Advancing Atmospheric Science with Blue Gene/L

Dr. Richard Loft
Computational Science Section
Scientific Computing Division
National Center for Atmospheric Research
Boulder, CO USA

NCAR

Summary of remarks…

- Blue Gene/L seems to us to be an important architecture for many reasons.
  - Power/space & “fuel” efficiency
  - $/Tflops sustained
  - Fast reduction network
- 80 km, 20 level explicit HOMME model
  - We see 587 Gflops on 1944 nodes of BG/L.
  - 13.7 simulated years/day (useful climate rate)
- 1/10 degree “eddy permitting” POP ocean model would be the next interesting thing to look at. (barotropic CG needs fast global sums)

NCAR

BlueGene/L Architecture

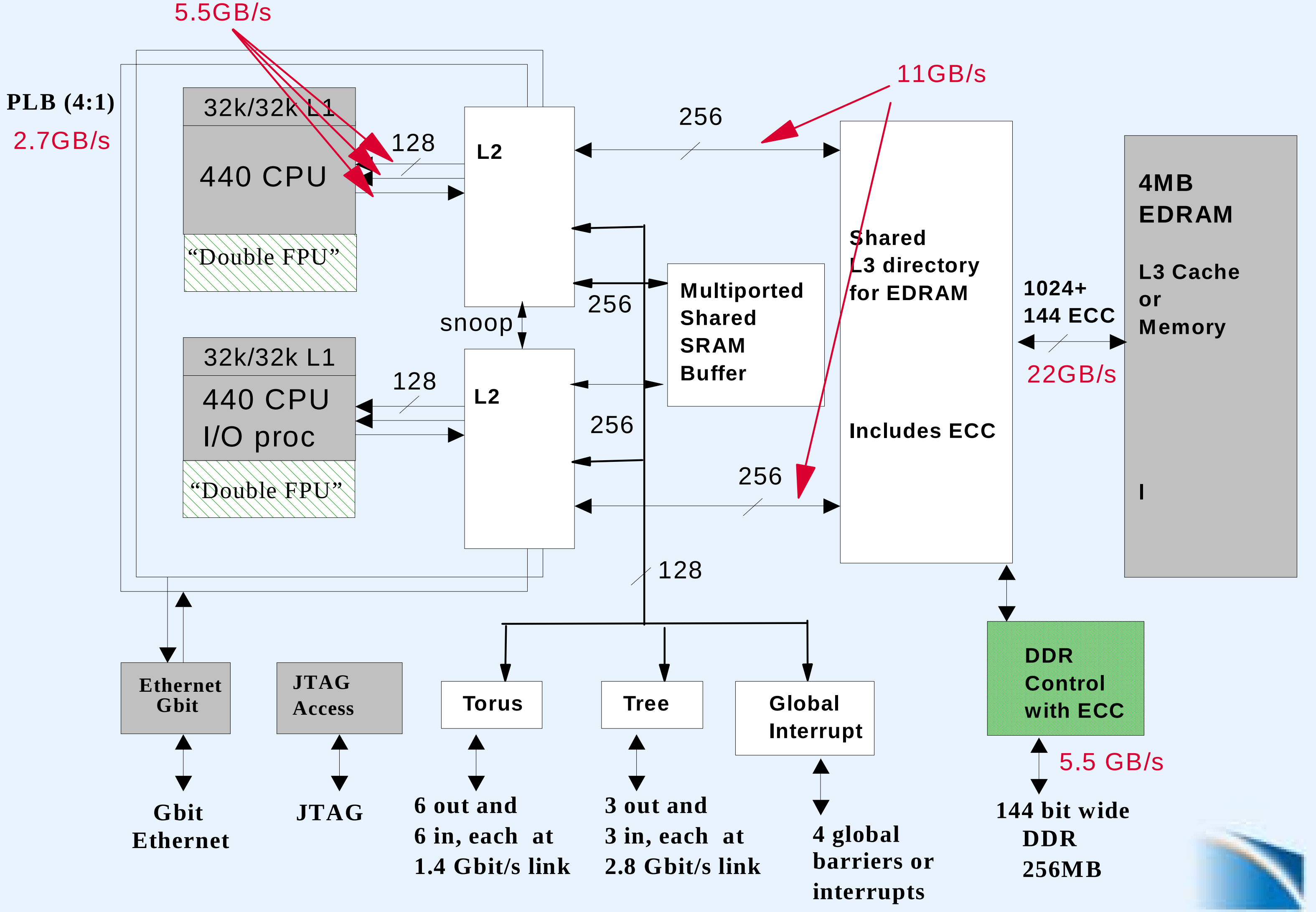
NCAR

Why Blue Gene/L is Attractive

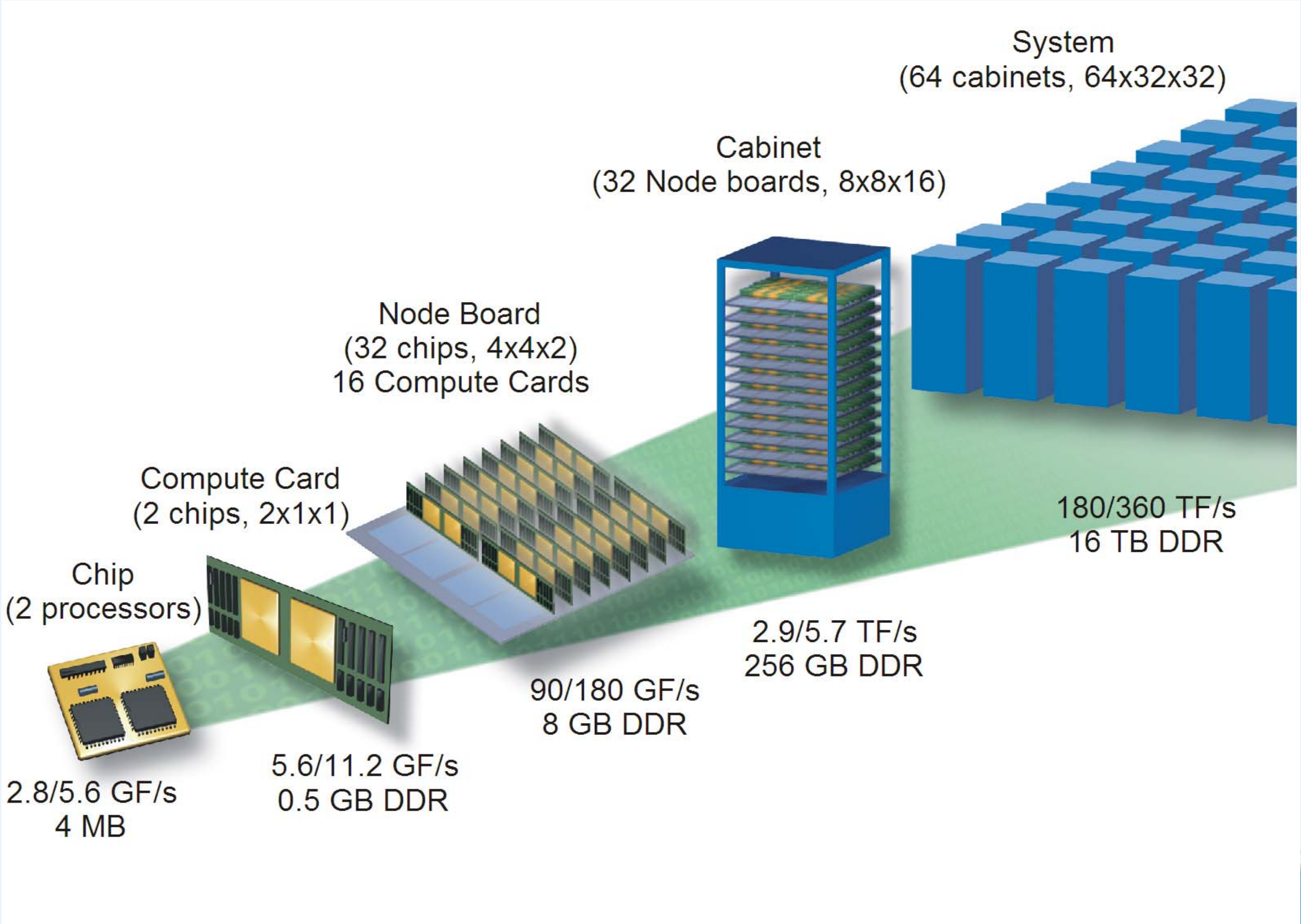
- **Pros**
  - Achieves high packaging density.
  - Lower power density.
  - Dedicated reduction network.
  - Puts network interfaces on chip.
  - Low cost per sustained FLOPs.
- **Questions**
  - High reliability?
  - Applications for 100k processors?
  - System robustness: I/O, scheduling flexibility.

