# Sandia's Programs in Supercomputing and Nanotechnology

October 23, 2007

Sudip Dosanjh
Computer and Software Systems
Sandia National Laboratories
sudip@sandia.gov

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

# Science and Engineering Apps

- **Continuum**
  - Computational fluid dynamics
  - Shock physics (CTH)
  - Arbitrary Lagrangian Eulerian (Alegra)
  - Structural mechanics
  - Combustion
  - Device simulations
  - E&M
- **Radiation**
  - Enclosure radiation
- **DAE**
  - Circuit Modeling
- **Particles**
  - Molecular dynamics (LAMMPS)
  - Particle-in-cell

# Informatics is an Emerging App

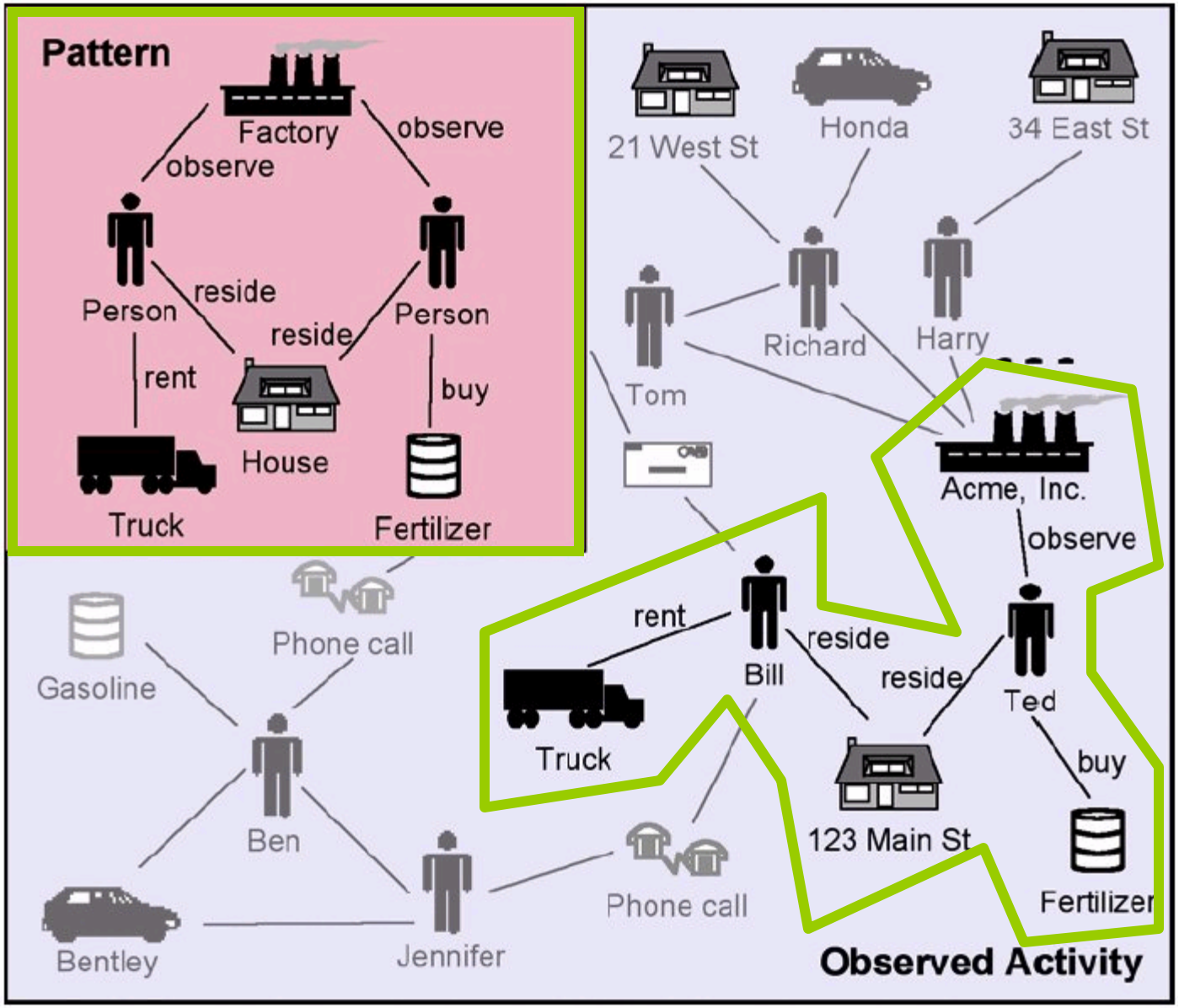



Image Source:
T. Coffman,
S. Greenblatt,
S. Marcus,
*Graph-based technologies for intelligence analysis*,
CACM, 47
(3, March 2004):
pp 45-47

Sandia National Laboratories

# Red Storm

## Before Upgrade

- 10,880 2.0 GHz single-core AMD Opteron CPUs
  - 43.52 TF/s peak
- SeaStar 1.2
- 2-4 GB per socket
- #9 on June 2006 Top 500 list
- Catamount LWK

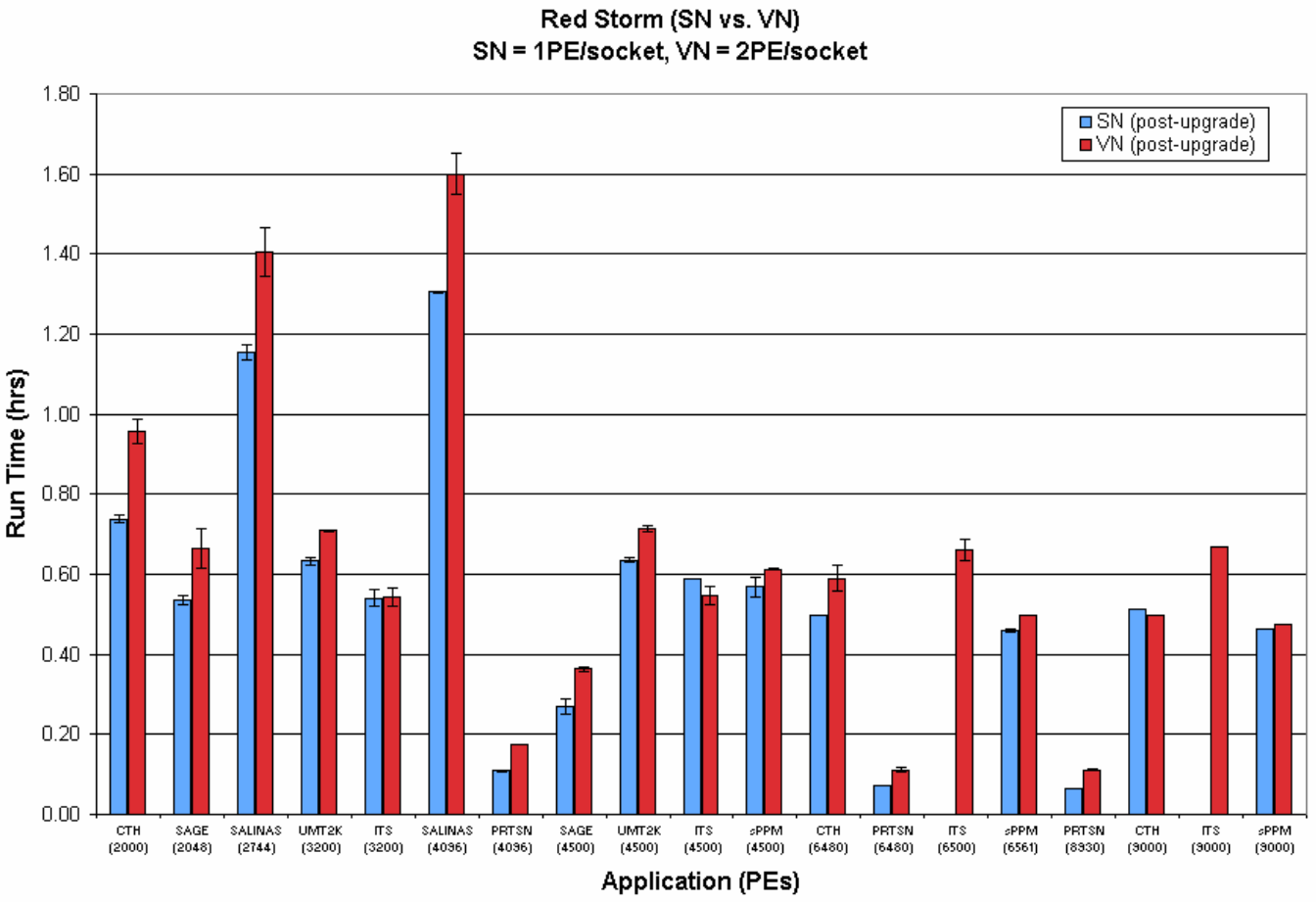
## After Upgrade

- 13,600 2.4 GHz dual-core AMD Opteron CPUs
  - 130.56 TF/s peak
- SeaStar 2.1 network
  - Doubled NIC bandwidth
- 2-4 GB per socket
- #3 on current Top 500 list
- Catamount LWK with virtual node mode support

