Argonne NATIONAL LABORATORY

Overview of the Argonne Leadership Computing Facility

Scott Parker

THE UNIVERSITY OF CHICAGO

Office of Science U.S. DEPARTMENT OF ENERGY

Argonne National Laboratory is managed by The University of Chicago for the U.S. Department of Energy

Argonne Leadership Computing Facility

- ALCF was established in 2006 at Argonne to provide the computational science community with a leading-edge computing capability dedicated to breakthrough science and engineering
- One of two DOE national Leadership Computing Facilities (the other is the National Center for Computational Sciences at Oak Ridge National Laboratory)
- Supports the primary mission of DOE's Office of Science Advanced Scientific Computing Research (ASCR) program to discover, develop, and deploy the computational and networking tools that enable researchers in the scientific disciplines to analyze, model, simulate, and predict complex phenomena important to DOE.

DOE INCITE Program

Innovative and Novel Computational Impact on Theory and Experiment

- **Solicits large computationally intensive research projects**
  - To enable high-impact scientific advances
  - Call for proposal opened once per year (call closed 7/1/2009)
  - INCITE Program web site: www.er.doe.gov/ascr/incite
- **Open to all scientific researchers and organizations**
  - Scientific Discipline Peer Review
  - Computational Readiness Review
- **Provides large computer time & data storage allocations**
  - To a small number of projects for 1-3 years
  - Academic, Federal Lab and Industry, with DOE or other support
- **Primary vehicle for selecting principal science projects for the Leadership Computing Facilities**
- **In 2009, 29 INCITE projects allocated ~400M CPU hours at the ALCF**

Discretionary Allocations

- Time is available for projects without INCITE allocations!
- ALCF Discretionary allocations provide time for:
  - Porting, scaling, and tuning applications
  - Benchmarking codes and preparing INCITE proposals
  - Preliminary science runs prior to an INCITE award
- To apply go to the ALCF allocations page
  - www.alcf.anl.gov/support/gettingstarted

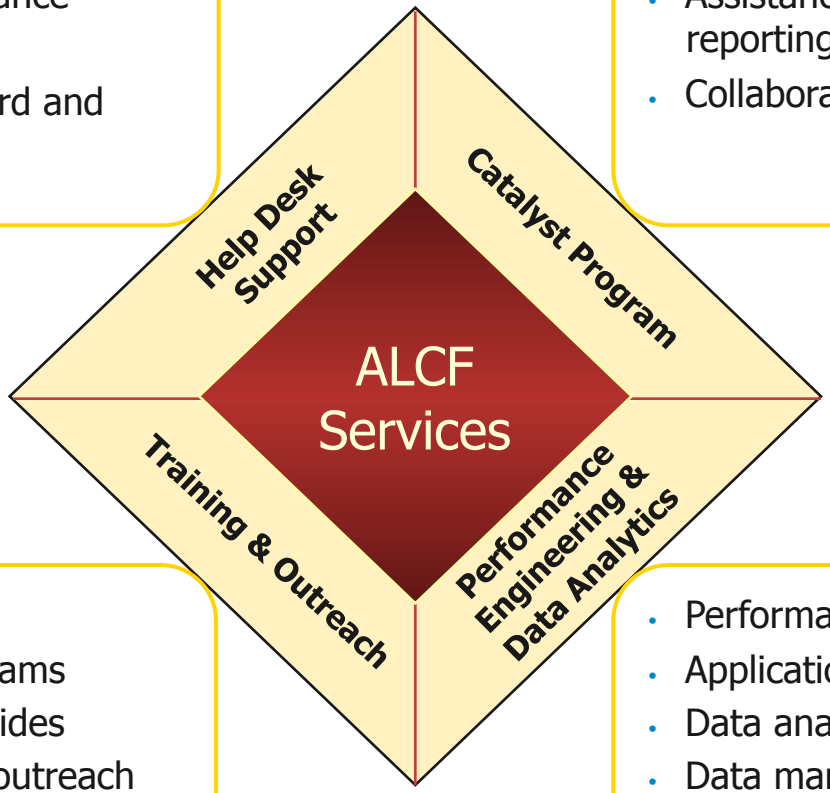
Get An Account!

- If you don't have an account on an ALCF BG/P system (*Intrepid or Surveyor)* you can apply for a workshop account
- To apply:
  - Go to the URL: https://accounts.alcf.anl.gov/accounts/request.php
  - Select "Proceed with Account Request" at the bottom of the page
  - Select the project 'CScADS'
  - Foreign nationals require 593 forms which can take a while
- Running under the 'CScADS' project
  - User with account can run using the 'CScADS' project
    - *qsub –A CScADS …*

ALCF Service Offerings

- Startup assistance
- User administration assistance
- Job management services
- Technical support (Standard and Emergency)

- ALCF science liaison
- Assistance with proposals, planning, reporting
- Collaboration within science domains



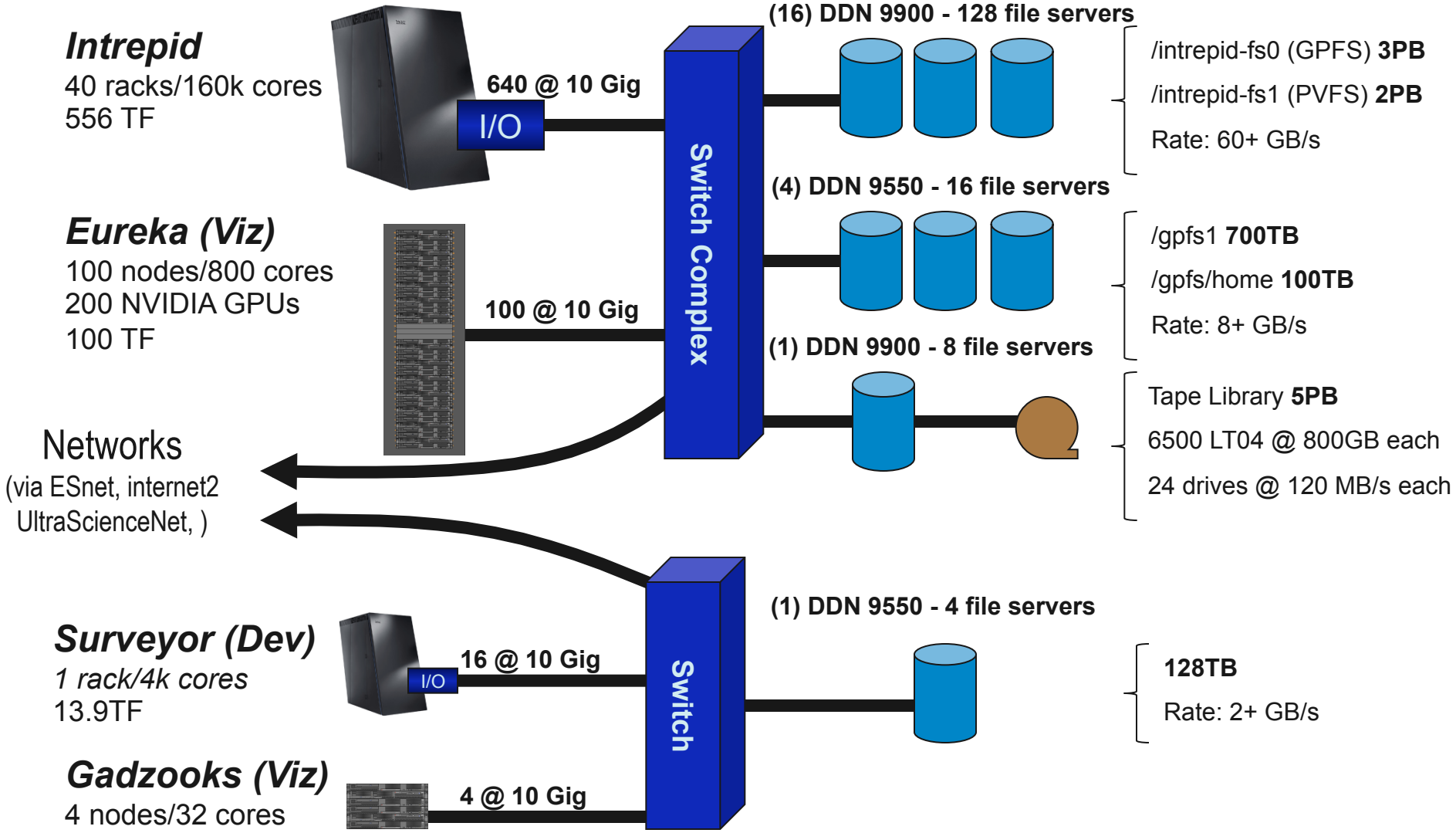
- Workshops & seminars
- Customized training programs
- On-line content & user guides
- Educational and industry outreach programs

- Performance engineering
- Application tuning
- Data analytics
- Data management services

Argonne Leadership Computing Facility

- *Intrepid* - ALCF Blue Gene/P System:
  - 40,960 nodes / 163,840 PPC cores
  - 80 Terabytes of memory
  - Peak flop rate: 557 Teraflops
  - Linpack flop rate: 450.3
  - #7 on the Top500 list
- *Eureka* - ALCF Visualization System:
  - 100 nodes / 800 2.0 GHz Xeon cores
  - 3.2 Terabytes of memory
  - 200 NVIDIA FX5600 GPUs
  - Peak flop rate: 100 Teraflops
- Storage:
  - 8 Petabytes of disk storage with an I/O rate of 80 GB/s
  - 8 Petabytes of archival storage (10,000 volume tape archive)

ALCF Resources - Overview



Blue Gene/P at ALCF

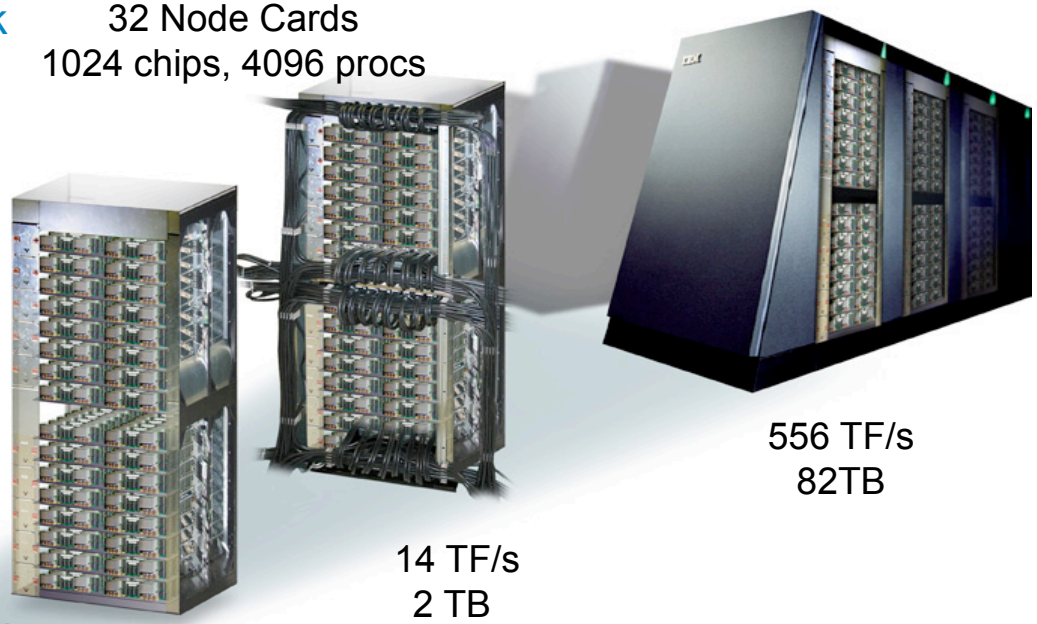
BG/P: Covers removed

BlueGene/P Overview

- 4 850Mhz PowerPC cores per chip
- 1 chip, 2 GB of DDR SDRAM, 5 network interfaces per compute node
- 32 compute nodes per node card
- 32 node cards per rack
- 1,024 nodes total per rack
- 40 rack on Intrepid