NCAR

BlueGene/L ASIC

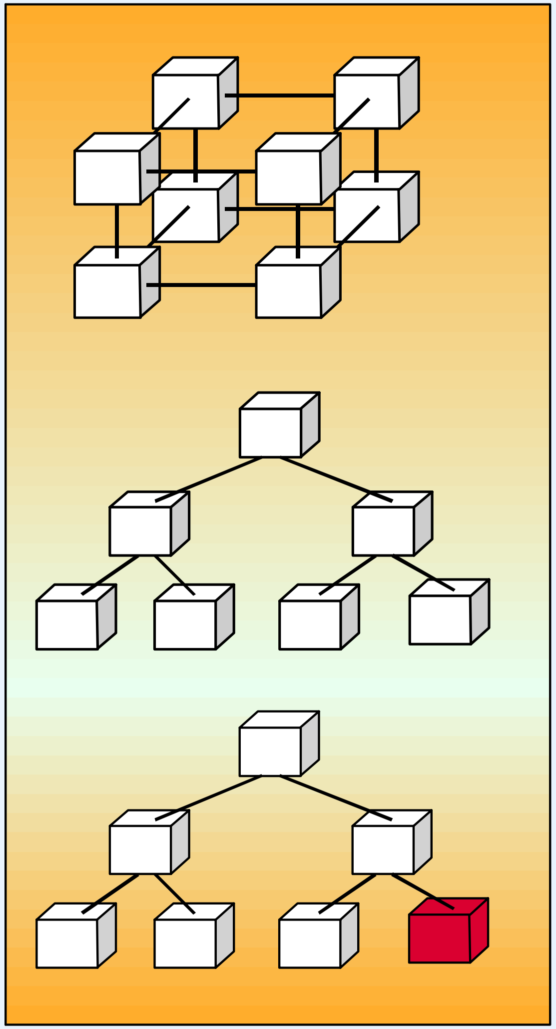


The Blue Gene/L Architecture



NCAR

BlueGene/L Has Five Networks



3-Dimensional Torus

– interconnects all compute nodes (65,536)

Global Tree

– point-to-point, one-to-all broadcast, reduction functionality

Global Interrupts

– AND/OR operations for global barriers
– 1.5 microseconds latency (64K system)

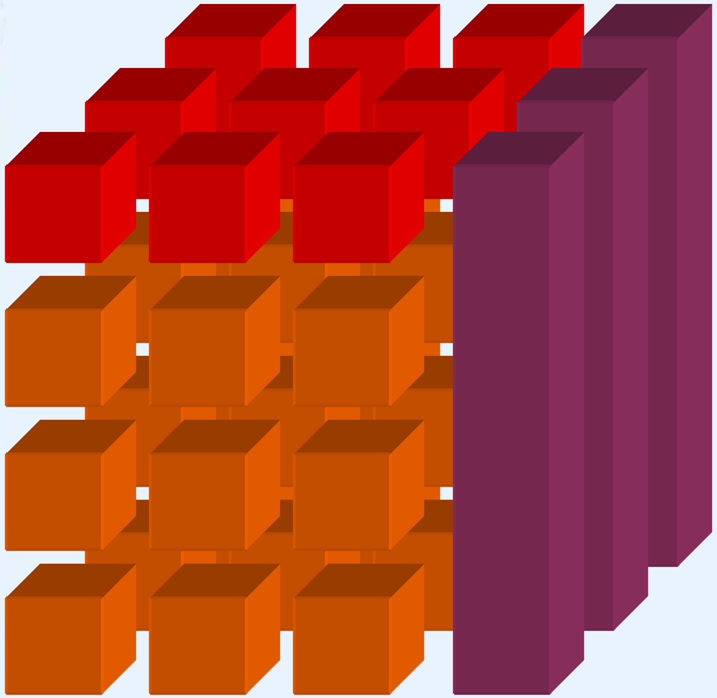
Ethernet

– incorporated into every node ASIC
– active in the I/O nodes (1:64 in LLNL configuration)
  - 1K 1Gbit links
– all external comm. (file I/O, control, user interaction, etc.)

JTAG (Control)

NCAR

BlueGene/L System Software Architecture



- User applications execute exclusively in the **compute nodes**
  - avoid asynchronous events (e.g., daemons, interrupts)
- The outside world interacts only with the **I/O nodes**, an offload engine
  - standard solution: Linux
- Machine monitoring and control also offloaded to **service nodes**: large SP system or Linux cluster.

