SAND2011-6687C

Mini-Applications: Tools for Co-Design

Richard F. Barrett
Center for Computing Research
Sandia National Laboratories

SAND 2011-6687C

VNIIEF XIII International Workshop
Supercomputing and Mathematical Modeling
Sarov, Russia

October 3-7, 2011

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.

Co-Design

A model for cooperative development

• Detailed interactions at each boundary.

• Higher level discussions among colored areas

Reference: Geist, A. and Dosanjh, S., "IESP Exascale Challenge: Co-Design of Architectures and Algorithms", *International Journal of High Performance Computing, 2009*

Application
Methods
Algorithms
Basic Computations
Programming Model
Programming Mechanisms
Libraries
Compiler
Runtime System
Operating System
System Architecture
Node: Processors, Memory
Inter-node: Interconnect
Enabling Technologies
Fundamental Technology Capabilities

Sandia National Laboratories

Glossary

- Skeleton Application:
  - Communication accurate, computation fake.
- Compact Application:
  - A small version of a real application.
  - Attempting some tie to physics.
- Benchmarks
  - HPCC, NAS, SPEC, HPL

Mini-application specs

Intent	Provides a context for discussion across the Co-design space
Focus	Proxy for key app perf issue
Developer & owner	Application team
Scope of Change	Any and all
Size	O(1k) lines of code
Availability	Open Source
Life span	Until its no longer useful

Mantevo Project

Greek for "predict"

Participants:

National Laboratories, Universities, and Industry

We welcome your participation

Mini-Applications:

- **HPCCG:** HPC Conjugate Gradient.
- **miniFE**: unstructured implicit FEM/FVM.
- **phdMesh**: explicit FEM, contact detection.
- **MiniMD**: Molecular Dynamics Force computations.
- **MiniXyce**: Circuit RC ladder.
- **MiniALE**: ALE remap step
- **MiniGhost**: Structured Eulerian
- **MiniAero**: aerospace engineering application; tbd.

Under development

http://software.sandia.gov/Mantevo

Sandia National Laboratories

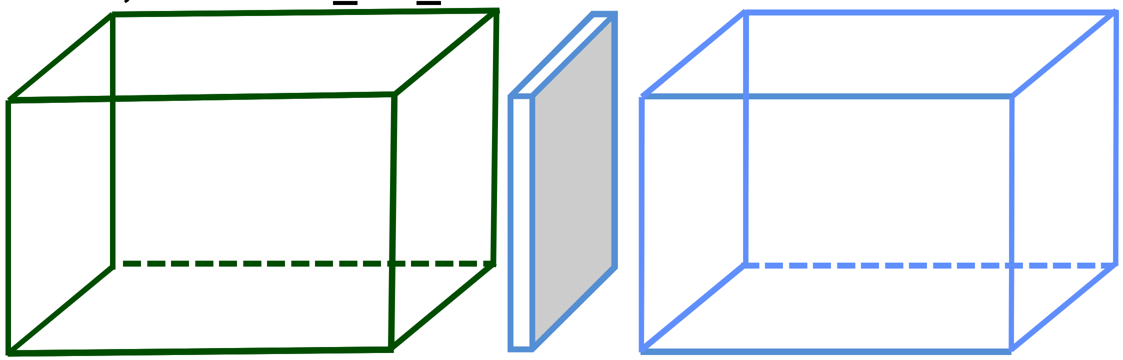
Two Critical Issues for a Mini-Application

- **What does it represent?**
- **What does it *not* represent?**

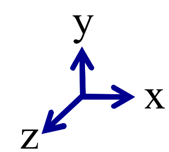
Sandia National Laboratories

Mini-Application: MiniGhost
Exploring Bulk Synchronous Parallel (BSP) model With Message Aggregation

