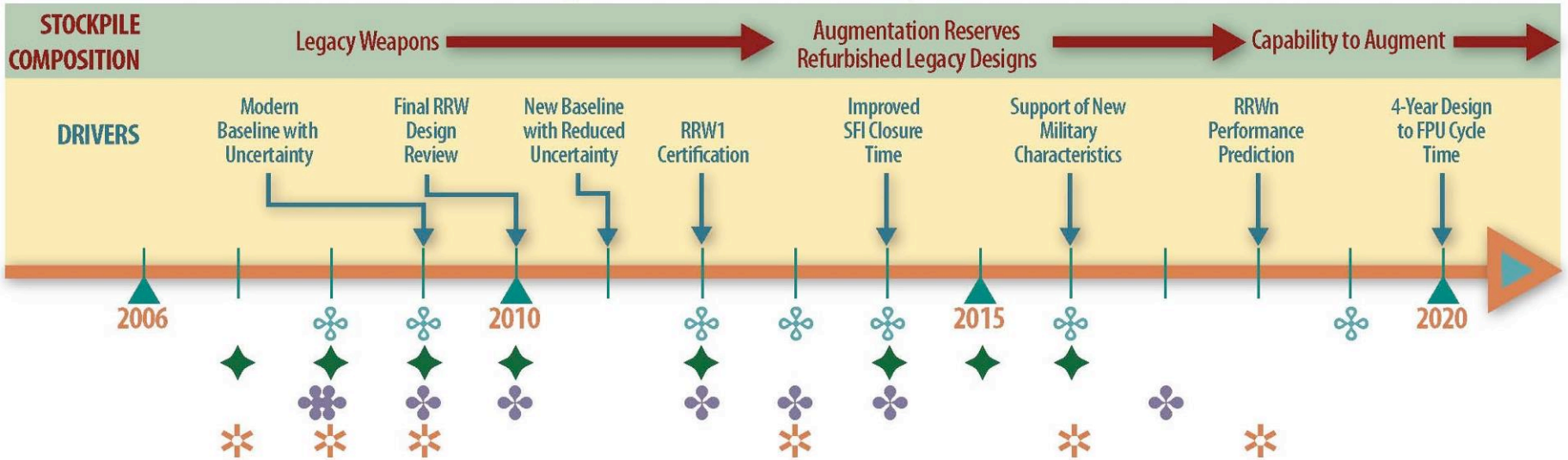
Sandia National Laboratories

www.nnsa.doe.gov/ASC

ASC Roadmap

Computational Weapons Science and Simulation:

Targets to address Nuclear Weapons Issues



ASC Targets

FOCUS AREA 1: ADDRESS NATIONAL SECURITY SIMULATION NEEDS

- 2008: National code strategy
- 2009: Modular physics and engineering packages for national weapons codes
- 2012: Tested capability to address emerging threats, effects, and attribution
- 2013: 50% improvement in setup-to-solution time for SFI simulations (with respect to 2006)
- 2014: Full-system engineering and physics simulation capability
- 2016: Capability to certify fire safety for an unfielded weapon
- 2019: 50% improvement in setup-to-solution time for SFI simulations (with respect to 2013)

FOCUS AREA 2: ESTABLISH A VALIDATED PREDICTIVE CAPABILITY FOR KEY PHYSICAL PHENOMENA

- 2007: Launch Thermonuclear Burn Initiative (TBI) collaboration
- 2008: Realistic plutonium aging simulations
- 2009: Science-based replacement for Knob #1
- 2010: Science-based models for neutron tube simulations
- 2012: Validated science-based replacement for Knob #2
- 2014: NIF-validated simulations supporting replacement of knob #3
- 2015: Science-based models for fire excitation simulations
- 2016: Predictive model for Knob #4

FOCUS AREA 3: QUANTIFY AND AGGREGATE UNCERTAINTIES IN SIMULATION TOOLS

- 2008: National verification & validation strategy
- 2008: Assessment of major simulation uncertainties
- 2009: Shared weapons physical databases
- 2010: Uncertainty Quantification (UQ) methodology for QMU
- 2012: 20% reduction in overall prediction error bars (with respect to 2006)
- 2013: Re-assessment of major simulation uncertainties
- 2014: Demonstrated uncertainty aggregation for QMU
- 2017: 20% Reduction in overall prediction error bars (with respect to 2012)

FOCUS AREA 4: PROVIDE MISSION-RESPONSIVE COMPUTATIONAL ENVIRONMENTS

- 2007: Initiate new National User Facility model for capability supercomputing
- 2008: Seamless user environments for capacity computing
- 2009: Petascale computing
- 2013: Seamless user environments for capability computing
- 2016: 100x petascale computing
- 2018: Exascale computing

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 BW 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

