COMPUTER, COMPUTATIONAL & STATISTICAL SCIENCES

LA-UR 08-06343

ASC™

Driving a Hybrid in the Fast-lane: The Petascale Roadrunner System at Los Alamos

Darren J. Kerbyson

Performance and Architecture Laboratory (PAL)
http://www.c3.lanl.gov/pal
Computer, Computational & Statistical Sciences Division
Los Alamos National Laboratory
New Mexico, USA

NNSA

Los Alamos NATIONAL LABORATORY

PAL

Cores

Complexity

Constraints

C

Los Alamos
NATIONAL LABORATORY
EST. 1943

PAL

“The future will be like the present only more so”

Groucho Marx

Giga-flops 10^9

Tera-flops 10^{12}

Peta-flops 10^{15}

Exa-flops 10^{18}

Don’t mattera-flops 10^{21}

Matt Reilly, SiCortex

Los Alamos NATIONAL LABORATORY EST. 1943

Background

- **PAL is the performance analysis team at Los Alamos**
- **Large-scale:**
  - Large-scale Applications + Large-scale System = Large-scale performance
- **Applications of interest to Los Alamos, Dept. of Energy, NSF**
- **Systems:**
  - ASCI (Q, purple, red-storm), ORNL (Jaguar), IBM BG/L, BG/P
  - PERCS (-> Blue Waters), Zia, Sequoia …
- **Early processor analysis: Barcelona, PowerXCell-8i, Nehalem**

- **Analysis presented was undertaken by PAL @ Los Alamos**
  - Kevin Barker, Kei Davis, Adolfy Hoisie, Mike Lang, Scott Pakin, Jose Sancho
- **Many other Los Alamos Roadrunner people including**
  - Andy White, Ken Koch, John Turner
- **Further researchers acknowledged during this talk**

Why Roadrunner ?

- State bird of New Mexico

Los Alamos NATIONAL LABORATORY EST. 1943

Why Roadrunner ?

千兆 速 計算機

Thousand Trillion **Fast** **Computer**

Los Alamos NATIONAL LABORATORY EST. 1943

Roadrunner is a first …

- **1st general purpose HPC system to break the petaflop barrier:**

1.38 PF peak

1.026PF sustained on Linpack benchmark

(*Kistler, Gunnels, Benton, Brokenshire*)

First #1 Infiniband machine

First #1 Linux machine

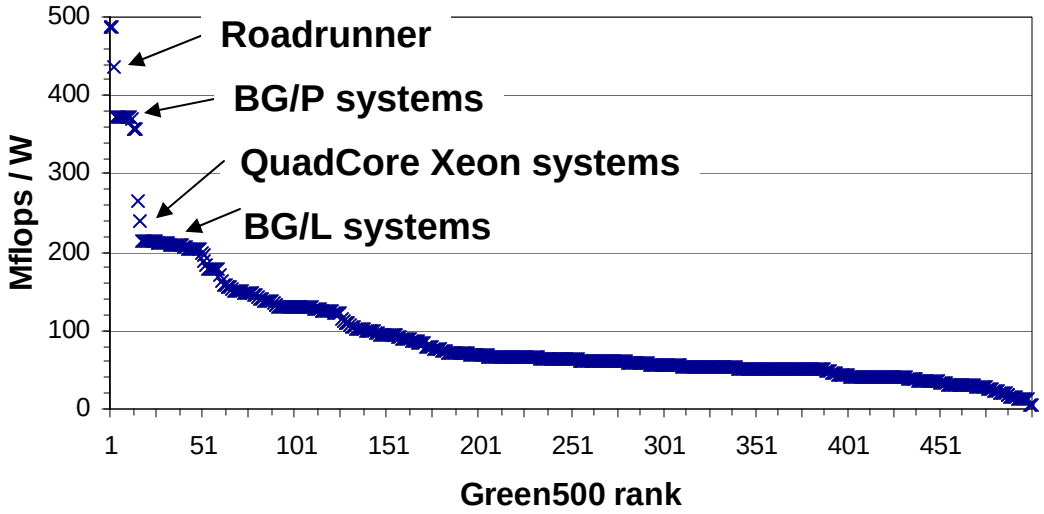
First #1 heterogeneous machine

First #1 commodity Cluster

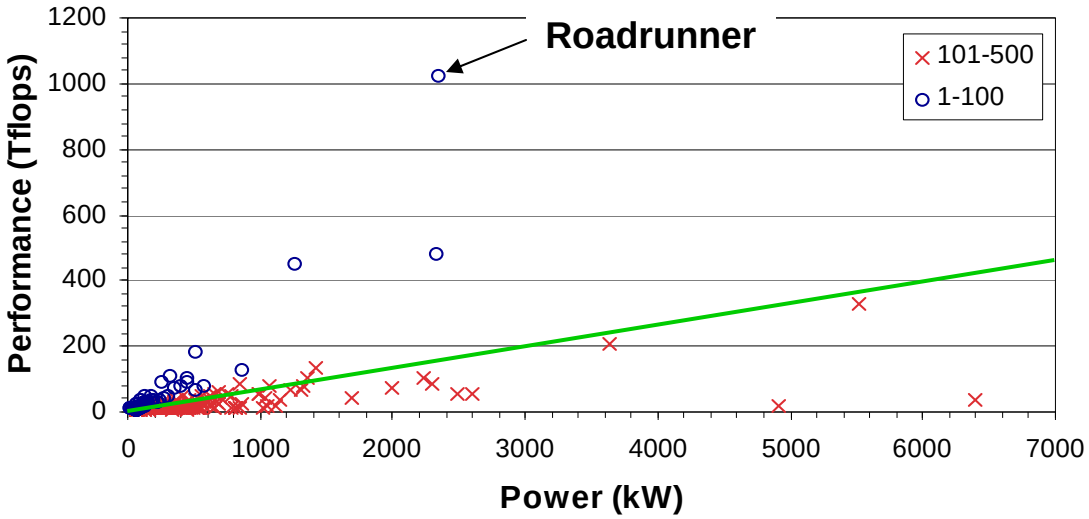
- **Already shown to be highly productive for applications**
  - Accelerating science

Roadrunner has a high power efficiency

- **Data taken from Green500, June '08**



- **3rd on Green500**
  - 437 MFlops/W
  - 1st & 2nd ranked are small Cell only clusters



- **~2350 MW measured on Linpack**
  - Not the most power hungry system !
- **The *Green-line* =**
  - 62.8 Mflops / W

Los Alamos NATIONAL LABORATORY EST. 1943

Outline of talk

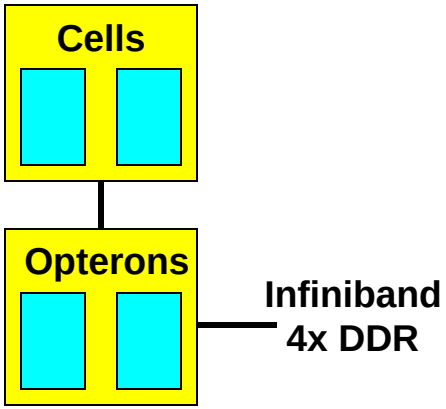
- **A brief history of Roadrunner**
  - pre-Roadrunner analysis of systems with accelerators
- **Architecture**
  - Chips, nodes, connectivity
- **Low-level Performance Characteristics**
- **Application Performance**

Note:

- Data is preliminary (machine not yet in production)

Roadrunner and Accelerators

- **Roadrunner @ Los Alamos**
  - Hybrid system containing
    - » **traditional cluster (Opterons)**
    - » **accelerators - IBM Cell**
  - It is not a “Cluster of Cells”



- **All *standard* services provided by the Opterons**
- **High compute potential on PowerXCell 8i processors**
- **Application challenge:**
  - Efficient utilization of the Hybrid system
- **Performance Modeling was important for Roadrunner**
  - examine potential performance of accelerated systems (Pre-RR)
  - compare performance across vendor systems (Pre-IBM)
  - predict performance of final system & compare with other systems
  - provide a performance baseline during system acceptance

pre-RR

'02 to '05 Dark Horse (Cell based), Architecture studies

2nd half '05 Modeling of systems with Accelerators

Analysis general but case-study was Clearspeed CSX600

Apr '06 Roadrunner RFP available

mid '06 IBM chosen as vendor

Phase 1

Jan '07 Base-system Operational

76TF Opteron cluster (replaced ASCI Q, Cells to be added later)

Phase 2

During '07 Final configuration developed

Integrated Tri-blade (Opterons & Cell eDP), & Cell-BE evaluation

Jun '07 First Cell eDP available for testing

Oct '07 Assessment of final configuration

Phase 3

Mar '08 First triblades available

Apr '08 Construction of 1st Roadrunner CU

May '08 All 17 CUs constructed

1.026 PF on Linpack benchmark, available for early science runs

Production

Early '09 Roadrunner enters production

Los Alamos
NATIONAL LABORATORY
EST. 1943

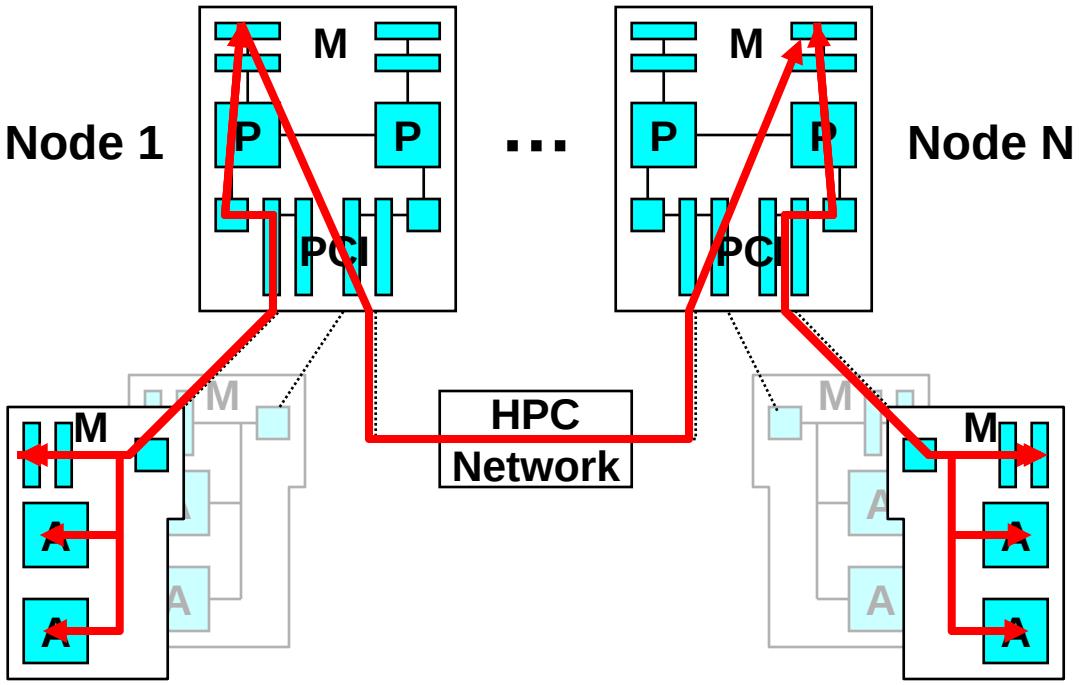
Life before Roadrunner @ Los Alamos

- **Large systems were Homogeneous**
  - Single-core, dual-core (more recently quad-core)
  - Mainly Opterons, some Xeons, main system until ’07 was Alpha
- **Many experimental systems**
  - GPUs (from viz.)
  - FPGAs (from embedded),
  - Clearspeed
- **Early acquisition of a small Cell system**
  - First hardware delivered outside of IBM
  - Primarily for analysis of viz. (ray tracing)

- **Mid 2005, question asked:**
  - Could accelerators be useful in large-scale systems ?
  - Example of Titech TSUBAME with Clearspeed

Two-level Heterogenous System analysis
Pre-Roadrunner (mid 2005)

- Compute nodes (e.g. with 2-sockets)
- HPC interconnection network (e.g. Infiniband)
- Accelerators placed in each node (e.g. PCI based)

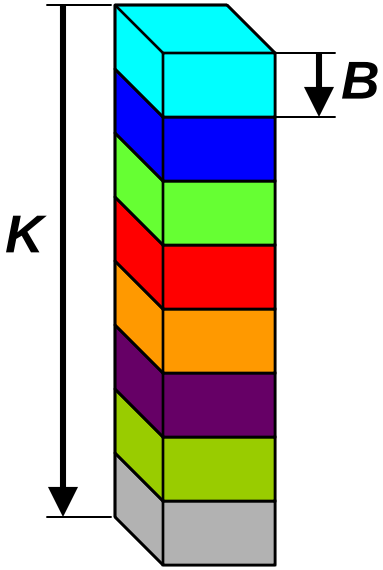


1) Start-up
   Node -> Accelerator
2) Process on accelerator
3) Inter-node communication
   Accelerator -> Node ->
   HPC Network ->
   Node -> Accelerator
4) Repeat 2 (& 3)
5) Finalize
   Accelerator -> Node