Intrepid System
40 Racks

Rack
32 Node Cards
1024 chips, 4096 procs



556 TF/s
82TB

14 TF/s
2 TB

Node Card
(32 chips 4x4x2)
32 compute, 0-2 IO cards

435 GF/s
64 GB

Compute Card
1 chip, 20 DRAMs

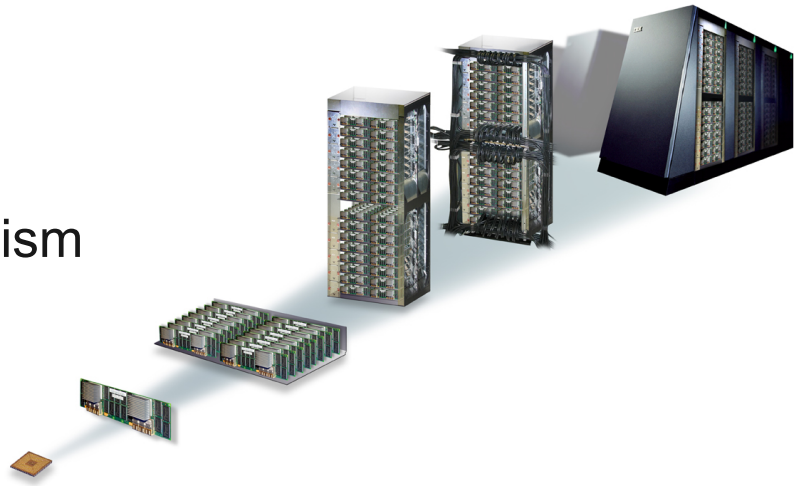
13.6 GF/s
2.0 GB DDR
Supports 4-way SMP

Chip
4 processors

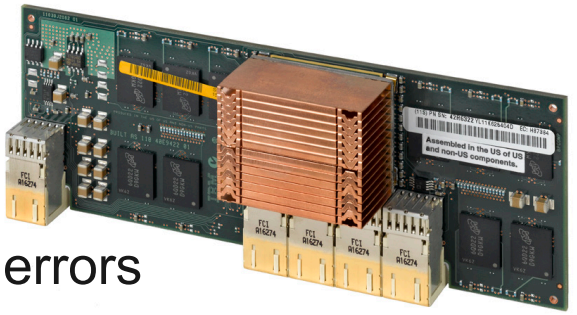
850 MHz
8 MB EDRAM

Front End Node / Service Node
System p Servers
Linux SLES10

Blue Gene DNA



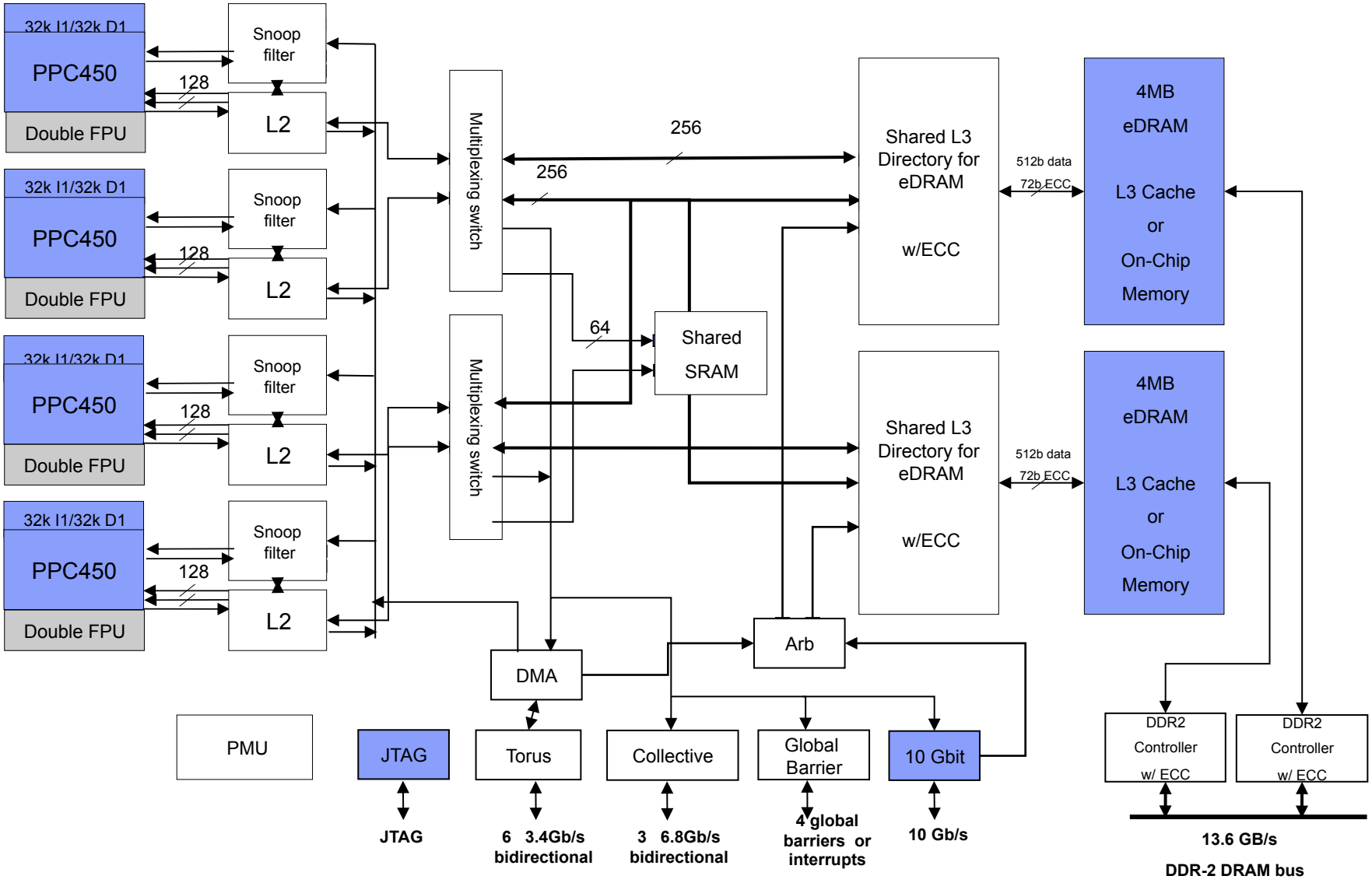
- Low power design ➡ massive parallelism
  - The leader in Green Computing
- System on a Chip (SoC)
  - Improves Price / Performance
  - Reduces system complexity & power
- Custom designed ASIC
  - Reducing overall part count, reducing errors
  - Permits tweaking CPU design to reduce soft errors



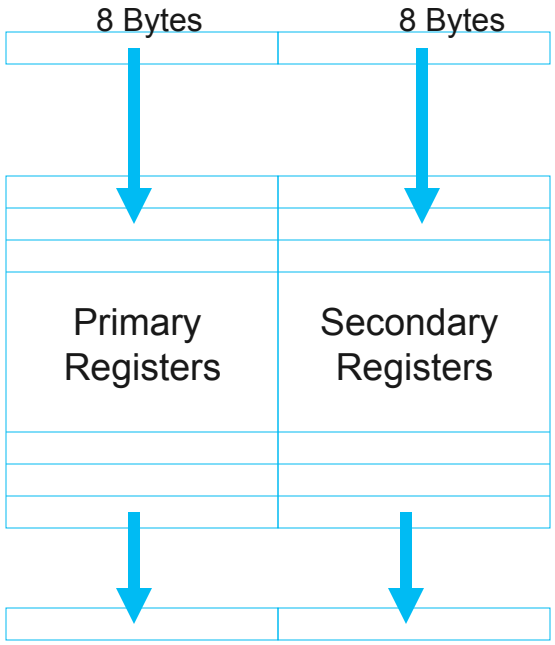
- Dense packaging
- Fast communication network(s)
- Sophisticated RAS (reliability, availability, and serviceability)
- Dynamic software provisioning and configuration
  - **Key feature for linking computer science with applications**