DO I = 1, NUMBER_OF_VARIABLES

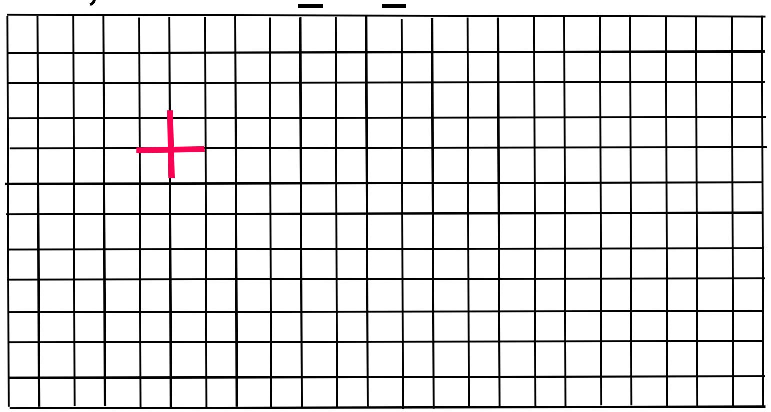


Exchange boundary data



END DO

DO I = 1, NUMBER_OF_VARIABLES

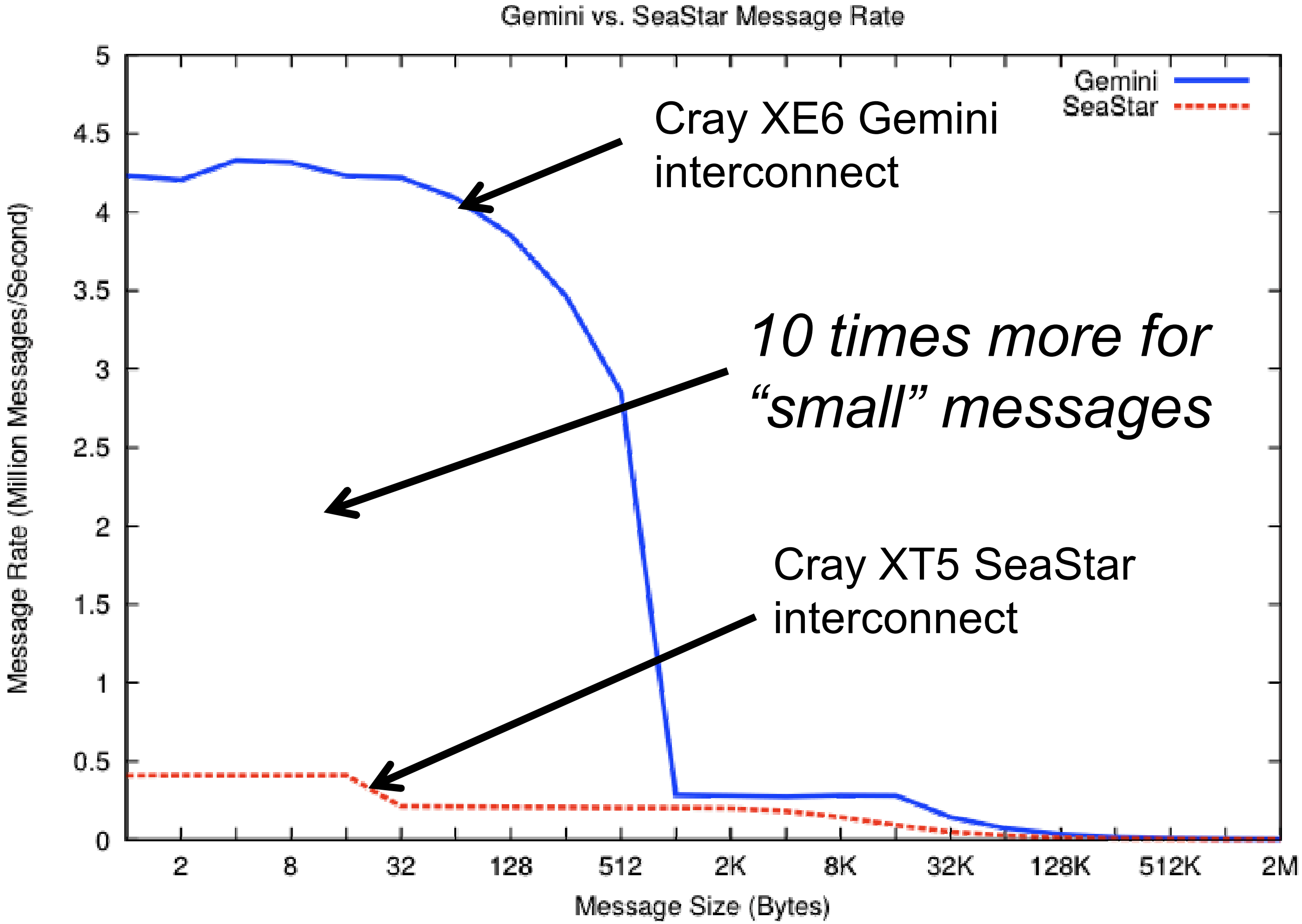


Computation

END DO

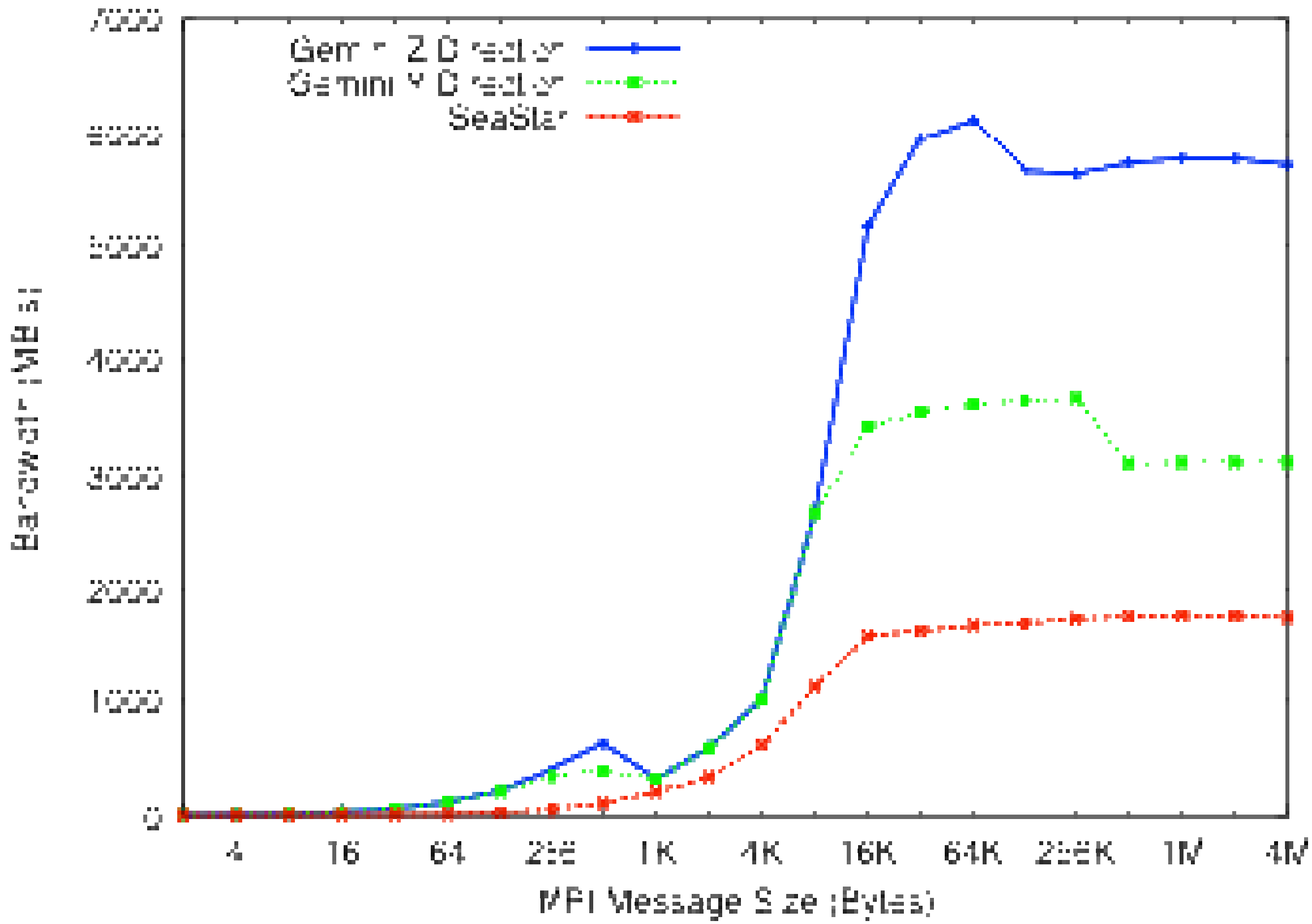
Sandia National Laboratories

Cielo Cray XE6 Gemini node inter-connect *Message Inject Rate*



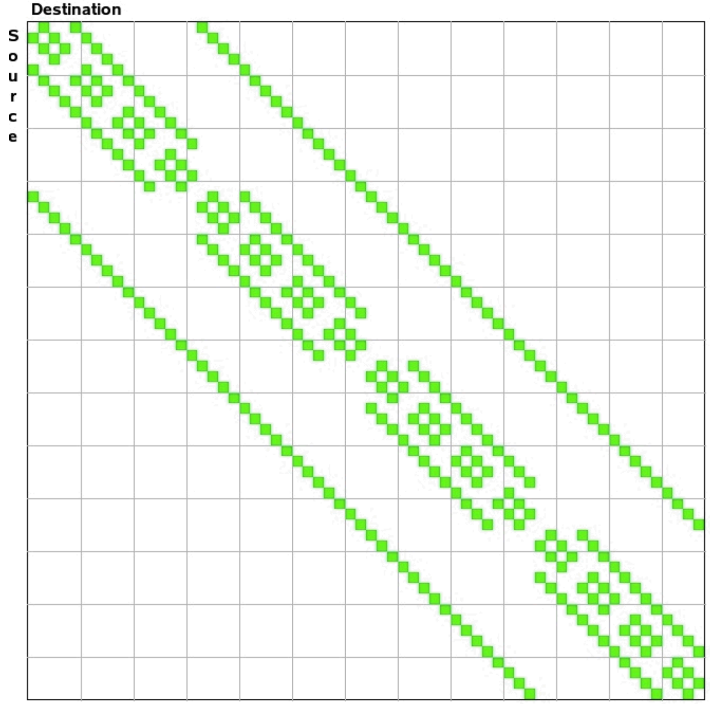
