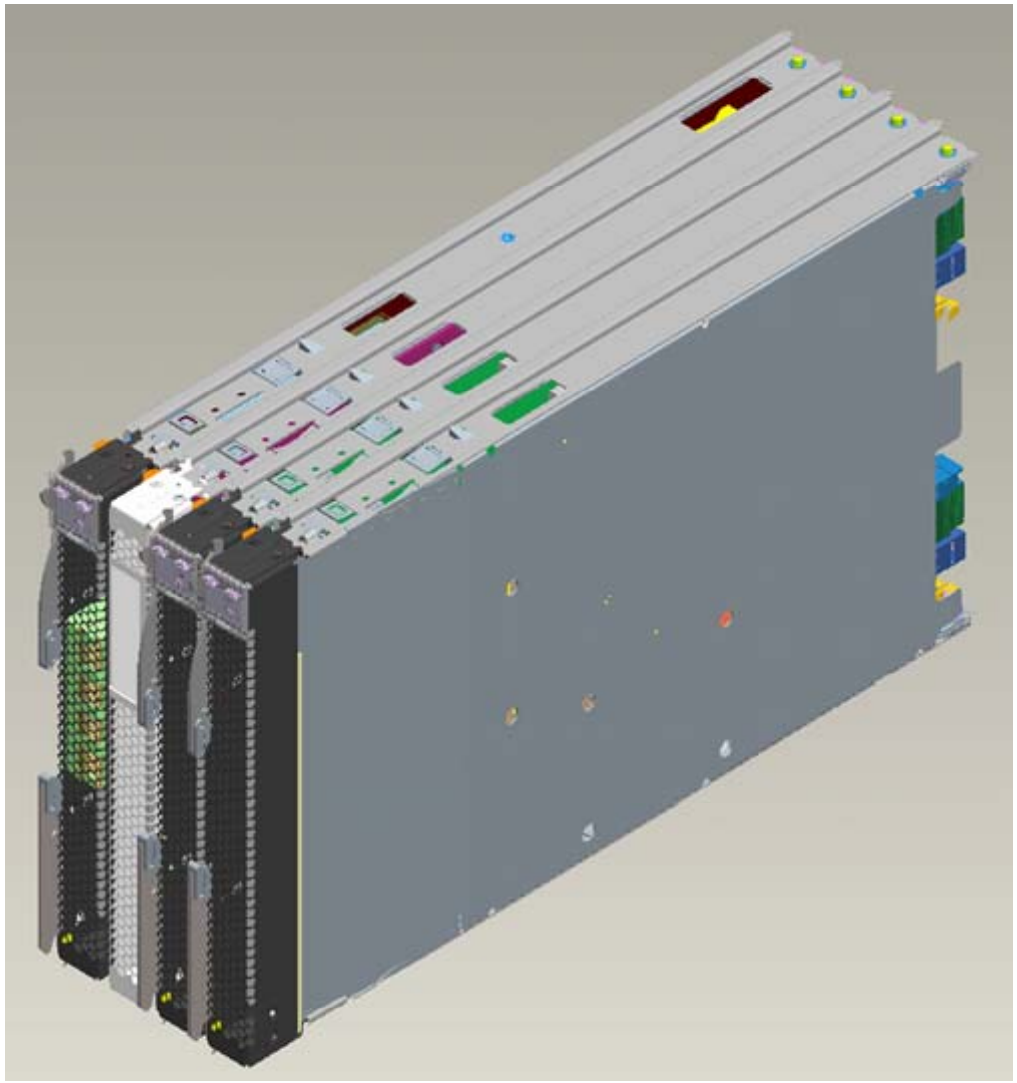
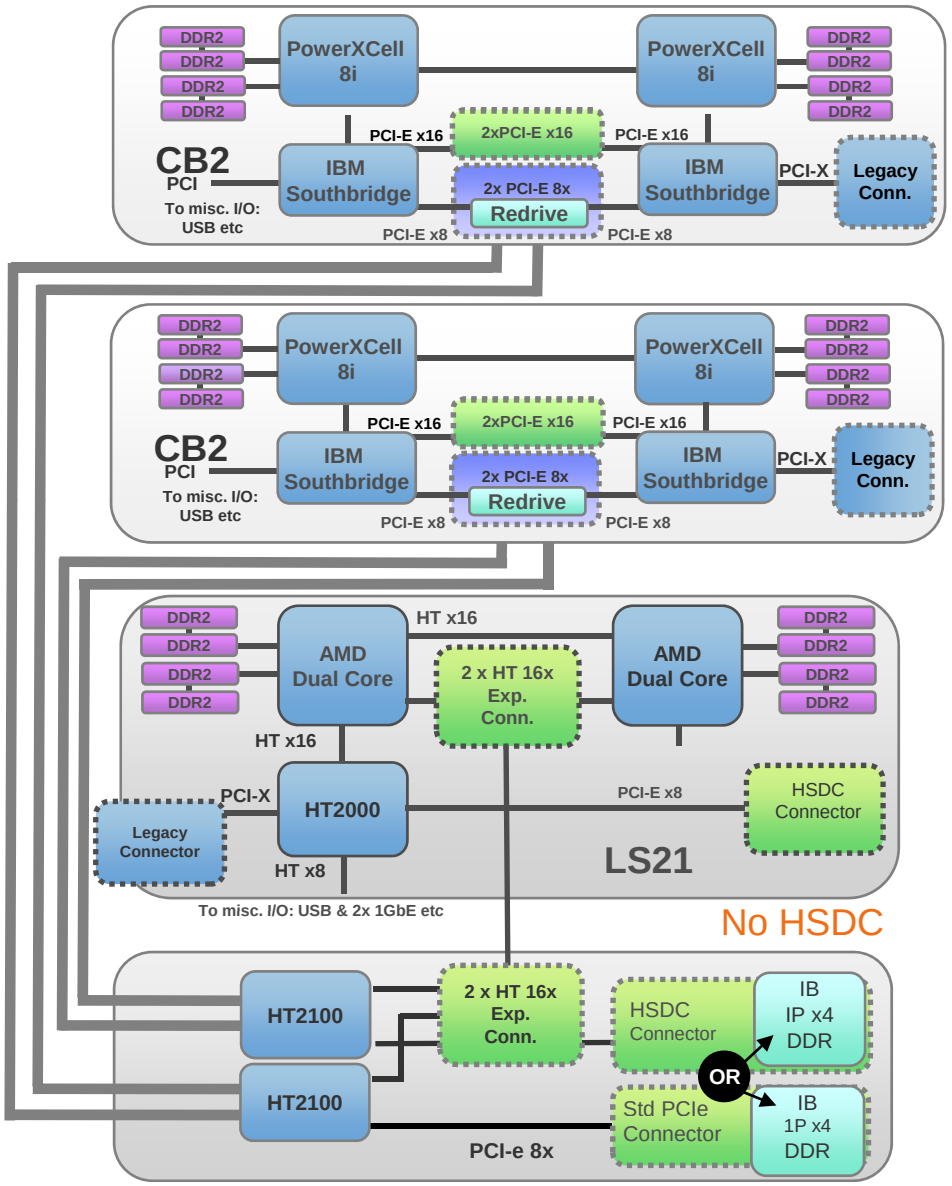
Roadrunner - Integrated Hybrid Node