**Link bandwidth/flop is still reasonable (approx. 1)**
**Some concerns about memory bandwidth/flop**

Sandia National Laboratories

# Catamount Virtual Node LWK Performs Well on 7X Applications



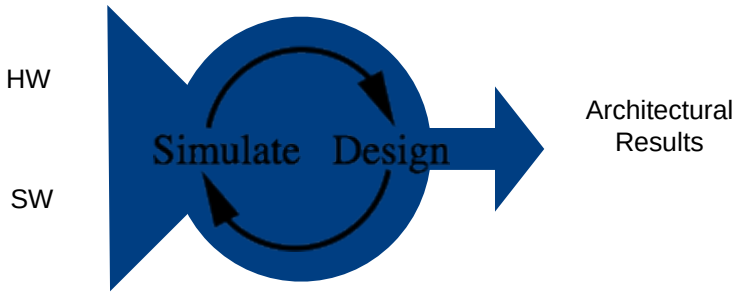

Sandia National Laboratories

# Need Better Modeling

- Better prediction of application performance on new architectures
- Trade-off studies to determine sensitivities to key parameters
  - Improved investment of NRE
- Design of future supercomputers

# Structural Simulation Toolkit (SST)

- Motivation
  - Currently developing a simulation environment to ...
    - Provide validated baseline for future exploration
    - Answer “What If” questions to guide future design efforts
    - Understand complex system-level interactions

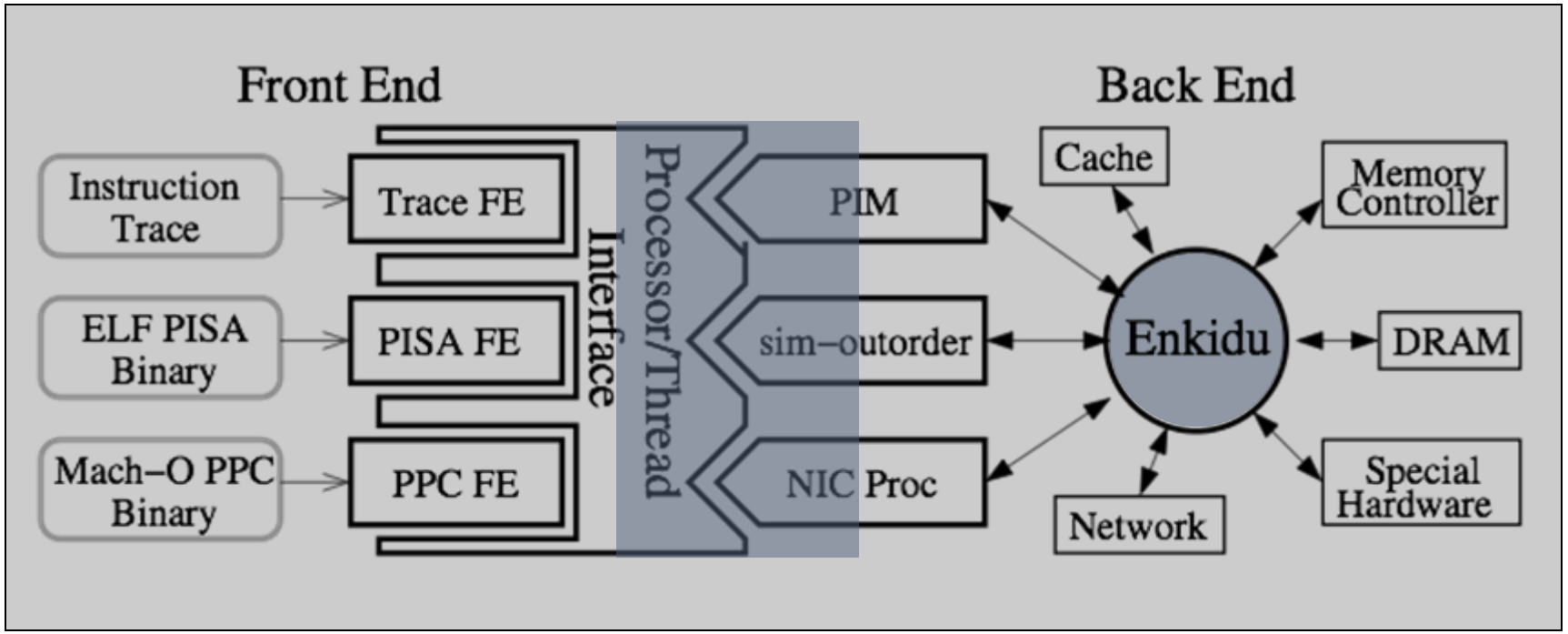



- Goals
  - Focus on parallel systems: HW & SW
  - Quick turnaround
  - Flexibility
    - Multiple front-ends
      - Execution driven
      - Trace driven
    - Multiple back-ends
      - Explore novel architectures (e.g. Multi-core, NIC, Memory)
      - Support conventional architectures (e.g. Single core, DDR)
  - Reusable, Extensible, & Parallelizable

- Customers
  - Micro-architects
  - System Architects
  - Application Performance Analysis

Sandia National Laboratories

# SST: Structure



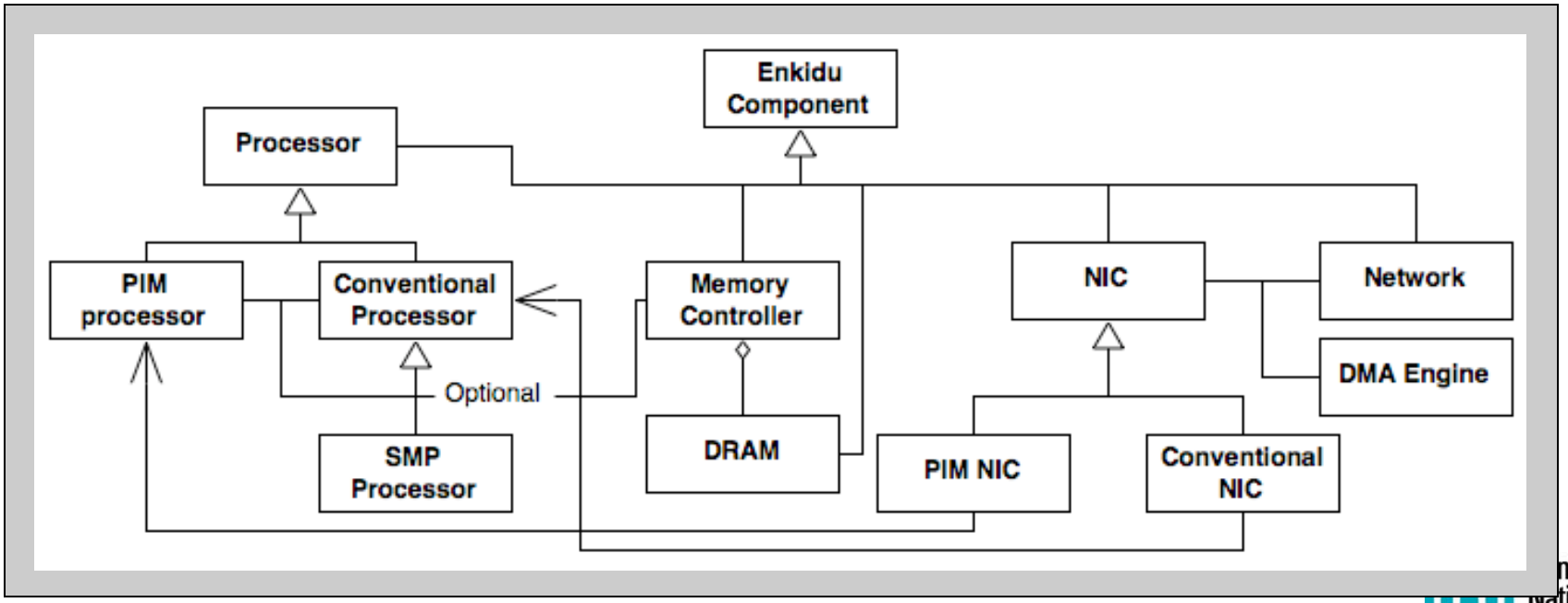

- Front-Ends & Back-Ends Joined by Processor/Thread Interface
- Enkidu “glues” back-end components