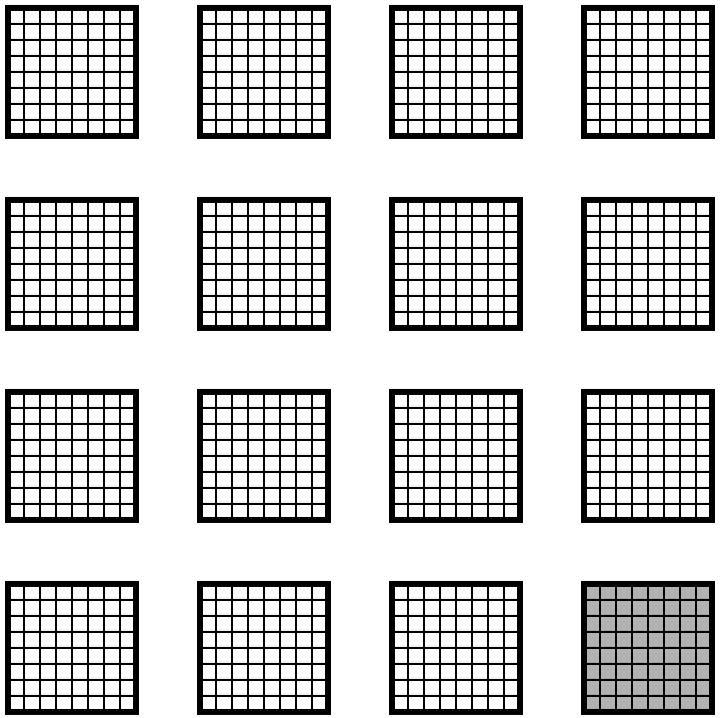
Wavefront Algorithms

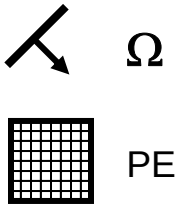
- **Dependency in processing order of grid-points**
  - Each grid-point requires boundary data from *upstream* neighbor
- **3-D grid is typically parallelized in 2-D**
  - Column (without blocking) gives poor efficiency
  - Blocking used to increase parallel efficiency (c.f. blocking for cache)

Sub-grid (1PE)



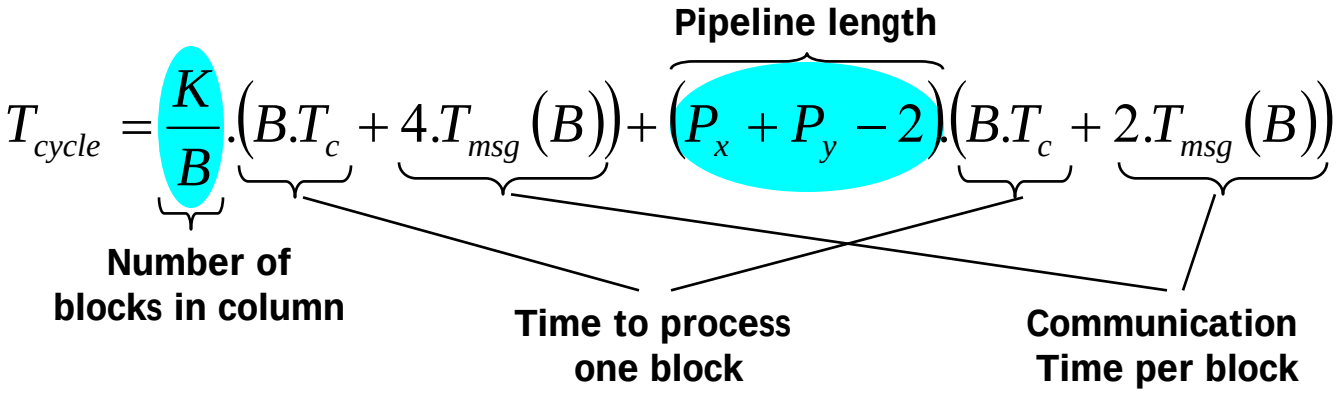
4x4 processors (top-view)





Characteristics of wavefronts

- **A pipeline in several dimensions, with 2-D parallelization:**
  - pipeline length = $P_x + P_y - 2$ (for one direction)
- **Blocking factor, *B = K-planes per block***
  - increases parallel efficiency
- **Basic performance model uses pipeline length and blocking:**



- **Pipeline effect minimized when B = 1**
- **Parallel overhead (message time) minimized when B = K, and**
  - In general, Block size decreases with scale

Wavefronts and accelerators

- **Each Block processed on the accelerator**
  - To process a block we have a pipeline on the accelerator

$$B.T_c = \frac{B}{B'}.\left(B'T_{AC} + 4.T_{msgAC}\left(B'\right)\right) + \left(P_x^{'} + P_y^{'} - 2\right).\left(B'.T_{AC} + 2.T_{msgAC}\left(B'\right)\right)$$

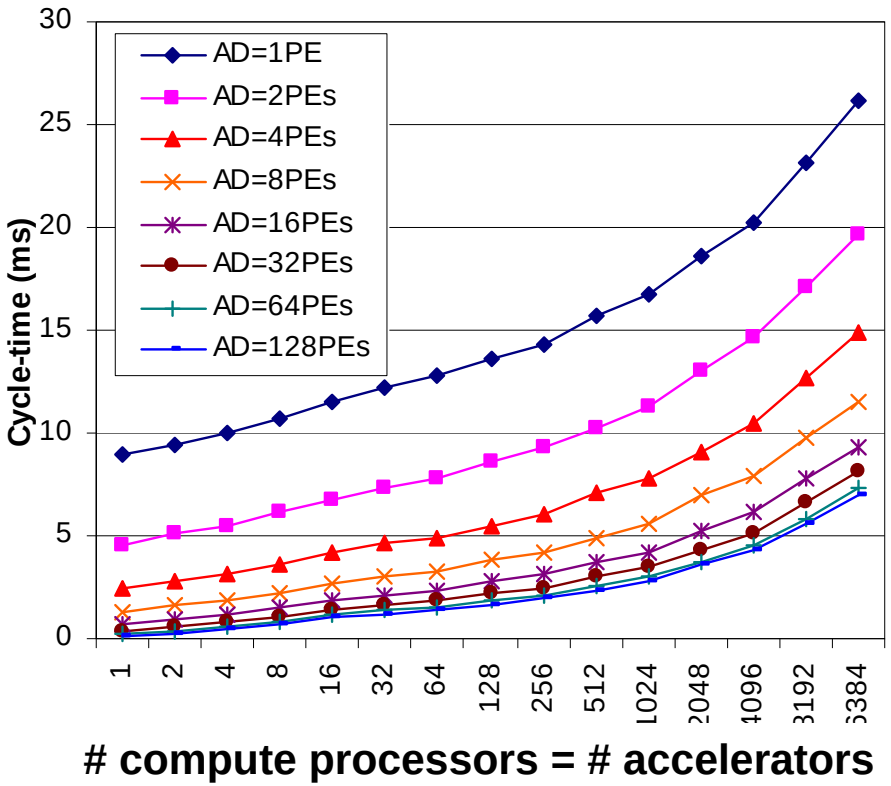
  - $(P`_x+P`_y-2)$ is the pipeline length of the accelerator
  - $B`$ is the blocking factor on the accelerator (usually 1)
  - T_{AC} is the compute time on the accelerator

- **Increases the pipeline length by factor $(P`_x+P`_y-2)$**

- **Effect of accelerator pipeline is minimized when *B* is large**

- **But at large-scale want *B* to be small**

Expected Performance (iteration-time)



- **Assumptions (hypothetical system):**
  - Weak-scaling
  - 16x8x1000 sub-grids
  - Processing time per cell = 70ns
  - Inter-PE (on Accelerator)
    - » Bandwidth =1GB/s,
    - » Latency = 50ns
  - Inter-node (MPI)
    - » Bandwidth = 1.6GB/s,
    - » Latency = 4µs

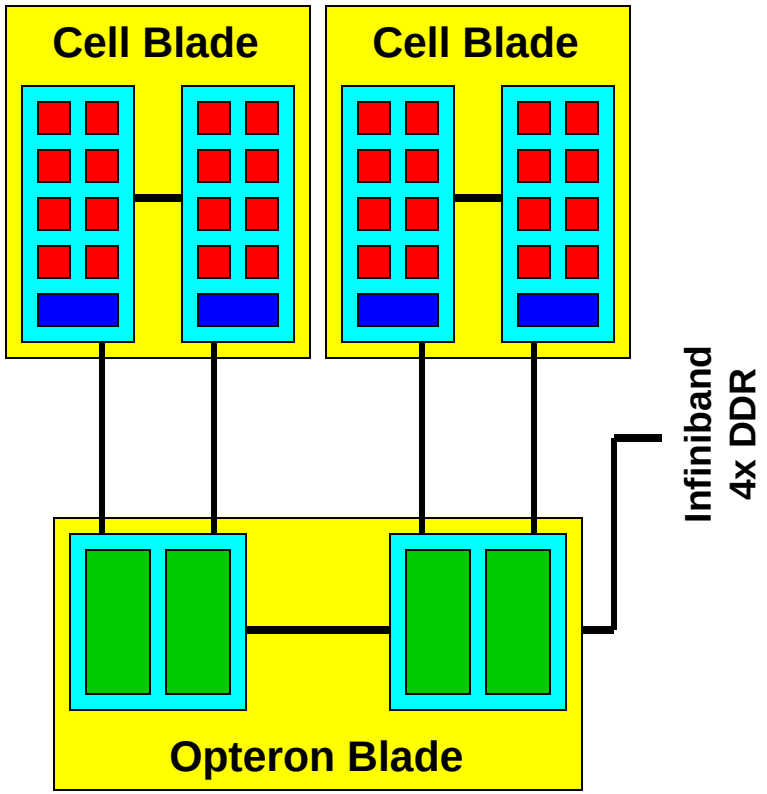
- **At largest scale, 16,384 compute processors & 16,384 accelerators**
  - Performance improvement is ~3.5x using Accelerators with 128x more PEs

"Analysis of Wavefront Alorithms on Large-scale Two-Level Heterogeneous Processing Systems", D.J. Kerbyson, A. Hoisie, presented at workshop on Unique Chips and Systems, Austin, Tx, Mar '06

Roadrunner @ Los Alamos

- **Heterogeneous**
  - AMD Opterons blades
  - IBM PowerXCell 8i (Cell)
  - One-to-one: Opteron-to-cell
- **Compute node**
  - 1x Opteron blade (2x dual-cores)
  - 2x Cell blades (each: 2x PowerXCell8i)
- **Each Cell blade connected to Opteron blade via PCIe x8**
- **Compute-node interconnected using Infiniband:**
  - 180 Nodes form a CU (full fat-tree)
  - Full system = 17CUs (=3,060 compute-nodes), reduced fat-tree

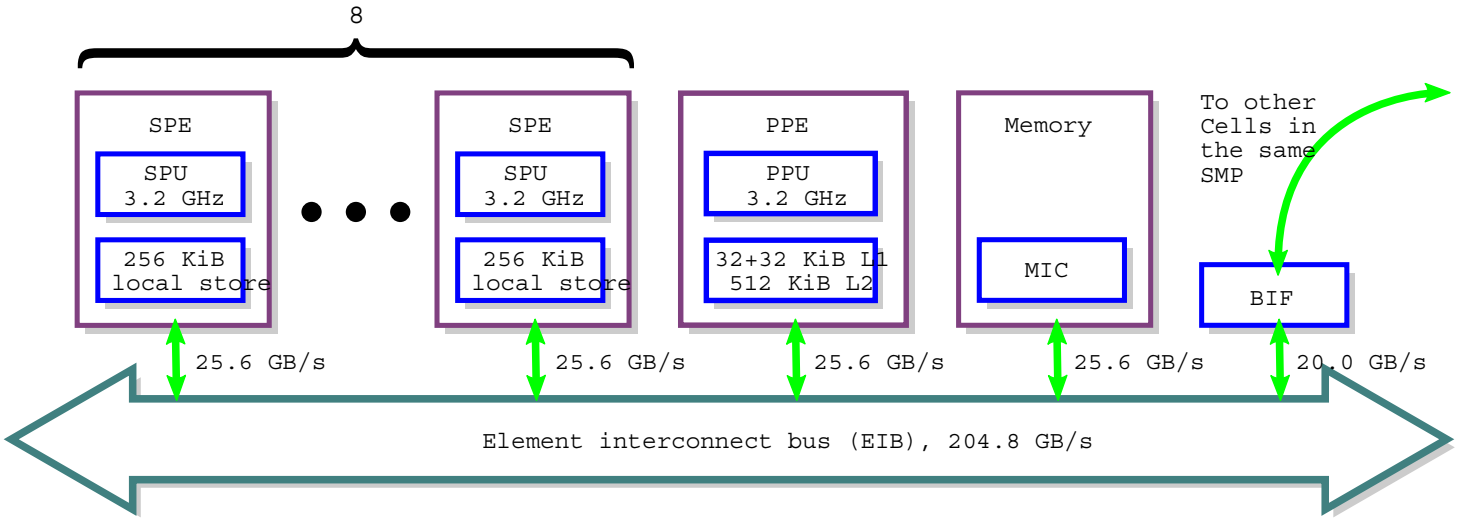
A Compute-node *Triblade*



Los Alamos NATIONAL LABORATORY EST. 1943

PowerXCell 8i Processor:
Implementation of Cell Broadband-Engine Architecture

- **8 Synergistic Processing Elements (SPE)**
  - 4 DP (or 8 SP) flops per cycle
  - 256KB local-store (software managed)
- **1 Power Processing Element (PPE)**
  - General purpose, runs O/S
- **EIB interconnects SPEs, PPE & memory interface**