OAK RIDGE National Laboratory

Sandia National Laboratories

Industry Trends

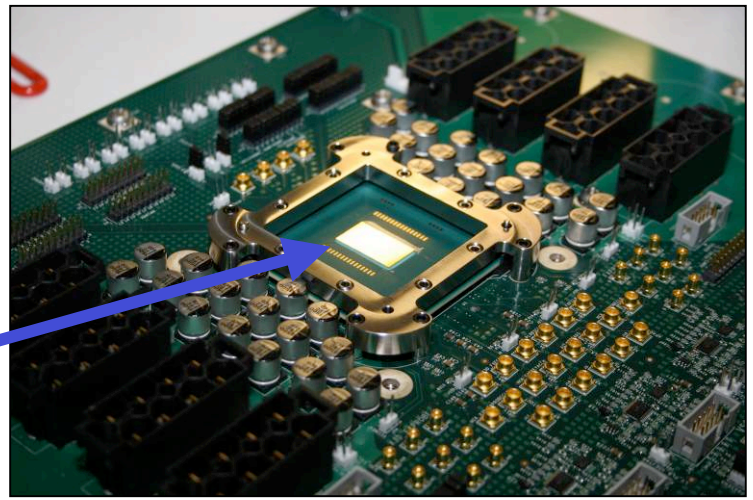
Existing industry trends not going to meet HPC application needs

- **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**
  - **Supercomputer architectures must be designed with an understanding of the applications they are intended to run**
  - **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**