# SST: Capabilities and Components

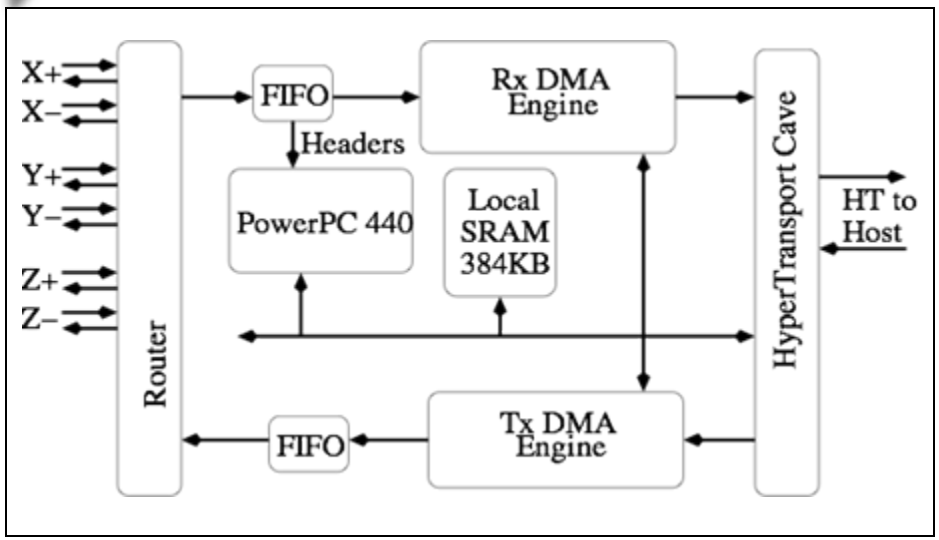
**Processor-in-Memory**
**Multithreaded Processor**
**EDRAM**
**DRAM**
**FBDIMM Channels**
**PIM Network Interface**
**Memory Controller**

**Conventional Processor**
**SMP/CMP Processors**
**Heterogeneous Proc**

**Programmable NIC**
**Simple Network**
**2D/3D Mesh Router**
**PIM NIC Processor**
**DMA Engine**
**NIC BUS**



# Applying SST: Red Storm SeaStar NIC



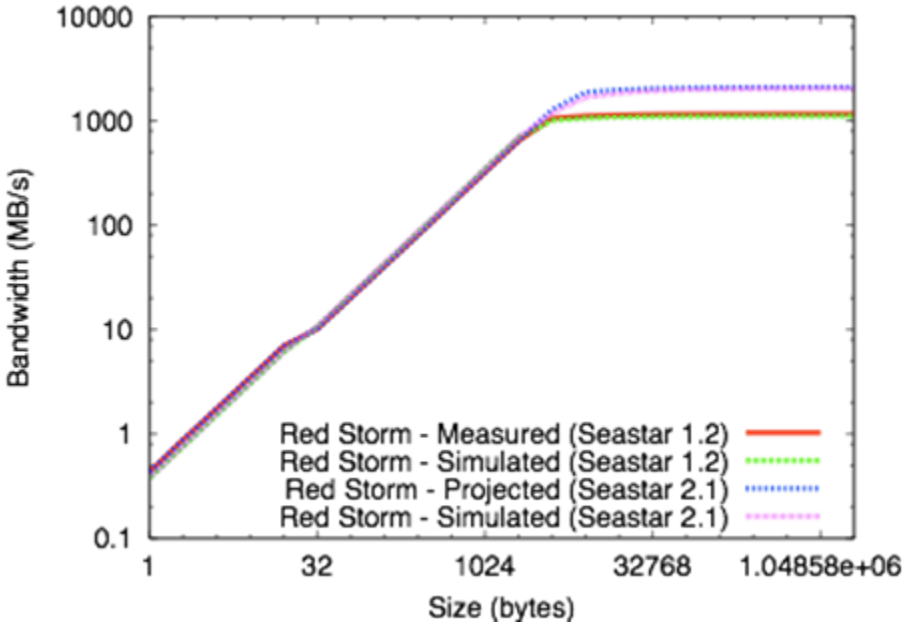

- Architectural Features
  - Embedded 500 Mhz PPC440, local SRAM, DMA Engines, NIC Bus
  - High speed network interface to 3D mesh router
  - 800Mhz HyperTransport interface to CPU
  - Host/NIC communicate through memory

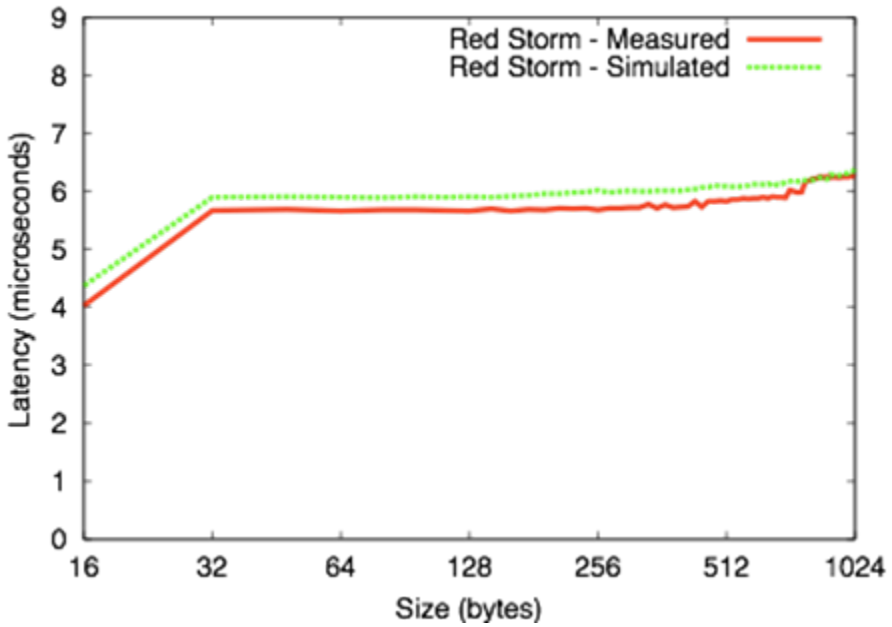
- HyperTransport Modeling
  - HyperTransport connection modeled at two components
  - HTLink models latency
  - HTLink_bw models link bandwidth
    - Models contention
    - Tracks backlog of requests w/ simple BW counting scheme
    - Implements flow-control with finite request queue
    - Queue depth set to cover round-trip times and allow full bandwidth

- NIC Modeling
  - PPC 440: Used SimpleScalar
  - Local SRAM: Existing SST component
  - Tx/Rx DMA engines
    - Existing component
    - Respond to same commands at RS DMA
    - Flow controlled
  - HT Interface
    - Connects CPU/NIC
  - NIC Bus
    - Connects internal NIC components (PPC, SRAM, etc...)