3 watts per sustained gigaflops

Blue Gene/P ASIC



Double Hummer Floating Point Unit



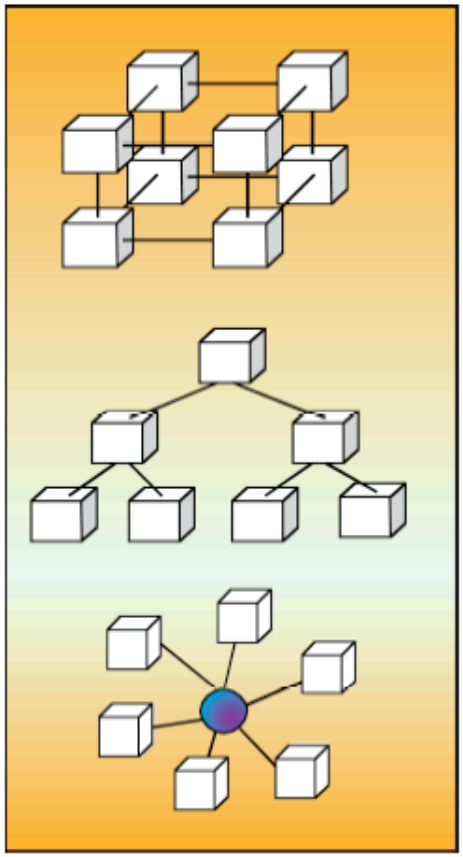
Quad word Load
16 Bytes per instruction

Full range of parallel and cross SIMD floating-point instructions

Quad word Store
16 Bytes per instruction

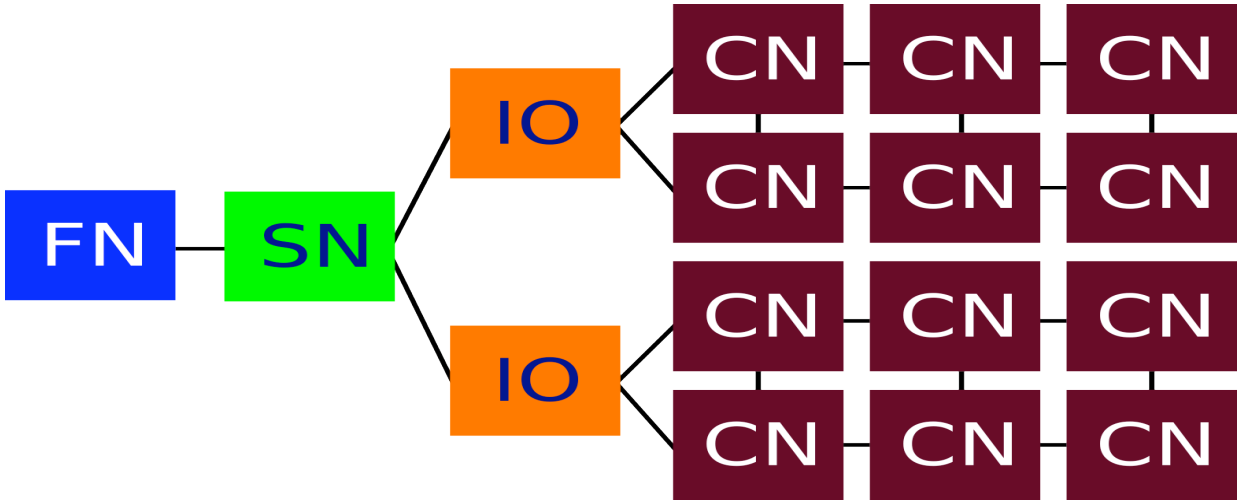
Quad word load/store operations require data aligned on 16-Byte boundaries.

Blue Gene/P Interconnection Networks



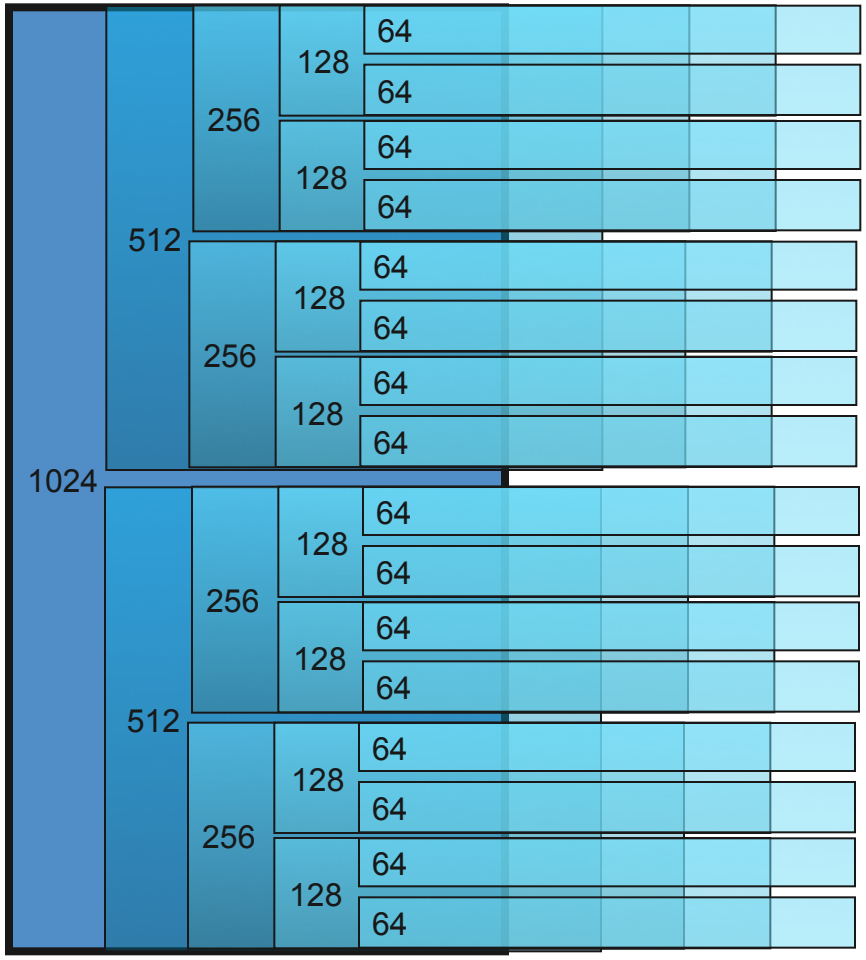
- 3 Dimensional Torus
  - Interconnects all compute nodes
  - Communications backbone for point-to-point
  - 3.4 Gb/s on all 12 node links (5.1 GB/s per node)
  - 0.5 µs latency between nearest neighbors, 5 µs to the farthest
  - MPI: 3 µs latency for one hop, 10 µs to the farthest
  - *Requires half-rack or larger partition*
- Collective Network
  - One-to-all broadcast functionality
  - Reduction operations for integers and doubles
  - 6.8 Gb/s of bandwidth per link per direction
  - Latency of one way tree traversal 1.3 µs, MPI 5 µs
  - Interconnects all compute nodes and I/O nodes
- Low Latency Global Barrier and Interrupt
  - Latency of one way to reach 72K nodes 0.65 µs, MPI 1.6 µs
- 10 Gb/s functional Ethernet
  - Disk I/O
- 1Gb private control (JTAG)
  - Service node/system management

Blue Gene/P Heterogeneity



- **Front-end nodes (FN)**: dedicated for user's to login, compile programs, submit jobs, query job status, debug applications, 2.5 GHz PowerPC 970, Linux OS
- **Service nodes (SN)**: perform system management services, create and monitoring processes, initialize and monitor hardware, configure partitions, control jobs, store statistics
- **I/O nodes (IO)**: provide a number of OS services, such as files, sockets, process management, debugging, 1 I/O node per 64 compute nodes
- **Compute nodes (CN)**: run user application, accessed only through qsub command, no shell, limited OS services, quad core 850 MHz PowerPC 450, CNK OS

BG/P Partitions



- *Intrepid* compute nodes are grouped into partitions ranging from 64 to 40,960 nodes in powers of 2
- Jobs run in smallest partition into which they fit
- Job makes entire partition unavailable
- Only 1 job may run in a partition
- Smaller partitions are enclosed inside of larger ones
- Minimum partition size is 64 nodes
  - *1 I/O node for each 64 compute nodes*
- Partition's networks electrically isolated, each partition is it's own torus/mesh
- Partitions <512 nodes form a mesh network, partitions >=512 nodes form a torus

Partitions on 1 rack of *Intrepid* showing number of nodes

Blue Gene/P Software

- System:
  - Linux on Login and I/O Nodes
  - Compute Node Kernel (CNK) O/S on Compute Nodes (Linux like)
  - XL compilers (C, C++, Fortran 77-90-95-2003)
  - Python
  - MPI/OpenMP
  - ESSL math libraries
- Management:
  - BlueGene/P Control system (IBM DB2 database)
  - Cobalt – resource manager(qsub, qstat, qdel, qalter)
  - Clusterbank – allocation management system
- Storage:
  - GPFS – parallel filesystem
  - PVFS – high performance parallel filesystem
  - Tape systems - HPSS, Amanda

Unique and Challenging Features

- Low power cores (850 MHz, 3.4 GFlop) but many of them (163,840)
  - High scalability is key to high total flop rate
  - Balance between CPU and network makes scalability easier
- Relatively low memory per CPU core (but very large aggregate).
  - BG/P: 2 Gig / 4 cores
  - True SMP is possible (sharing data structures)
- Single node optimization
  - Use of “double hummer” requires lots of hand tuning and compiler experimentation
  - Strategies: Good math libs, performance counters, code tools
- Scalable I/O strategy required
  - One file per process **strongly** discouraged
  - PnetCDF and HDF5 are good strategies for effectively using parallel storage system (GPFS, PVFS)
- Debugging at scale remains challenging
  - Tool groups are helping, but the issue is nevertheless hard

More about BlueGene/P

- Logical partitions, with complete electrical isolation
- Partition rebooted between jobs
- Only allowed one thread/process per core using one of three modes
  - SMP – 1 process with 4 threads - 1 thread per core
  - Dual – 2 processes with 2 thread
  - VN – 4 processes – 1 process per core
- Specialized IBM kernels – compute node kernel, I/O node kernel
  - Single-executable kernel for compute nodes
  - Usually runs MPI code, can also run in “HTC” mode
  - Stripped-down Linux for I/O nodes
  - The ciod daemon handles system calls
- But also can run custom kernels
    - *ZeptoOS / Compute Node Linux*
    - *ZeptoOS I/O node kernel*
    - *Plan 9 (INCITE project)*

Programming Models and Development Environment

- Languages:
  - Full language support with IBM XL and GNU compilers
  - Languages: Fortran, C, C++, Python
- MPI:
  - Based on MPICH2 1.0.x base code:
    - *MPI-IO supported*
    - *One-sided communication supported*
    - *No process management (MPI_Spawn(), MPI_Connect(), etc)*
  - Utilizes the 3 different BG/P networks for different MPI functions
- Threads:
  - OpenMP 2.5
  - NPTL Pthreads
- Linux development environment:
  - Compute Node Kernel provides look and feel of a Linux environment
    - *POSIX routines (with some restrictions: no fork() or system())*
    - *BG/P adds pthread support, additional socket support*
  - Supports statically and dynamically linked libraries
  - Cross compile since login nodes and compute nodes have different processor & OS

Restrictions and Complications

- SPMD model:
  - compute nodes run the same executable (changing)
- Space Sharing:
  - one parallel job per partition of machine
  - one process/thread per core in each compute node
    - *smp-mode, one MPI task/node, 4 threads/task, 2GB of RAM*
    - *dual-mode, two MPI tasks/node, 2 threads/task, 1GB of RAM*
    - *vn-mode, 4 single-threaded MPI tasks/node, 512MB of RAM*
- memory limited to physical memory (no virtual memory)
- restricted set of POSIX routines (no fork, system, …)
- Cross compiling required due to hardware & O/S differences between login and compute nodes

Performance and Debugging Tools

- IBM High Performance Computing Toolkit
  - MPI Profile and Tracing Library
  - HPM Library for hardware performance counters
  - Xprofiler visualization of gprof profiles
- Universal Performance Counters
  - Blue Gene/P provides 256 on chip counters for hardware events
  - UPC and HPM libraries provide access to counters
- TAU – Tuning and Analysis Utilities
- gprof – standard linux profiling tool
- Core files – lightweight core files, text format, no full memory dump
- Coreprocessor - Generates stack trace from core files
- GDB – gnu debugger
- TotalView - Debugging across ten of thousands of cores

