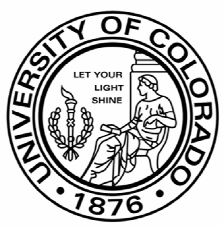
Blue Gene Experience at the National Center for Atmospheric Research

October 4, 2006

Theron Voran
voran@ucar.edu

Computer Science Section
National Center for Atmospheric Research

Department of Computer Science
University of Colorado at Boulder



NCAR

Why Blue Gene?

- Extreme scalability, balanced architecture, simple design
- Efficient energy usage
- A change from IBM Power systems at NCAR
- But familiar
  - Programming model
  - Chip (similar to Power4)
  - Linux on front-end and IO nodes
- Interesting research platform

Outline

- System Overview
- Applications
- In the Classroom
- Scheduler Development
- TeraGrid Integration
- Other Current Research Activities

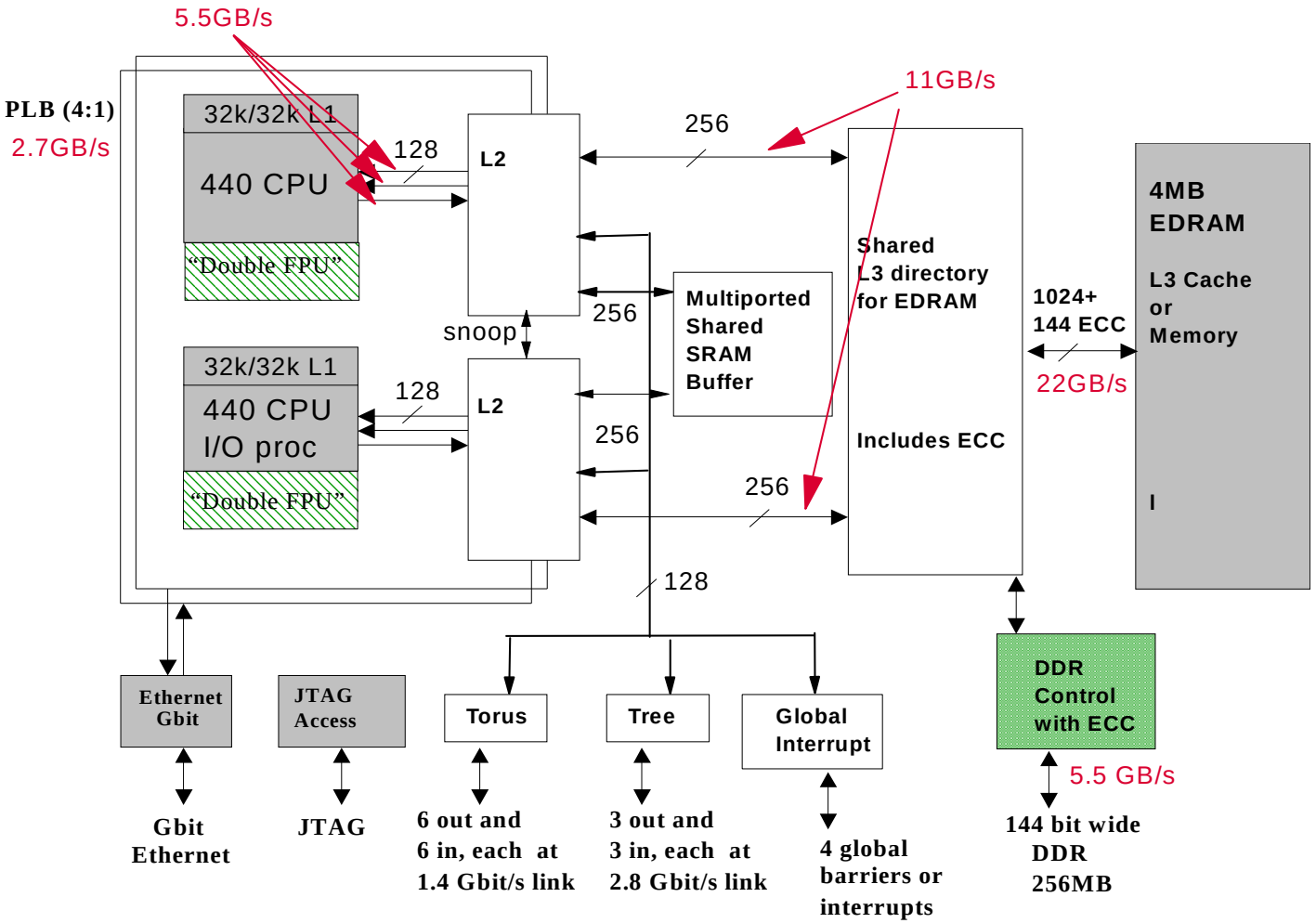
Frost Fun Facts

Henry Tufo and Rich Loft, with Frost

- Collaborative effort
  - Univ of Colorado at Boulder (CU)
  - NCAR
  - Univ of Colorado at Denver
- Debuted in June 2005, tied for 58th place on Top500
  - 5.73 Tflops peak – 4.71 sustained
- 25KW loaded power usage
- 4 front-ends, 1 service node
- 6TB usable storage

- Why is it leaning?