Sandia National Laboratories

# Validating SST: Latency & Bandwidth



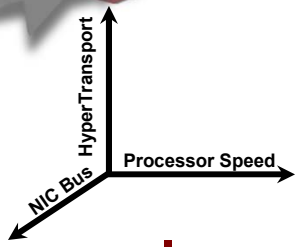

- Used MPI “ping-pong” and OSU streaming BW
- Compared with real Seastar 1.2 and 2.1 chips
- Latency, message rate, and bandwidth
  - within 5% for range of sizes

Sandia National Laboratories

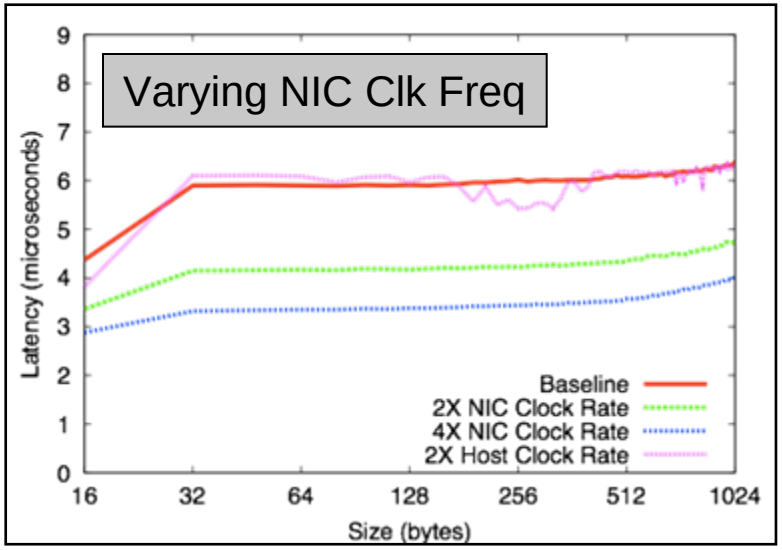
# Validating SST: Primitives

| Routine | Simulated | Actual |
|---|---|---|
| PUT Command | 0.486 | 0.592 |
| tx_complete USER | 0.196 | 0.154 |
| rx_message ACK | 0.959 | 1.002 |
| rx_complete ACK | 0.127 | 0.242 |
| POST Command | 0.477 | 0.442 |
| rx_message USER | 1.936 | 1.686 |
| tx_complete ACK | 0.114 | 0.118 |
| rx_complete USER | 0.230 | 0.378 |

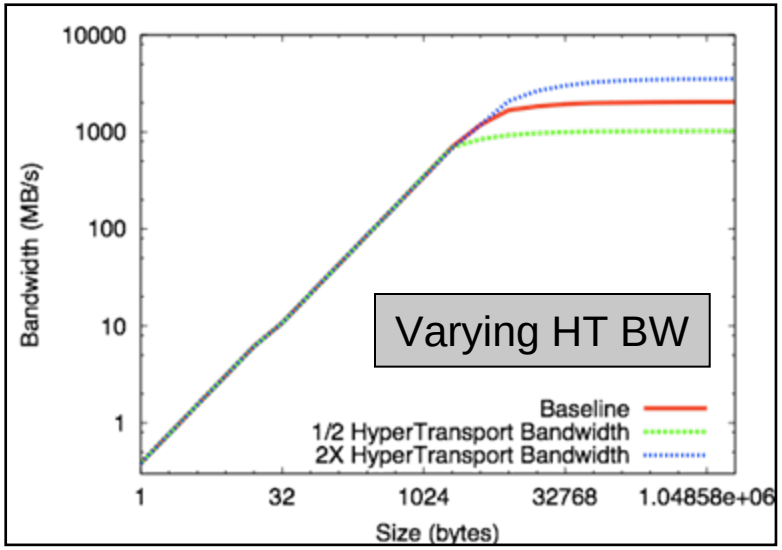
- Sources of Error
  - Small message optimization in Red Storm (<16 bytes)
  - Lack of cache-line invalidation instruction
  - Processor model?



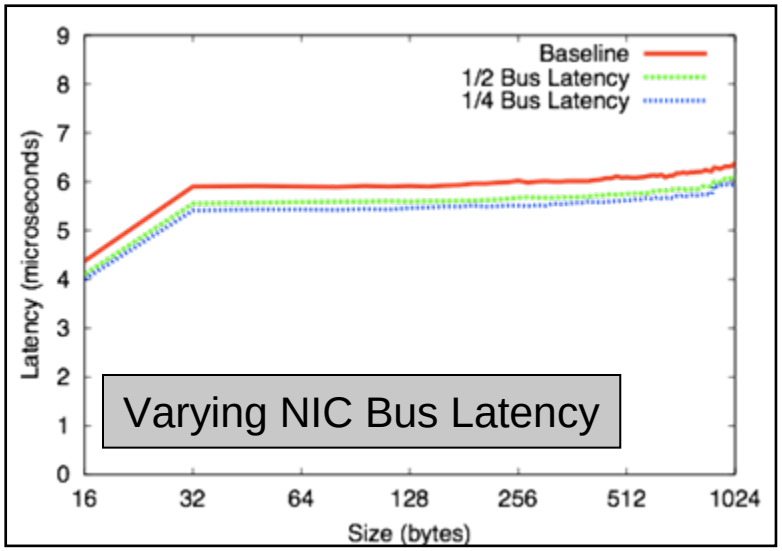

# Design Space Exploration



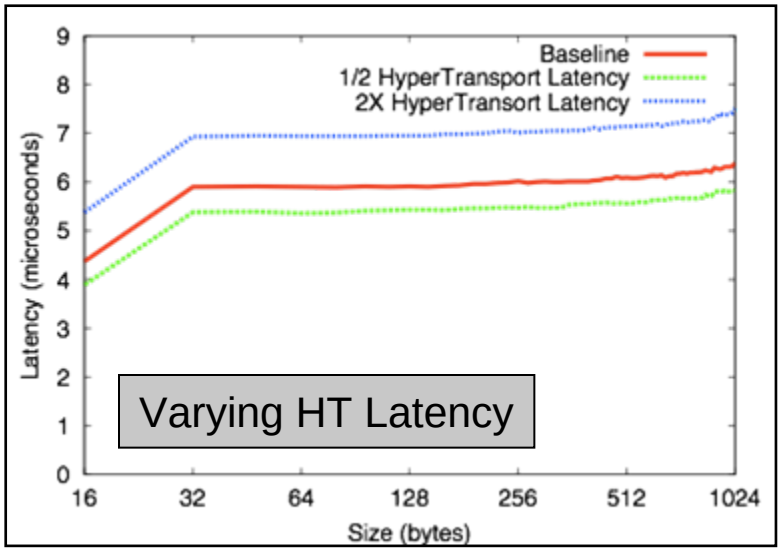

- Vary NIC Clock Rate Effect On Latency:
  - 2X NIC Clock -> 30% improvement in latency
  - 4X NIC Clock -> 50% improvement




- Vary HT Bus Effect On Bandwidth:
  - No effect to small messages
  - Does effect peak bandwidth