Sandia National Laboratories

Cielo Cray XE6 Gemini node inter-connect
Bandwidth

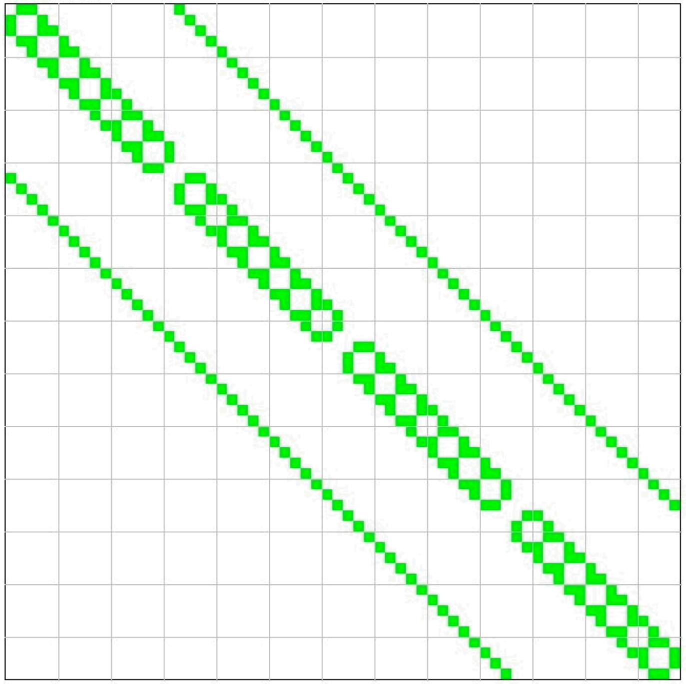


Inter-process Communication patterns

miniGhost

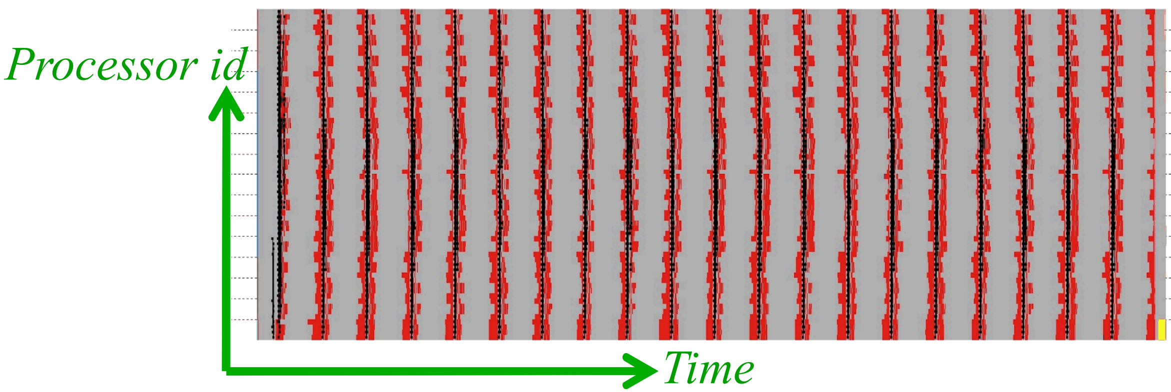


Represented Application

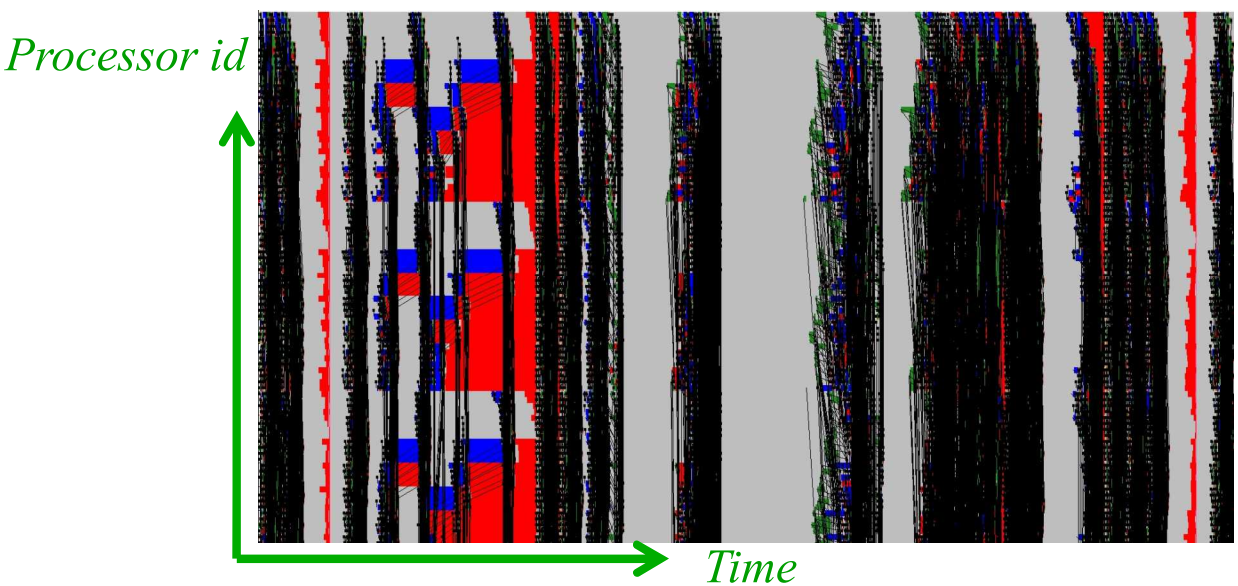


- Processor row *i* sends to processor column *j*
- Color indicates volume.