- LS21 AMD Host Blade
  - Dual socket dual core AMD Opteron
  - DDR2 direct attach DIMM
- Expansion Card
  - 2 HT2100 HT<->PCIe bridges
- CB2 Based Accelerator Blades
  - Dual Cell Sockets
  - DDR2 direct attach DIMM
- AMD Host to Cell eDP connectivity
  - Two x8 PCIe Host to QS22 links



PowerXCell 8i / AMD TriBlade (Dual Core Opteron, IB-DDR)

- AMD Host Blade + Expansion
  - Dual socket dual core AMD Opteron (2 x 7.2 GFLOPS)
    - LS21 + 2 by HT 16x connector
  - DDR2 direct attach DIMM channels
    - 8GB
    - 10.7 GB/s/socket (0.48 B/FLOP)
  - New Expansion Card
    - 2 HT2100 HT<->PCI-e bridges
- CB2 Accelerator Blade
  - Dual Cell Sockets
    - 204 GFLOPS @ 3.2Ghz (2 x 102 GFLOPS)
  - DDR2 direct attach DIMM channels
    - 8 GB
    - 25.6 GB/s per PowerXCell 8i chip* (0.25 B/FLOP)
  - AMD Host to PowerXCell 8i connectivity
    - Two x8 PCIe Host to CB2 links
    - ~2+2 GB/s/link → ~4+ 4 GB/s total POR