Blue Gene/L system overview

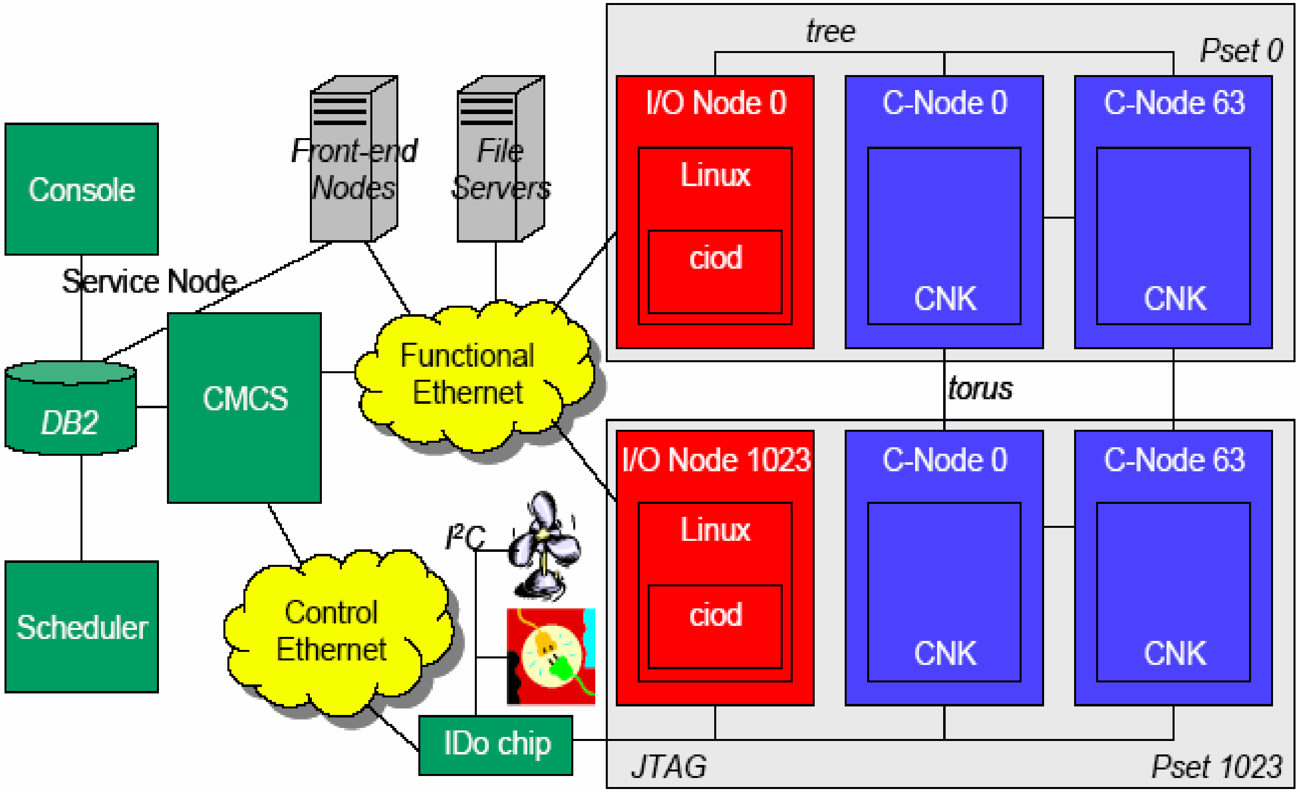


Figure 1: Overview of a complete system with BlueGene/L as the computational core.

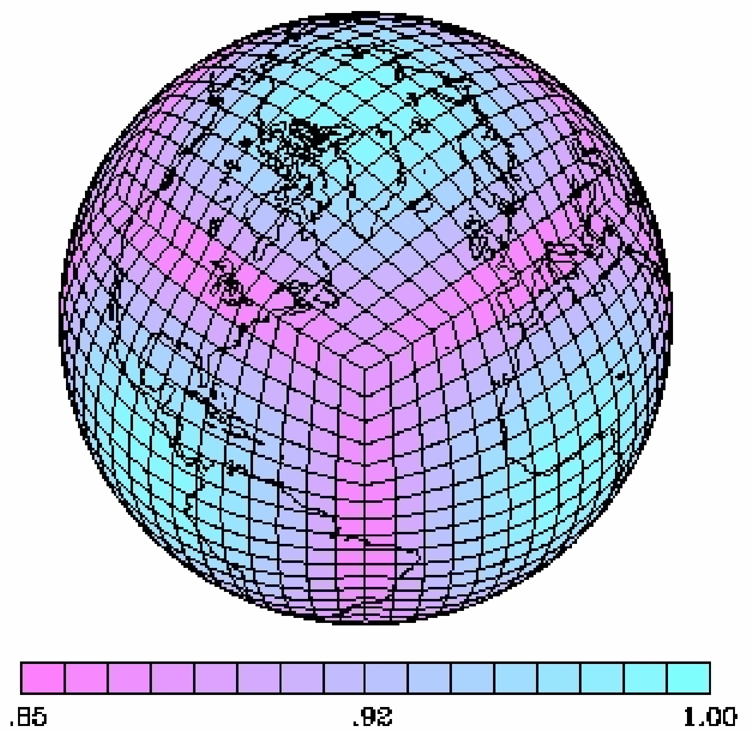
Scalable atmospheric dynamics…

Description of HOMME Atmospheric Dynamical Core

- High Order Method Modeling Environment (HOMME)
  - Solving the moist "primitive" equations on cube-sphere
  - CAM finite difference vertical discretization
  - Held-Suarez simplified test physics package
- Algorithmic Advantages of High Order Methods
  - h-p finite element method on quadrilaterals (Ne x N)
  - Exponential convergence in p (N)
- Computational Advantages of High Order Methods
  - Naturally cache-blocked N x N computations
  - Nearest-neighbor communication between elements
  - Well suited to parallel µprocessor systems
- Semi-Implicit Time Step to eliminate fast gravity wave modes
  - Crank-Nicholson time integration scheme
  - Use iterative CG solver for the resultant elliptic equation
  - Simple metric term/mass matrix preconditioner.
  - Acceleration 2.5x over explicit integration

NCAR

The Cube-Sphere

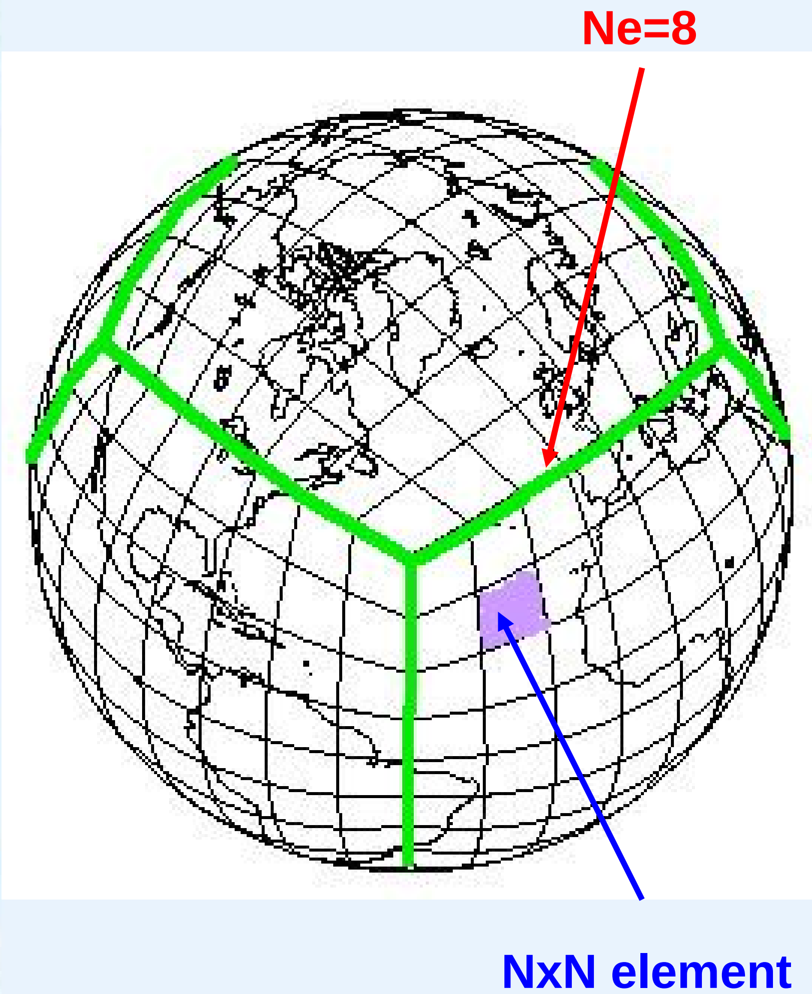


Ne=16 Degree of on-uniformity

- Equal angular grid
  - Rancic et al (1996)
- Avoids pole problems
- Quasi-uniform
- Curvilinear coordinates: metric terms
- Ne=16 shown above