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

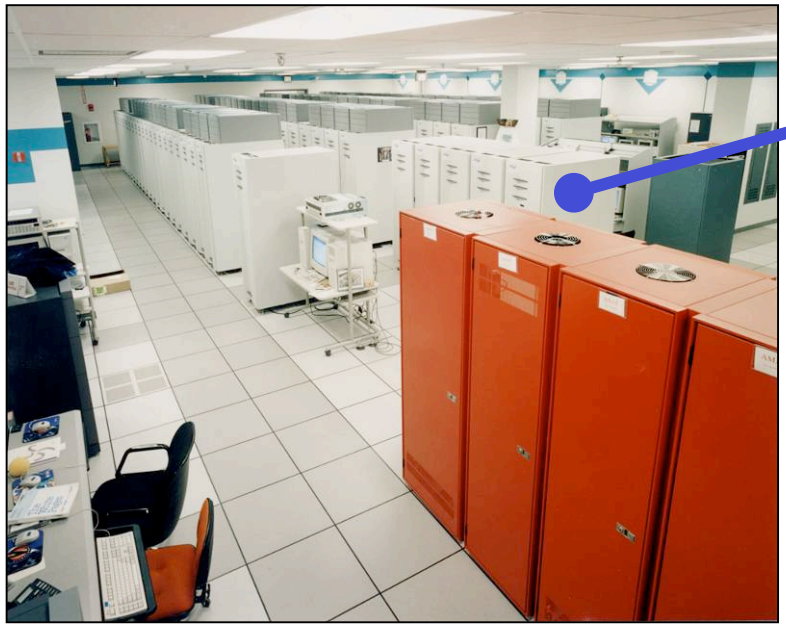
2007 - 1 TeraFLOPs on a chip

- **275 mm^2 (size of a dime) & 62 W**



1997 - 1 TeraFLOPs in a room

- **2,500 ft^2 & 500,000 W**



OAK RIDGE National Laboratory

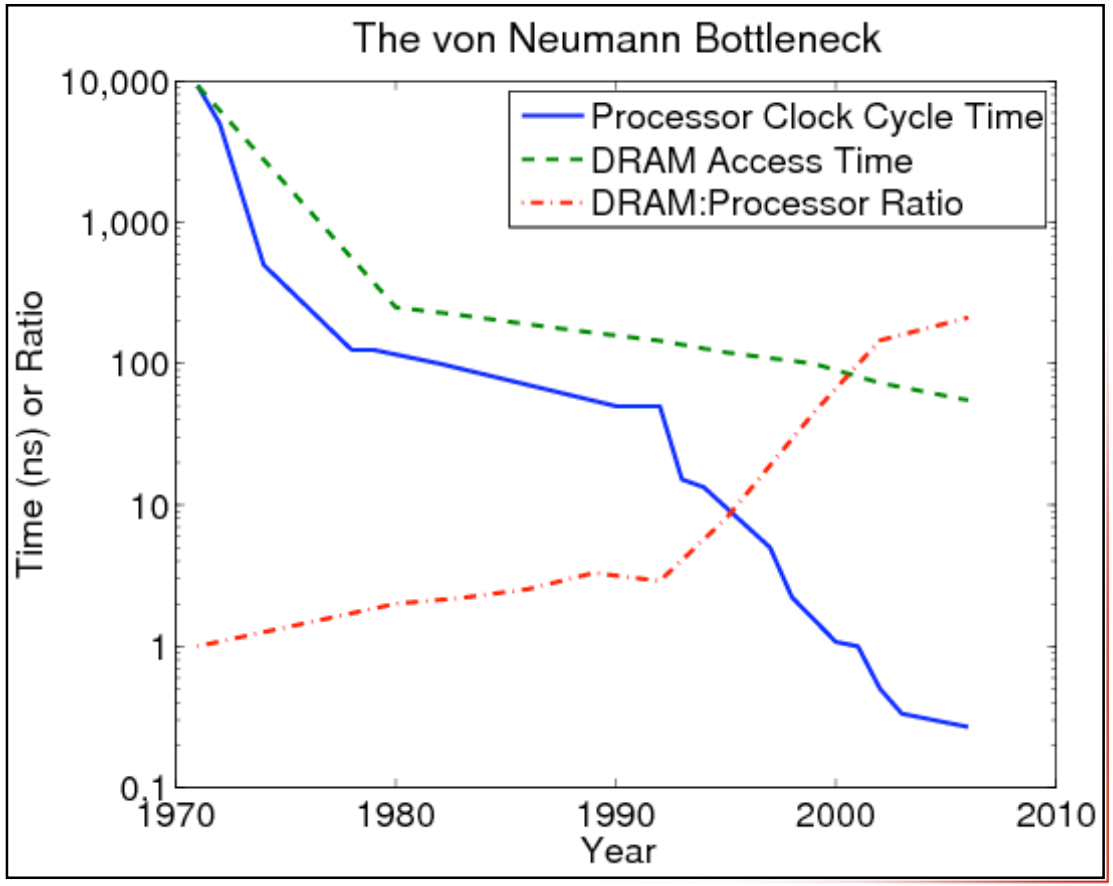
Sandia National Laboratories

And Then There's the Memory Wall

"FLOPS are 'free'. In most cases we can now compute on the data as fast as we can move it." - Doug Miles, The Portland Group

What we observe today:

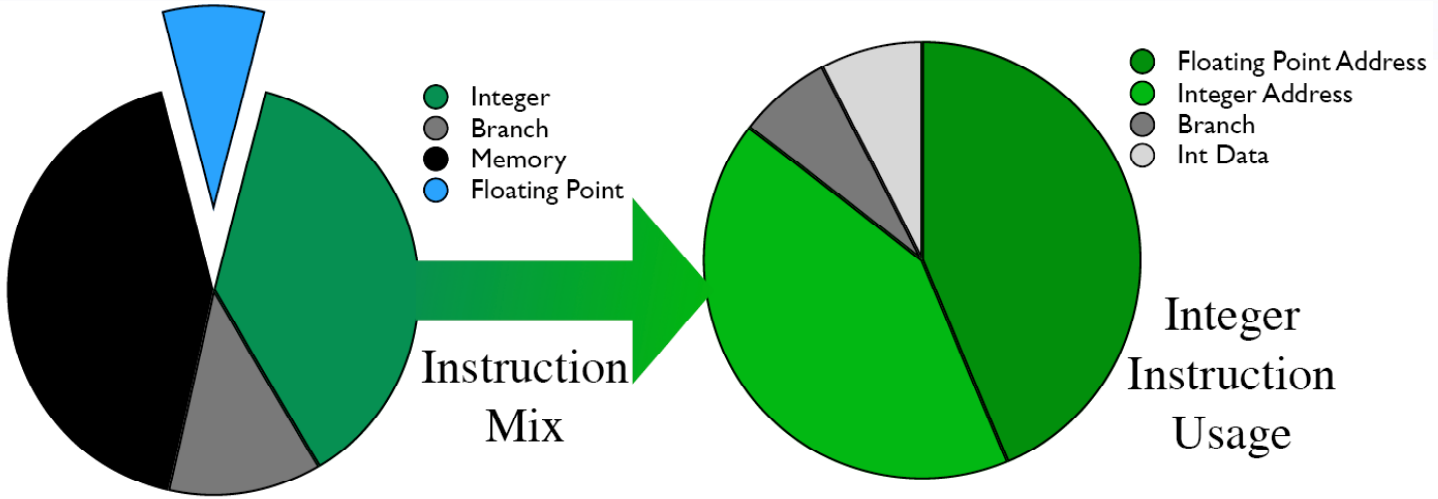
- Logic transistors are free
- The von Neumann architecture is a bottleneck
- Exponential increases in performance will come from increased concurrency not increased clock rates if the cores are not starved for data or instructions