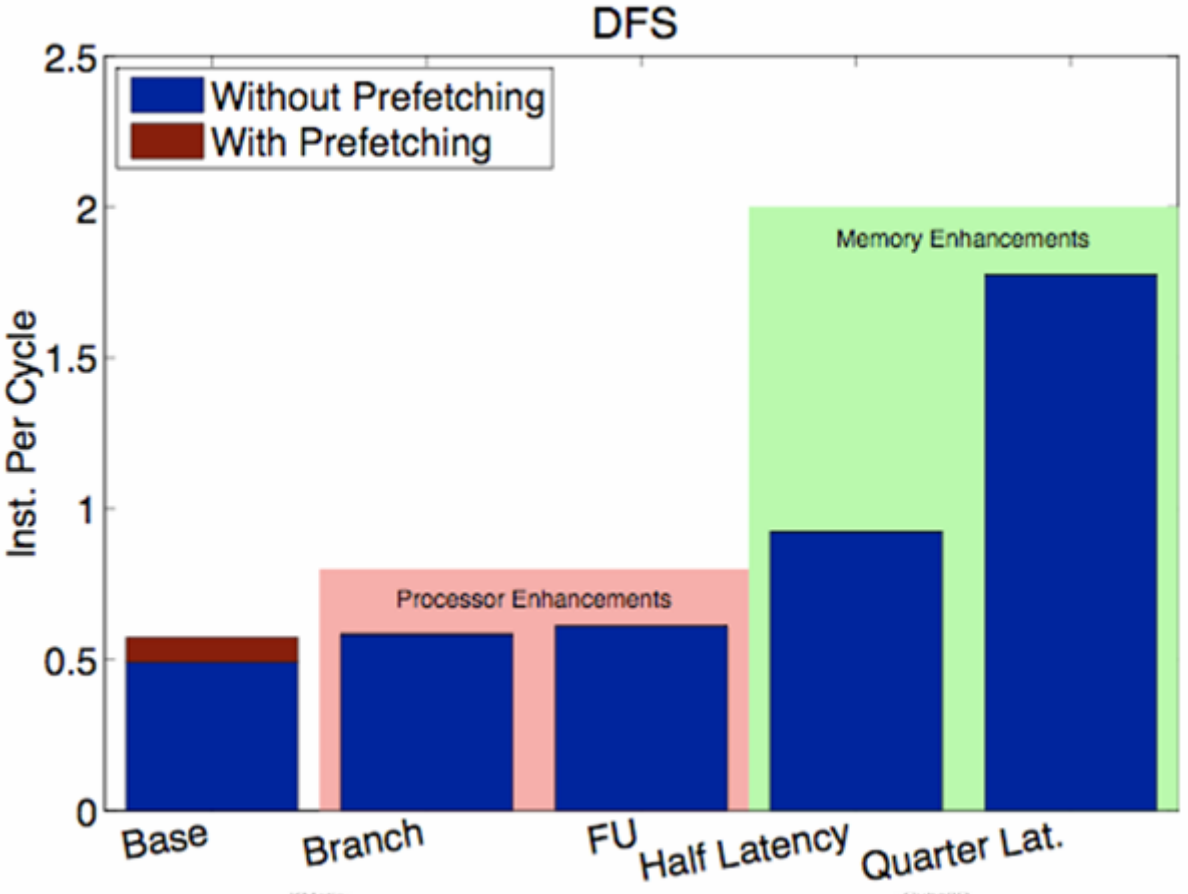

- Vary NIC Bus Effect On Latency:
  - 1/2 Bus Latency -> 8% latency reduction

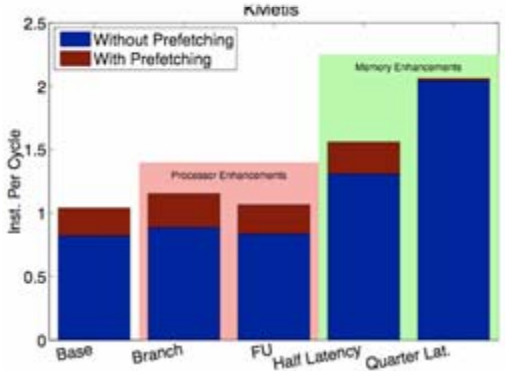


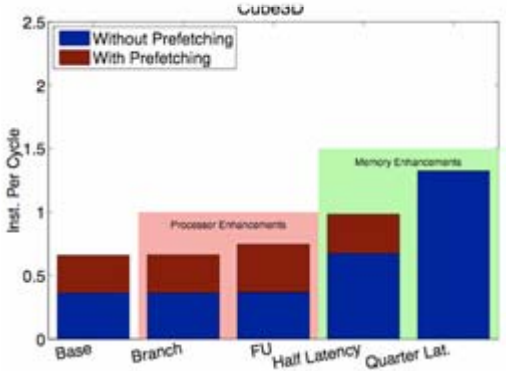

- Vary HT Bus Effect On Latency:
  - MPI latency increases linearly with HT latency
    - 4 HT transactions per MPI message

Sandia National Laboratories

# Finding the Bottleneck: Computation, Branches. or Memory



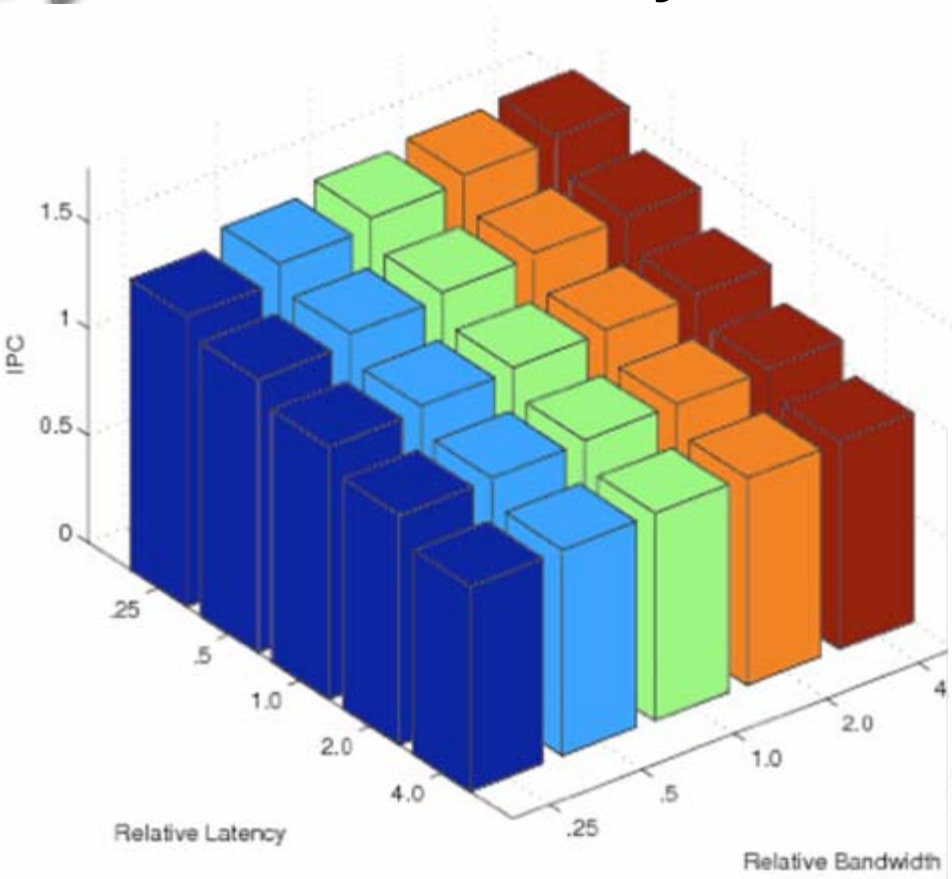




- In the node, Memory performance is key bottleneck
- Even perfect branch prediction and infinite FUs would be less valuable than improving memory latency.
- Prefetching, caches don't help emerging applications