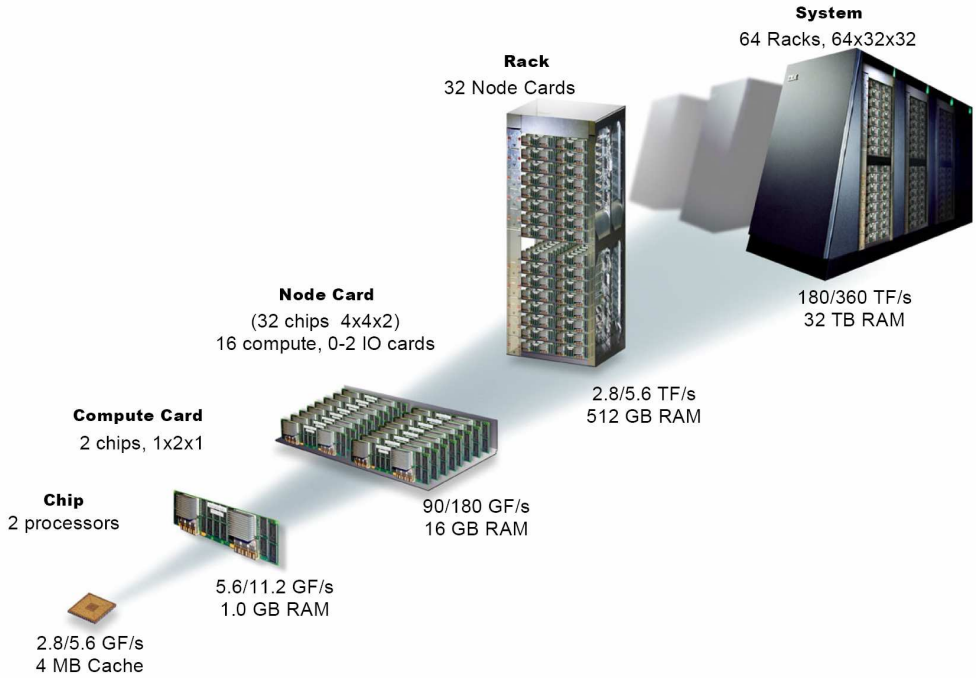
System Internals



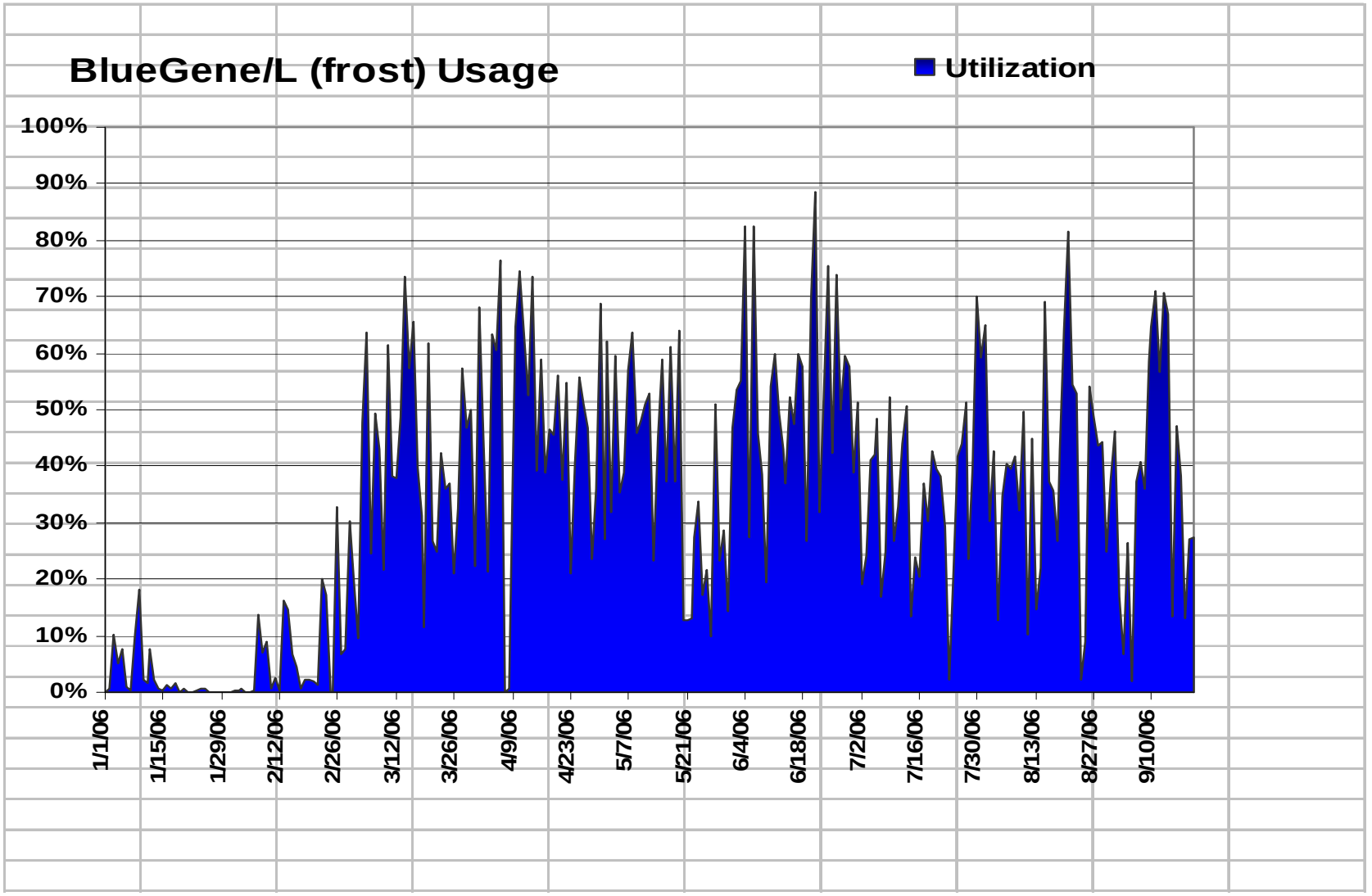
Blue Gene/L system on-a-chip

More Details

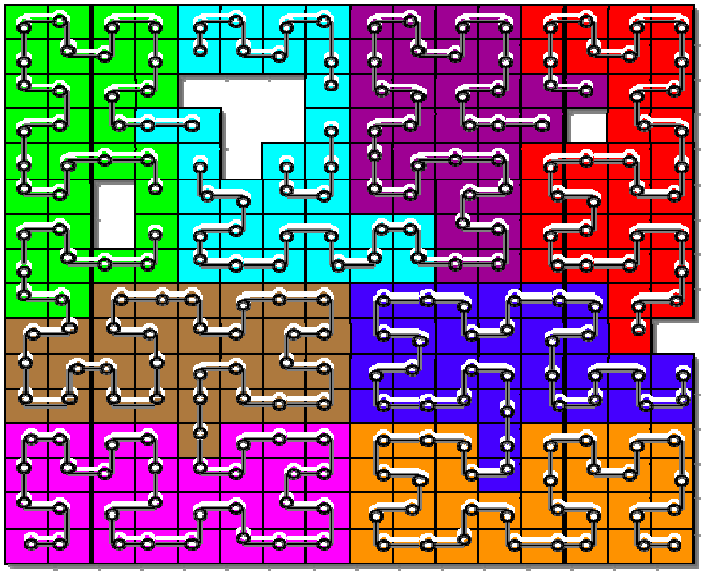
- Chips
  - PPC440 @700MHZ, 2 cores per node
  - 512 MB memory per node
  - Coprocessor vs Virtual Node
  - 1:32 IO to Compute ratio