Not the Cell-BE that is used in the Sony Playstation 3

Los Alamos NATIONAL LABORATORY EST. 1943

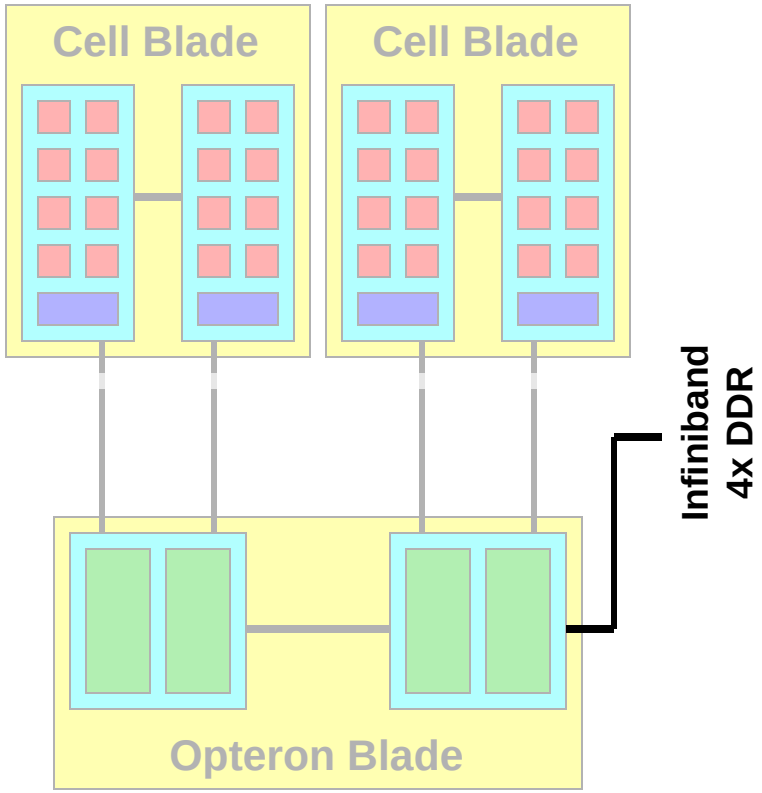
A comparison of Cells

	Cell-BE	PowerXCell 8i
Memory	**Rambus** **2GB limit** **25.6GB/s**	**DDR2** **16GB limit** **25.6GB/s**
DP Floating-point (SPEs)	**14.3GF/s (DP)** **6 cycle stall** **13 cycle latency** **No dual issue with DP**	**102.4 GF/s (DP)** **Fully pipelined** **9 cycle latency** **Dual issue with DP**

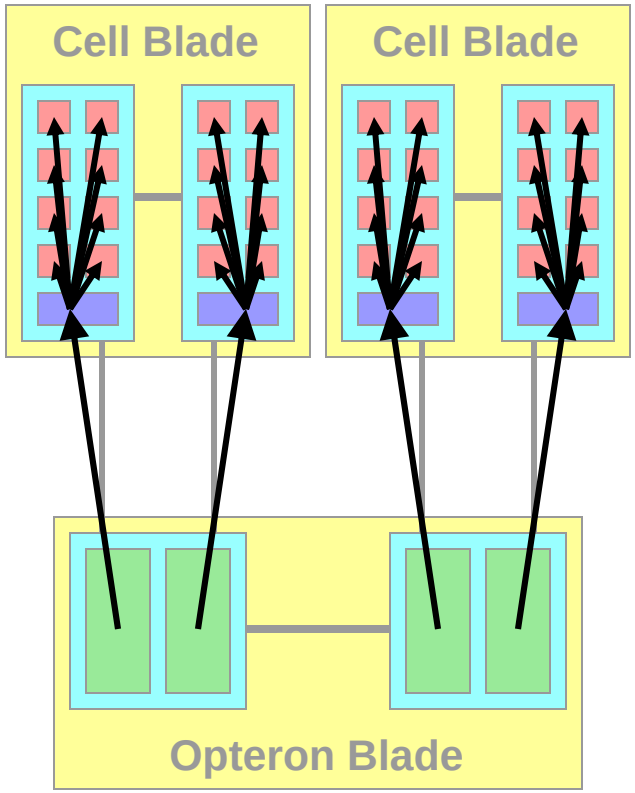
- **Need to efficiently use SPEs to obtain high performance**
  - Vectorization (2-way SIMD)
  - Two pipes (odd & even), 2 instructions/cycle
  - Heterogeneous functional units, in-order execution
  - Software managed Local-Store (256 KB)
- **High Speed DMA engine**
  - To other SPE's local stores & main memory
- **PPE under-powered**
  - typically 5x slower than an Opteron

Roadrunner: Usage

- **Non-hybrid (Opteron only)**
  - Codes run without modification
- **Hybrid (Opteron and Cell)**
  - Code **performance hotspots** ported to the Cell
  - Also incremental porting
- **Cell-centric (Cell only)**
  - Need support for communications between Cells
    - » **Between PPEs (e.g. MP Relay)**
    - » **Between SPEs (e.g. CML)**

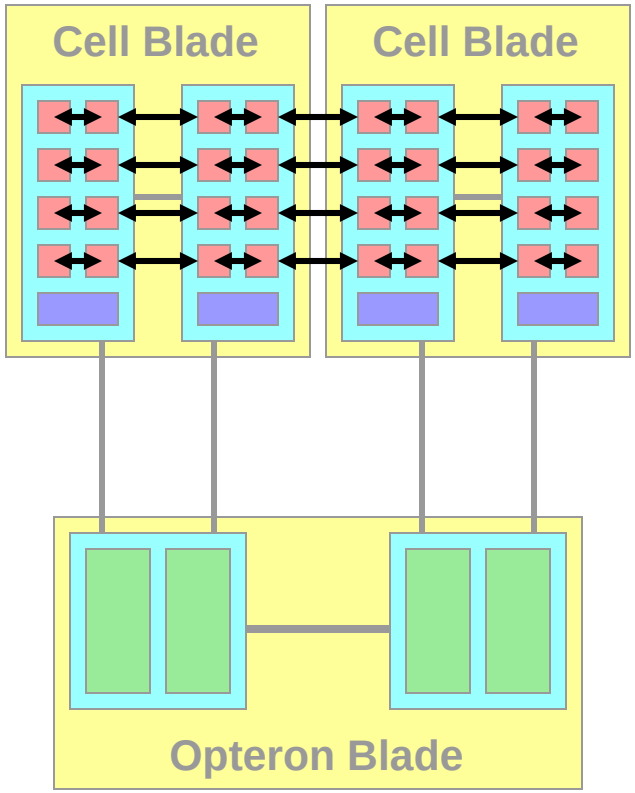


Hybrid (general accelerator approach)



- One MPI rank per Opteron
- SPE = accelerator
- Opterons see each other and their local SPEs
- Opteron pushes work (data) to SPEs and receives results
- Currently: DaCS
  - Data Communication and Synchronization for Opteron <-> Cell
  - Also looking at MPI
- libSPE (or ALF) for SPE work management

SPE-centric (Cell-Messaging-Layer)

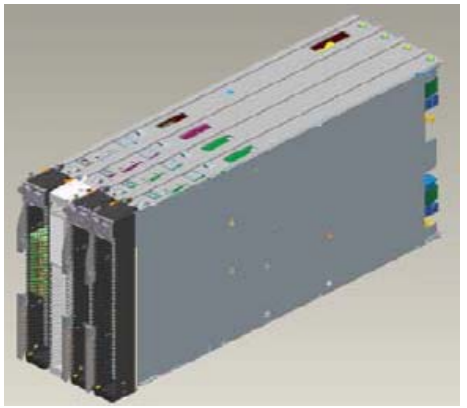


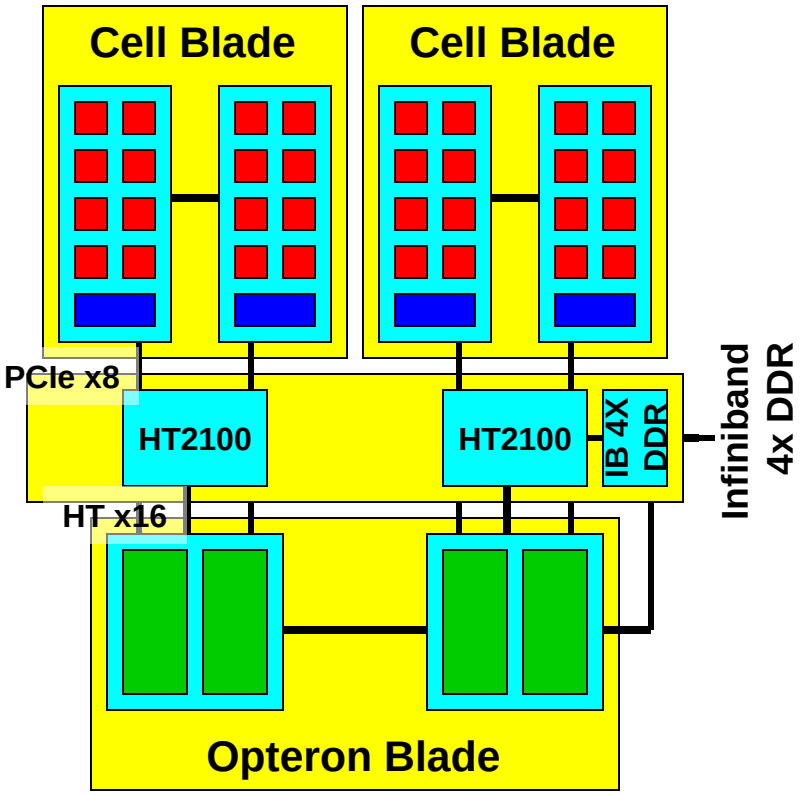
- **One MPI rank per SPE**
- **Opteron = NIC (& extra storage)**
- **SPEs see each other and their local Opteron**
  - SPEs communicate directly with other SPEs
  - PPE provides support
- **MPI subset, currently:**
  - blocking MPI pt2pt & collectives
  - No tags, no communicators
  - Small memory footprint
- **“Cluster of 100,000 SPEs”**

***Receiver-initiated Message Passing over RDMA Networks*.**
S. Pakin, IPDPS 2008, Miami, FL, April 2008

Los Alamos NATIONAL LABORATORY EST. 1943

Roadrunner: Compute Node (peak-performance)



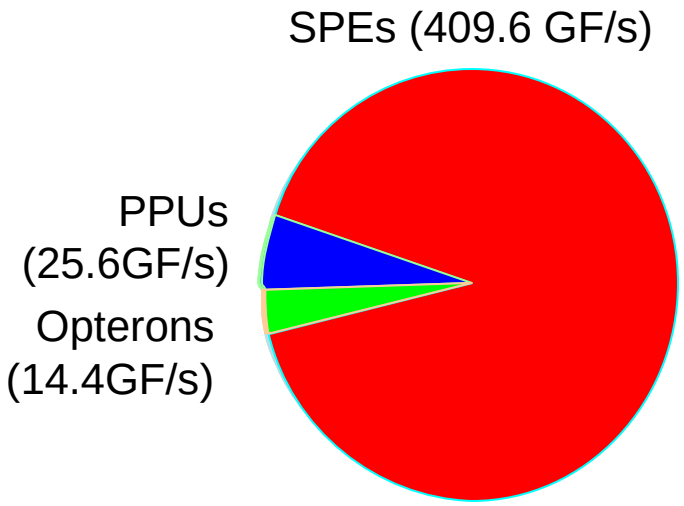


- **4x PowerXCell 8i (3.2GHz)**
  = 4x (PPU + 8 SPUs)
  - SPEs (per cell) = 102.4 Gflop/s (DP)
  - PPE (per cell) = 6.4 Gflops/s (DP)
- **4x AMD cores (1.8GHz)**
  - AMD = 3.6 Gflop/s (DP) / core
- **Cell <-> AMD**
  - Bandwidth = 2.0GB/s + 2.0GB/s
  - Latency ~1.5µs
- **AMD <-> AMD (inter-node)**
  - Bandwidth = 2.0GB/s + 2.0GB/s
  - Latency ~ 1.5µs