Supported Libraries and Programs

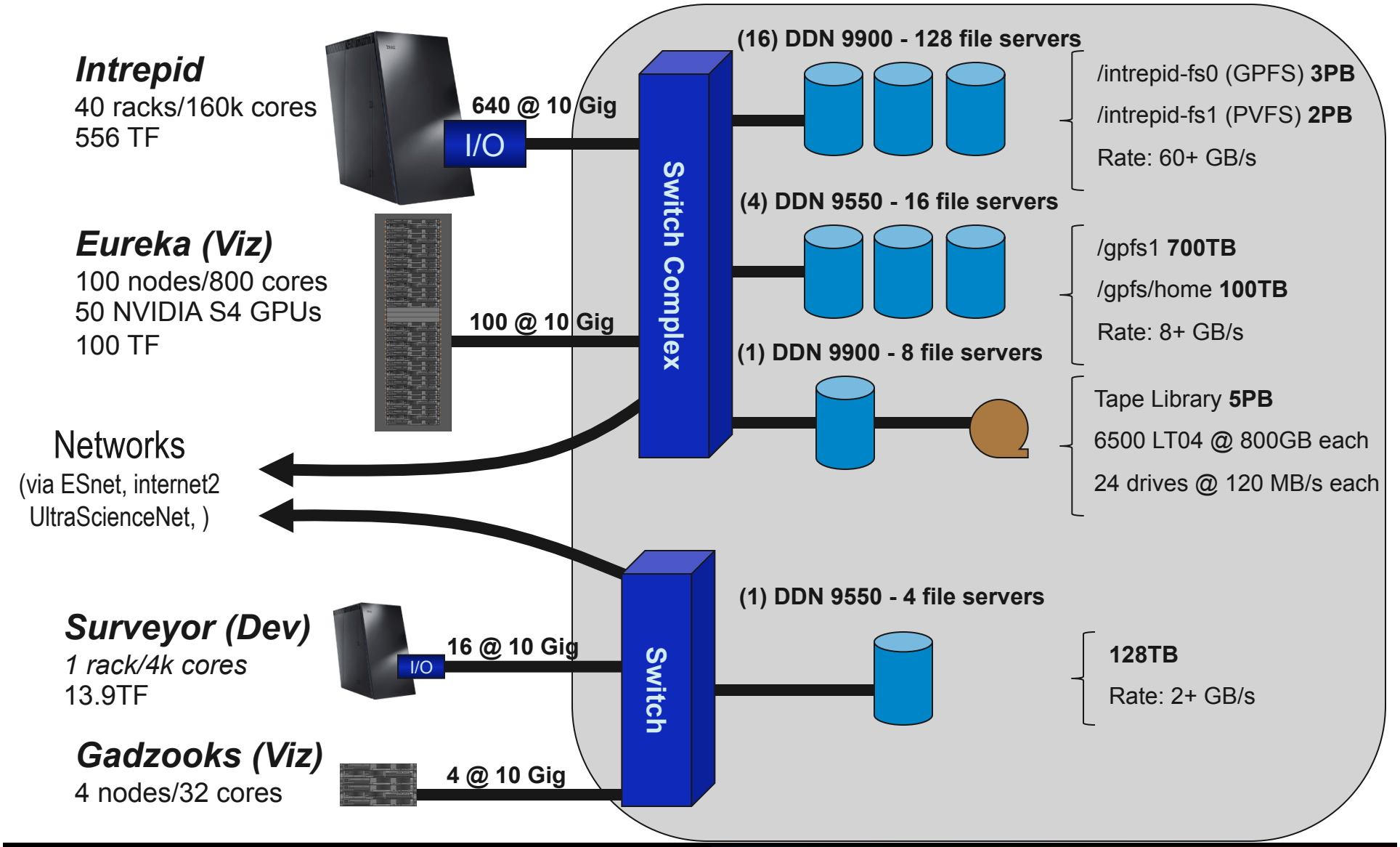
Library	Location	Description
BLAS, LAPACK	/soft/apps/blas-lapack-lib	Basic vector linear algebra subroutines.
ESSL	/soft/apps/ESSL-4.3	Mathematical subroutines designed to improve the performance of engineering and scientific applications on BlueGene
LIBGOTO	/soft/apps/LIBGOTO	Very efficient BLAS-1.2.3 implementation for BlueGene from Kazushige
SCALAPACK	/soft/apps/SCALAPACK	High-performance linear algebra routines for distributed-memory message-passing MIMD computers and networks of workstations.
PETSc	/soft/apps/petsc	A suite of data structures and routines that provide the building blocks for the implementation of large-scale application codes on serial and parallel computers.
fftw-2.1.5, fftw-3.1.2	/soft/apps/fftw-	A library for computing the discrete many-dimensional Fourier transform
p3dfft	/soft/apps/p3dfft-2.1beta-	Highly scalable parallel 3D Fast Fourier Transforms library.
hypre-2.0.0	/soft/apps/hypre-2.0.0	A library for solving large, sparse linear systems of equations on massively parallel computers.
SuperLU-3.0	/soft/apps/SuperLU_3.0	A library for the direct solution of large, sparse, non-symmetric systems of linear equations on high performance machines.
MUMPs-4.7.3	/soft/apps/MUMPS_4.7.3	A multifrontal massively parallel sparse direct solver.
spooles-2.2	/soft/apps/spooles-2.2	A library for solving sparse real and complex linear systems of equations.

Supported Libraries and Programs

Program	Location	Description
TotalView	/soft/apps/totalview-8.5.0-0	Multithreaded, multiprocess source code debugger for high performance computing.
Coreprocessor	/soft/apps/coreprocessor.pl	A tool to debug and provide postmortem analysis of dead applications.
TAU-2.17	/soft/apps/tau	A portable profiling and tracing toolkit for performance analysis of parallel programs written in Fortran, C++, and C
HPCT	/soft/apps/hpct_bgp	MPI profiling and tracing library, which collects profiling and tracing data for MPI programs.

Program	Location	Description
armci	/bgsys/drivers/ppcfloor/comm	The Aggregate Remote Memory Copy (ARMCI) library
HDF5	/soft/apps/hdf5-1.6.6	The Hierarchical Data Format (HDF) is a model for managing and storing data.
NetCDF	/soft/apps/netcdf-3.6.2	A set of software libraries and machine-independent data formats that supports the creation, access, and sharing of array-oriented scientific data.
Parallel NetCDF	/soft/apps/parallel-netcdf-1.0.2	A library providing high-performance I/O while still maintaining file-format compatibility with Unidata's NetCDF.
mercurial-0.9.5	/soft/apps/mercurial-0.9.5	A distributed version-control system
Scons	/soft/apps/scons-0.97	A cross-platform substitute for the classic Make utility
tcl-8.4.14	/soft/apps/tcl-8.4.14	A dynamic programming language,

File Systems

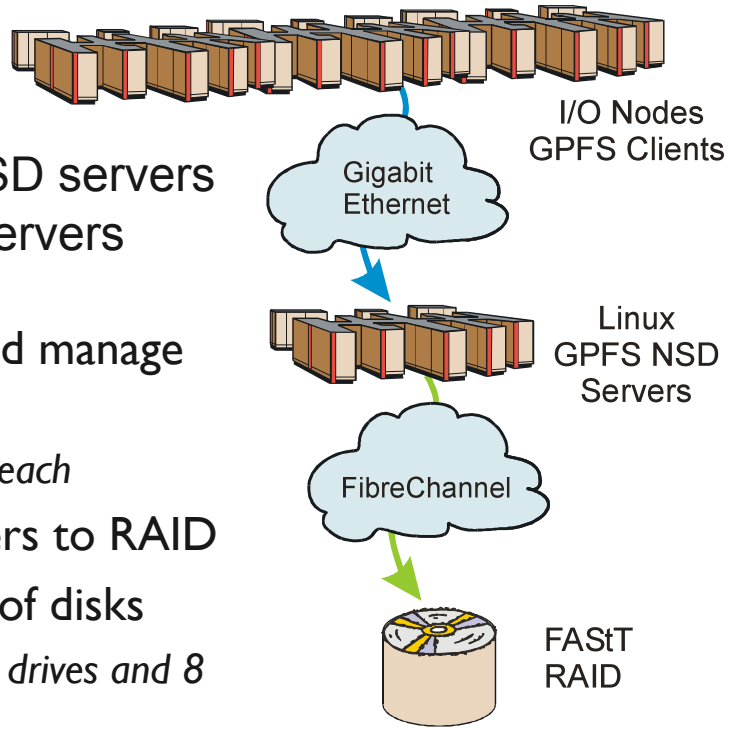


Intrepid File Systems

- /gpfs/home
  - GPS, 100 TB
  - Intended for storing source, configuration, and input files
  - Not intended for larger scale parallel I/O, or storing large files
  - Backup and snap-shot files
- /intrepid-fs0
  - GPFS, 3 PB, 60+ GB/s
  - Intended for very fast parallel IO, program input and output
  - Not backed up, but you can initiate archive via HPSS
- /intrepid-fs1
  - PVFS, 2 PB, 50+ GB/s
  - Intended for very fast parallel IO, program input and output
  - Not backed up, but you can initiate archive via HPSS
  - NOTE: Binaries can not be executed from PVFS

General Parallel File System (GPFS) on Blue Gene

- Blue Gene can generate enormous I/O demand
  - BG/P IO-rich has 640 I/O nodes at 10Gb/s
  - Requires a parallel file system (peak 78 GB/s)
- GPFS Setup:
  - **I/O nodes** run GPFS client that call external NSD servers
  - **10 GB/s Ethernet** connect I/O nodes to NSD servers
    - *900+ port 10 Gigabit Ethernet Myricom switch complex*
  - **NSD Servers** run parallel file system software and manage incoming FS traffic from I/O nodes
    - *136 two dual core Opteron servers with 8 Gbytes of RAM each*
  - **Infiniband/Fiber Channel** connect NSD servers to RAID
  - **Enterprise storage** controllers and large racks of disks
    - *17 DataDirect S2A9900 controller pairs with 480 1 Tbyte drives and 8 InfiniBand ports per pair*
- Brings traditional benefits of GPFS to Blue Gene
  - I/O parallelism
  - Cache consistent shared access
  - Aggressive read-ahead, write-behind



Visualization and Data Analytics

- *Eureka* makes data analytics and visualization at Intrepid's scale possible through the world's largest installation of NVIDIA S4 external GPUs
- The system consists of 100 servers with 200 Quadro FX5600 graphics engines.
- The system is attached directly to the core switch complex of the Blue Gene/P, providing very high throughput between the BG/P and the analysis nodes, and also to the parallel file system