Runtime profiles



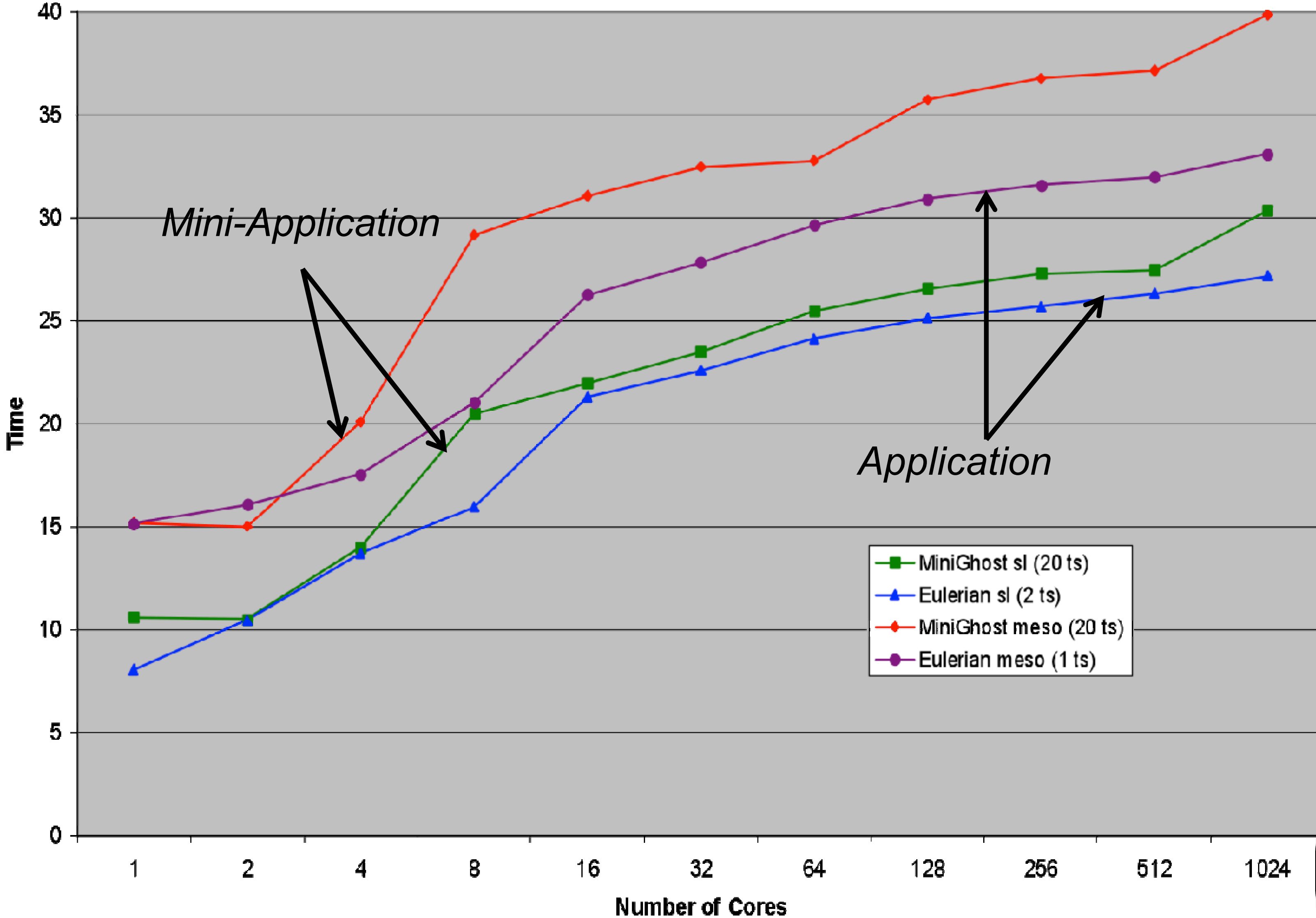
miniGhost
(many time steps)



Represented Application
(1 time step)

Gray is computation, black communication, red synchronization

Performance Comparison MiniGhost and Application (Two problem sets on Cray XT5)



Sandia National Laboratories

Mini-Application: HPCCG

- Solves sparse linear system of equations using the Conjugate Gradient (CG) Method.
- Found to not adequately create the context for the represented application we are studying.
- So…

Sandia National Laboratories

Mini-Application: MiniFE

- Domain: 3D box of finite elements
  - But structure not exploited, so “unstructured”
- Recursive Bisection of hexahedra elements
- Stiff system: Linear, symmetric positive definite matrix from 27-pt stencil, solved using CG
- Options:
  - Inject computational imbalance, MPI-overlap, threads (OpenMP, qthreads, Trilinos TPI), CUDA, Intel TBB.
- 1,500 lines of C++ code

MiniFE

Solves the element diffusion matrix for the steady conduction equation[1]

$$(K^e_{12})_{xy} = \int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} k_{xy}\left(J^*_{11}\frac{\partial\psi_1}{\partial\xi} + J^*_{12}\frac{\partial\psi_1}{\partial\nu} + J^*_{13}\frac{\partial\psi_1}{\partial\zeta}\right) \cdot$$

$$\left(J^*_{21}\frac{\partial\psi_2}{\partial\xi} + J^*_{22}\frac{\partial\psi_2}{\partial\nu} + J^*_{23}\frac{\partial\psi_1}{\partial\zeta}\right)|J|d\xi d\nu d\zeta$$

$$\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} F(\xi,\nu,\zeta)d\xi d\nu d\zeta \approx \Sigma_{I=1}^{M}\Sigma_{J=1}^{N}\Sigma_{K=1}^{P}F(\xi_I,\nu_J,\zeta_K)W_I,W_J,W_K$$

[1] *"The Finite Element Method in Heat Transfer and Fluid Dynamics, 2nd Edition", Reddy and Gartling, CRC Press, 2001.*