OAK RIDGE National Laboratory

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

The Need for HPC Innovation and Investment is Well Documented

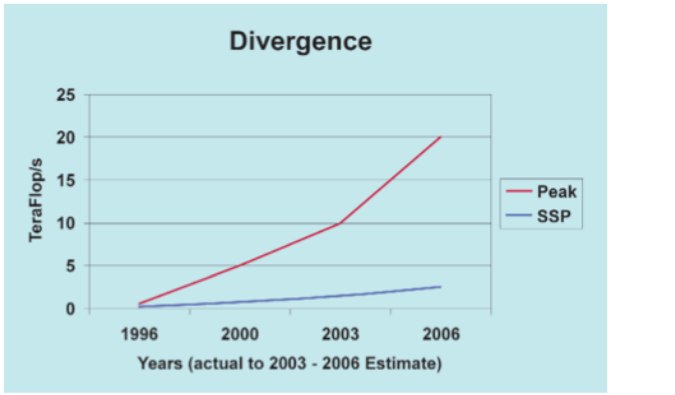
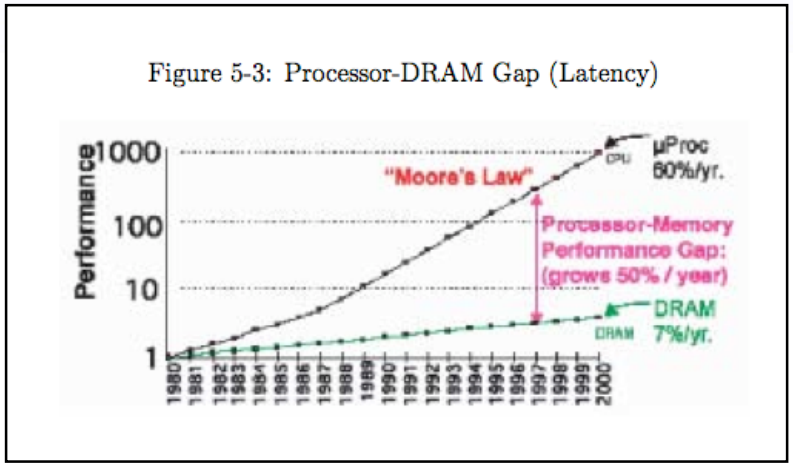


Figure 1: Divergence Problem for HEC Centers (SSP = Sustained System Performance)

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."

OAK RIDGE National Laboratory

Sandia National Laboratories

Impediments to Useful Exascale Computing

- Data Movement
  - Local
    - cache architectures
    - main memory architectures
  - Remote
    - Topology
    - Link BW
    - Injection MW
    - Messaging Rate
  - File I/O
    - Network Architectures
    - Parallel File Systems
    - Disk BW
    - Disk latency
    - Meta-data services
- Power Consumption
  - Do Nothing: 100 to 140 MW
- Scalability
  - 10,000,000 nodes
  - 1,000,000,000 cores
  - 10,000,000,000 threads
- Resilience
  - Perhaps a harder problem than all the others
  - Do Nothing: an MTBI of 10's of minutes
- Programming Environment
  - Data movement will drive new paradigms