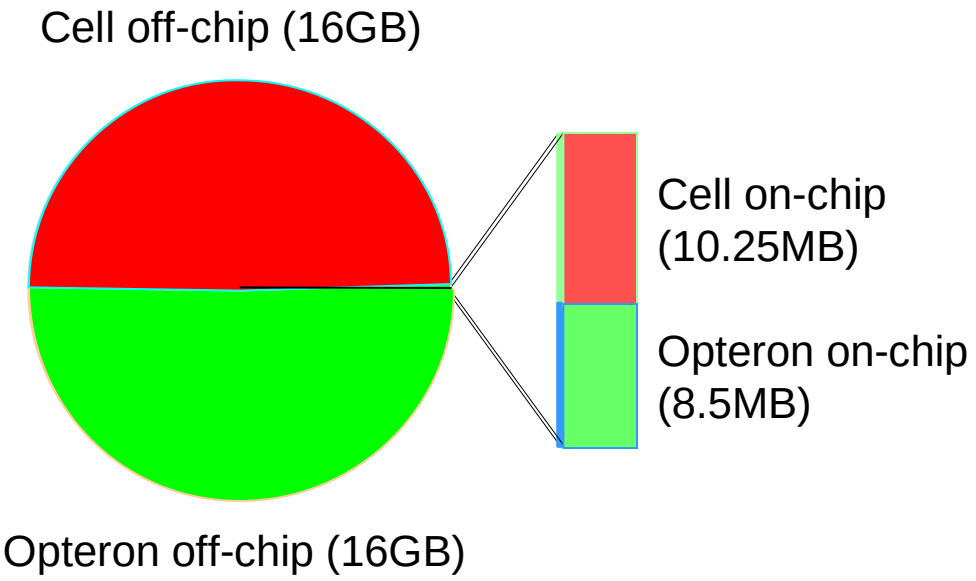
Relative capacities: Opterons & Cells

- **90% of the peak flops in the SPEs on the Cell**
- **Equal main memory between the Cells and Opterons**

Peak flops (DP) / node



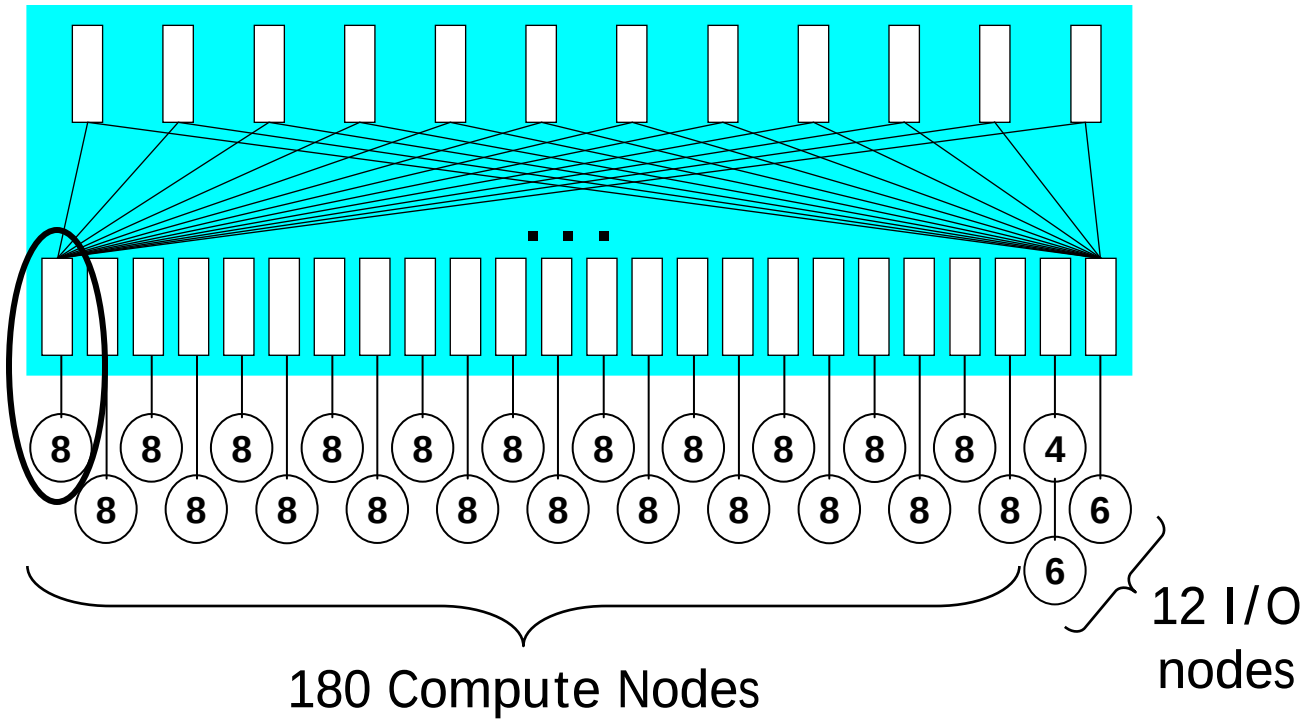
Memory / node

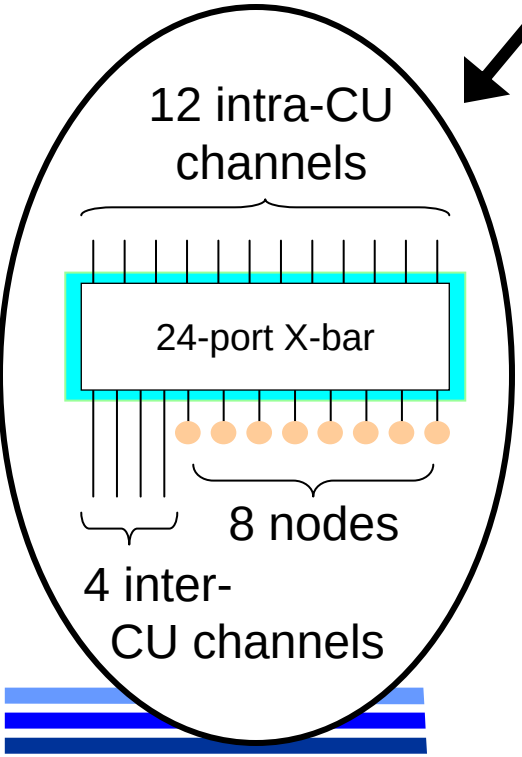


Roadrunner: Connected Unit (CU)

288 Port 4x DDR Infiniband Switch

36 individual 24-port xbar chips

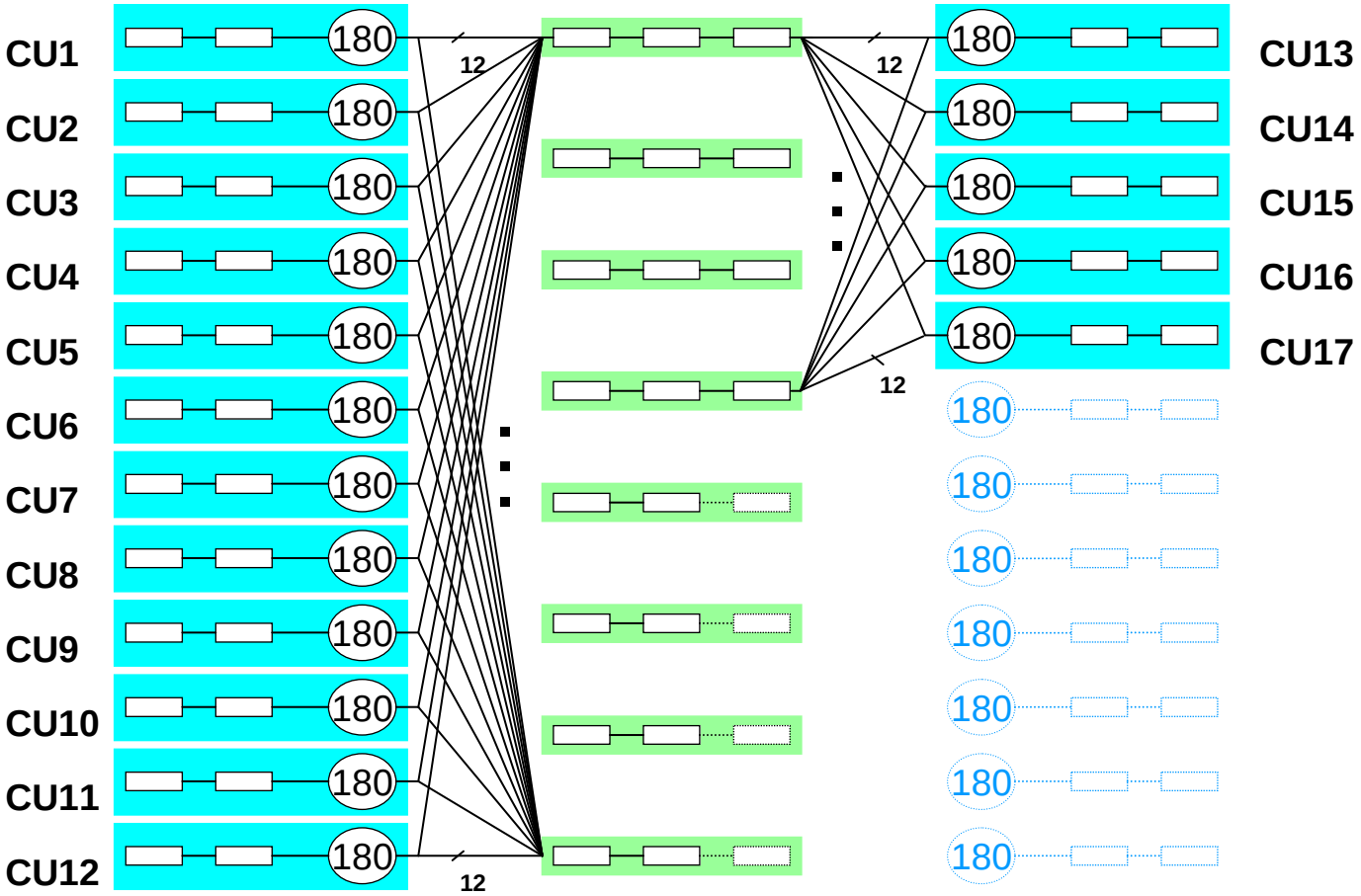




- **Full fat-tree within a CU**
  - 8 DOWN links to 12 UP links per Xbar
- **Reduced fat-tree between CUs**
  - Only 4 inter-CU links per 8 nodes (2:1 reduction)

Los Alamos NATIONAL LABORATORY EST. 1943

Interconnection of 17 Roadrunner CUs:

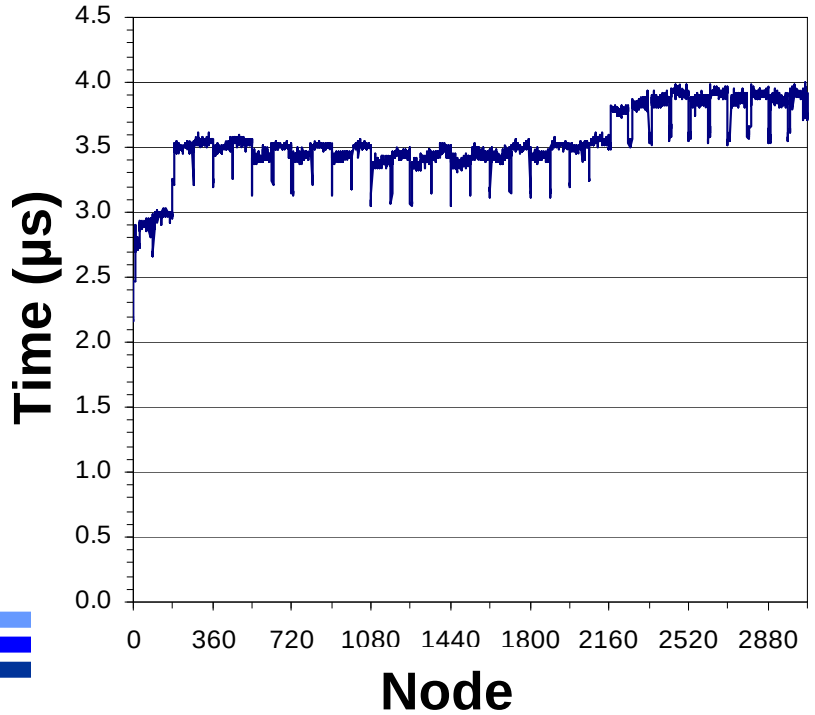


- **Eight 288-port switches interconnect 17CUs (some room for expansion)**
  - Each CU has 12 links to each inter-CU switch

Los Alamos NATIONAL LABORATORY EST. 1943

Non-standard interconnection lowers average hop-count (latency)

Destination node	No. of destinations	Hop count
Self	1	0
Within same crossbar	7	1
Within same CU	172	3
In CUs 2-12, same crossbar	88	3
In CUs 2-12, different crossbar	1892	5
In CUs 13-17, same crossbar	40	5
In CUs 13-17, different crossbar	860	7
Total	3060	5.38 (average)



- **Measured MPI latency (from node 0)**
  - Min = 2.5µs, Max = 4.0µs
- **Exposes network hierarchy**
  - Same CU,
  - 11 CUs (same-side inter-CU switches)
  - 5CUs on other-side
  - Lower latency every 90 nodes
    - » **Destination node at entry point to CU from inter-CU switch**

Summary of Roadrunner Characteristics

System	
CU count	17
Node count	3,060
Peak Performance (DP)	1.38 Pflops/s
(SP)	2.91 Pflops/s

Connected Unit (CU)	
Node count	180
Peak performance / CU (DP)	80.9 Tflops
(SP)	171.1 Tflops

Compute Node (triblade)	*1x Opteron blade*	*2x Cell blades*
Processor count	2	4
Processor-core count	4	4 PPEs, 32 SPEs
Clock Speed	1.8 GHz	3.2 GHz
Peak-performance/node (DP)	14.4 Gflops/s	435.2 Gflops/s
(SP)	28.8 Gflops.s	921.6 Gflops/s
Memory per processor	4 GB (667MHz DDR2)	4 GB (800MHz DDR2)