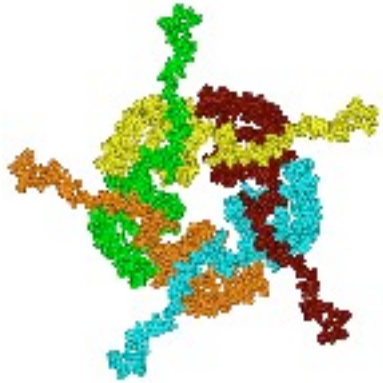
Eureka Visualization System

- 100 Nodes:
  - Two 2.0 GHz Quad Core Xeon (8 cores/node)
  - 32 GB RAM
- 50 NVidia S4 external GPUs
  - 200 Quadro FX5600 high end graphics cards
- Nodes connect to the S4 via a 16x PCIe V2.0 card
- Over 111 TF peak FLOP rating
  - *Includes the GPUs*
- 3.2 TB of RAM (5% of intrepid RAM)
- No local scratch space
  - Data access is all via the central parallel file system
  - Myricom 10G (10 Gbs) NIC

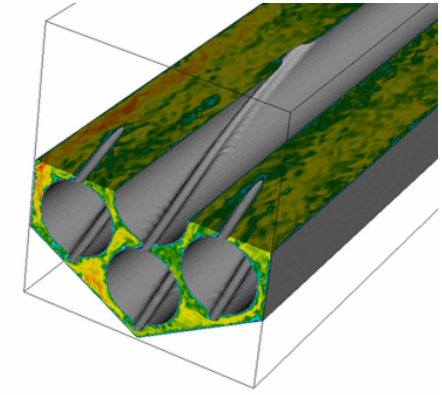
Software Resources

- Installed and Supported Software:
  - VisIT
  - ParaView
  - VTK
  - VMD
  - And good old gnuplot
- Software stack is driven by our users
  - So if we don't have what you need, please speak up
  - Note that we don't list any commercial apps
    - *No user driven demand to date*

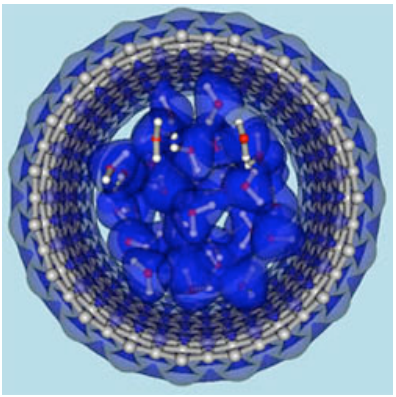
ALCF INCITE Projects Span Many Domains



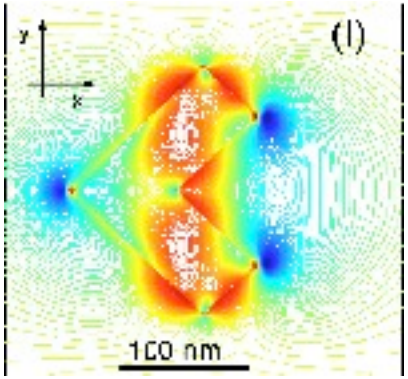
Life Sciences
U CA-San Diego



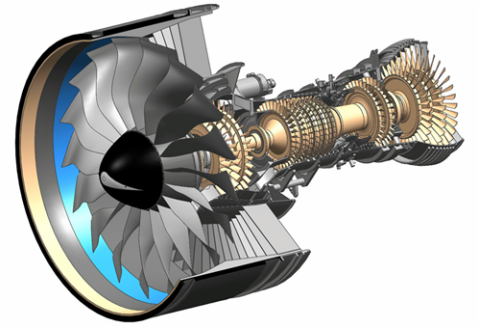
Applied Math
Argonne Nat'l Lab



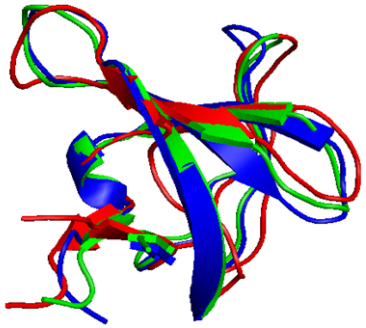
Physical Chemistry
U CA-Davis



Nanoscience
Northwestern U



Engineering Physics
Pratt & Whitney

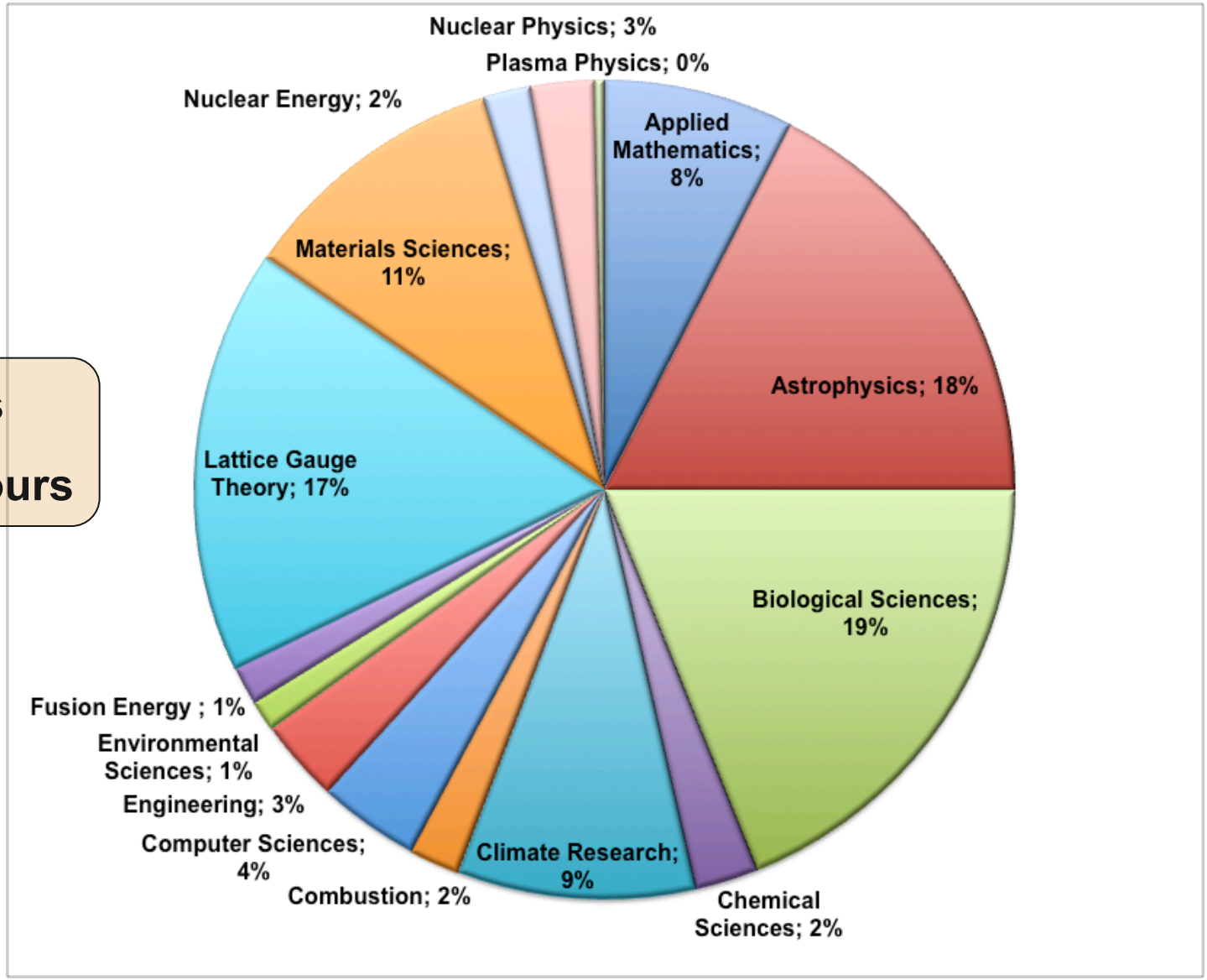


Biology
U Washington

2009 INCITE Allocations at ALCF

28 projects

400 M CPU Hours



Final Thoughts on Blue Gene

- General purpose architecture capable in virtually all areas of computational science
- Though having many special features it presents an essentially standard Linux/PowerPC programming environment
- Pursues performance through high levels of parallelism with good balance between processor and network speed
- Significant impact on HPC – 4 of the top 10 machines are currently Blue Gene systems
- Next Generation Blue Gene system in development – Lawrence Livermore announced acquisition of 20 Petaflop Blue Gene/Q
- Delivers excellent performance per watt, performance per square foot
- High reliability and availability
- Able to run applications consistently and with high performance across the entire system

If you want to know more…

- ALCF web site: www.alcf.anl.gov
  - Information on ALCF system and activities
  - Information on applying for accounts
- Getting Started Guide:
  - https://wiki.alcf.anl.gov/index.php/Quick_Reference_Guide
- ALCF Support Wiki: wiki.alcf.anl.gov
  - Documentation
  - FAQ
- Support email address: support@alcf.anl.gov
  - Any question: account, technical, etc.

Science Projects Using ALCF Resources

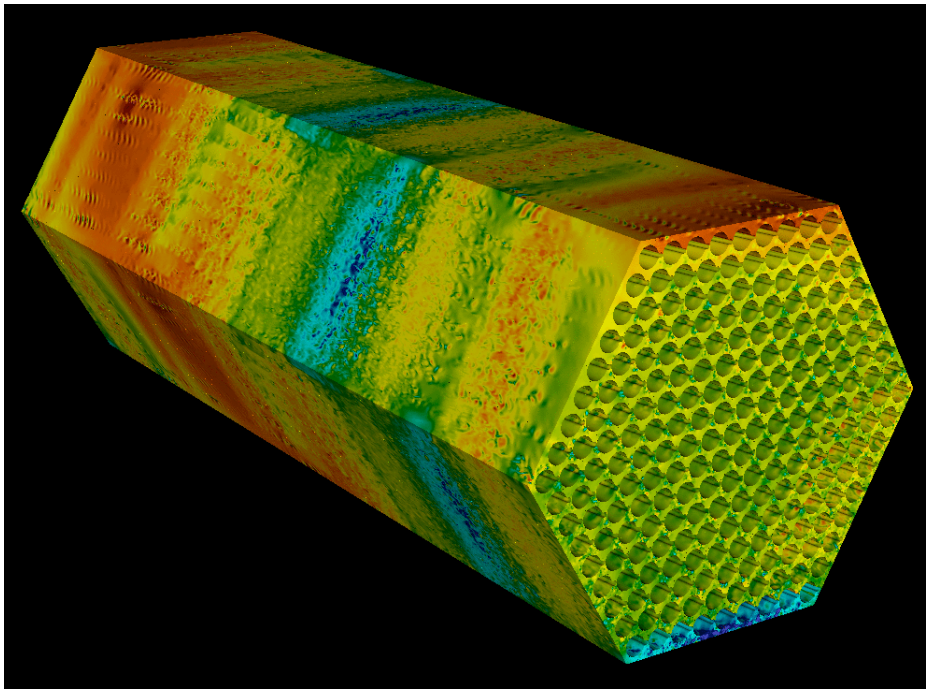
- Parkinson's Disease research
- Electronic/Atomic structure of liquid water
- Prediction of protein structure
- Simulation of aircraft engine combustion
- Heat/fluid dynamics for advanced burner reactor design
- Molecular mechanisms of bubble formation
- Validation of Type Ia supernova models
- Lattice Quantum Chromodynamics
- Designing nanostructured hydrogen storage materials
- Next-generation Community Climate System Model
- Study of turbulence
- Simulation of heart rhythm disorders