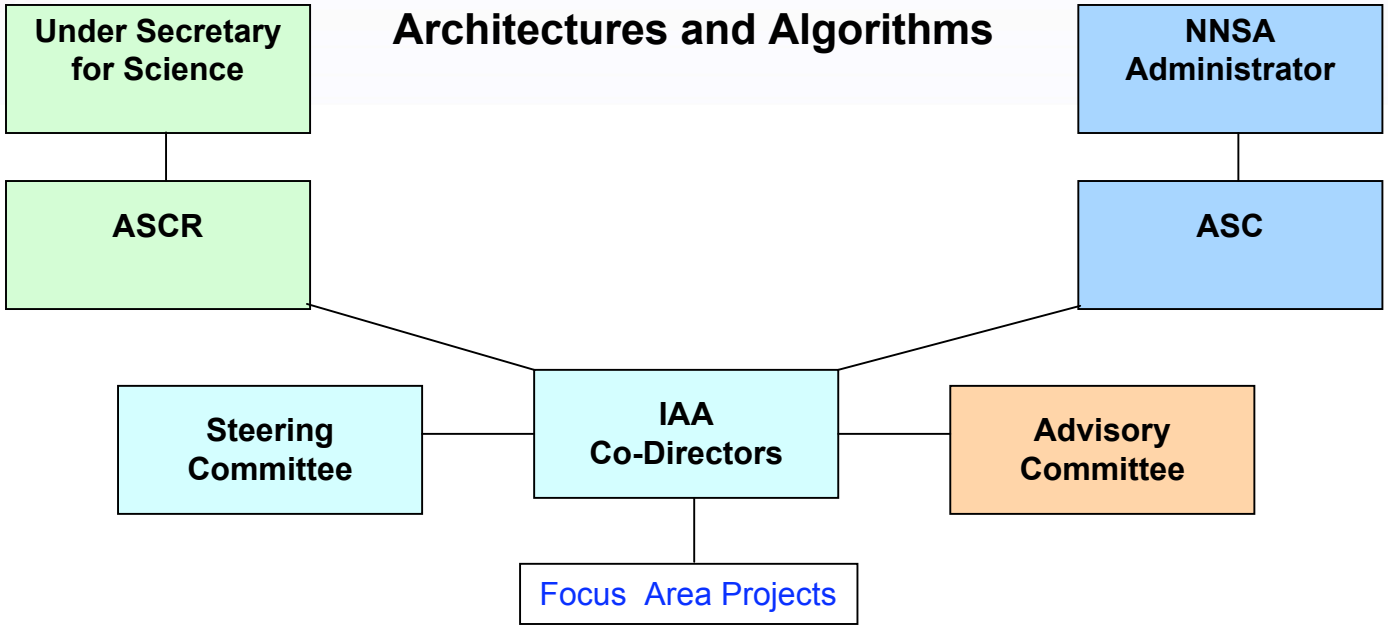
IAA Mission and Strategy

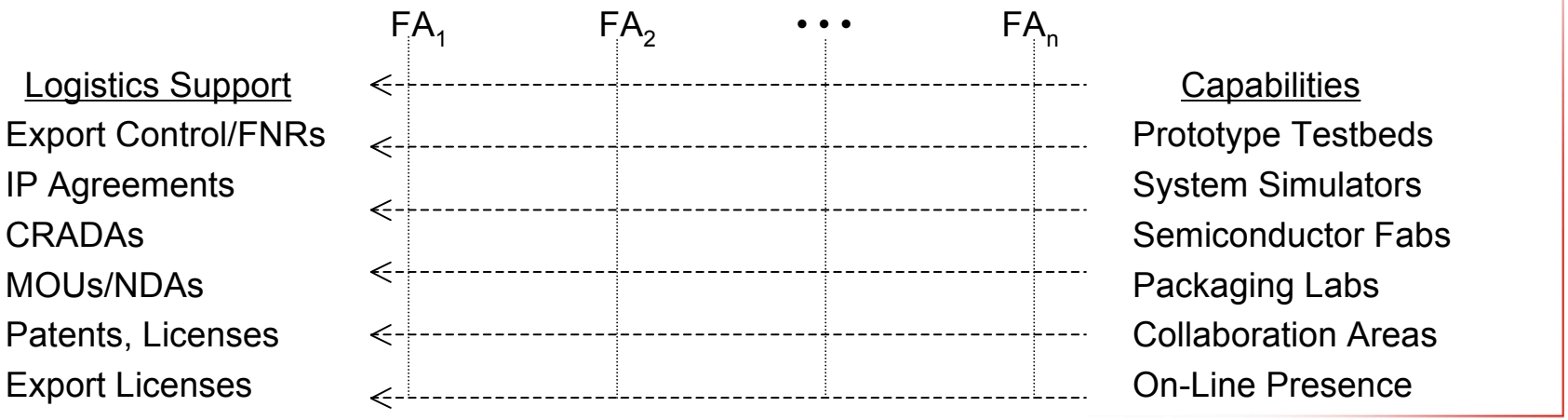
IAA is being proposed as the medium through which architectures and applications can be co-designed in order to create synergy in their respective evolutions.

- Focused R&D on key impediments to high performance in partnership with industry and academia
- 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

The Department of Energy

Institute for Advanced Architectures and Algorithms





OAK RIDGE National Laboratory

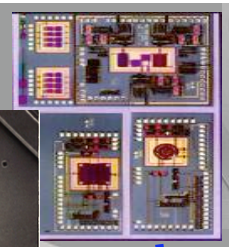
Sandia National Laboratories

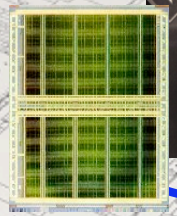
Uniqueness

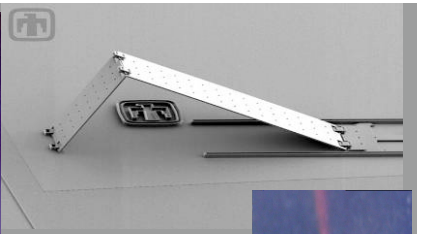
- Partnerships with industry, as opposed to contract management
- Cuts across DOE and other government agencies and laboratories
- A focus on impacting commercial product lines
  - National competitiveness
  - Impact on a broad spectrum of platform acquisitions
- A focus on problems of interest to DOE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

MESA Complex

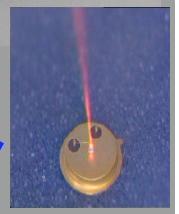
Components







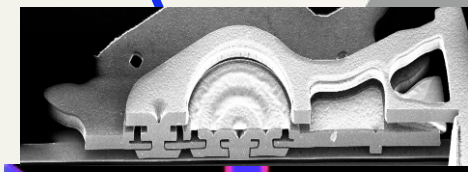
MicroFab

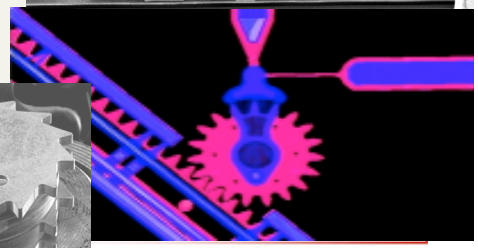


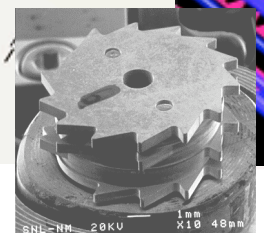
Integrated, Co-located Capability for Design, Fabrication, Packaging