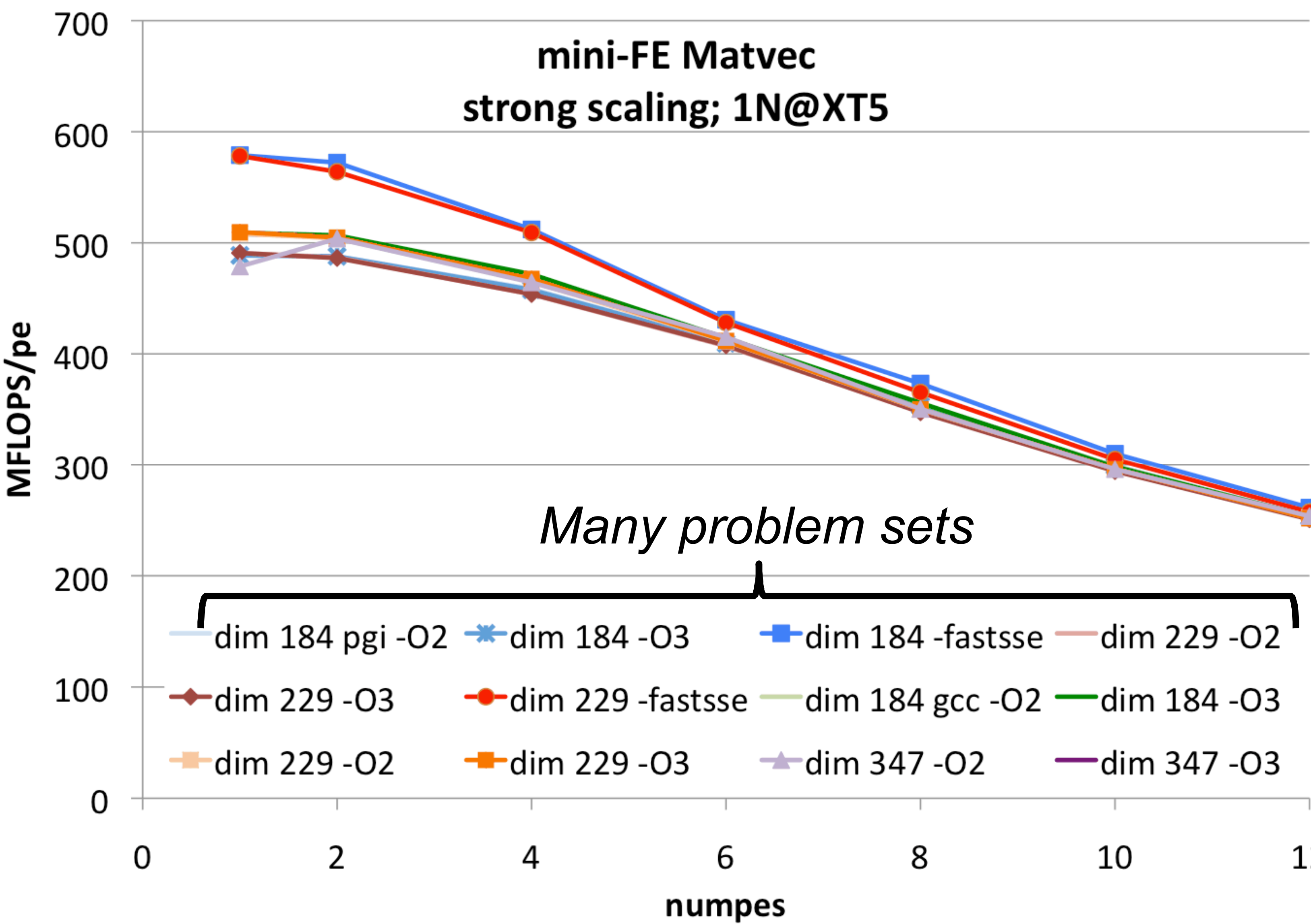
Sandia National Laboratories

Represented application uses only –O2 optimization Should it go higher?

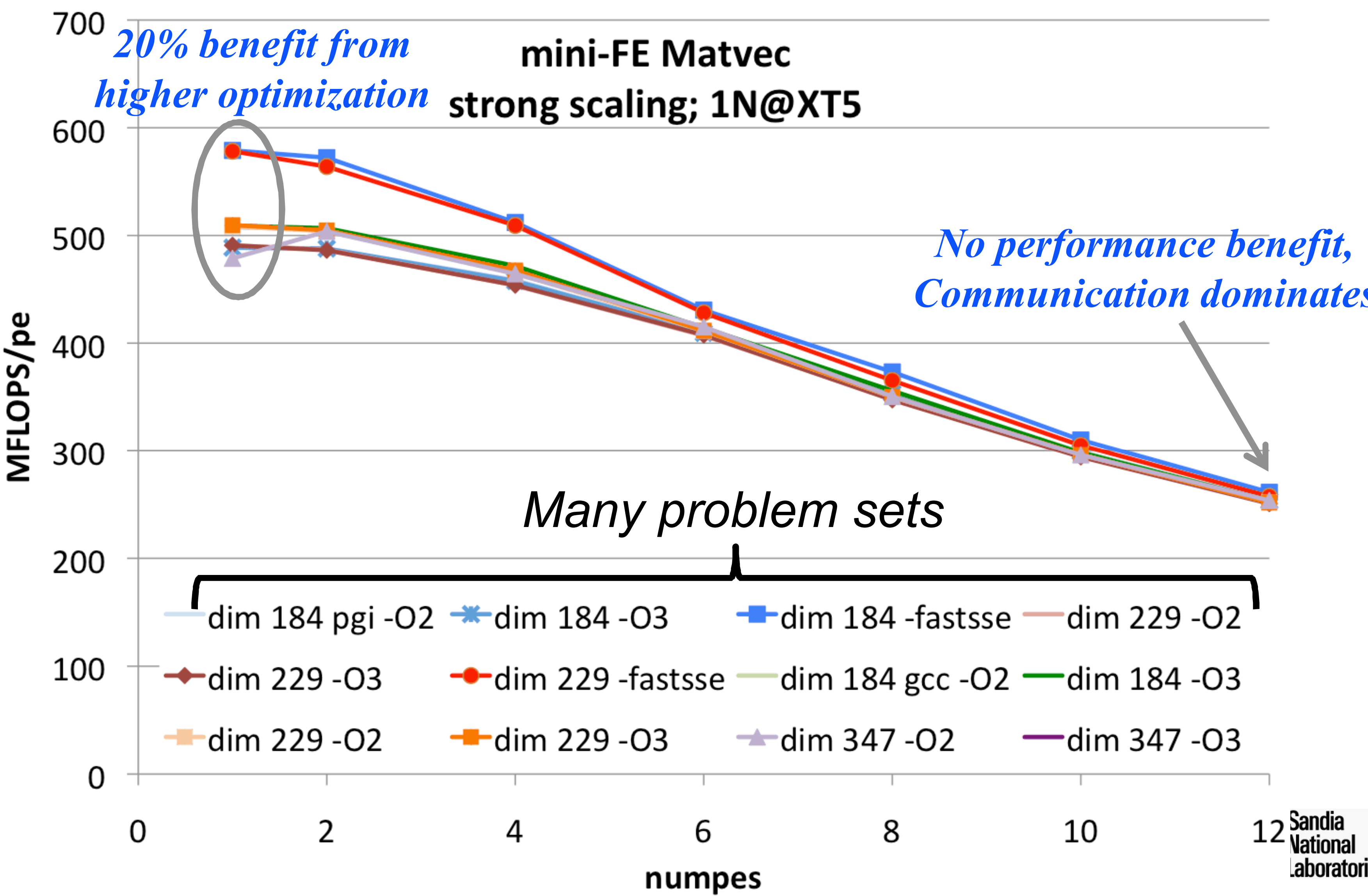
- Experiment using miniFE with different compilers and optimization levels

Sandia National Laboratories

Full application uses only –O2 optimization Should it go higher?



Full application uses only –O2 optimization Should it go higher?