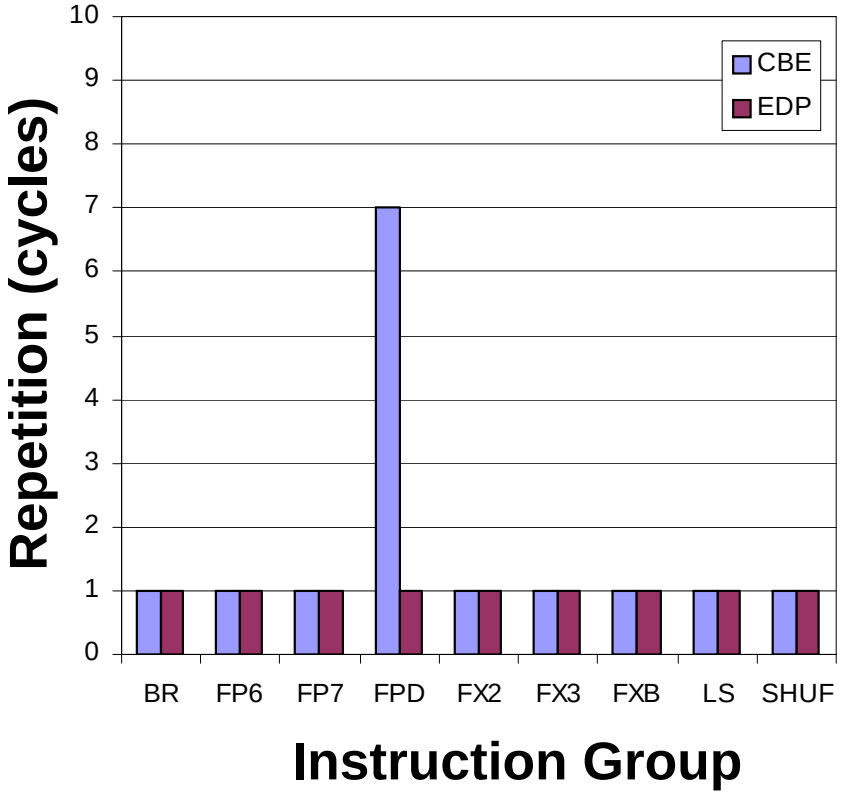
278 racks, 5200 ft², 500,000 lbs., 55 miles of IB cables

Los Alamos NATIONAL LABORATORY EST. 1943

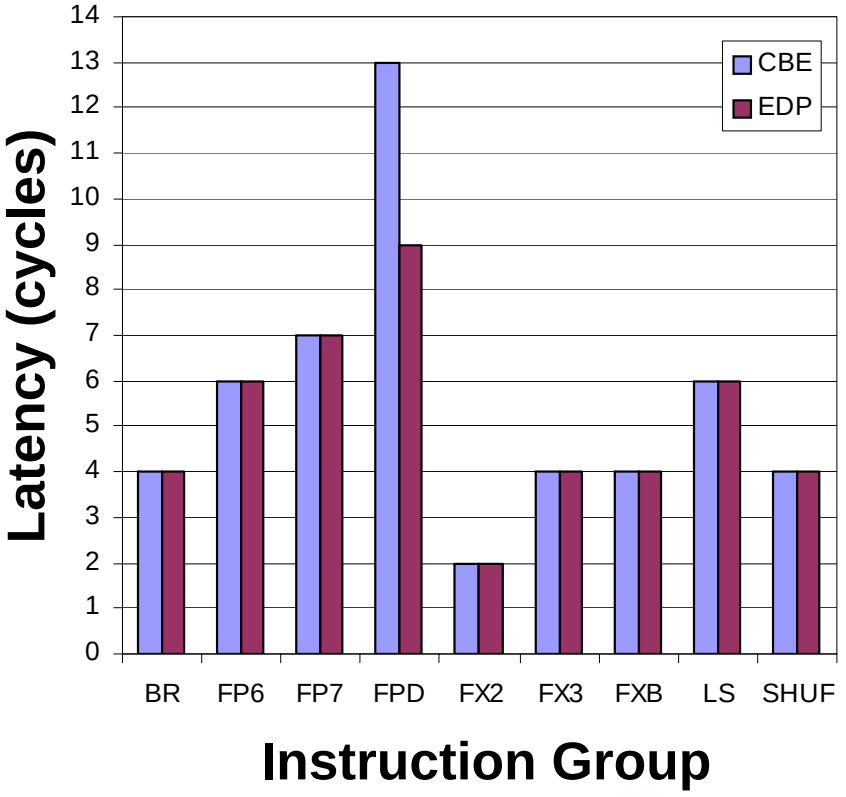
PowerXCell 8i vs Cell-BE: low-level performance

- Micro-benchmarks used to measure instruction characteristics

Cycles between instruction issues



Instruction pipeline latency



Los Alamos NATIONAL LABORATORY EST. 1943

PowerXCell 8i vs Cell-BE application performance

- **PowerXCell 8i increased peak DP flops by 7x**
- **First testing in July ’07:**
  - summary of testing with two DDR2 memory speeds (667MHz and 800MHz)

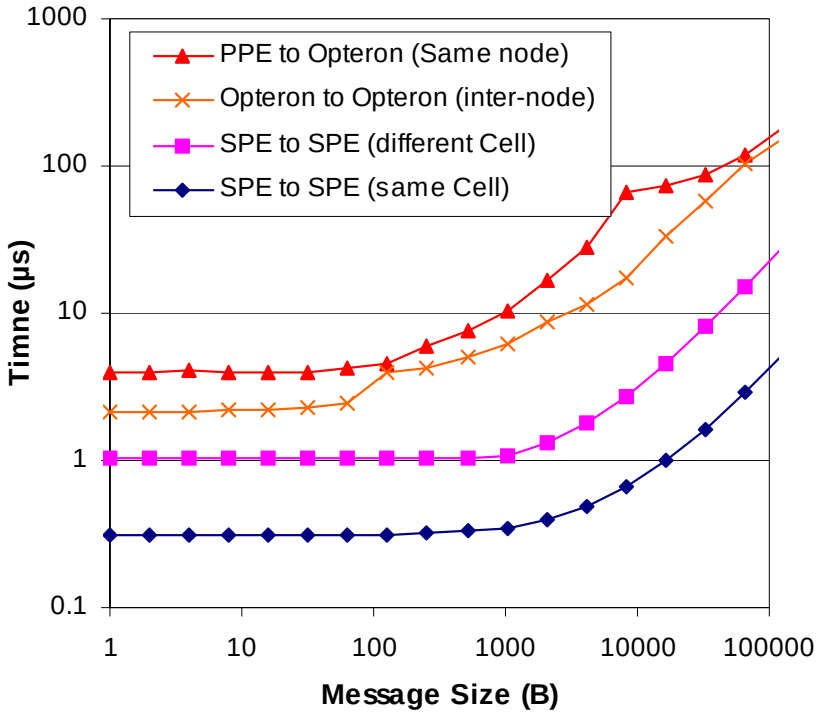
	PowerXCell 8i (667MHz DDR2) vs. CBE	PowerXCell 8i (800MHz DDR2) vs. CBE	
VPIC	1.01x	1.01x	Single-Precision
CellMD	1.50x	1.50x	Core of SPaSM
Hybrid-IMC	1.50x	1.50x	Monte-carlo particle
Sweep3D	1.72x	1.77x	Wavefront

Los Alamos NATIONAL LABORATORY EST. 1943

Communication Performance

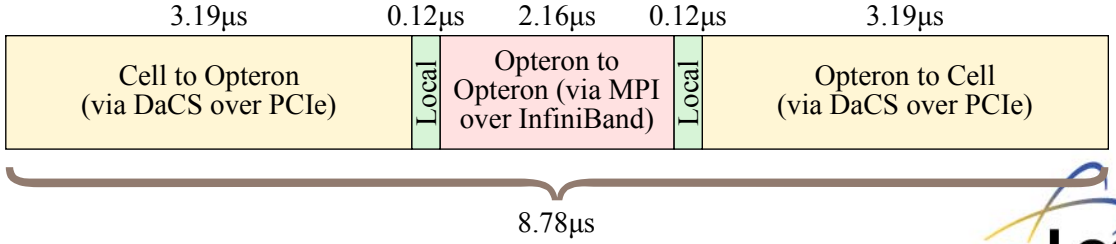
- **Hierarchy of channels, e.g.**
  - Intra-chip: SPE to SPE
  - Inter-chip: SPE to SPE
  - Intra-node: PPE to Opteron
  - Inter-node: Opteron to Opteron

	Latency 0-B, µs	Bandwidth 128-KB, GB/s
Intra-chip SPE->SPE	**0.3**	**23.9**
Inter-chip SPE->SPE	**0.8**	**4.5**
Intra-node PPE->Opteron	**3.2**	**0.7**
Inter-node Opteron->Opteron	**2.1**	**0.8**



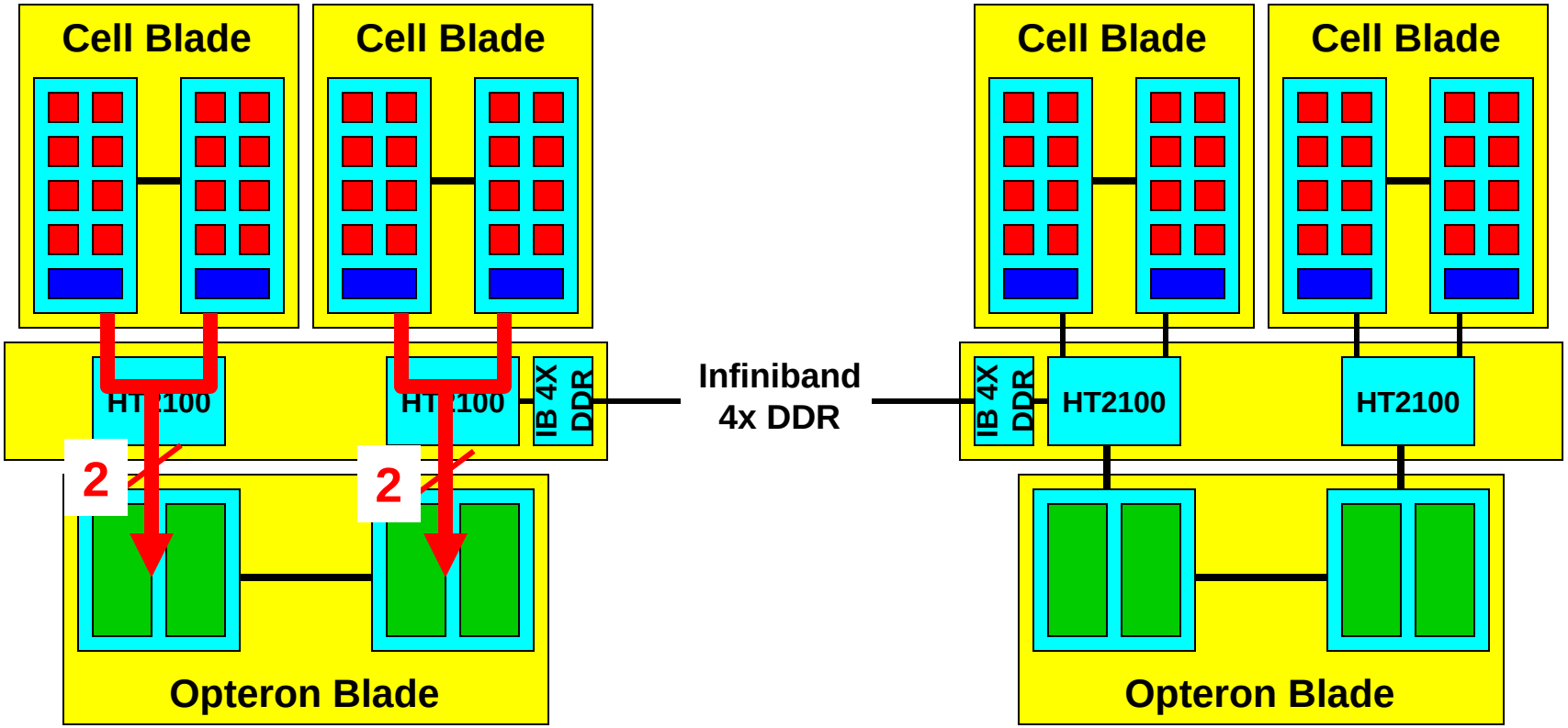
- **Inter node communication between Cells requires multiple steps**