NCAR

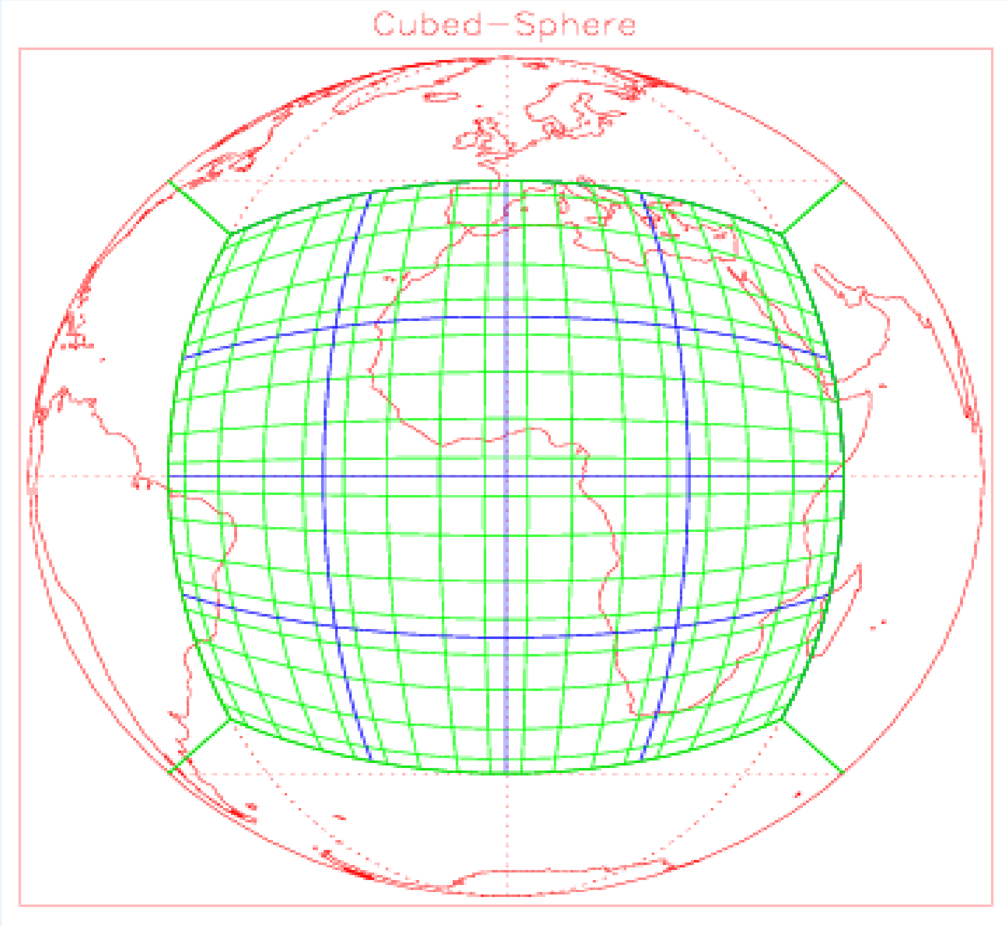
The Spectral Element Computational Mesh: the "Cube-Sphere"



- Spectral Elements:
  - A quadrilateral "patch" of gridpoints **N x N**
  - Gauss-Lobbattto Grid
  - N=8 is optimal (Taylor)
- Cube
  - **Ne** = Elements on an edge
  - 6*Ne*Ne elements total
- Cube Partitioning
  - Metis
  - Space filling curve partitioning algorithm
- Ne=8 shown ~180 km