- Interconnects
  - 3D Torus (154 MB/s one direction)
  - Tree (354 MB/s)
  - Global Interrupt
  - GigE
  - JTAG/IDO
- Storage
  - 4 Power5 systems as GPFS cluster
  - NFS export to BGL IO nodes



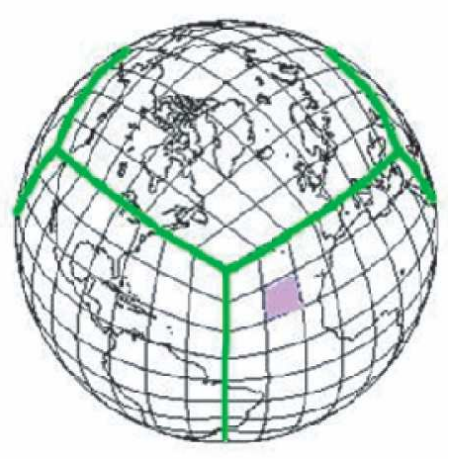
Frost Utilization



HOMME



- High Order Method Modeling Environment
- Spectral element dynamical core
- Proved scalable on other platforms
- Cubed-sphere topology
- Space-filling curves



HOMME Performance

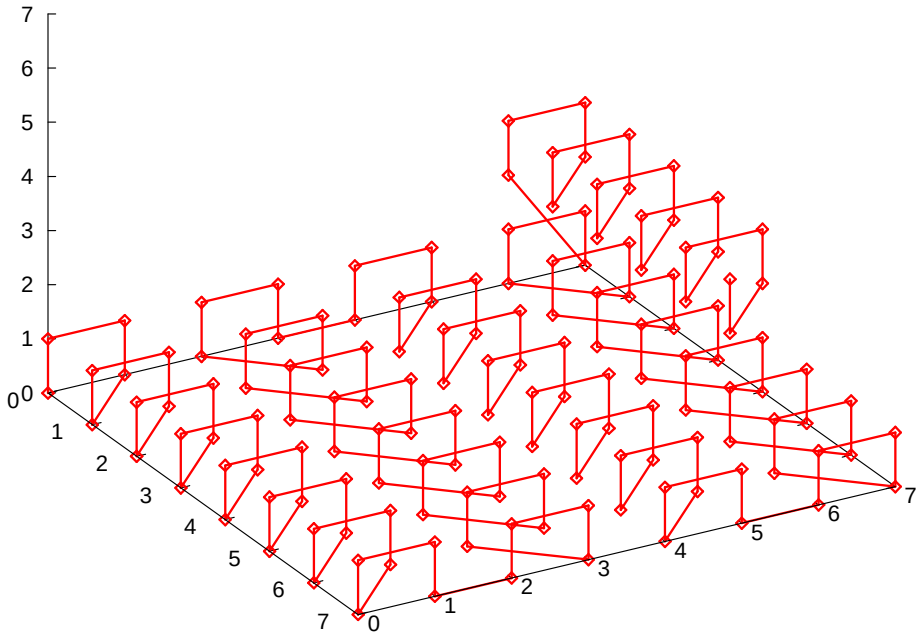
- Ported in 2004 on BG/L prototype at TJ Watson, with eventual goal of Gordon Bell submission in 2005

Serial and parallel obstacles:

- SIMD instructions
- Eager vs Adaptive routing
- Mapping strategies

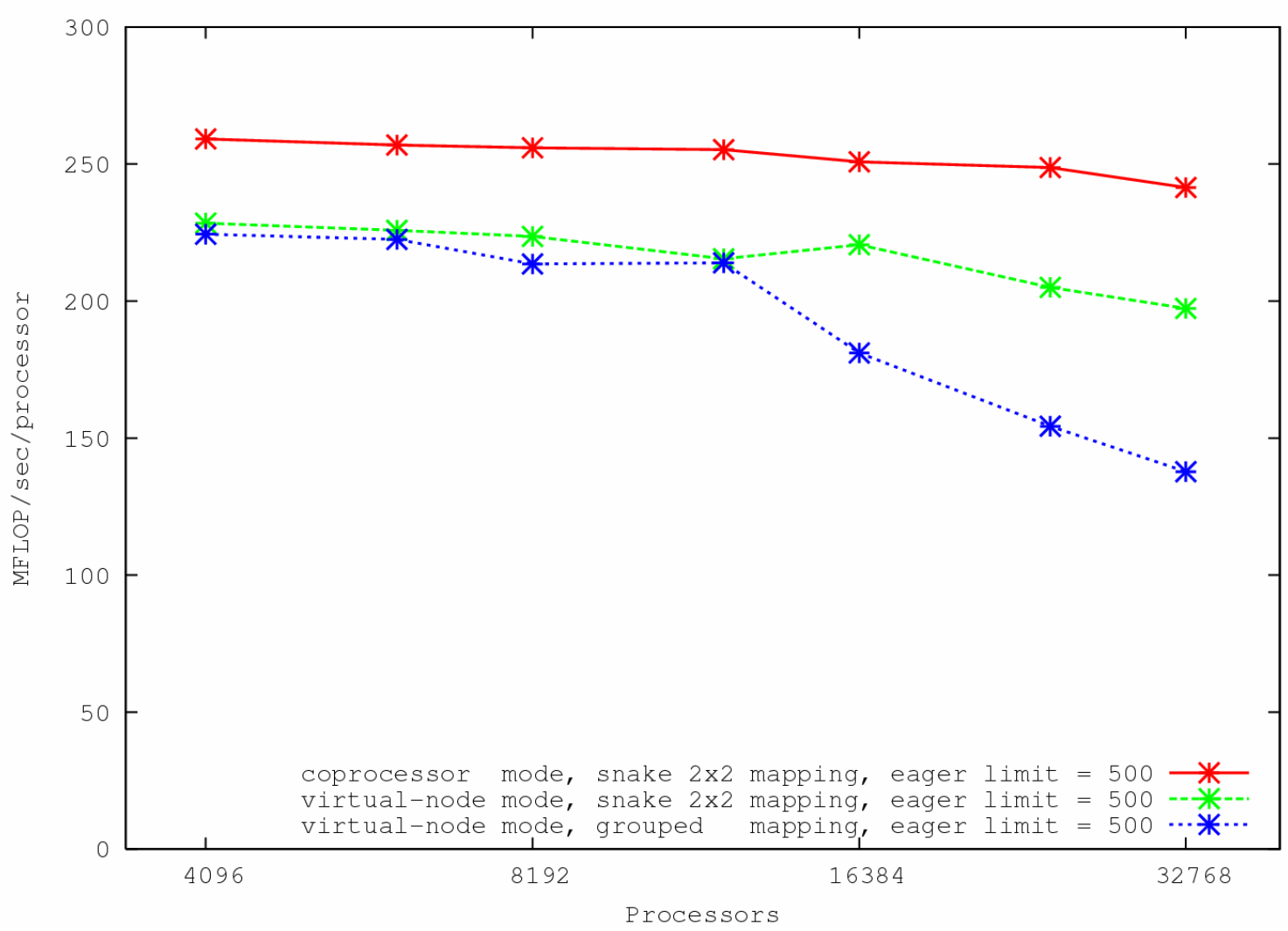
Result:

- Good scalability out to 32,768 processors (3 elements per processor)