Roadrunner is a petascale system in 2008

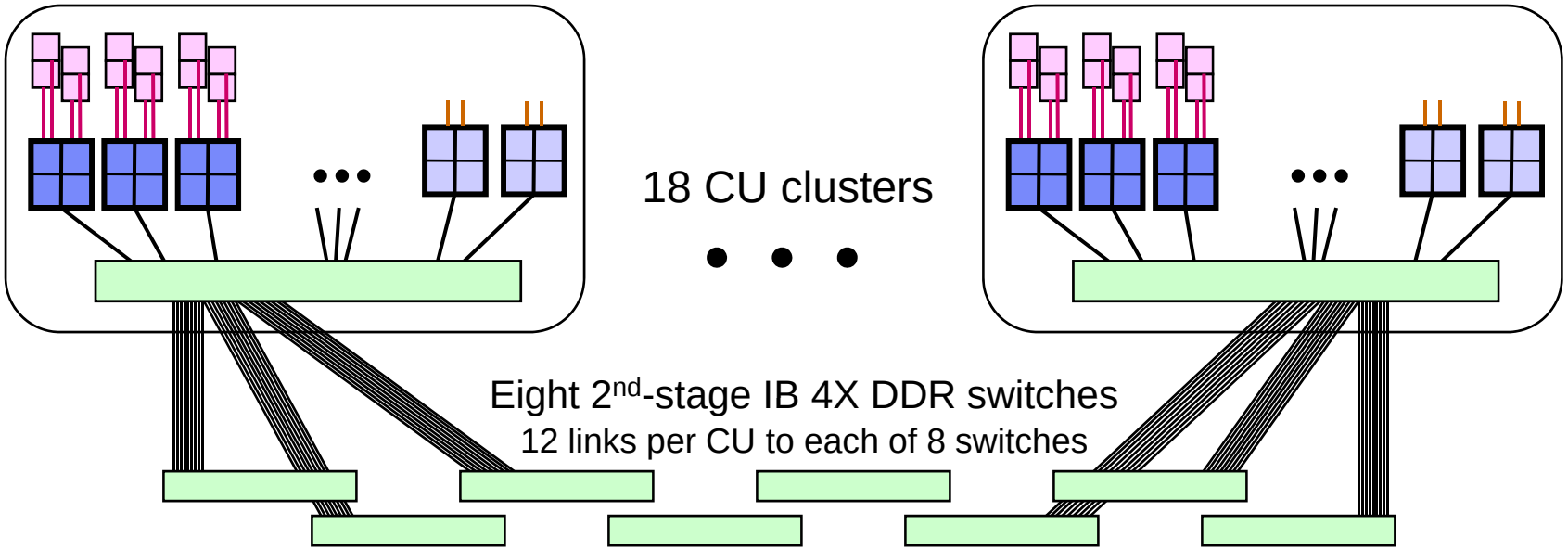
Full Roadrunner Specifications:

6,912 dual-core Opterons
- **49.8 TF DP peak Opteron**
- **27.6 TB Opteron memory**

12,960 Cell eDP chips
- **1.33 PF DP peak Cell eDP**
- **2.65 PF SP peak Cell eDP**
- **51.8 TB Cell memory**
- **277 TB/s Cell memory BW**

3,456 nodes on 2-stage IB 4X DDR
- **13.8 TB/s aggregate BW (bi-dir) (1st stage)**
- **6.9 TB/s aggregate BW (bi-dir) (2nd stage)**
- **3.5 TB/s bi-section BW (bi-dir) (2nd stage)**