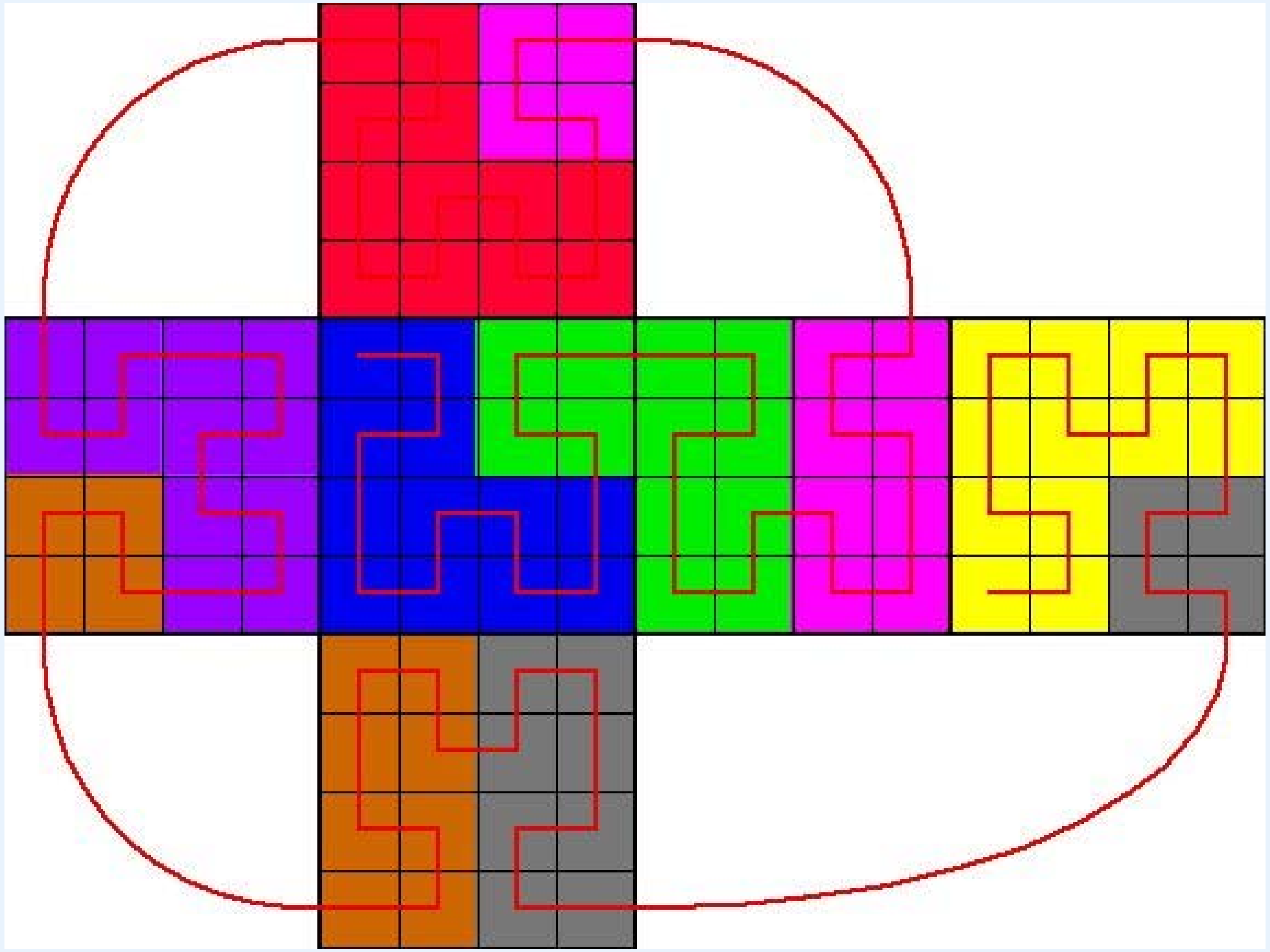
NCAR

6th Order Spectral Elements on the Ne=4 Cube Sphere



NCAR

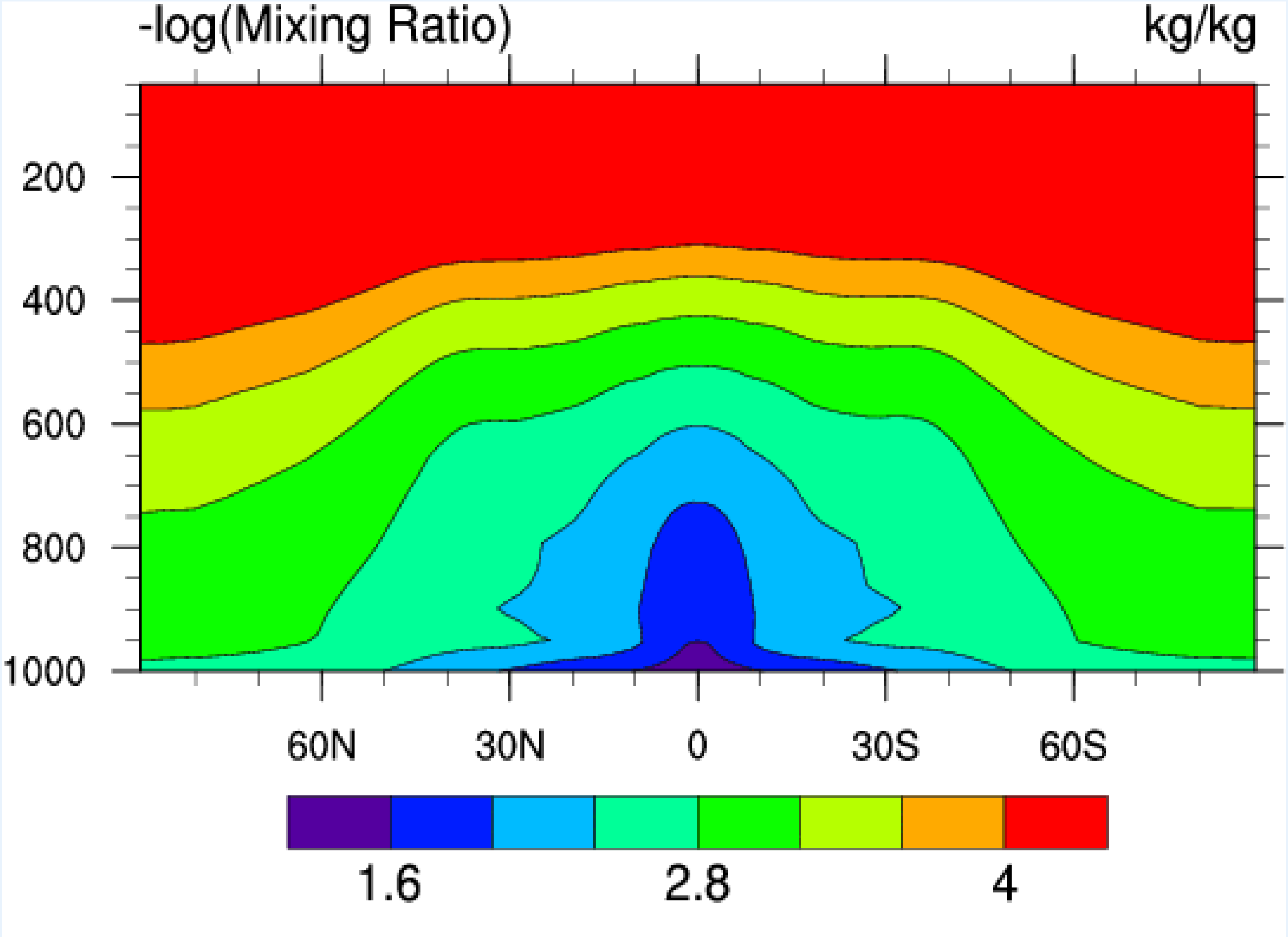
Partitioning a cubed-sphere on 8 processors