Computational Nuclear Engineering

Paul Fischer
ANL

Science

- Improve the Safety, Efficiency and Cost of Liquid-Metal-Cooled Fast Reactors through computation
- Simulation requires full pin assembly, cannot use symmetries.
  - Requires leadership class scale resources



Methods and Challenges

- Developed new algebraic multigrid algorithm to scale to 64K+ CPUs
- Resolved outstanding community question confirming validity of periodic boundary questions
  - Huge benefit to time to solution

217-pin configuration

Breakthrough calculation
•2.95M spectral elements
•1B grid points

Underway ➡

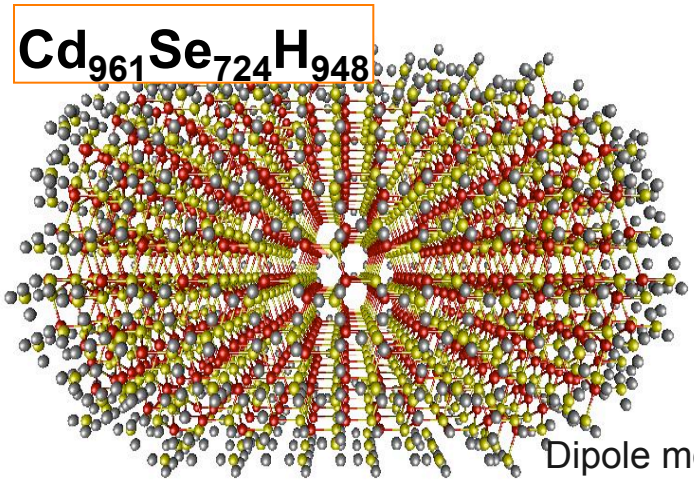
# Pins	Target # CPUs
✔ 7	2K
✔ 19	16K
217	130K

Thousand Atom Nanostructures

Lin-Wang Wang
LBL

Science

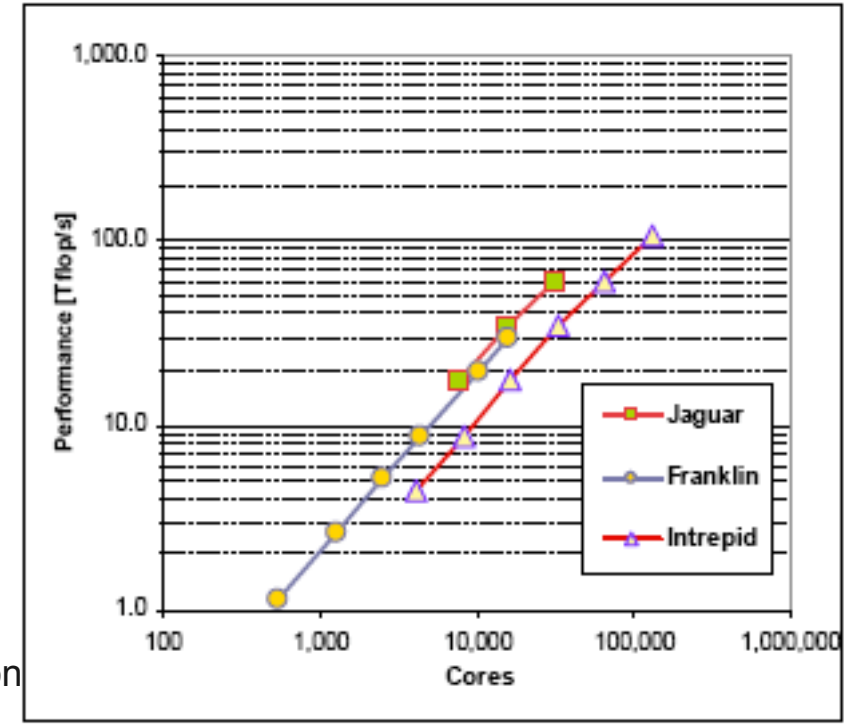
- Design better materials for products including solar cells
- *Ab initio* electronic structure calculations
- Lin-Wang Wang, B. Lee, H. Shan, Z. Zhao, J. Meza, E. Strohmaier, D. Bailey, “Linear Scaling Divide-and-conquer Electronic Structure Calculations for Thousand Atom Nanostructures,” SC08, to appear.



Dipole moment calculated on 2633 atom quantum rod

Methods and Challenges

- Novel divide & conquer approach to solve DFT but reducing $O(n^3)$ to $O(n)$
  - Many months to 30 hours
  - Direct DFT impractica
- Mapping critical
  - Linear scaling to 160K cores and a 10% improvement in per-core performance

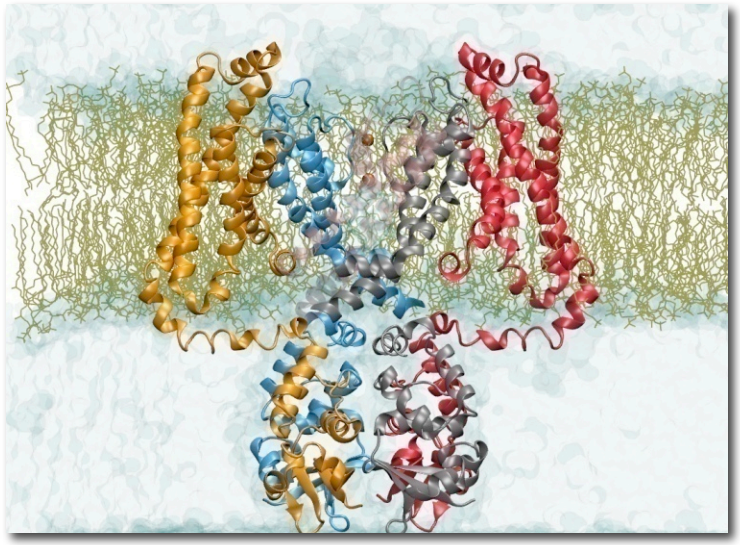


Gating Mechanism of Membrane Proteins

Benoit Roux
ANL, University of Chicago

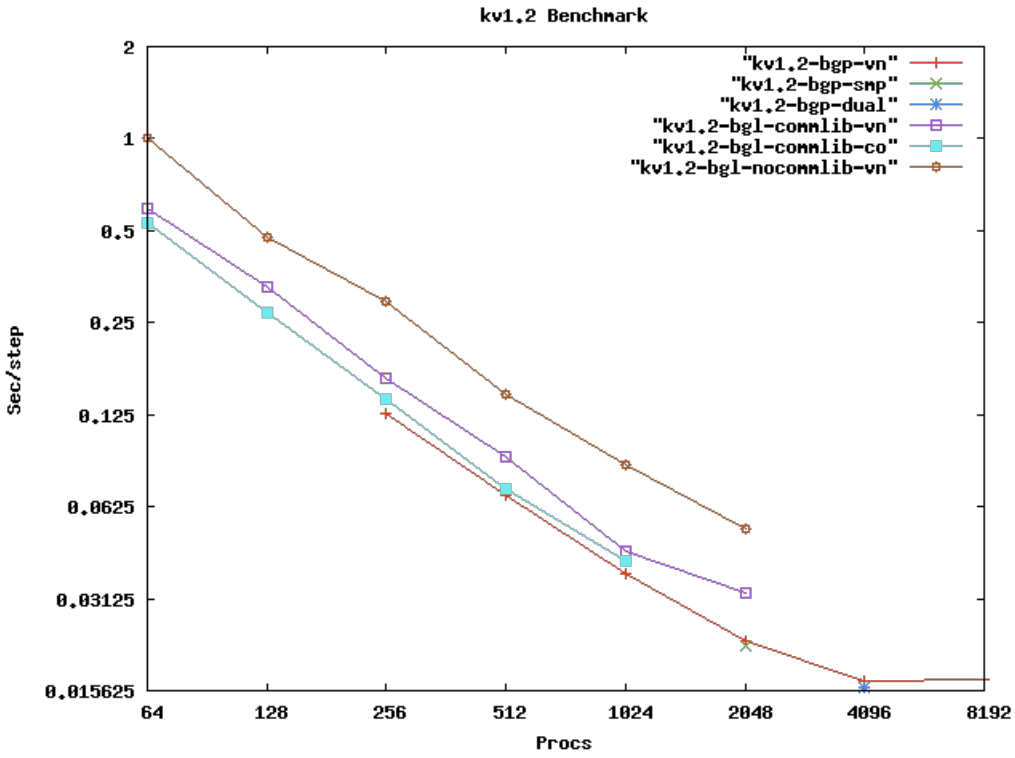
Science

- Understand how proteins work so we can alter them to change their function
- Validated the atomic models of Kv1.2 and first to calculate the gating charge in the two functional states



Methods and Challenges

- NAMD with periodicity and particle-mesh Ewald method



kv1.2 Benchmark (352K atoms)

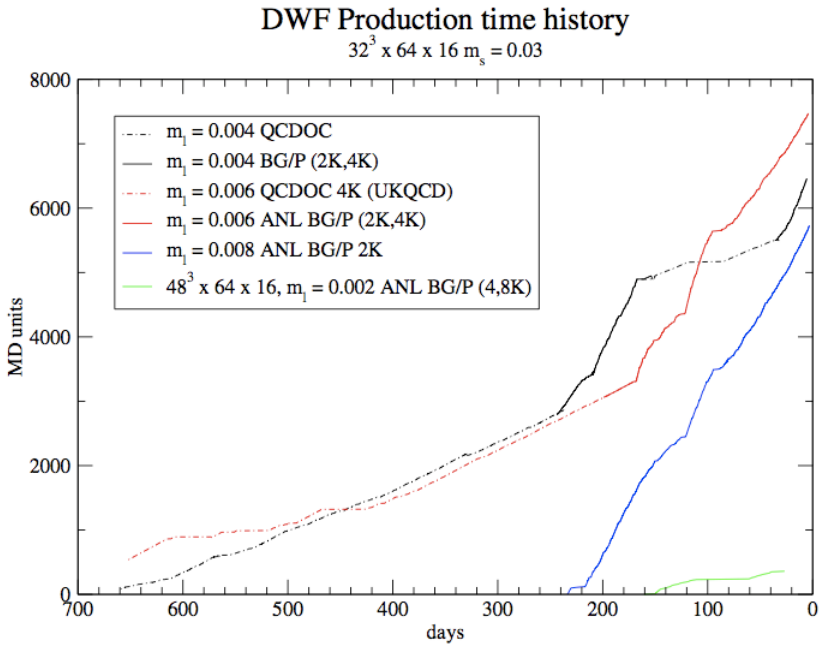
15-20% gain over BG/L customized ("commlib") version.