Cielo Cray XE6 Node Architecture

NUMA node: 4 cores share memory and L3 cache

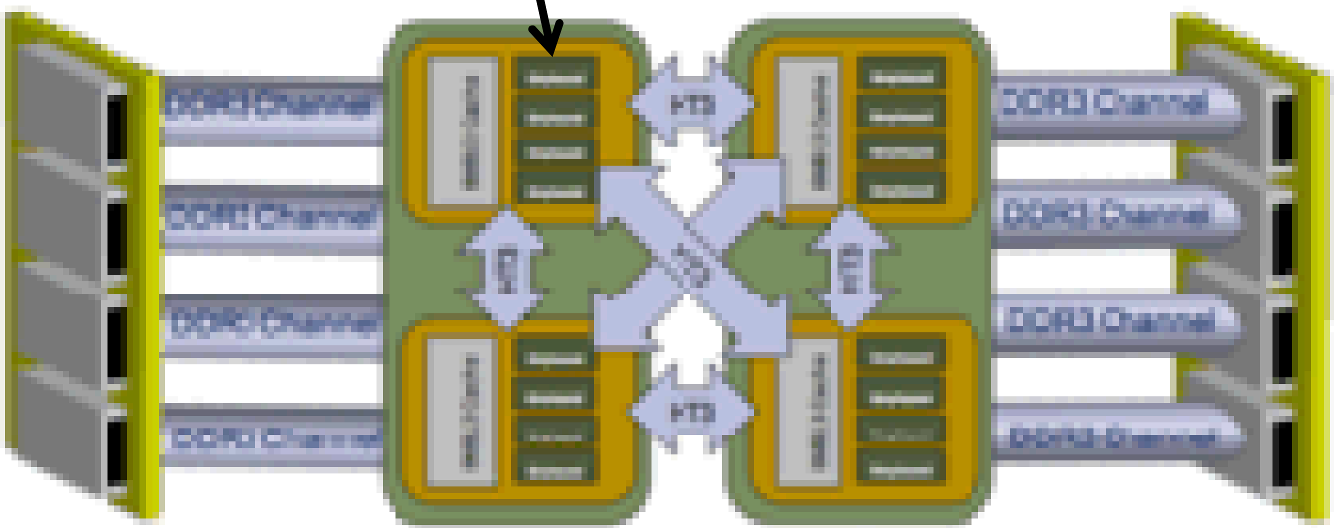


Image courtesy Cray, Inc.

Cielo Cray XE6 Node Architecture

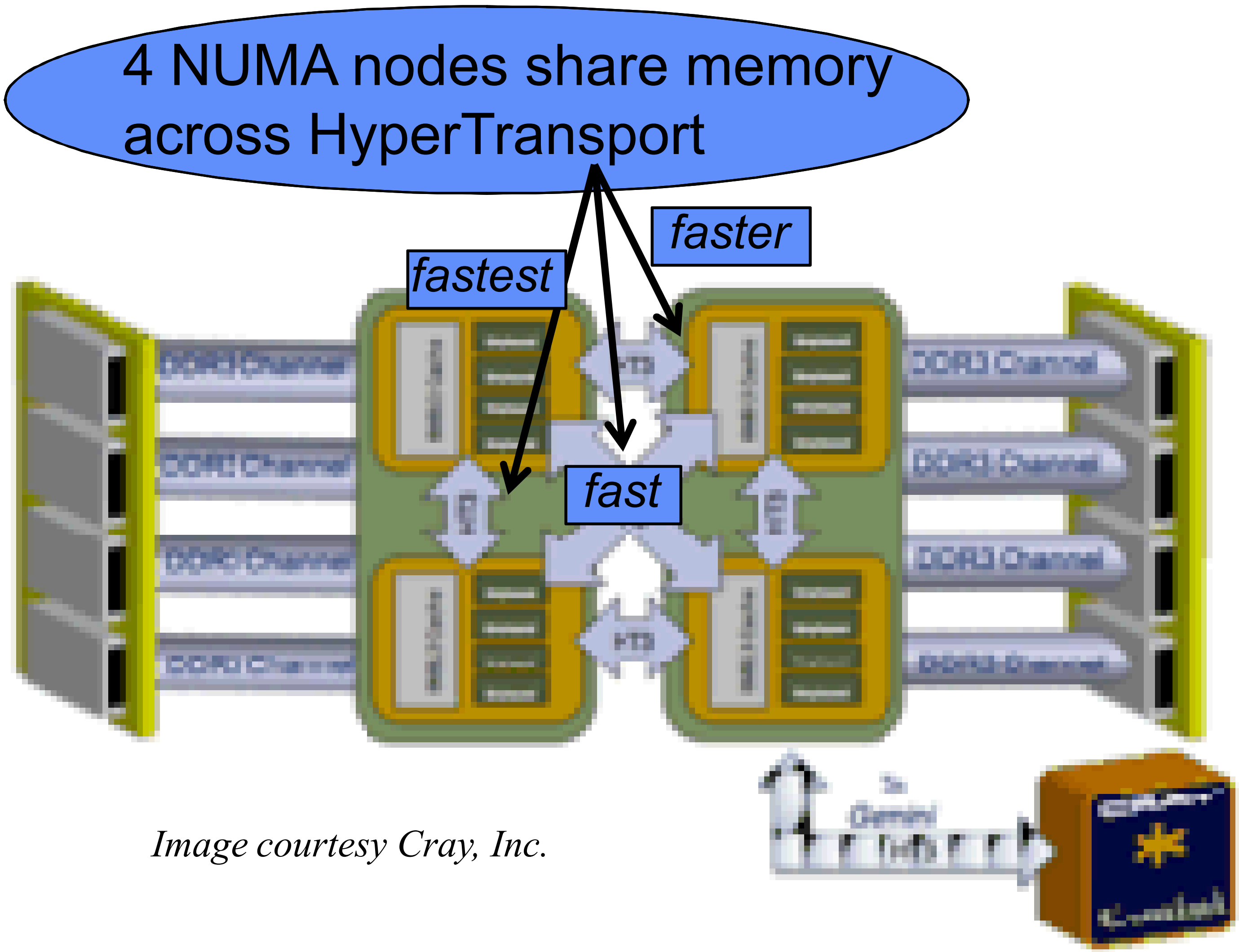
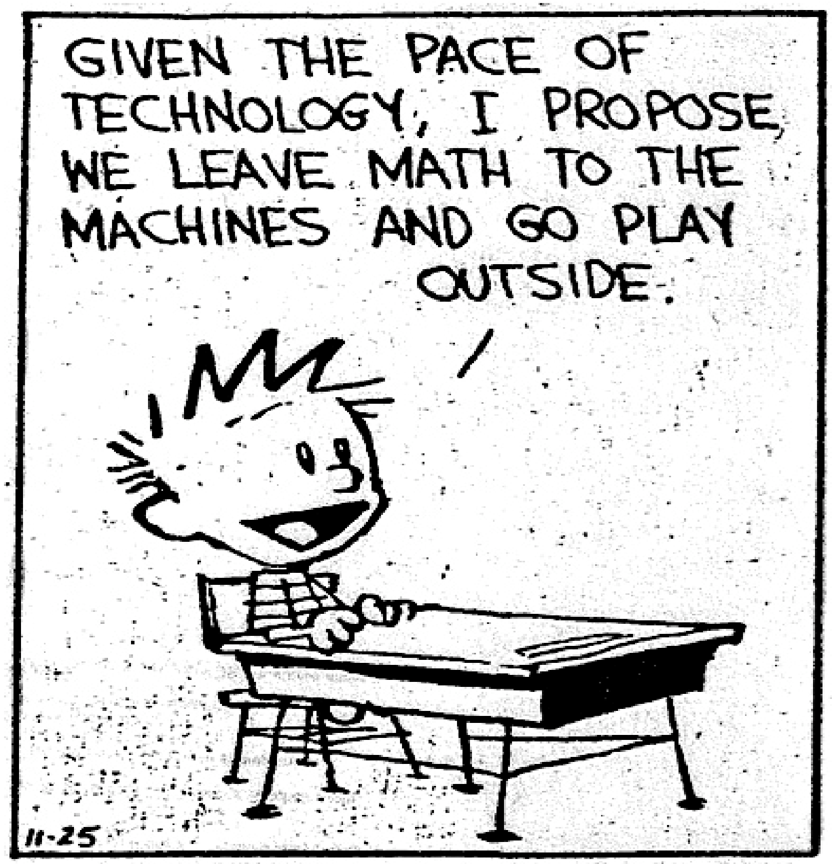