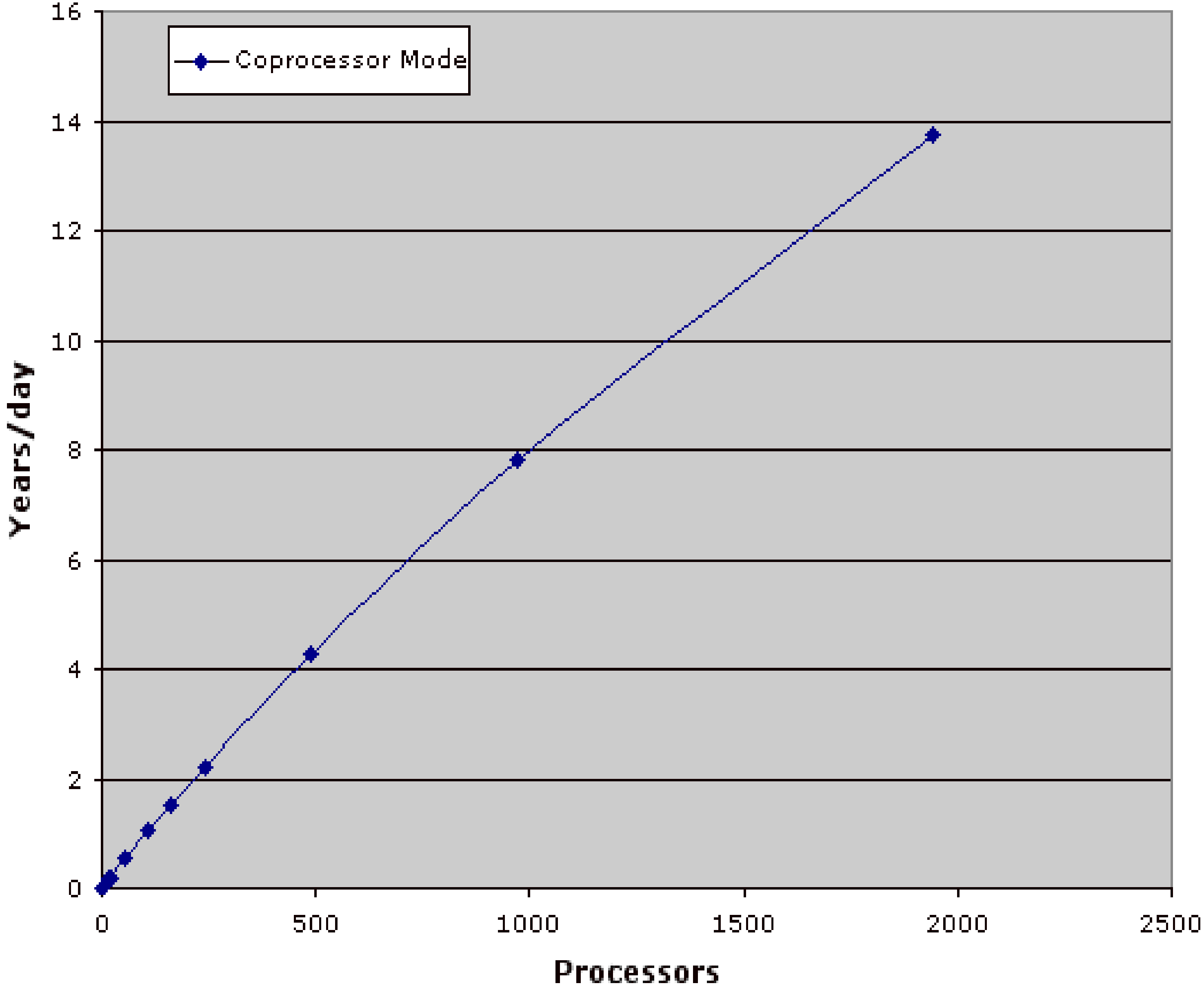
HOMME on Blue Gene/L…

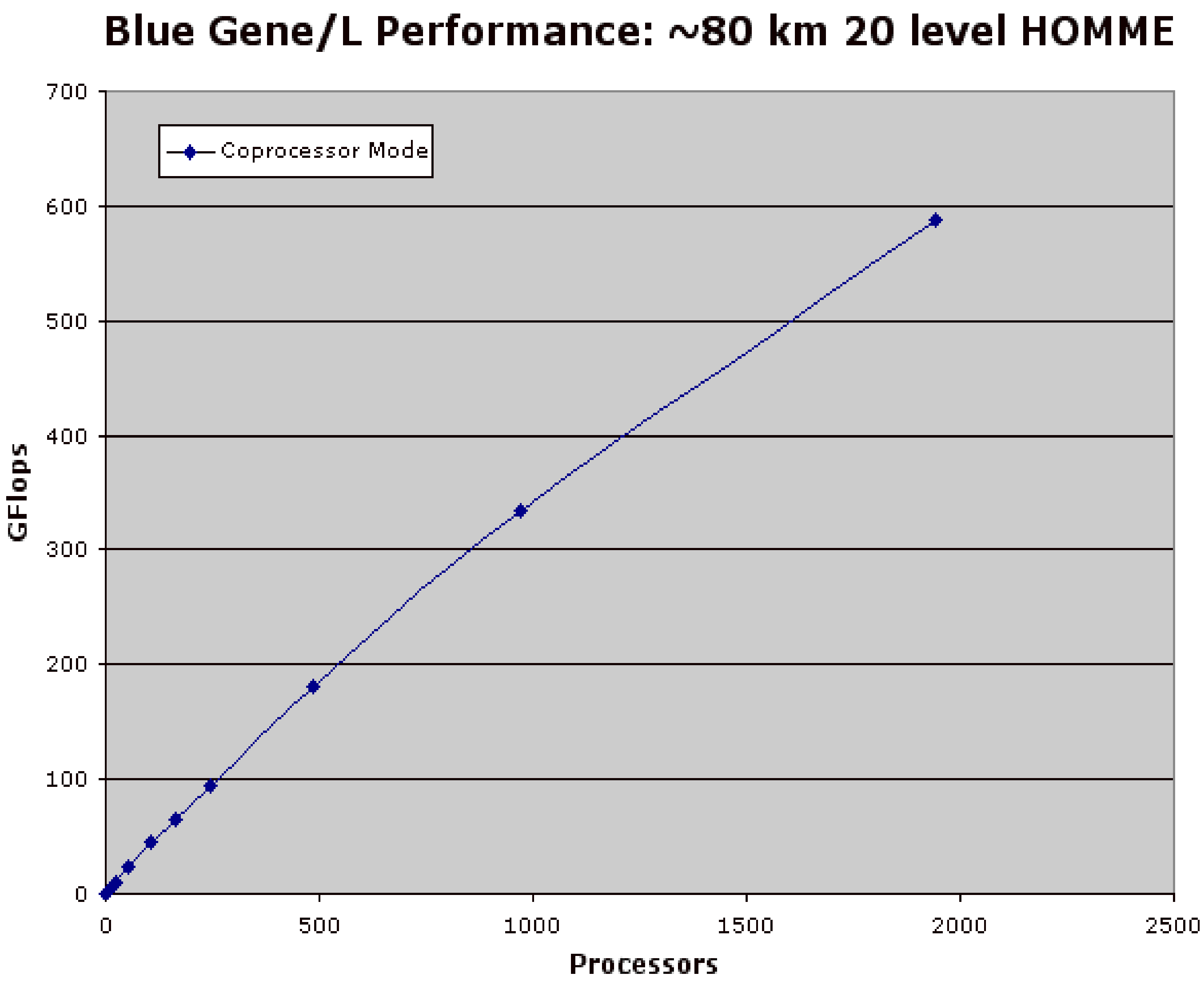
NCAR

Moist Held-Suarez Test Case: Temporally and Zonally Averaged Mixing Ratio



Explicit HOMME Integration Rate: 80 km, 20 levels



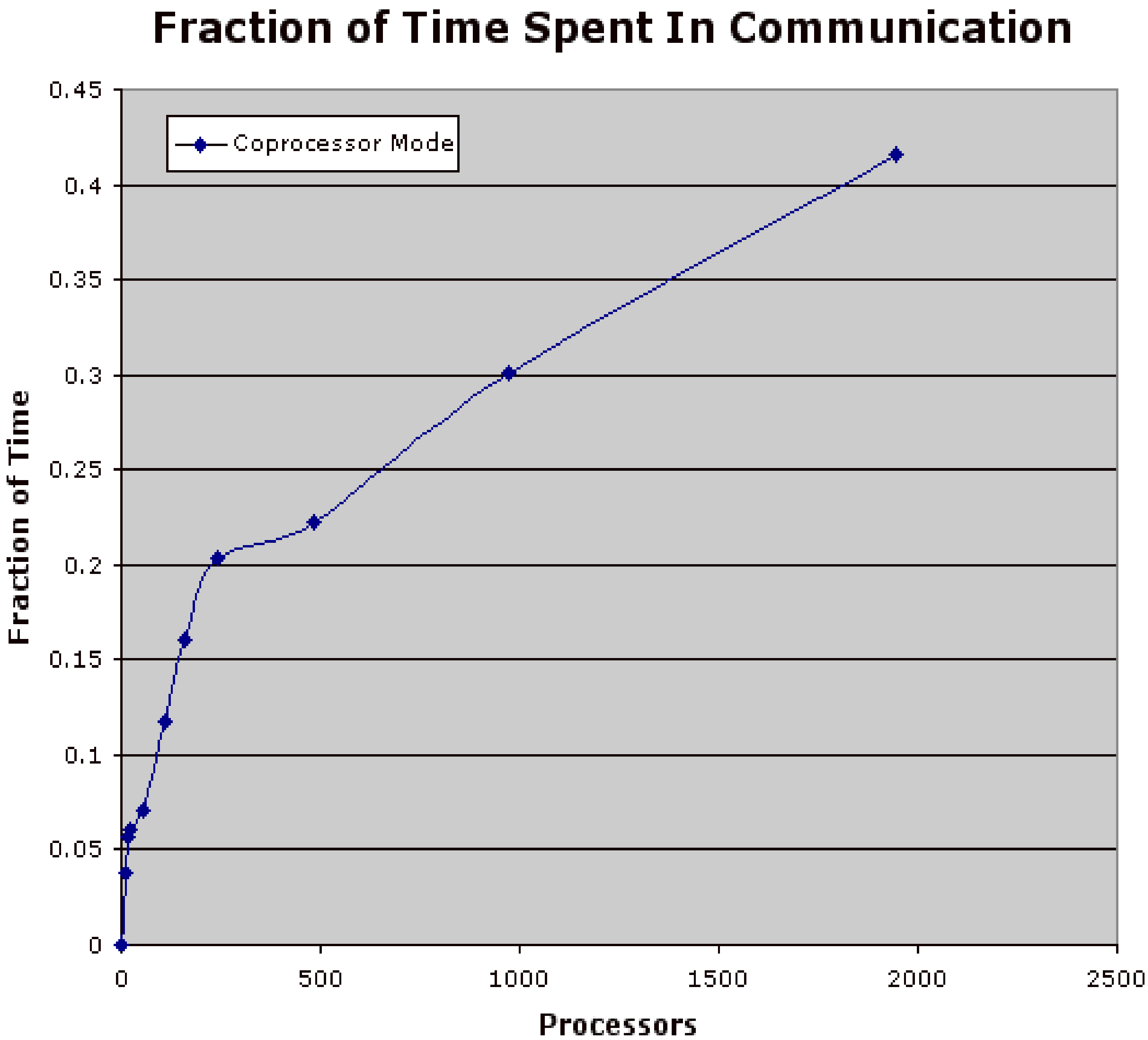