Lattice QCD

Bob Sugar and US-QCD

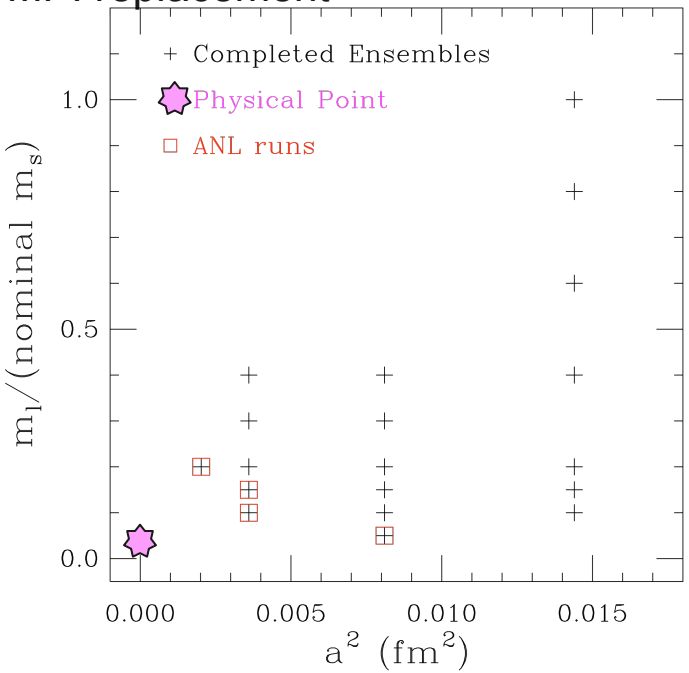
Science

- Addresses fundamental questions in high energy and nuclear physics
- Directly related to major experimental programs
- Determine parameters for Standard Model, including quark mass



Methods and Challenges

- Rational Hybrid Monte Carlo
- For scalability and performance developed
  - QLA : 3x3 matrix linear algebra operations
  - QMP : low-level routines, partial MPI replacement

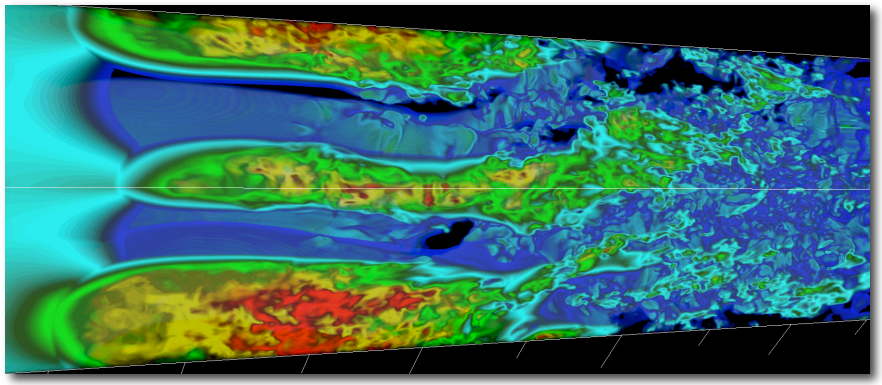


FLASH Project on Intrepid

Don Lamb, University of Chicago

Science

- Answered critical question on critical process in Type Ia supernovae
  - First simulated buoyancy-driven turbulent nuclear combustion in the fully-developed turbulent regime while also simultaneously resolving the Gibson scale
  - Reveals complex structure
  - Requires higher resolution studies



Methods and Challenges

- Operator split, multi-physics
- Block structured adaptive mesh
- Multi-pole and multi-grid gravity solves
- Load balancing with smaller memory footprint per core
- Single CPU performance optimizations

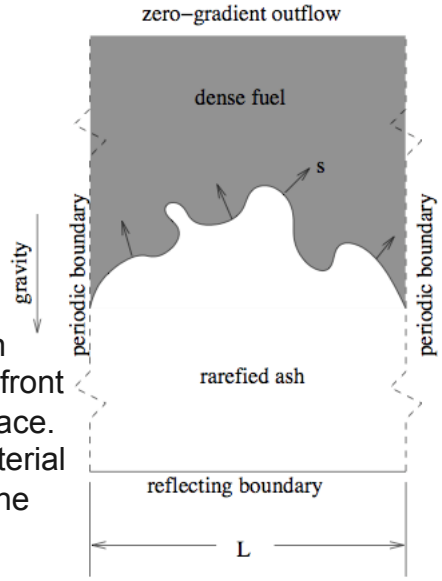


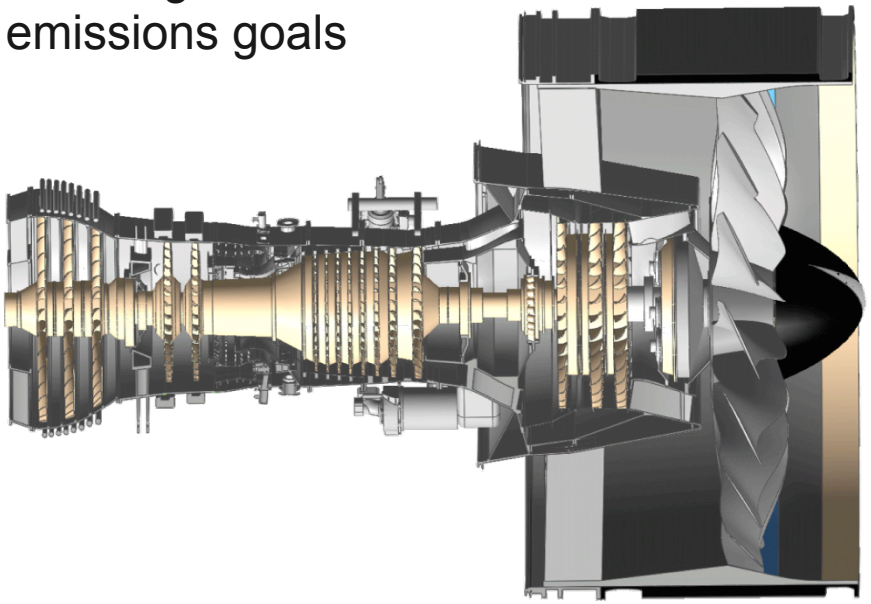
Diagram of simulation configuration. The flame front forms a propagating surface. Lower density burned material rises and accelerates the burning.

Faster Design of Better Jet Engines

Peter Bradly
Pratt&Whitney

Science

- Save cost and time by designing engines through simulation rather than building models
- Technologies from simulations now being applied to next generation high-efficiency low-emission engines
- A key enabler for the depth of understanding needed to needed to meet emissions goals



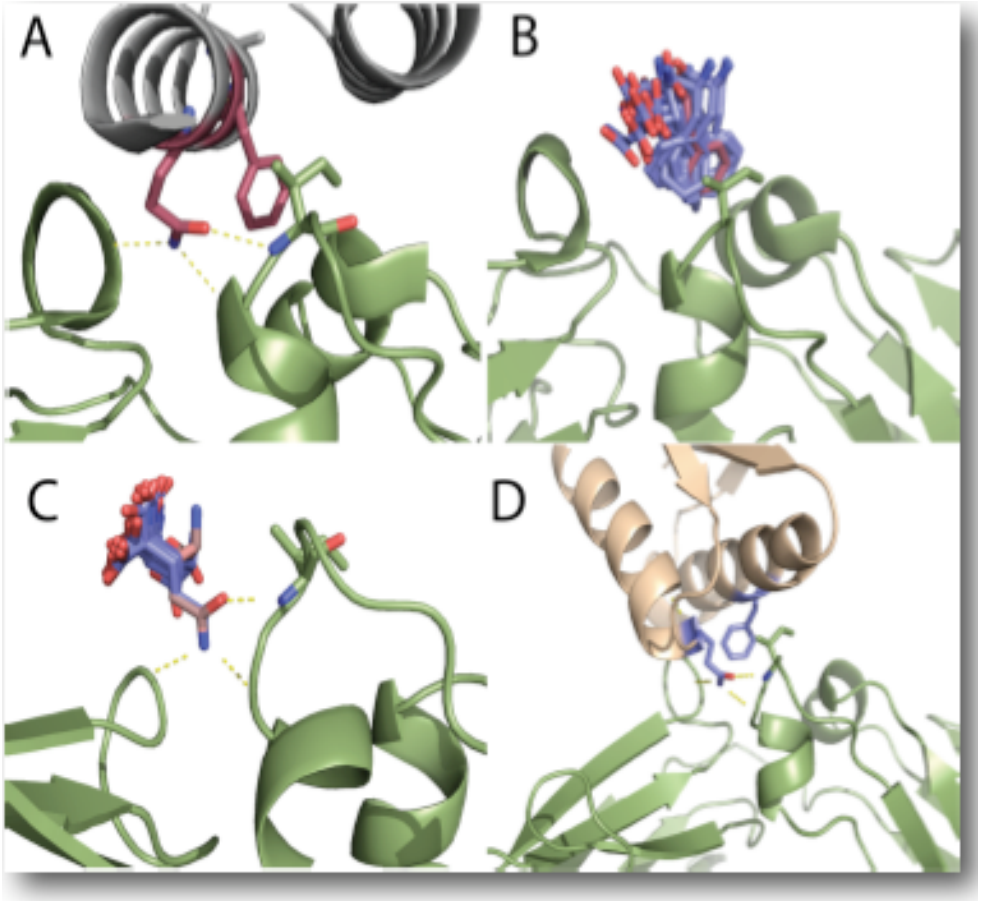
Challenges

- I/O algorithm redesign speeds up simulations by 3x

Computational Protein Structure Prediction and Protein Design

David Baker
University of Washington

- Computationally design protein-based inhibitors towards pathogens like H1N1
- Rapid turn around of huge campaigns on ALCF reinvented how the science is done and enables new research
- Rapidly determine an accurate, high-resolution structure of any protein sequence up to 150-200 residues
- Incorporating sparse experimental NMR data into Rosetta to allow larger proteins