Snake mapping on 8x8x8 3D torus

HOMME Scalability on 32 Racks

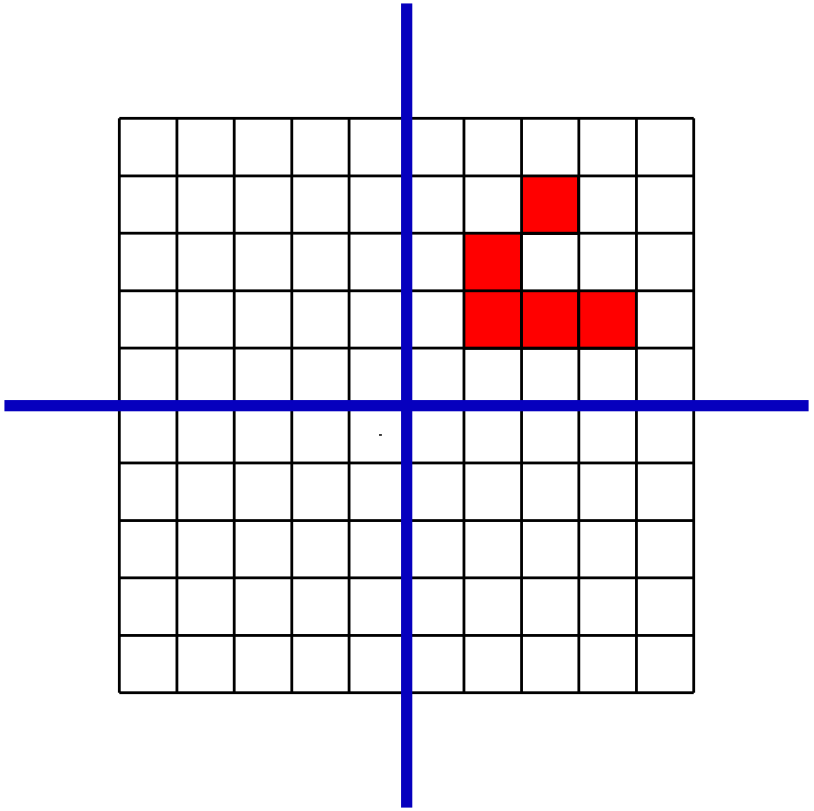


Other Applications

- Popular codes on Frost
  - WRF
  - CAM, POP, CICE
  - MPIKAIA
  - EULAG
  - BOB
  - PETSc
- Used as a scalability test bed, in preparation for runs on 20-rack BG/W system

Classroom Access

- Henry Tufo's 'High Performance Scientific Computing' course at University of Colorado
- Let students loose on 2048 processors
  - Thinking BIG
  - Throughput and latency studies
  - Scalability tests - Conway's Game of Life
  - Final projects
- Feedback from 'novice' HPC users