Blue Gene/L Performance: ~80 km 20 level HOMME
Coprocessor Mode
GFlops
Processors
0
100
200
300
400
500
600
700
0
500
1000
1500
2000
2500

Blue Gene/L Per-Processor Performance

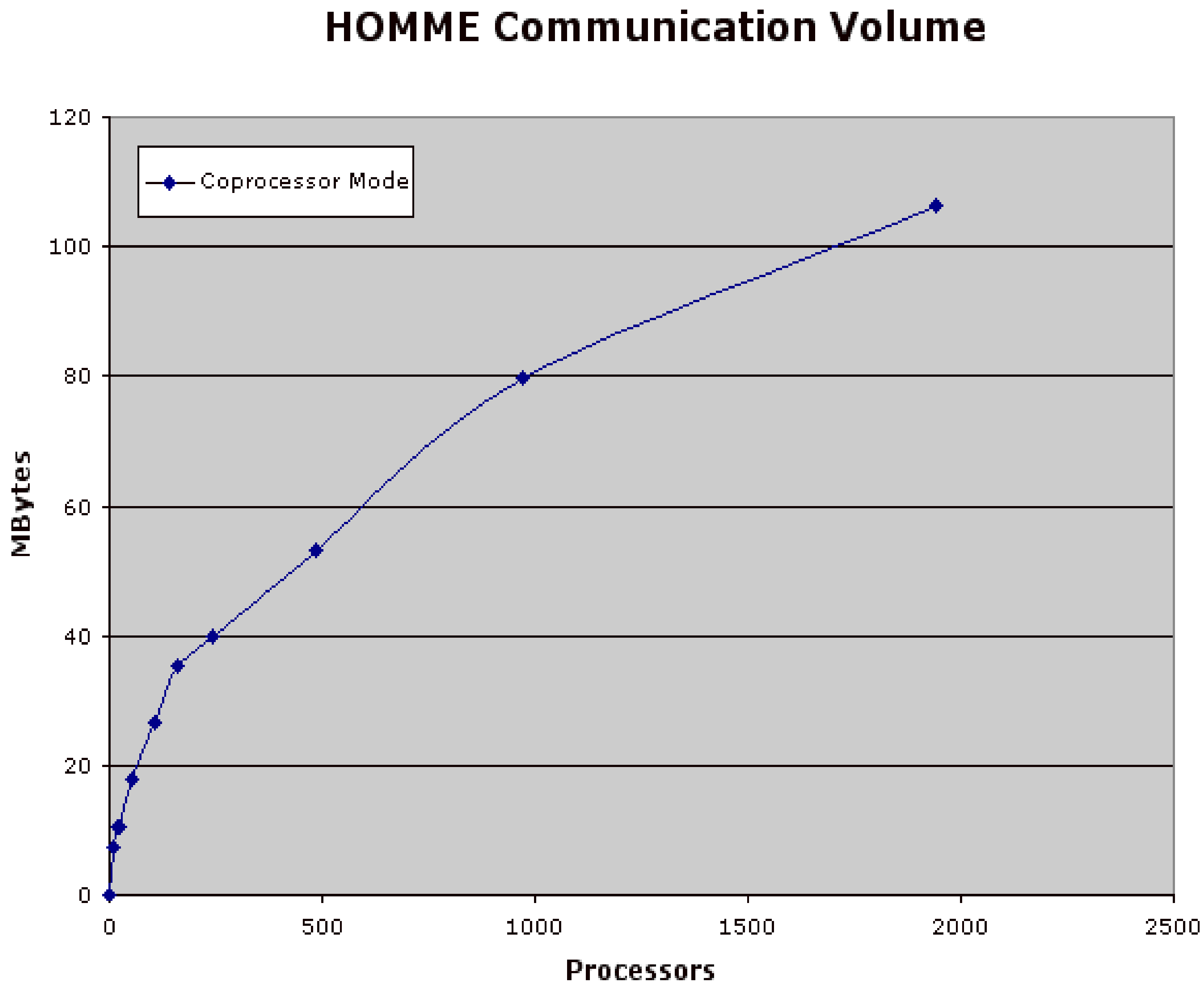
QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