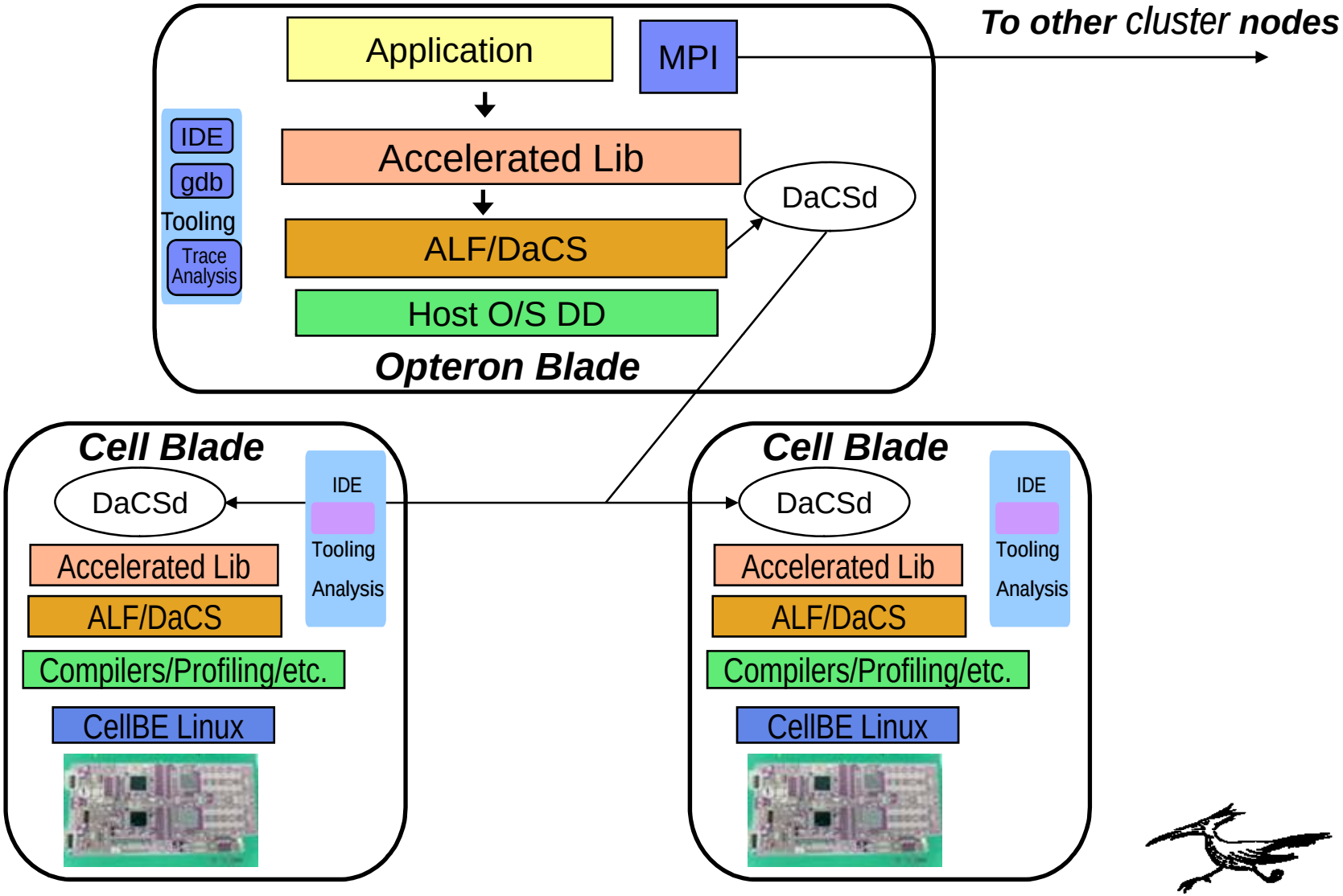
432 10 GigE I/O links on 216 I/O nodes
- **432 GB/s aggregate I/O BW (uni-dir)**
- **(IB limited)**