Science

MicroLab







OAK RIDGE National Laboratory

Sandia National Laboratories

Execution Plan

- Project Planning
  - Joint SNL/ORNL meetings
- Workshops
  - Work with industry and academia to define thrust areas
  - "Memory Opportunities for High-Performance Computing", Jan 2008 in Albuquerque (Fred Johnson and Bob Meisner were on the program committee)
  - Planning started for an Interconnect Workshop, Summer 2008
  - Planning started for an Algorithm Workshop, Fall 2008
  - Training
    - Fellowships, summer internships, and interactions with academia to help train the next generation of HPC experts.

- Define and prioritize focus areas
  - High-speed interconnects *
  - Memory subsystems *
  - Power
  - Processor microarchitecture
  - RAS/Resiliency
  - System Software
  - Scalable I/O
  - Hierarchical algorithms *
  - System simulators *
  - Application performance modeling
  - Programming models
  - Tools

* FY '08 Project Starts

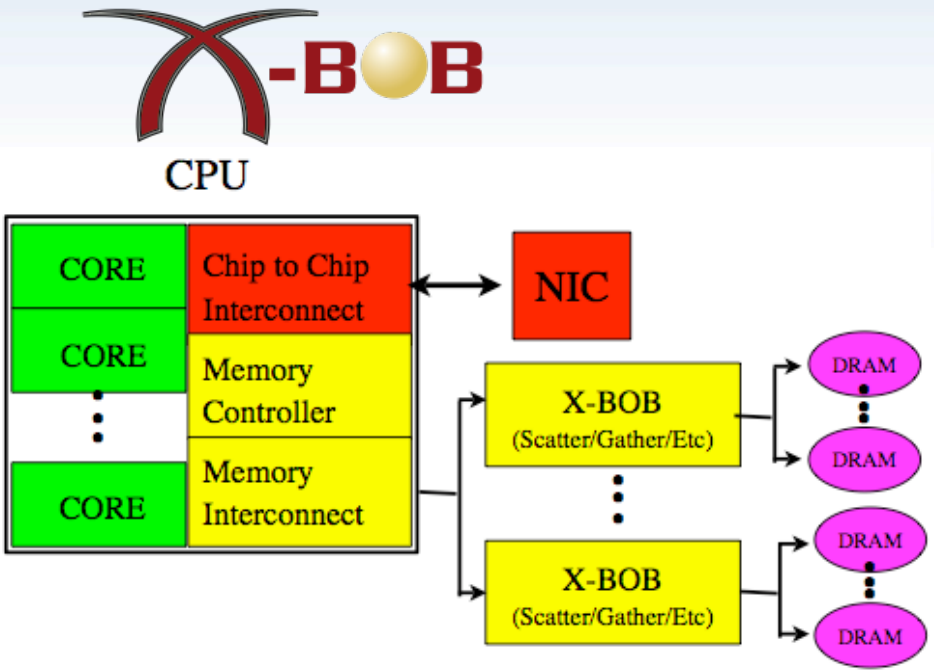
OAK RIDGE National Laboratory

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.



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), AMD, SUN, Intel, Cray, IBM
- Academia: USC/ISI (Draper/Hall), LSU (Sterling)