e.g. 0-byte latency



Los Alamos NATIONAL LABORATORY EST. 1943

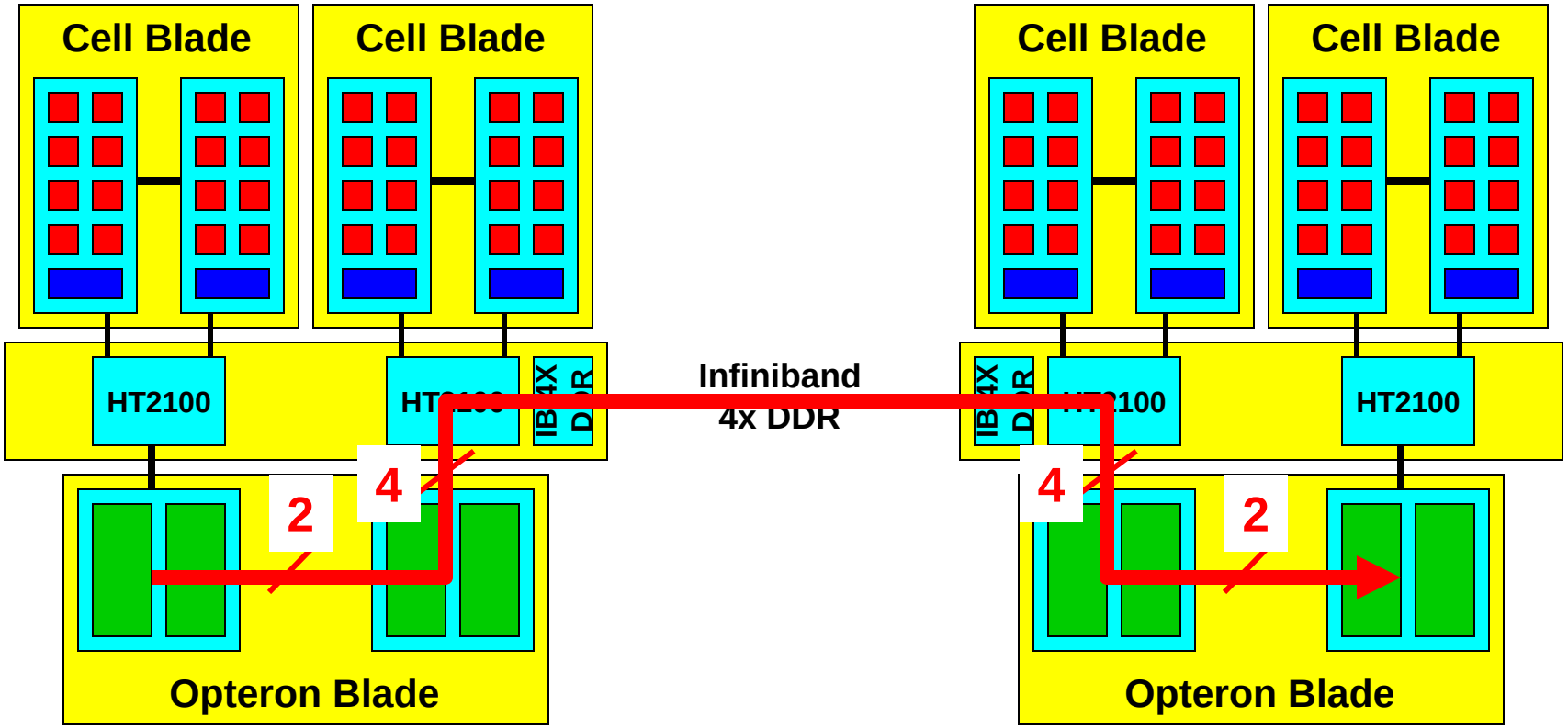
Example: Communication steps



1) Cells (Node 1) -> Opterons (Node 1)
2) Opterons (Node 1) -> Opterons (Node 2)
3) Opterons (Node 2) -> Cells (Node 2)

Los Alamos NATIONAL LABORATORY EST. 1943

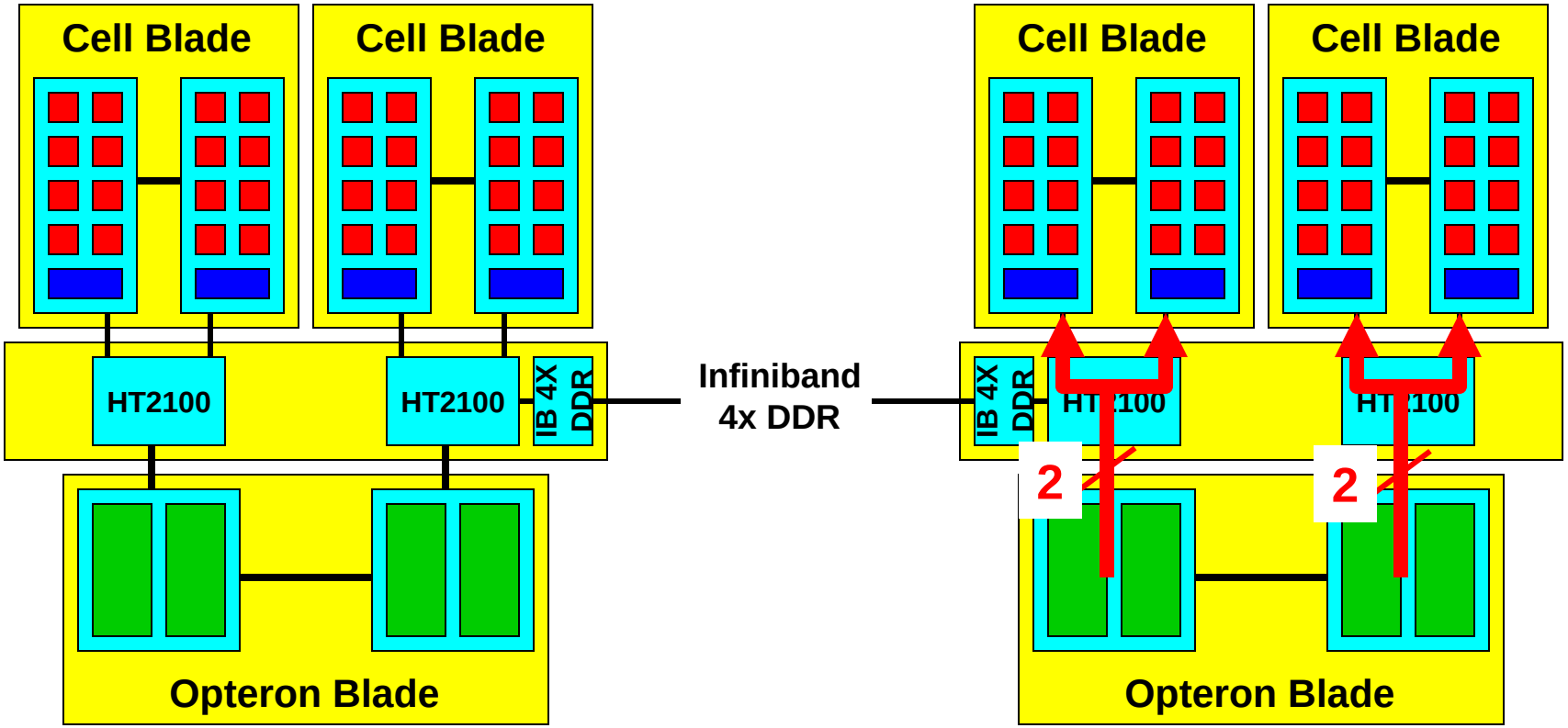
Example: Communication steps



1) Cells (Node 1) -> Opterons (Node 1)
2) Opterons (Node 1) -> Opterons (Node 2)
3) Opterons (Node 2) -> Cells (Node 2)

Los Alamos NATIONAL LABORATORY EST. 1943

Example: Communication steps



1) Cells (Node 1) -> Opterons (Node 1)
2) Opterons (Node 1) -> Opterons (Node 2)
3) Opterons (Node 2) -> Cells (Node 2)

Los Alamos NATIONAL LABORATORY EST. 1943

Application performance on Roadrunner

- **Performance measured for some applications**
  - Early system access for scientific runs
- **Performance modeled for many application/system configurations**
  - Does performance line up with expectations?
    - » **Is the system going to meet the performance expectations laid out in the assessment criteria?**
  - What performance boost is achievable when using the accelerators?
    - » **hybrid (Cell + Opteron) vs. non-hybrid (Opteron only)**
  - Historical perspective: comparison of Roadrunner to previous machines?
    - » **Roadrunner vs. ASC Q**
  - How does Roadrunner compare to other current-technology machines?
    - » **Roadrunner vs. Quad-core systems**

Diversity of Applications

- **Four applications initially analyzed and ported to Roadrunner:**

  1) VPIC – Particle-in-Cell (*K. Bowers, B. Albright, B. Bergen*)
  - Cell-centric, Opterons used only for Message relay

  2) Sweep3D – Deterministic transport (*M. Lang, G. Johnson, O. Lubeck*)
  - Cell-centric, Opterons used only for Message relay

  3) SPaSM – Molecular Dynamics (*T. German, K. Kadau, S. Srinivasan*)
  - Hybrid, Both Cell and Opterons do useful work

  4) Milagro - non-deterministic transport (*T. Kelley, P. Henning*)
  - 2 versions

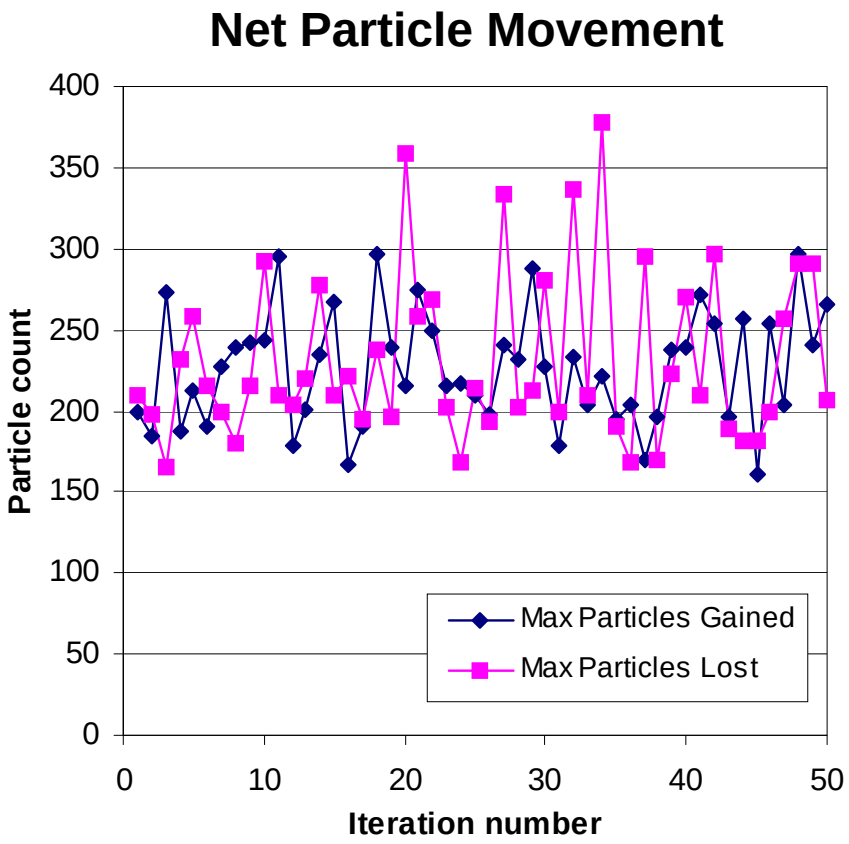
- **Others currently being explored**
  - Phylogenetic analysis of evolution of acute HIV infection
  - Molecular simulation of cellulose breakdown for biofuels
  - Cosmological simulation of large-scale Universe structure
  - Long-term evolution of formation & de-formation of metallic nanowires

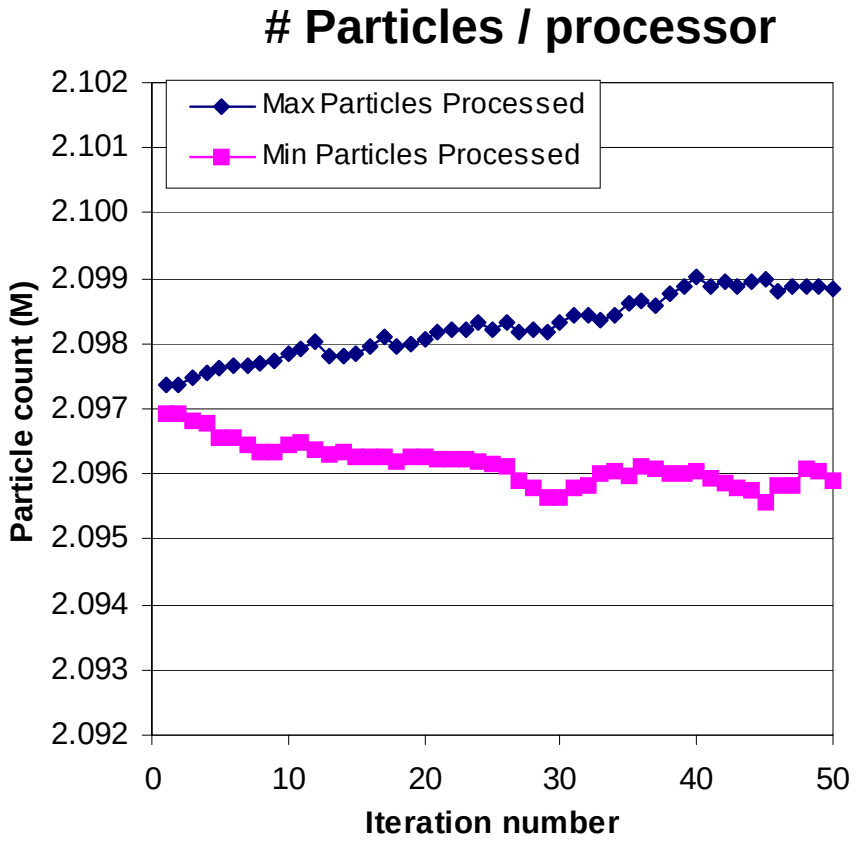
Example: VPIC

- **Plasma Particle-in-Cell**
  - 3-D volume containing multiple particle species (ions and electrons)
    - » **Split into Voxels, each contain ~equal # of ions & electrons**
  - ions and electrons can move between voxels
    - » **some communication per iteration (small-medium sized messages)**
  - Parallel Decomposition: in 1-D, 2-D or 3-D
  - Weak-scaling: constant work per processor
  - Periodic particle sorting to aid data layout in memory
- **Roadrunner implementation is Cell-centric**
  - PPE farms out work to SPEs
  - Opterons used for message relay between Cells
- **Benchmark:**
  - 16x16x16 voxels per processor, voxel contains 512 particles / species (2)
- **Laser Plasma Interaction:**
  - 13x14x14 voxels per processor, voxel contains 6420 particles / species (3)