Sandia National Laboratories

# Latency/Bandwidth Sensitivity

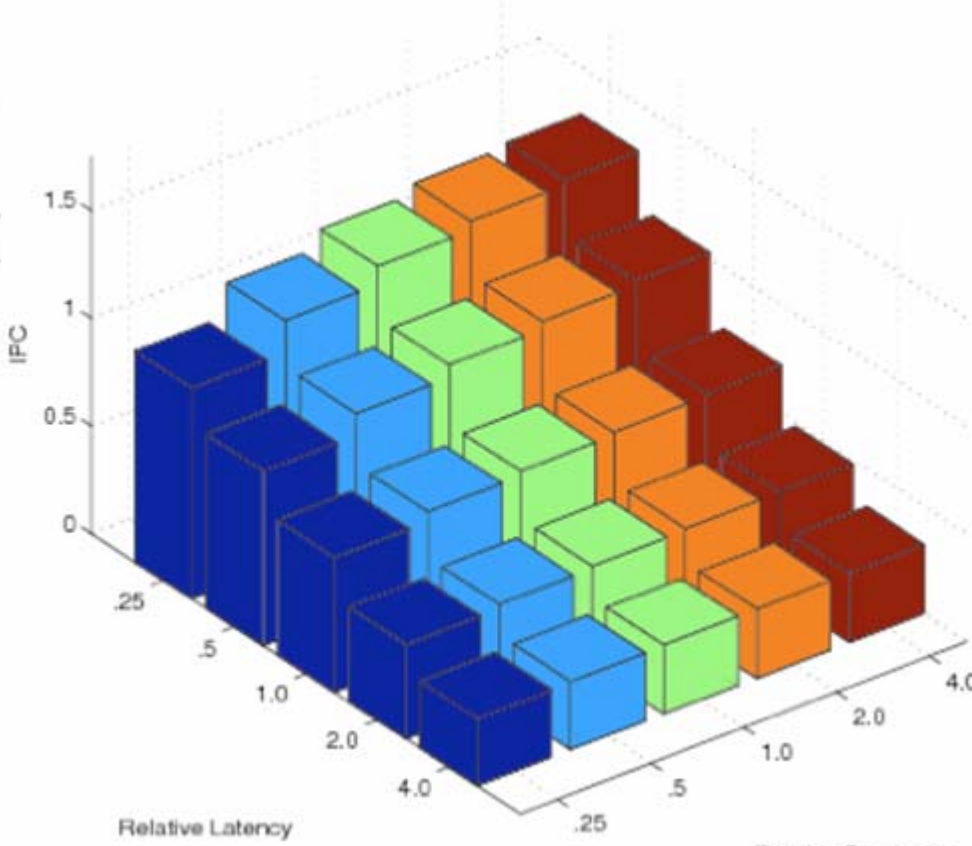



Scientific Applications

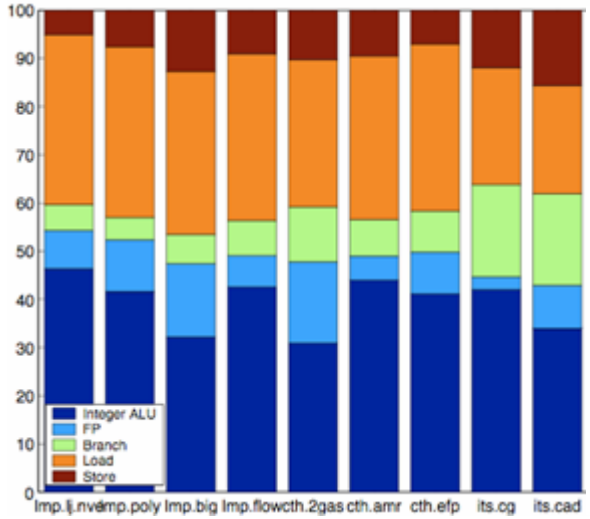
Emerging applications more sensitive to Latency and Bandwidth

Latency & Bandwidth are **both** constraining performance

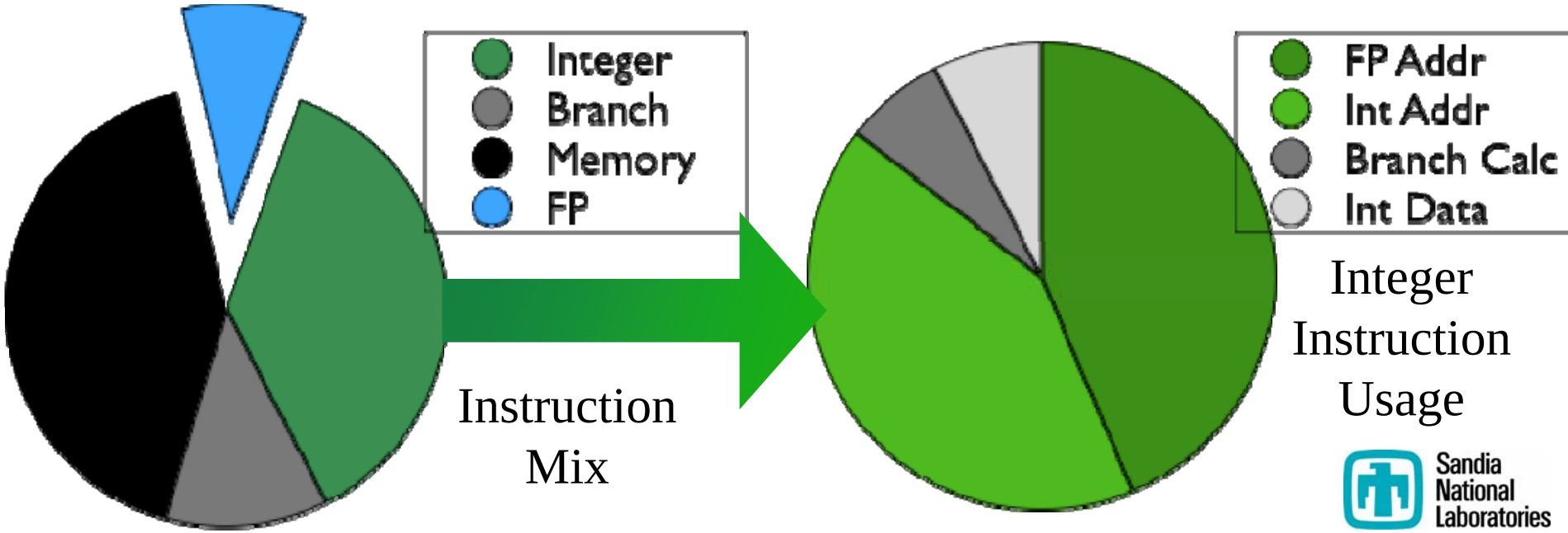
Informatics Applications



# Memory Operations Dominate

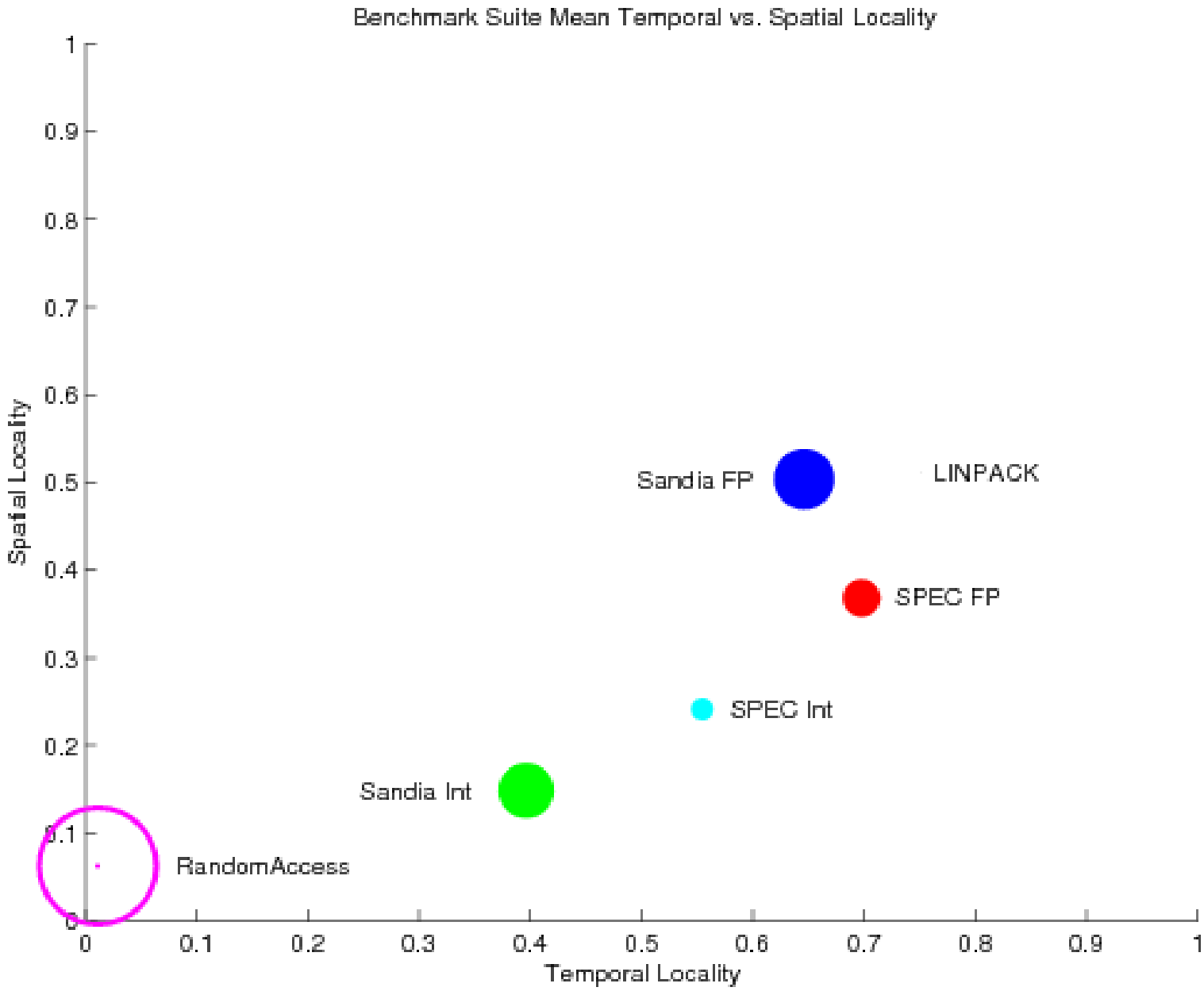



- FP ops ("Real work") < 10% of Sandia codes
- Several Integer calculations, loads for each FP load
- Memory and Integer Ops dominate
  - ...and most integer ops are computing memory addresses
- Theme: processing is now cheap, data movement is expensive