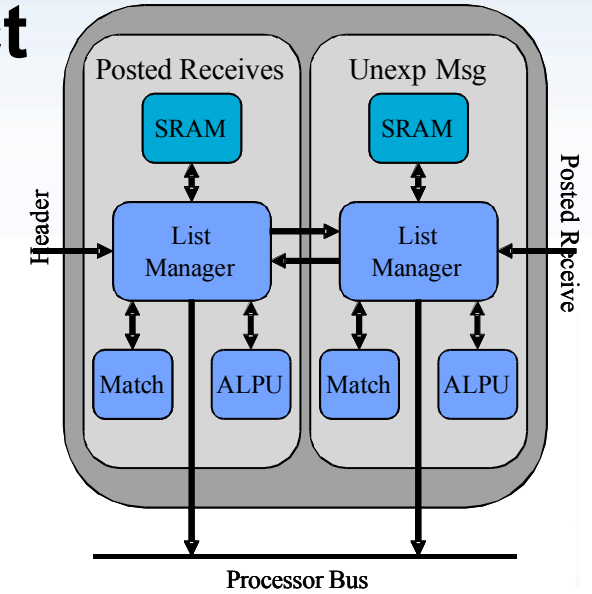
OAK RIDGE National Laboratory

Sandia National Laboratories

Interconnect Project

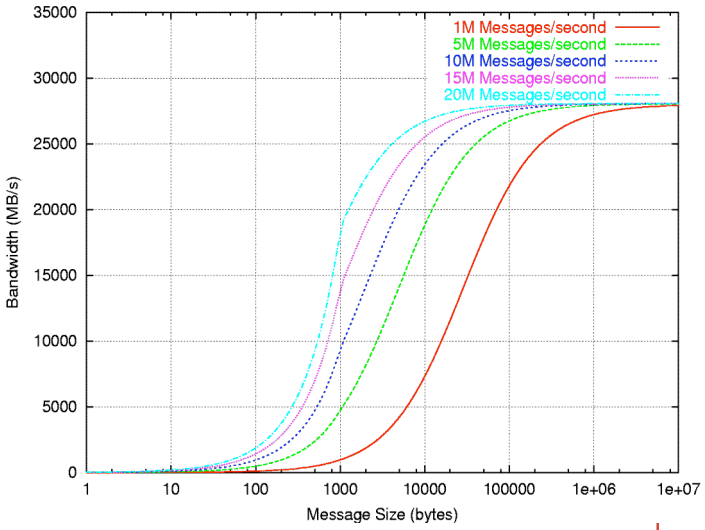
Vision: Ensure next generation interconnects satisfy HPC needs

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



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/st
- *Low Latency*: Maintain ~1us latency across system
- *High Reliability*: $<10^{-23}$ unrecovered bit error rate



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

Potential Collaborators:

- Academic: S. Yalamanchili (parallel simulation), B. Dally (topologies, routing), K. Bergman (optics)
- Industry: Intel, Cray

OAK RIDGE National Laboratory

Sandia National Laboratories

Simulator Project

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

Long Term Goals

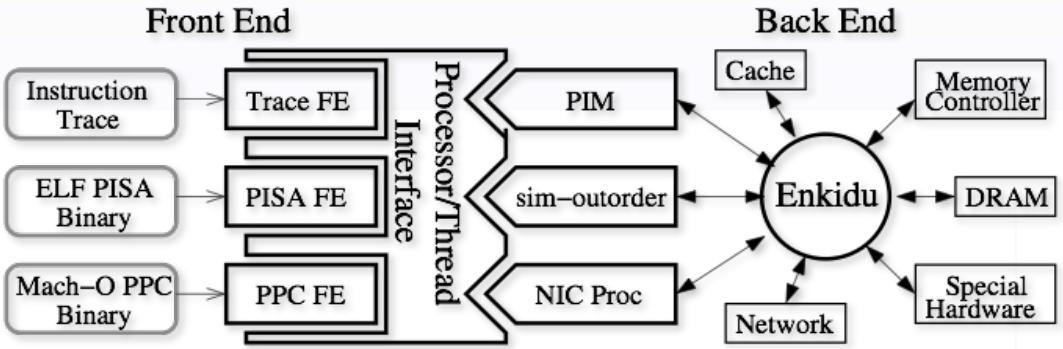
- Highly scalable parallel simulator
- Multi-scale simulation
- Technology model interface

Near Term-Goals

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

Potential 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

Current and future SST user sites

OAK RIDGE National Laboratory

Sandia National Laboratories

Algorithms Project

Long Term Vision

- Close application-architecture performance gap

Long Term Goals

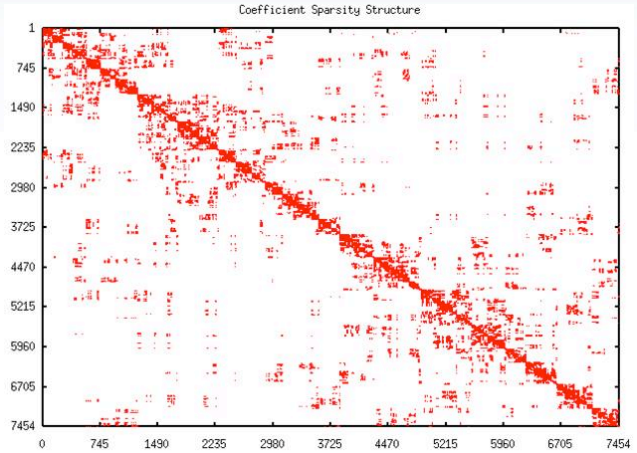
- Architecture-aware algorithms for scalable performance on hierarchical architectures
- Influence & constrain architecture design
- Performance modeling & characterization
- Deployment through libraries and frameworks

Near Term-Goals

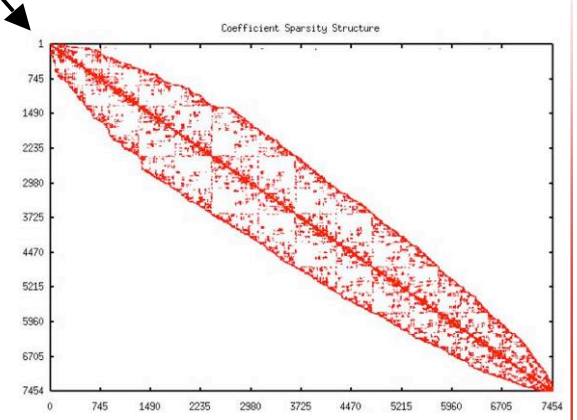
- Develop “microapps” to characterize and study application performance
- Optimize current solvers for hierarchical architectures
- Develop new algorithms kernels with scalable performance on “many-core” processors
- Explore new sources of parallelism & investigate storage/compute tradeoffs
- Investigate different programming models