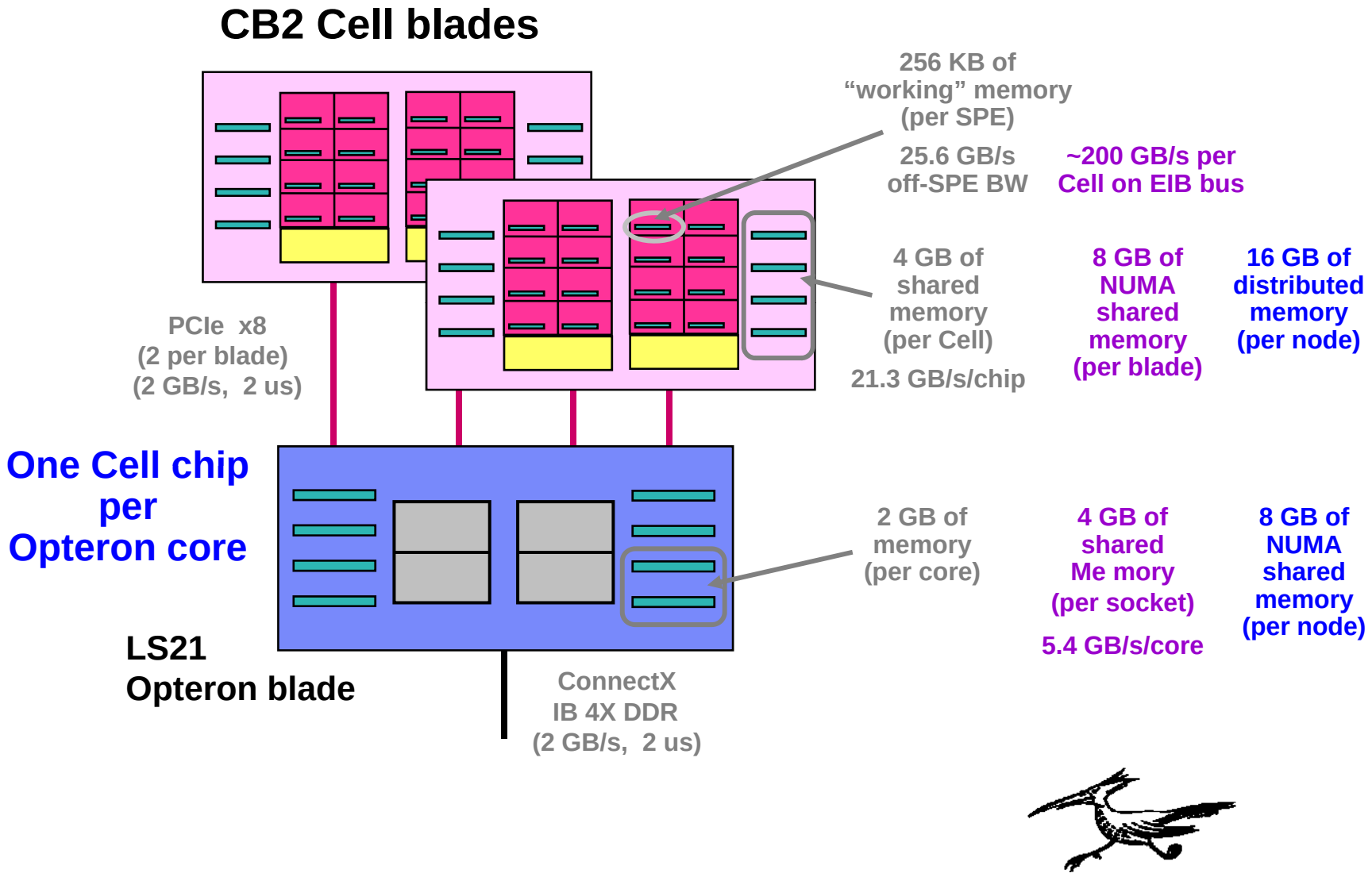
Roadrunner at a glance

- **Cluster of 18 Connected Units**
  - 6,912 AMD dual-core Opterons
  - 12,960 IBM Cell eDP accelerators
  - 49.8 Teraflops peak (Opteron)
  - 1.33 Petaflops peak (Cell eDP)
  - 1PF sustained Linpack
- **InfiniBand 4x DDR fabric**
  - 2-stage fat-tree; all-optical cables
  - Full bi-section BW within each CU
    - 384 GB/s (bi-directional)
  - Half bi-section BW among CUs
    - 3.45 TB/s (bi-directional)
  - Non-disruptive expansion to 24 CUs
- **80 TB aggregate memory**
  - 28 TB Opteron
  - 52 TB Cell
- **216 GB/s sustained File System I/O:**
  - 216x2 10G Ethernets to Panasas
- **RHEL & Fedora Linux**
- **SDK for Multicore Acceleration**
- **xCAT Cluster Management**
  - System-wide GigEnet network
- **3.9 MW Power:**
  - 0.35 GF/Watt
- **Area:**
  - 296 racks
  - 5500 ft^2

Hybrid Node System Software Stack



Roadrunner nodes have a memory hierarchy



Three types of processors work together.

- Parallel computing on Cell
  - data partitioning & work queue pipelining
  - process management & synchronization
- Remote communication to/from Cell
  - data communication & synchronization
  - process management & synchronization
  - computationally-intense offload
- MPI remains as the foundation