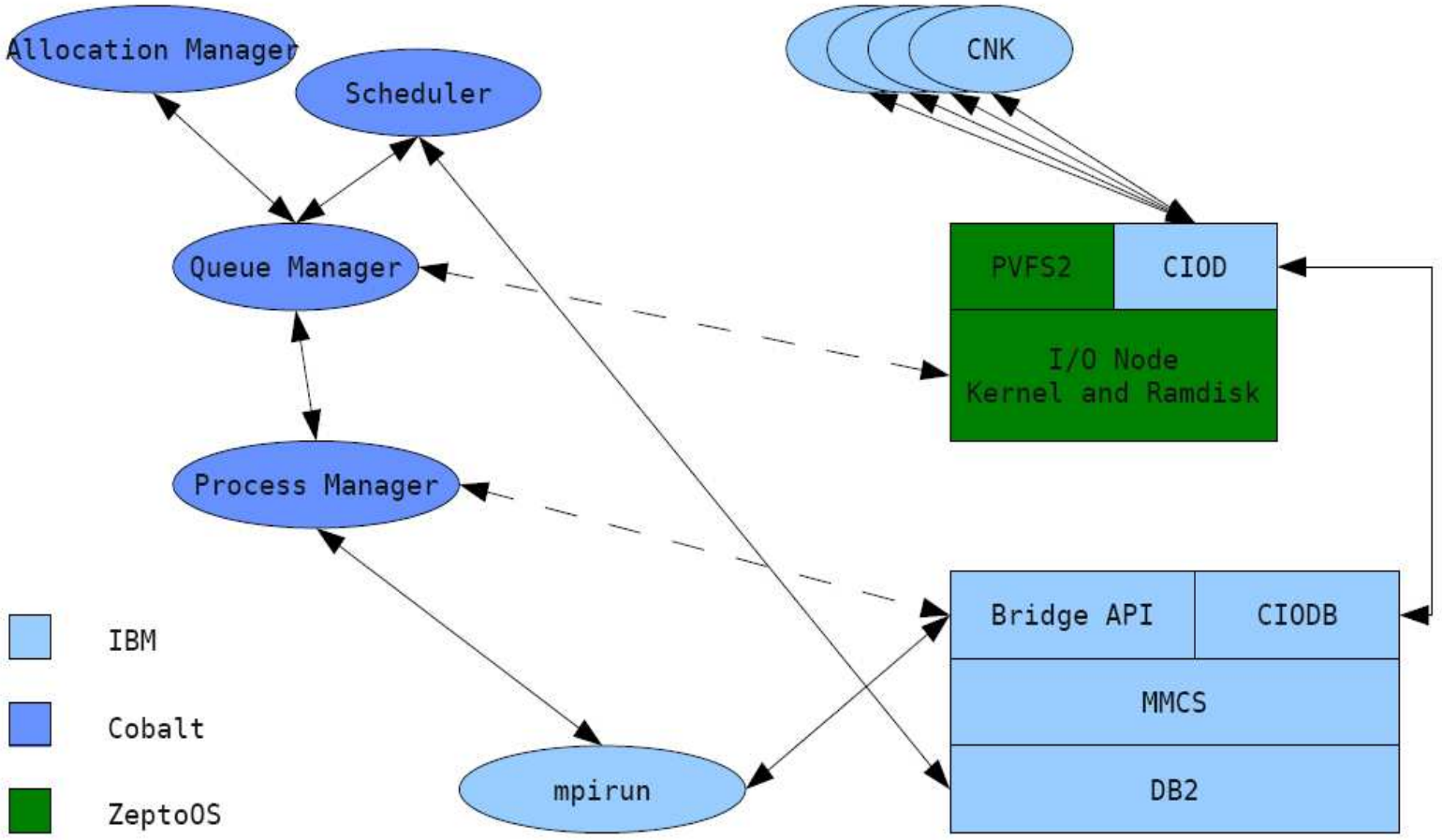
Cobalt

- Component-Based Lightweight Toolkit
- Open source resource manager and scheduler
- Developed by ANL along with NCAR/CU

- Component Architecture
  - Communication via XML-RPC
  - Process manager, queue manager, scheduler
- ~3000 lines of python code
- Manages traditional clusters also