Image courtesy Cray, Inc.

Ultimate Goal of Computational Science



http://www.gocomics.com/calvinandhobbes

Sandia National Laboratories

Summary

- Mantevo mini-applications:
  - Completely open process: LGPL, validation.
  - Highly collaborative tool for co-design.
- Challenges:
  - Engaging already-busy apps developers.
  - Maintaining relevance over time.
- SC'11 meeting (BOF, 16 November, 5:30-7:00)
- SIAM PP'12 set of 3 mini-symposia (12 talks)
- IPDPS'12 Workshop (under review)

Supplemental Slides

Dominant Issue: Scatter/Gather

A (B (I)) = C (D (I))

Will the next programming model be an incremental change or a revolutionary change?

Yes.

It will (mostly) be what we should have been doing (and wanted to do) with SCOTS.

Like early days of message passing, will probably require evolutionary changes wrt programming mechanisms (eg CUDA, OpenCL, HMPP, PGI accel, XYZ, …, and MPI.)

Do we need to completely rethink our applications or will incremental approaches suffice?

Perhaps will inspire new algorithms/applications?

Sandia National Laboratories

Programming Model of the Future
(prediction, not a preference)

- **SPMD MPI between nodes**
- **On-node: multiple "views" of the data structure; eg SIMD, SIMT, MIMD.**
- **C/C++/Fortran**
  - **With "helper" syntax/semantics, mechanisms, & libraries**

So said I, 8 June 2011, and again July 27, 2011.

AMG2006*

Platform: Jaguar

Architecture: XT4

CPU: AMD Quad

P-states (Frequency States)

- **P0: 2.1 GHz , 1.25V**
- **P1: 2.1 GHz , 1.25V**
- **P2: 1.7 GHz, 1.1625V**
- **P3: 1.4 GHz, 1.125V**
- **P4: 1.1 GHz, 1.1V**

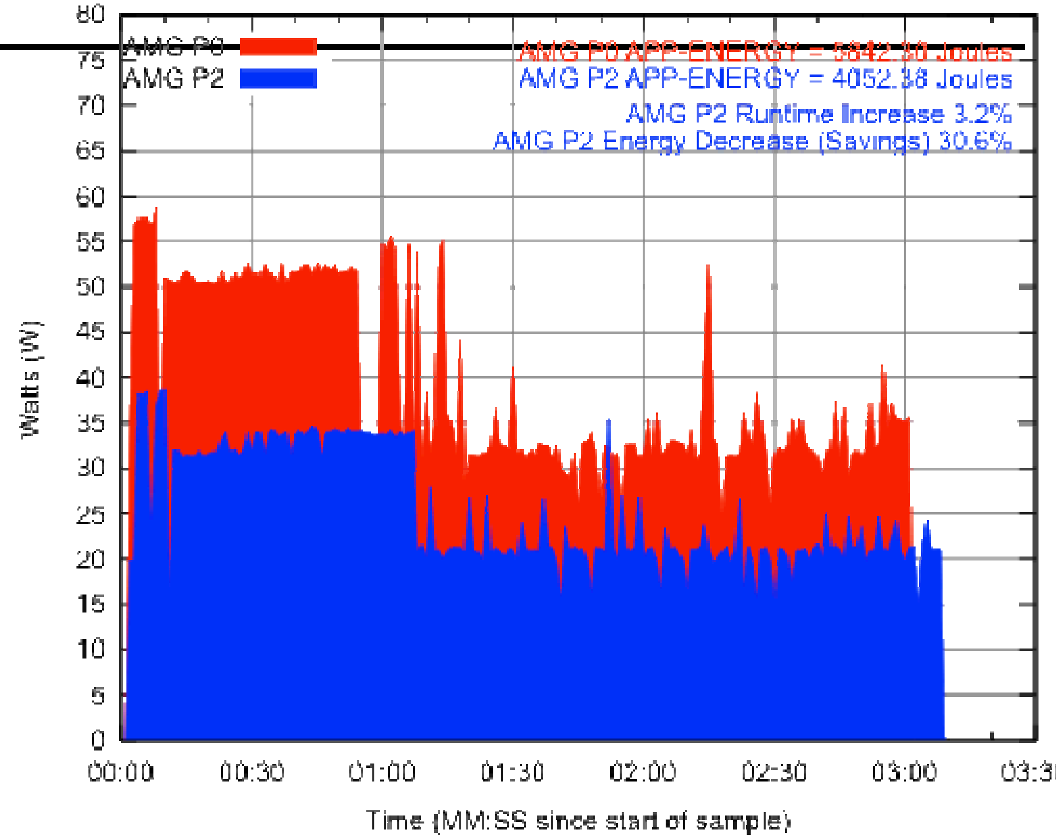
Nodes: 6144

Runtime Increase: 3.2%

Energy Decrease (Savings): 30.6%

<u>Order of magnitude energy savings vs. performance impact!</u>

Two application runs, same physical nodes, statically altering CPU frequency (P-state) allows lowering input voltage to chip resulting in larger energy savings.