The interfaces of protein-protein complexes often exhibit a handful of key interactions, termed hot-spots. At right, the original protein (A) is replaced by an easy-to-manufacture custom scaffold (D)

nking of all CASP8 groups

Climate Simulations on Blue Gene/P

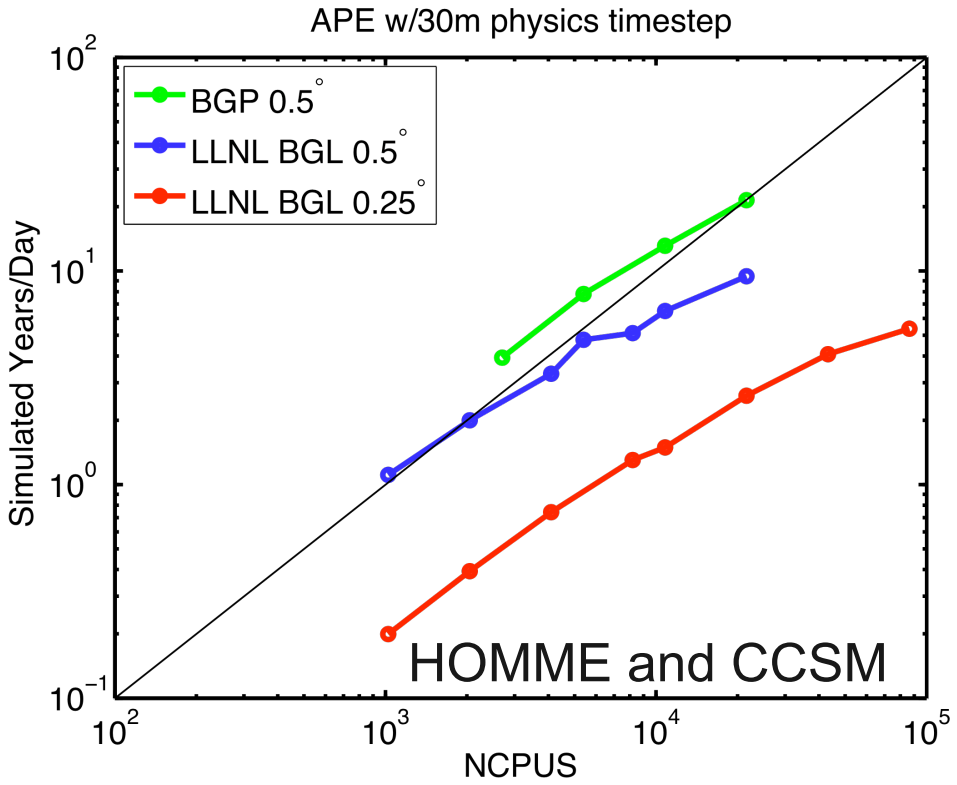
Warren Washington
NCAR

Science

- CCSM is a climate simulation code used by the DOE and NSF climate change experiments
- Moving from CCSM3.5 to CCSM4
- Aqua planet experiment runs
  - Full physics, no land model
  - BG/P is 2x faster on 20,000 cores than BG/L
  - BG/P, at 0.50 degree, achieves an integration rate of over 20 Simulated Years Per Day (SYPD)

Methods and Challenges

- Complicated compilation, parallel model and memory footprint for the Blue Gene/P architecture



Identifying Potential Drug Targets

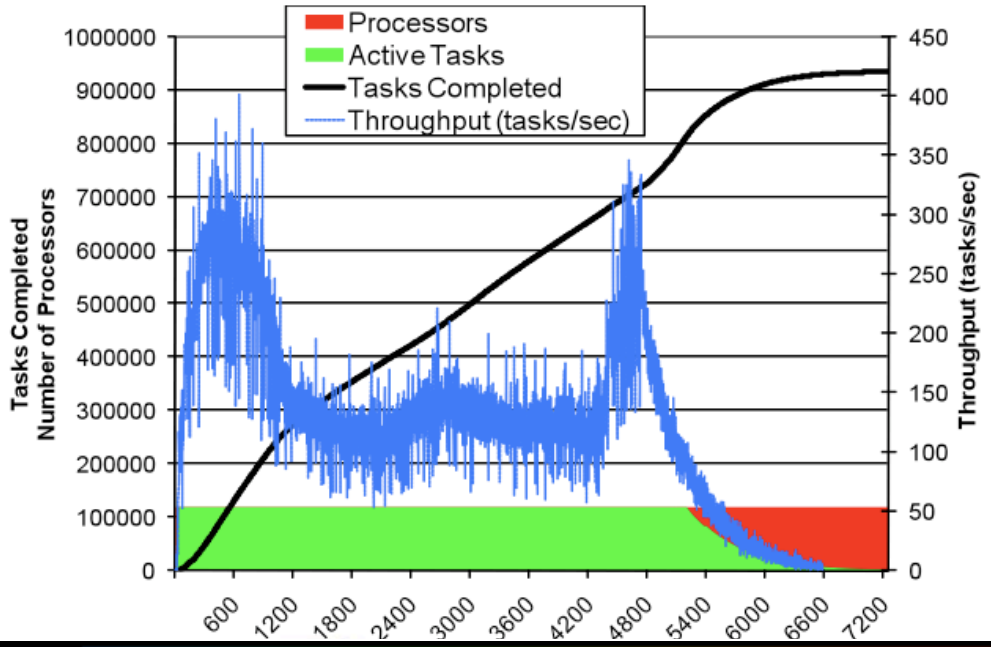
Michael Wilde
ANL

Science

- Reduce dead ends in antibiotics and anticancer drugs with DOCK5 and DOCK6
  - 9 enzymatic proteins in core metabolism of bacteria and humans screened against 15,351 natural compounds and existing drugs
  - Study correlations and re-prioritize proteins for further study
- Able to complete 21.43 CPU-years of analysis in 2.01 wall-hours

Methods and Challenges

- Port of framework, Falkon, to manage run
- Falkon requires non-standard BG/P kernel (ZeptoOS)
- Huge demand on I/O system as each core is controlling multiple files
- 118,000 cores were used running nearly one million tasks

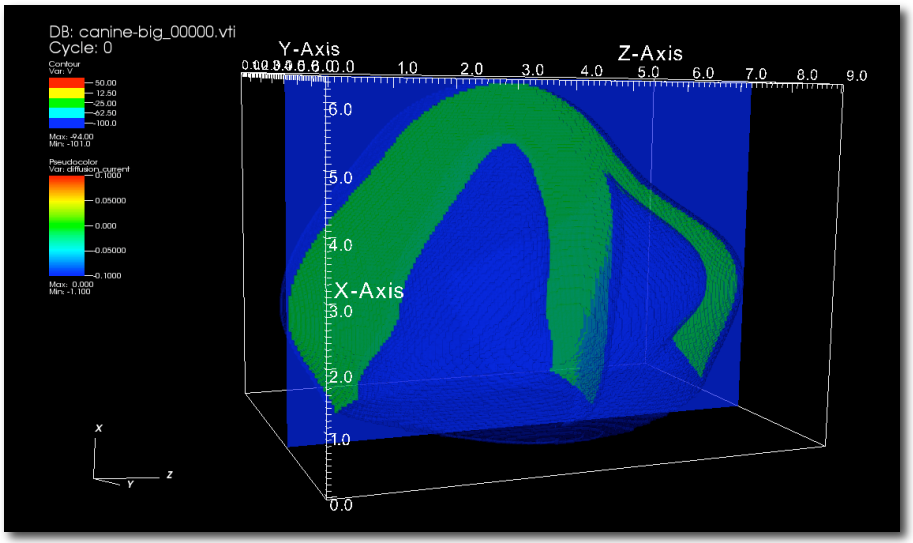


Cardiac Rhythm Disorders

Jeffrey Fox
Gene Network Science

Science

- Cardiac rhythm disorders are a leading cause of death
- Simulations of the canine heart at 250um made possible by performance improvements
- New hypotheses proposed for the wave break mechanism in the heart



Methods and Challenges

- Finite difference with homogeneous Neumann boundary conditions
- Exact voltage conservation
- Custom mapping for load-balancing and communication trade offs
- I/O and data analysis methods were unable to scale
  - 300% I/O improvement
  - 600x speedup from Restructure of data file and analysis algorithm

Insight into Parkinson's Disease

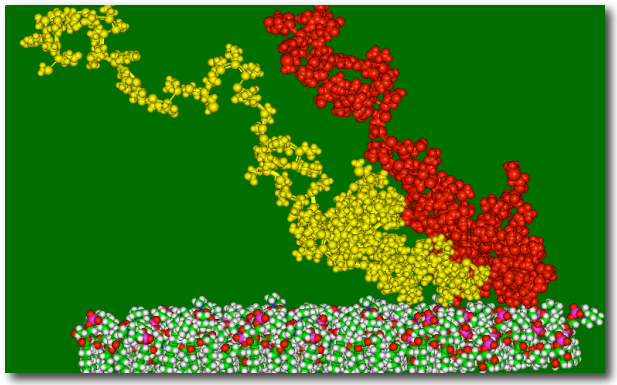
Igor Tsingelny
University of California, San Diego

Science

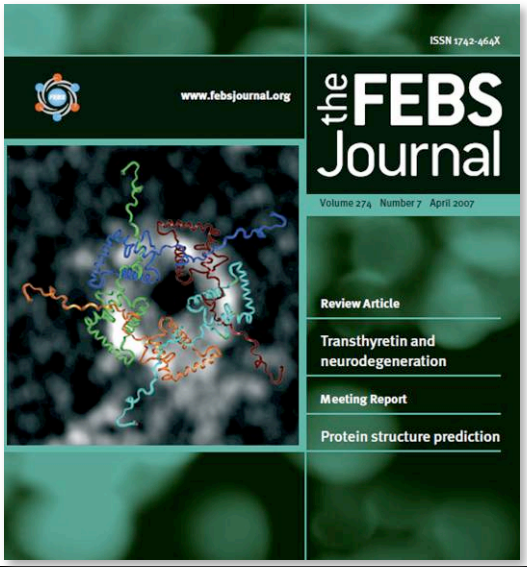
- Parkinson's Disease is the 2nd most common adult neurological disease
- Increased aggregation of *alpha-synuclein* protein is thought to lead to harmful pore-like structures in human membranes
- UCSD - SDSC team used molecular modeling and molecular dynamics simulations in combination with biochemical and ultrastructural analysis to show that *alpha-synuclein* can lead to the formation of pore-like structures on membranes

Methods and Challenges

- Using NAMD and MAPAS on Blue Gene at ALCF and SDSC



alpha-synuclein forming a dimer (above) and a completed pentamer (below) attached to a membrane



Understanding Water Structure

Guilia Galli
University of California, Davis

Science

- Structure of water in all phases is critical to *many* research fields
- Water confined at nanometer scale is less well understood

- Completed ab-initio calculations of Infrared Spectra of confined water
- Completed calculations of the phase diagram of water under pressure

Methods and Challenges

- *Ab-initio* simulations within Density Functional Theory
- Improve quantum theory coupling current code with quantum monte carlo code
- Mapping of processes to physical machine critical to performance