VPIC: Compute Considerations

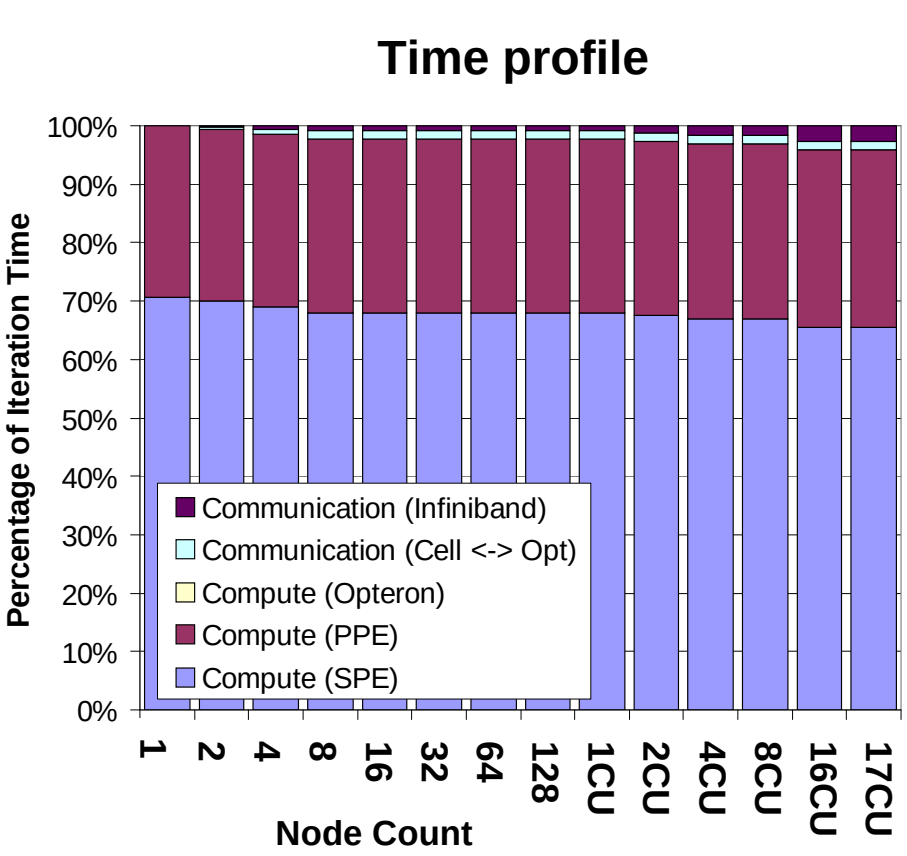
- **# particles per processor can vary over iterations**
  - Input deck dependent

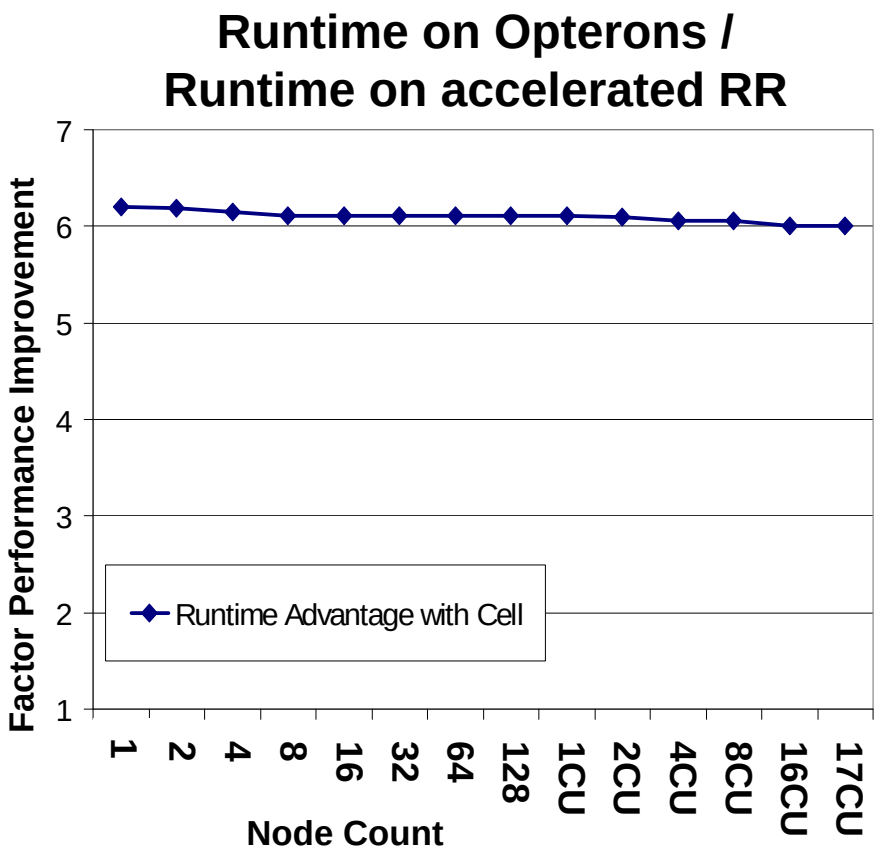




Los Alamos NATIONAL LABORATORY EST. 1943

VPIC: Initial Performance predictions



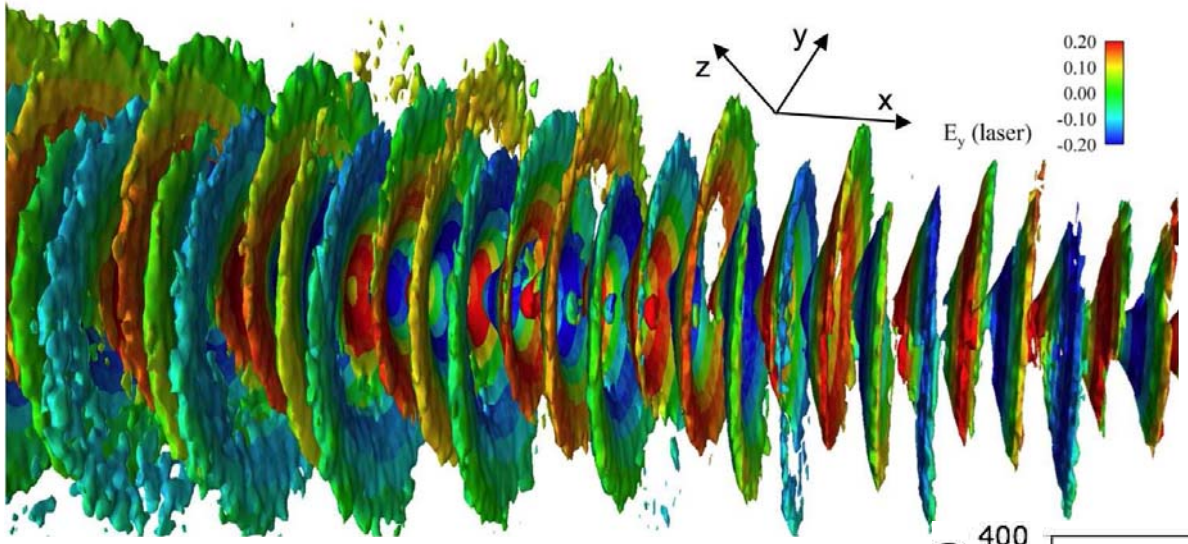


- **compute bound**
  - ~65% SPU, ~31% PPU
- **Very little communication overheads**
  - ~1% Cell <-> Opteron, ~3% Infiniband

- **Very Good scaling**
- **~6x better performance using Cell (benchmark input)**

Los Alamos NATIONAL LABORATORY EST. 1943

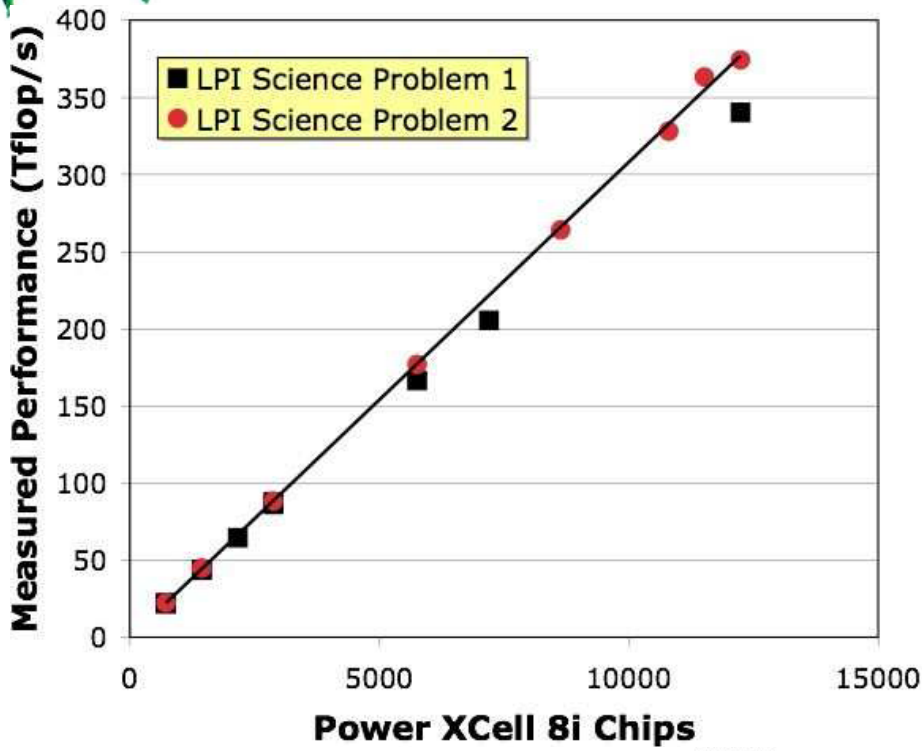
0.374 Pflops on Laser Plasma Interaction



3D laser plasma interaction

Isosurfaces of Electric Field
Color indicates laser field

- **Several large-runs on pre-production Roadrunner**
- **Almost linear scaling observed**
  - Overhead of communications is low