Sandia National Laboratories

# Application Characteristics




Sandia National Laboratories

# Viewgraph from Portland Group

## We Need a Change of Mindset

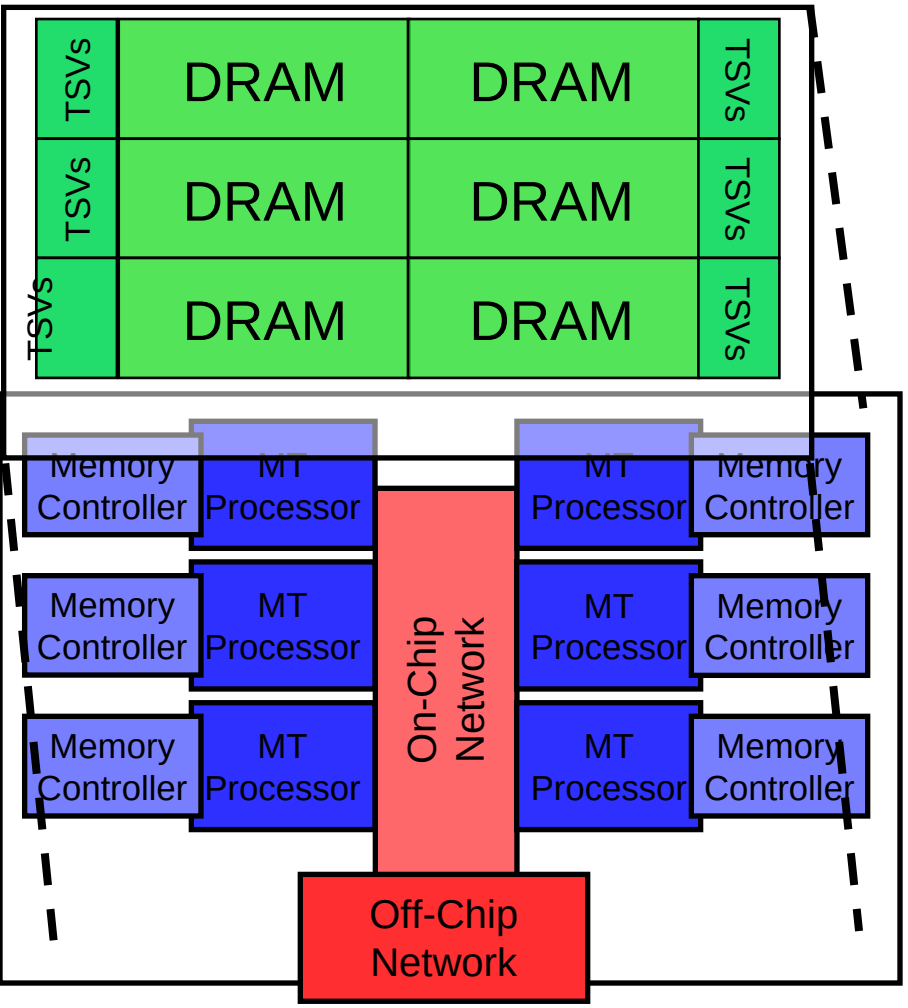
- FLOPS are “free”.  In most cases we can now compute on the data as fast as we can move it.
- CPUs (cores) must be optimized for efficient coordinated data movement.
- Compilers/tools must enable applications to benefit from multi-core CPUs
- Applications should be designed to minimize data movement.

# Issues

- Opportunity cost associated with building such a machine
- Industry interest in investigating different packaging technologies at Sandia

Sandia National Laboratories

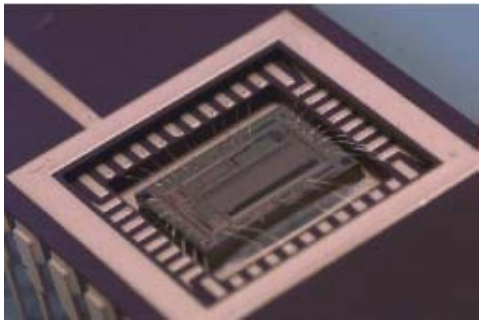
# Example Prototype Machine




- 3D Stacked Homogeneous Processing-In-Memory (PIM) Array
  - Hardware support for multithreading/thread migration
  - Enhanced Synchronization
  - Low latency/high bandwidth 3D stacked memory system
  - Highly scalable
    - Tight integration with network
  - Short vector processing
- Small Array (10's-100's of chips, 100's of GBs of memory), boards, software
- Industry collaboration for the memory system

Sandia National Laboratories

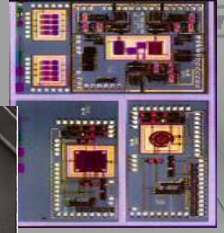
# Technical Challenges

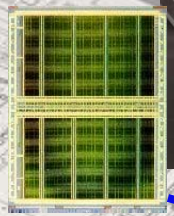


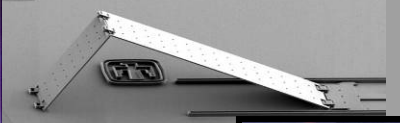
- Architecture
  - New Multithreaded Architecture
  - New Synchronization Mechanisms
  - New ISA
- System Software
  - Thread and Global Address Space Management
- VLSI Implementation
  - New (but simple!) architecture, power, validation
- Fabrication and Packaging
  - 3D integration, network implementation (SERDES or optics)
- Algorithms and Applications
  - Mapping to new architecture/programming model
  - New Application Classes (e.g., informatics)
- Compilers and Programming Models
  - Expressing multilevel parallelism and synchronization
  - Lack of easy infrastructure for targeting new architectures
- System Integration
  - Actually bringing a machine up in the lab

# MESA Complex

## Components







## System Engineering

Weapons Integration Facility
374 people
162,000 GSF

Construction: $77M
Equipment: $16M

MicroFab
0 people
98,000 GSF

Construction: $114M
Equipment: $139M





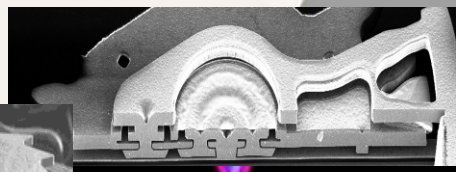


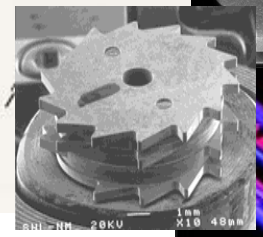
**Integrated, Co-located Capability for Design, Fabrication, Packaging**