Potential Partners

- UC Berkeley (Demmel, sparsity & complexity)
- Indiana (Lumsdaine, libraries)
- Tennessee (Dongarra, hybrid algorithms)
- Notre Dame (Kogge, memory systems)
- GT (Vuduc, sparse algorithms)
- Minnesota (Numrich, programming models)
- Illinois (Gropp, scalability)
- PGI (compiler optimization)
- ASCR/SciDAC program (Application integration)



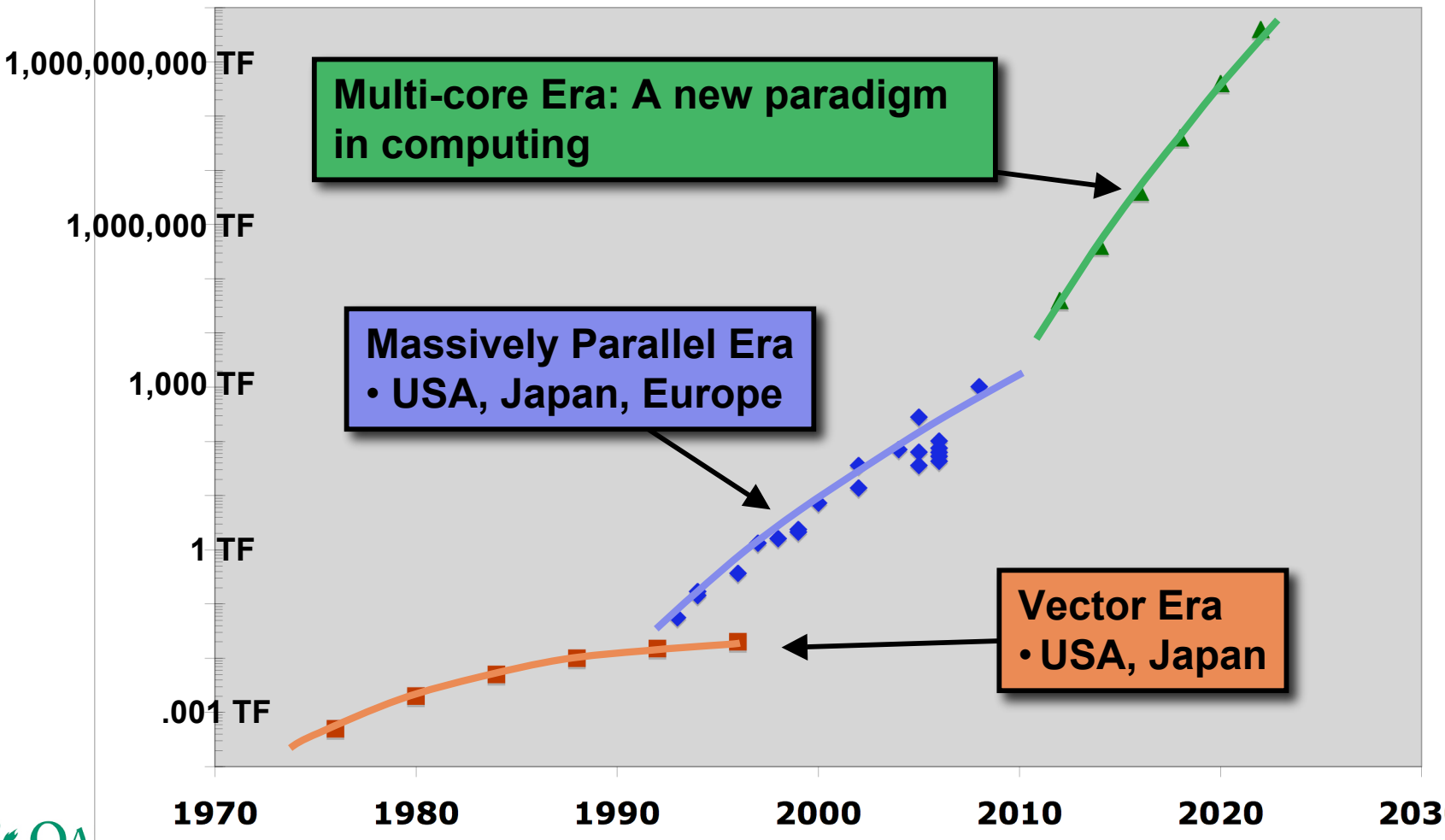
Reverse Cuthill-McKee Algorithm



OAK RIDGE National Laboratory

Sandia National Laboratories

IAA will help prepare DOE and the nation for a new era of computing in collaboration with industry and academia



OAK RIDGE National Laboratory

Sandia National Laboratories