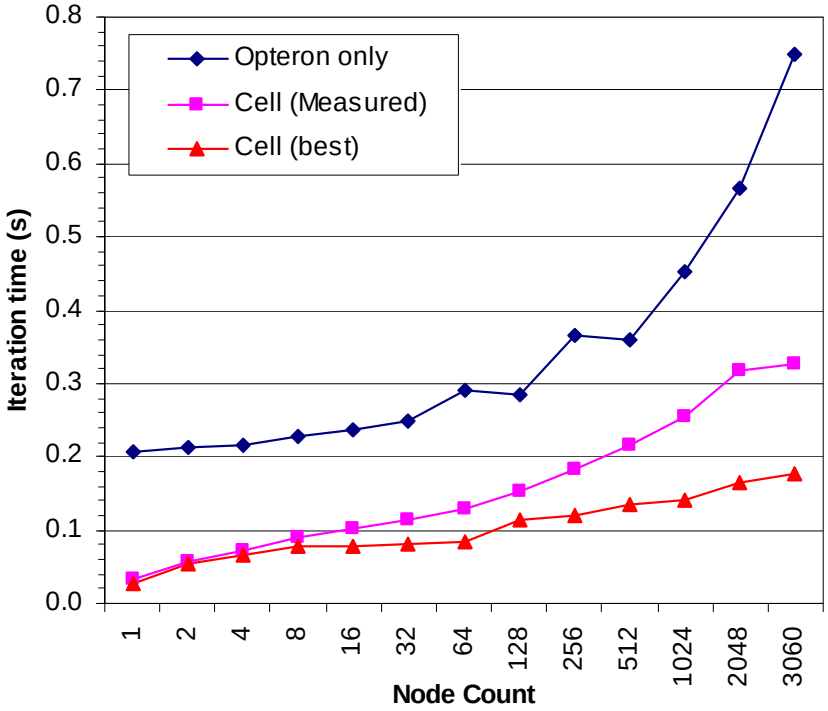


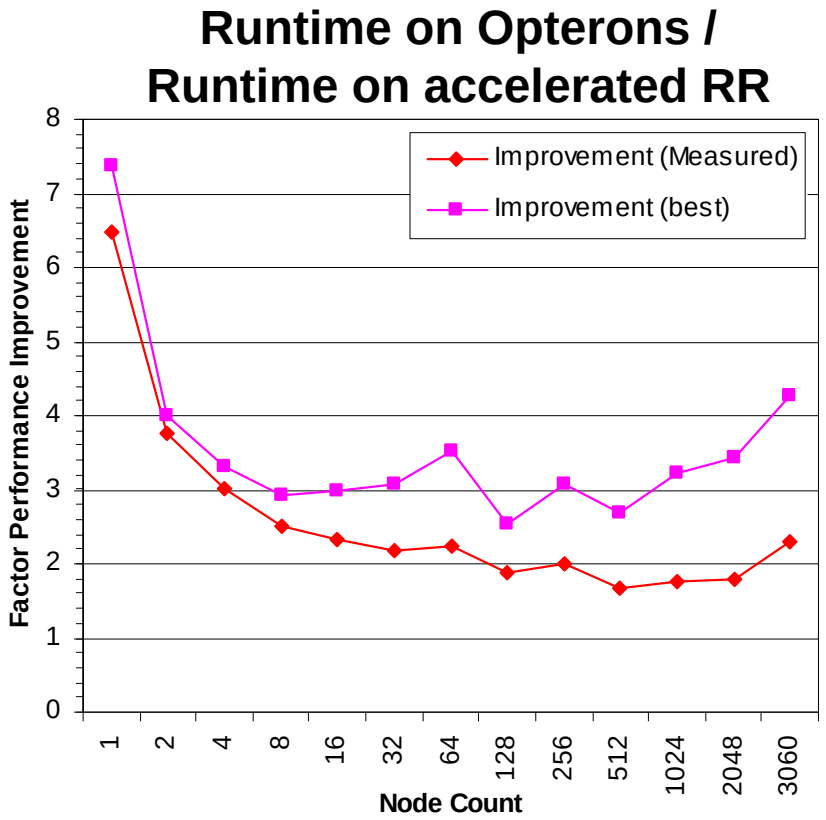
"0.374 Pflop/s Trillion-particle Particle-in-cell Modeling of Laser Plasma Interactions on Roadrunner", Bowers, Albright, Bergen et. Al., Gordon Bell Finalist, SC'08

Example: Sweep3D

- **Deterministic S_N transport**
  - Wavefront Algorithm
- **3-D problem decomposed across a logical 2-D processor array**
- **A key parameter is the computational block size**
  - Angles per block fixed at 6 (for high SPE compute efficiency)
  - K-planes per block is variable (decreases with scale for high parallel efficiency)
- **Weak-scaling**
  - Subgrid per processor 5x5x400
- **Fine-grained communications:**
  - 2 messages sent per SPE per block per cycle
  - Sizes depend on block size, 240B -> 4,800B
- **SPE centric (uses Cell-Messaging-Layer)**
  - Each SPE has an MPI rank, PPE & Opterons support MPI messaging
- **At small-scale performance is compute-bound**
- **At large-scale performance is impacted by both message latency and pipeline length**

Sweep3D: Performance on Roadrunner





- **Sweep3D sensitive to communication latency**
  - Increased due to Cell <-> Opteron
- **Current performance may improve (shown as *best* above)**
  - Looking at optimization of Cell <-> Opteron communications

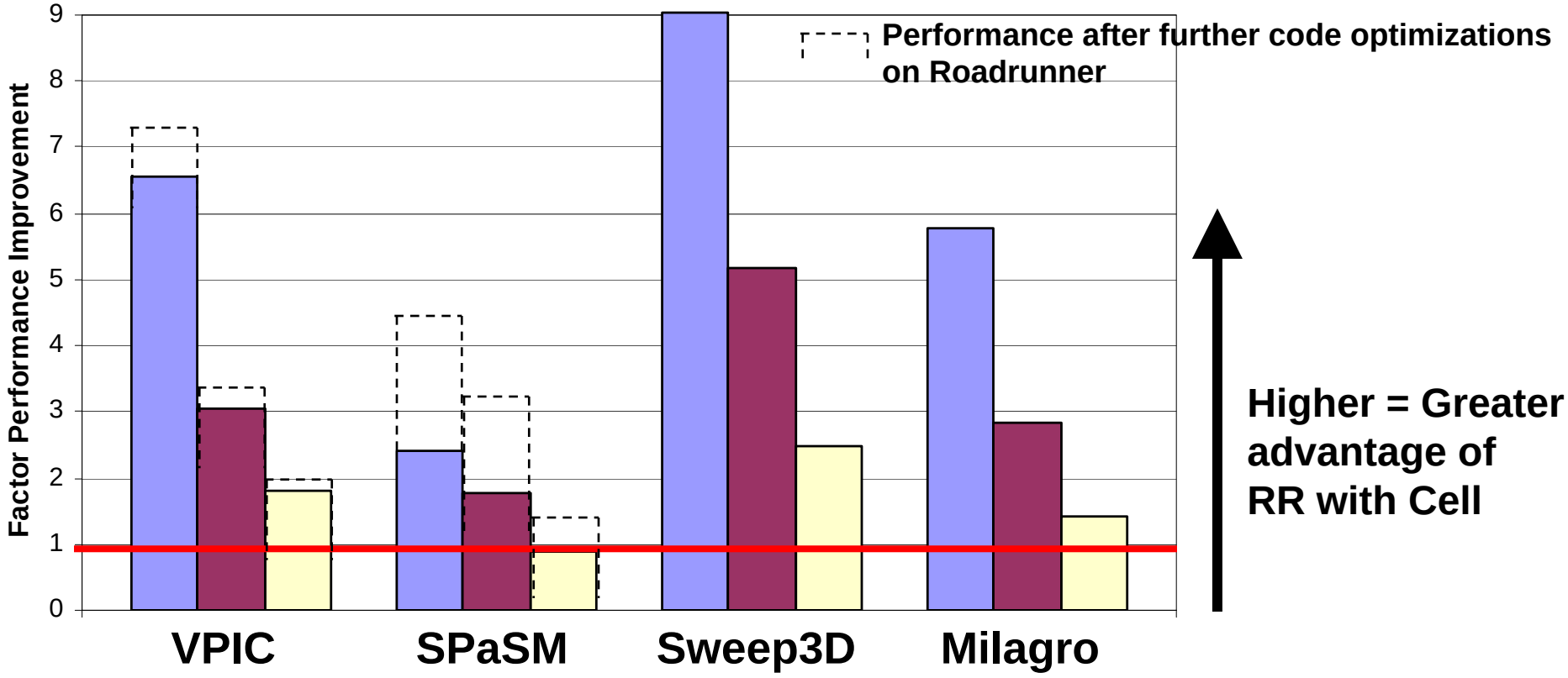
Los Alamos NATIONAL LABORATORY EST. 1943

Roadrunner Performance (Modeled)
Relative to other possible Systems

- **Nodes used for comparison:**
  - Triblade (4x cell-eDP, and AMD 2-socket x 2-core) [Roadrunner]
  - AMD Barcelona 2-socket x 4-core
  - AMD Barcelona 4-socket x 4-core

- **Fixed problem size per node**
  - when comparing node performance

Single Node Performance Comparison



RR without Cell (2-socket x 2-core) vs. RR with Cells

Barcelona (2s x 4c) vs. RR with Cells

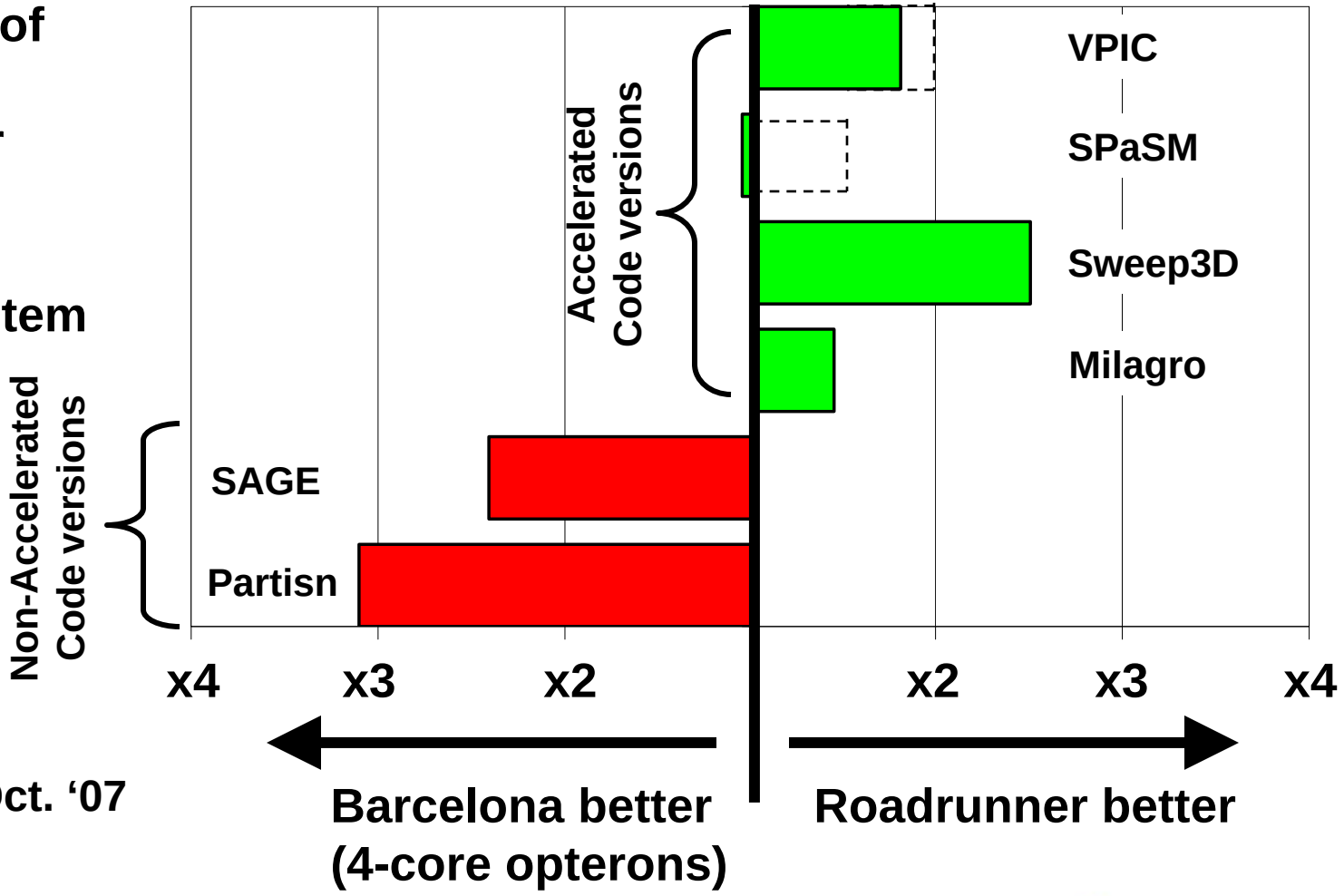
Barcelona (4s x 4c) vs. RR with Cells

Los Alamos NATIONAL LABORATORY EST. 1943

Roadrunner has a significant performance advantage

Performance of

Roadrunner
vs.
equivalent
Quad-core System



Note:
Code versions, Oct. '07

Los Alamos NATIONAL LABORATORY EST. 1943

So far with Roadrunner…

- **Currently system going through acceptance testing @ Los Alamos**
- **Pre-production scientific application runs on 17 CUs**
  - VPIC, Sweep3D, SPaSM, PetaVision
- **2 Gordon Bell finalists**
  - VPIC measured 374 Tflop/s
  - SPaSM measured 369 Tflop/s
- **Other applications being prepared for runs, including**
  - Petavision: goal is synthetic visual cognition
  - Already demonstrated 1.144 PF (SP) on kernel application
- **1.026 PF/s LINPACK in May 2008**
- **Roadrunner has shown that**
  - A hybrid opteron + cell system can be efficiently utilized, and
  - hybrid system can significantly accelerate application performance

Advancing Architectures

- **Technology:**
  - Heterogeneity, accelerators , GPUs
  - Clusters on a chip (cores++, networks)
  - Integrating processors on top of memory, or
  - Integrating memory on top of processors
  - Silicon Photonics
  - Hierarchical Connectivity (many levels of networks)
- **Workload:**
  - Programming models
  - Code optimizations
    - » **Overlap: communicate and compute**
    - » **Overlap: memory and compute (SW prefetching)**
  - Software managed memories
  - Weak -> Strong scaling
- **All of the above ?**

Further Roadrunner Resources

- **http://www.lanl.gov/roadrunner**
  - More details, slides, presentations, some documentation
- **The Cell Messaging Layer (CML)**
  - http://www.sourceforge.net/projects/cellmessaging
- **@ SC'08 in Austin:**
  - Technical Paper "**Entering the Petaflop Era: The Architecture and Performance of Roadrunner**"
  - Gordon Bell Finalist "**0.374 Pflop/s Trillion-particle Particle-in-cell Modeling of Laser Plasma Interactions on Roadrunner**"
  - Gordon Bell Finalist "**369 Tflop/s Molecular Dynamics Simulations on the Roadrunner General-purpose Heterogeneous Supercomputer**"
  - Birds of a Feather "**Roadrunner: First to a Petaflop, First of a New Breed**"
  - Los Alamos Booth in Exhibition Hall

Los Alamos
NATIONAL LABORATORY
EST.1943