http://www.mcs.anl.gov/cobalt

Cobalt Architecture

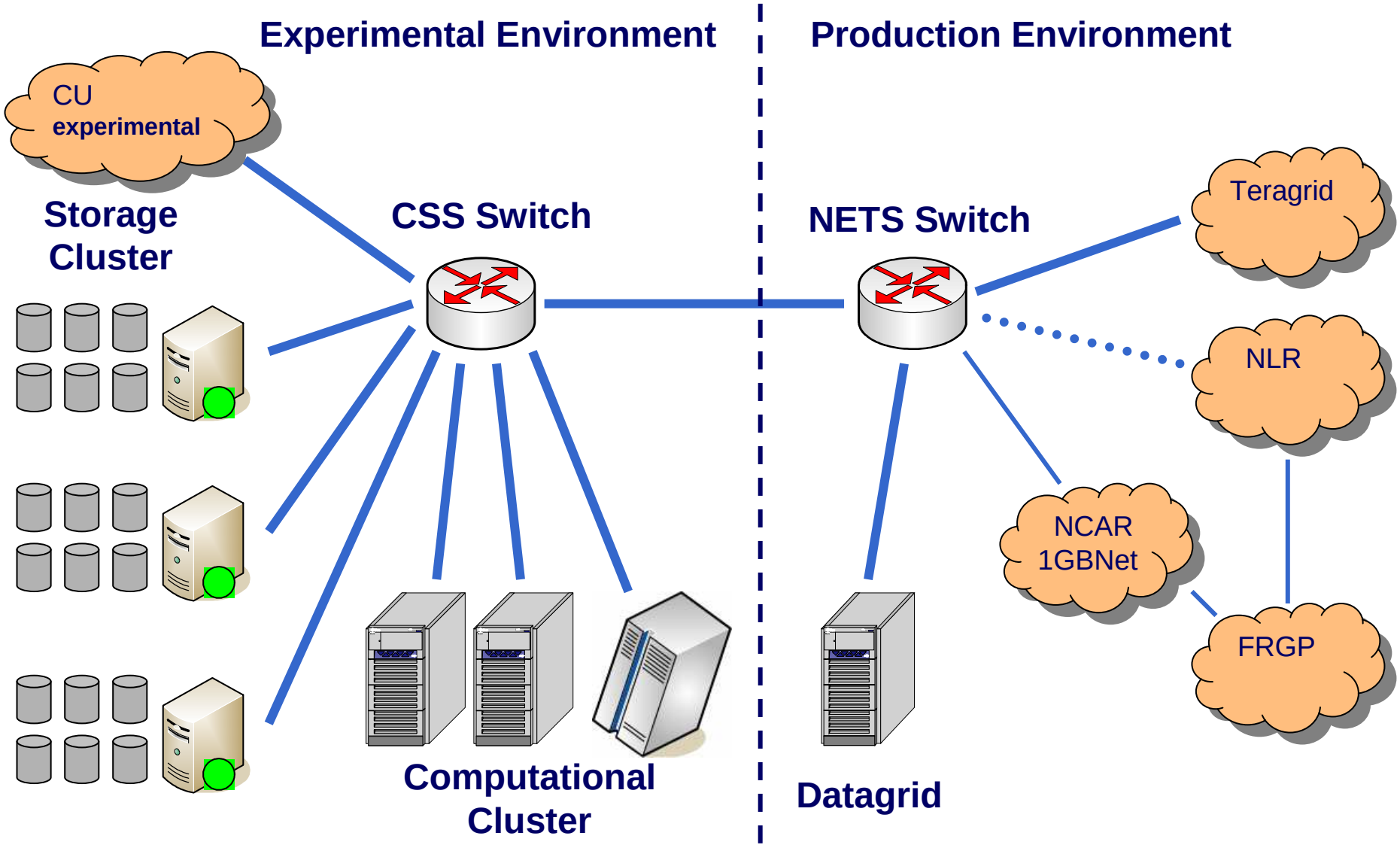


Cobalt Development Areas

- Scheduler improvements
  - Efficient packing
  - Multi-rack challenges
  - Simulation ability
  - Tunable scheduling parameters
- Visualization
  - Aid in scheduler development
  - Give users (and admins) better understanding of machine allocation
- Accounting / project management and logging
- Blue Gene/P
- TeraGrid integration

NCAR joins the TeraGrid, June 2006

TeraGrid Testbed



TeraGrid Activities

- Grid-enabling Frost
  - Common TeraGrid Software Stack (CTSS)
  - Grid Resource Allocation Manager (GRAM) and Cobalt interoperability
  - Security infrastructure
- Storage Cluster
  - 16 OSTs, 50-100 TB usable storage
  - 10G connectivity
  - GPFS-WAN
  - Lustre-WAN

Other Current Research Activities

- Scalability of CCSM components
  - POP
  - CICE
- Scalable solver experiments
- Efficient communication mapping
  - Coupled climate models
  - Petascale parallelism
- Meta-scheduling
  - Across sites
  - Cobalt vs other schedulers
- Storage
  - PVFS2 + ZeptoOS
  - Lustre

Frost has been a success as a …

- Research experiment
  - Utilization rates
- Educational tool
  - Classroom
  - Fertile ground for grad students
- Development platform
  - Petascale problems
  - Systems work

Questions?

voran@ucar.edu
https://wiki.cs.colorado.edu/BlueGeneWiki