## Science

MicroLab
274 people
131,000 GSF

Construction: $55M
Equipment: $13M



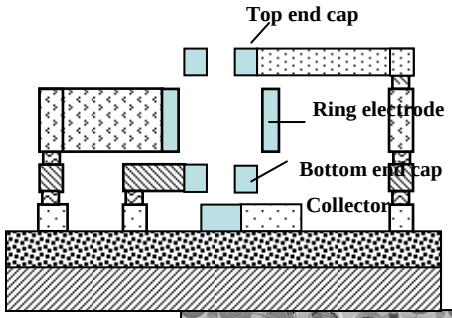


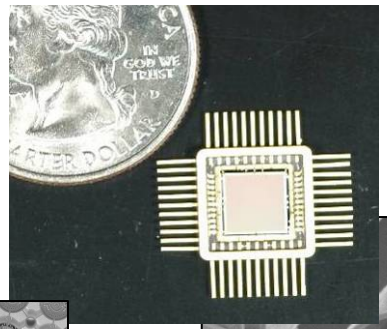
| | | | |
|---|---|---|---|
| TOTALS: | 391,000 GSF | 648 People | |
| Construction: | $246M | Equipment: | $168M |
| Contingency: | $48M | TEC: | $462M |

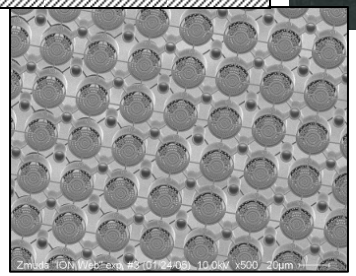
# Relevant Sandia Capabilities

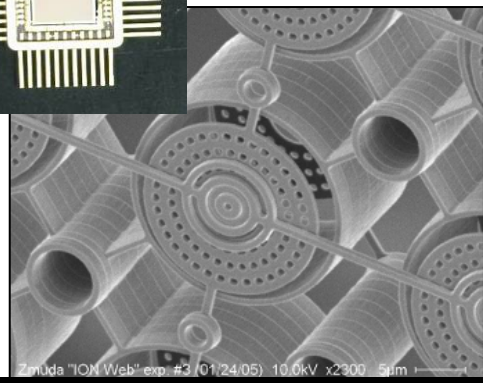
## Micro Ion Trap Fabrication

- Design of micro ion traps
- Microfabrication of MEMS-based micro ion traps
- Simulation of ion trap potentials and ion trajectories
- Robust packaging of micro ion trap arrays



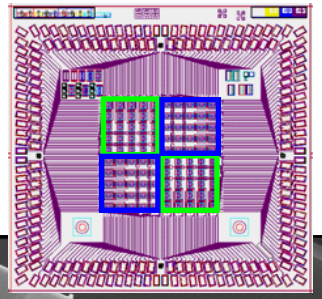



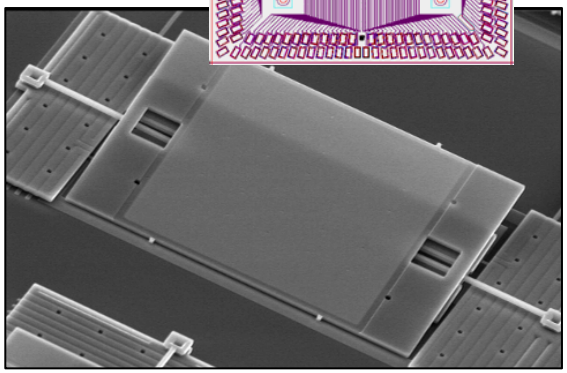


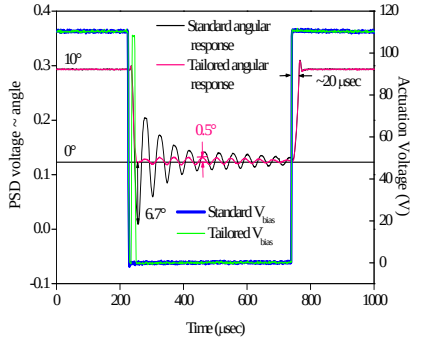


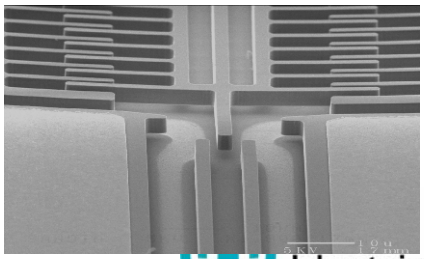

## Integrated Micro-optic Elements

- Design, modeling, and fabrication of MEMS-based micro mirrors for micro-optic applications
- Integration of micro mirrors and solid-state waveguides
- Control algorithms for micro-mirror operation







Laboratories

# Center for Integrated Nanotechnologies

Sandia National Laboratories • Los Alamos National Laboratory

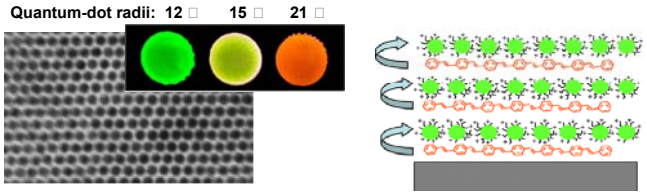
***"One scientific community focused on nanoscience integration"***


- **World-class scientific staff**
- **Vibrant user community**
- **State-of-the-art facilities**
- **A focused attack on nanoscience integration challenges**
- **Leveraging Laboratories' capabilities**
- **Developing & deploying innovative approaches to nanoscale integration**
- **Discovery through application with a diverse portfolio of customers**

Sandia National Laboratories

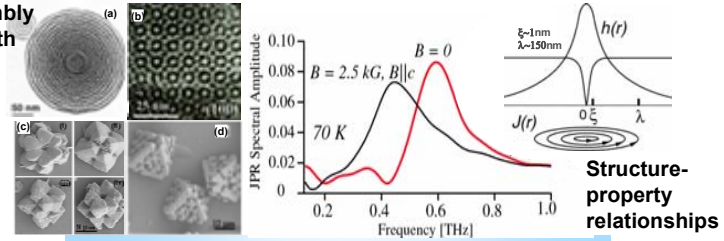
# CINT Thrust Areas provide broad base of expertise

## Nanoelectronics & Nanophotonics:
**Precise control of electronic and photonic wavefunctions**



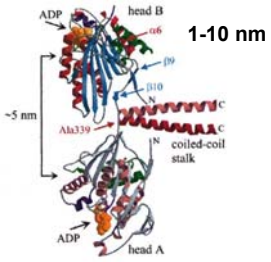

## Complex FunctionalNanomaterials:
**Relationships between synthesis, structure and complex and emergent properties**

Self-assembly on all length scales

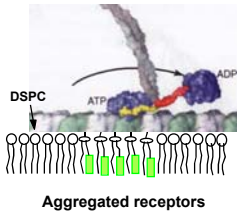



Structure-property relationships

## Nano-Bio-Micro Interfaces:
**Biological principles & functions imported into artificial bio-mimetic systems**



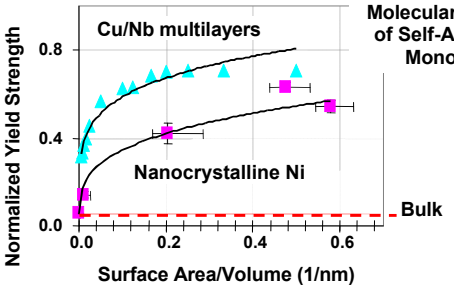

1-10 nm

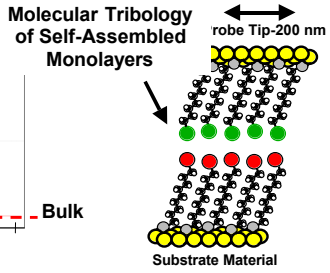



Aggregated receptors

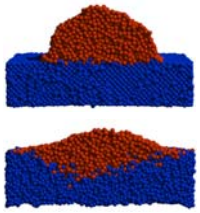
10 -100 nm

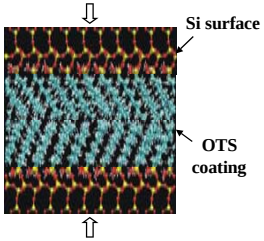
## Nanomechanics:
**Understanding the mechanical behavior of nanostructured materials**

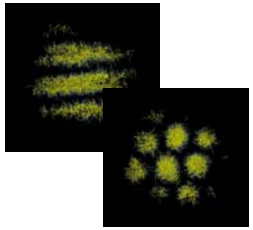



## Theory & Simulation:
**Theoretical, modeling and simulation techniques for multiple length and time scales and functionality**





Sandia National Laboratories

# Future challenges

- Data locality on chip
- Impact of programming models
- Accelerators