Single node capture of watts over time for each run of AMG2006, varying P-states

Sandia National Laboratories

LAMMPS*

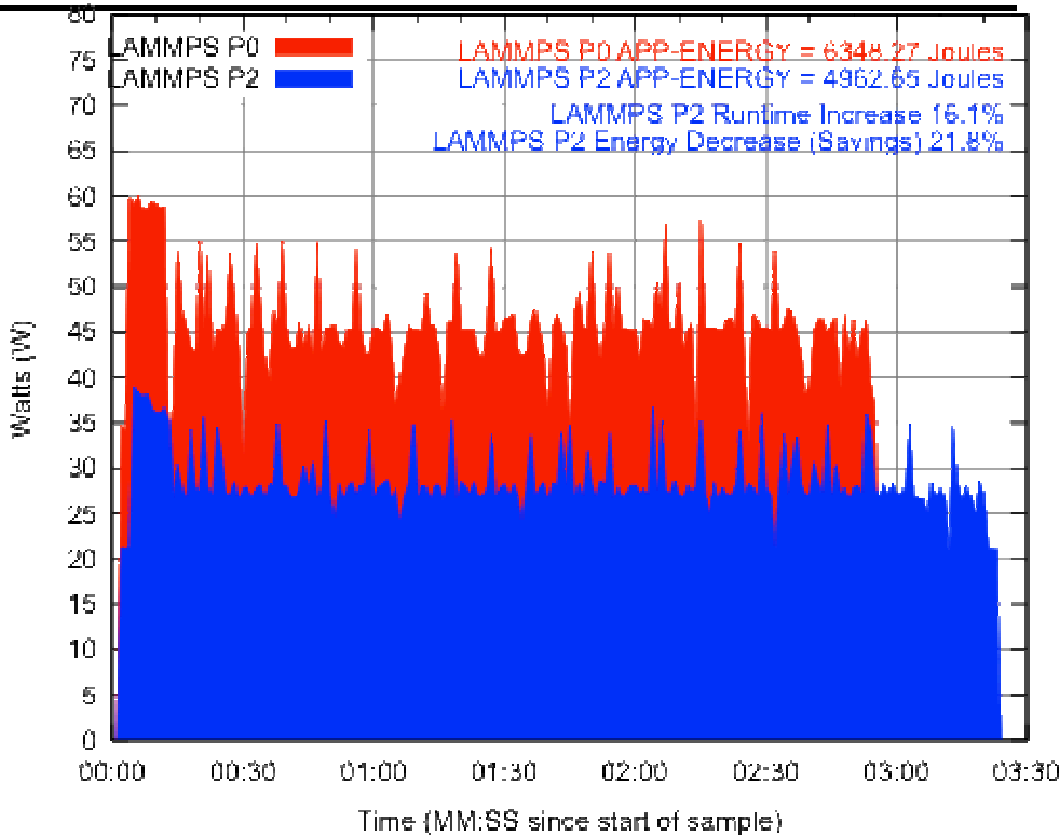
Platform: Jaguar
Architecture: XT4
CPU: AMD Quad
P-states (Frequency States)

- **P0: 2.1 GHz , 1.25V**
- **P1: 2.1 GHz , 1.25V**
- **P2: 1.7 GHz, 1.1625V**
- **P3: 1.4 GHz, 1.125V**
- **P4: 1.1 GHz, 1.1V**

Nodes: 4096
Runtime Increase: 16.1%
Energy Decrease (Savings): 21.8%

Compute intensive application, still observe significant energy savings. Illustrates which applications can expect most benefit.

Two application runs, same physical nodes, statically altering CPU frequency (P-state) allows lowering input voltage to chip resulting in larger energy savings.

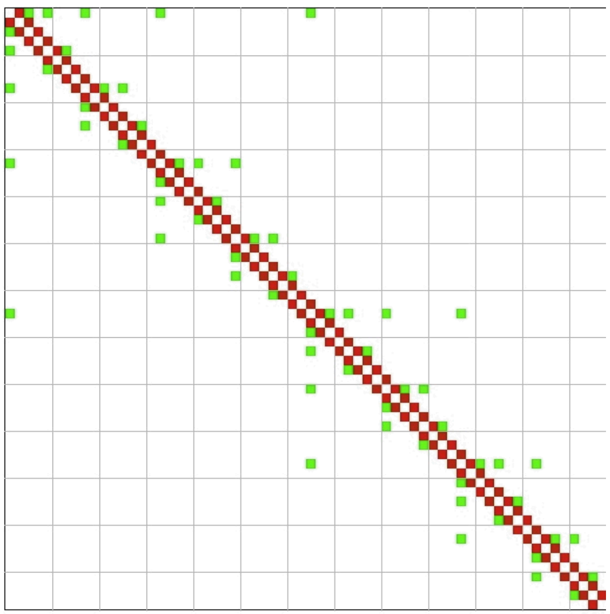


Single node capture of watts over time for each run of LAMMPS, varying P-states

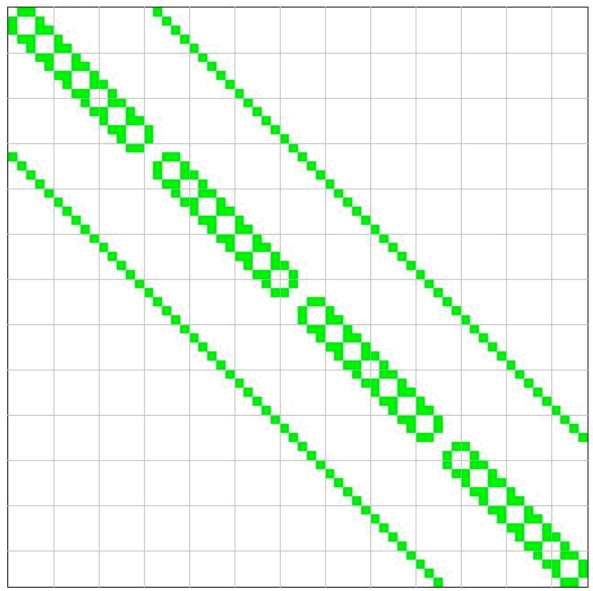
Sandia National Laboratories

Communication patterns

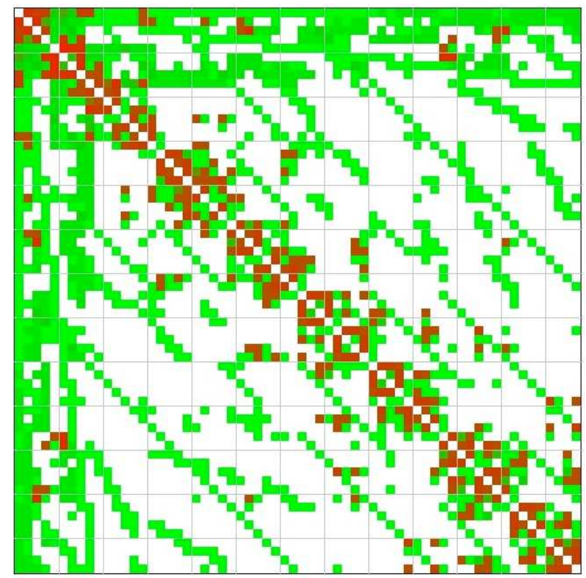
AMG



Eulerian



Newton-Krylov



Sandia National Laboratories