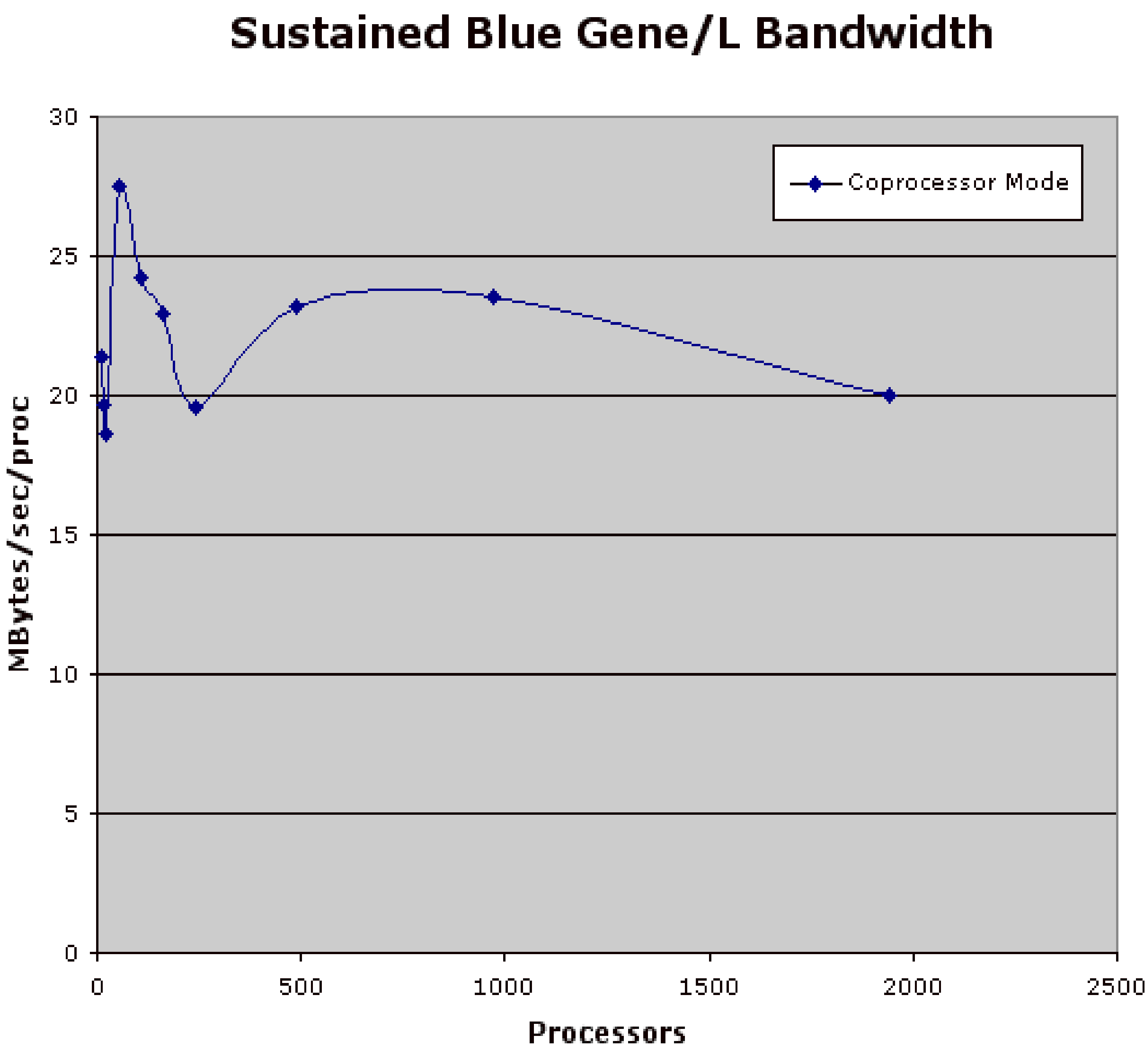
- **Ne=18, 20 level data**
- **Explicit primitive equations**
- **“Coprocessor” Mode**
- **16% of 2.8 Gflops peak**
- **BG/L vector intrinsics**
- **Not clear how many double FMA’s are being generated.**

NCAR

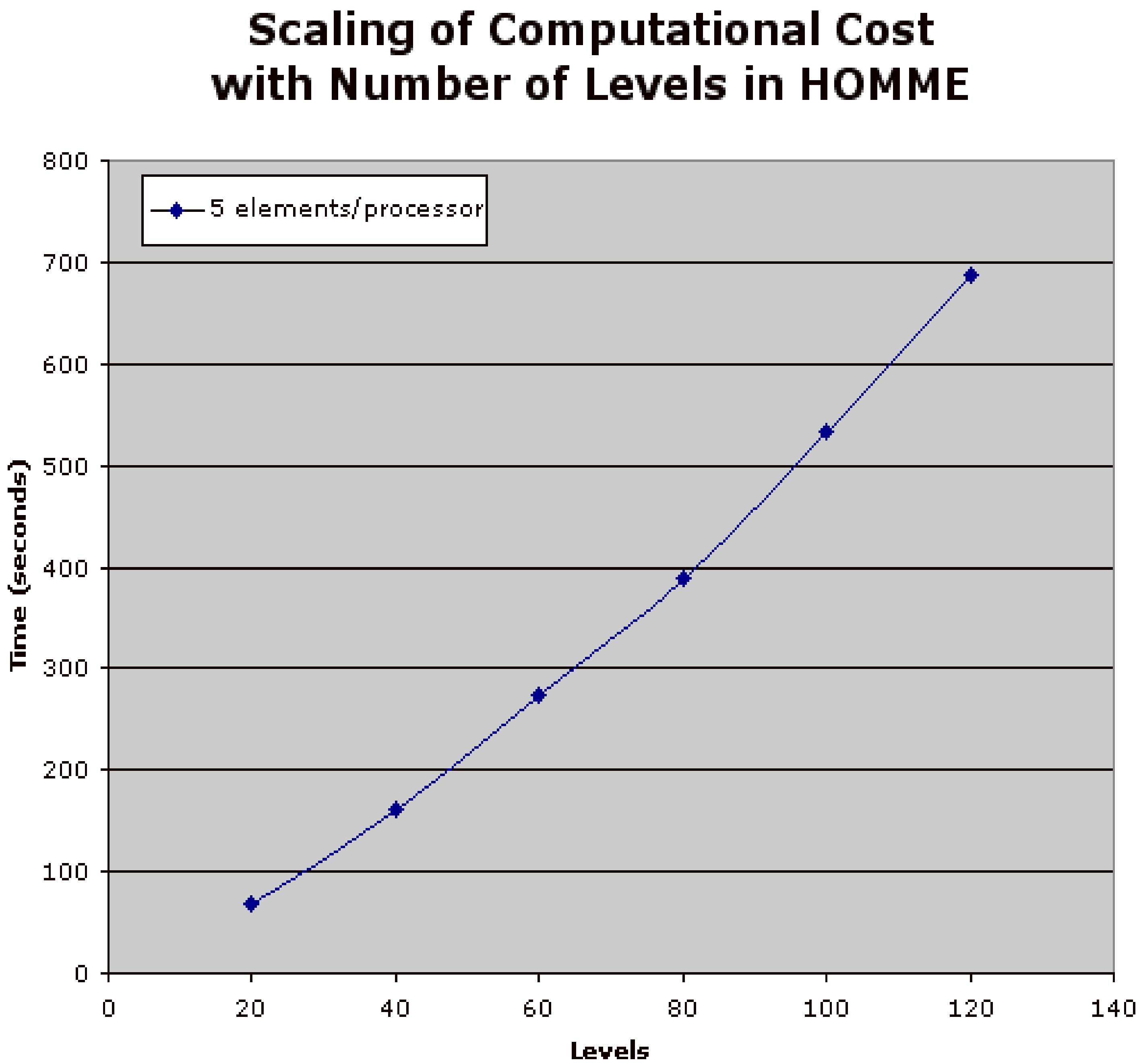


Fraction of Time Spent In Communication
Coprocessor Mode
Fraction of Time
Processors
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
500
1000
1500
2000
2500


HOMME Communication Volume
Coprocessor Mode
MBytes
0
20
40
60
80
100
120
Processors
0
500
1000
1500
2000
2500



Sustained Blue Gene/L Bandwidth
Coprocessor Mode
MBytes/sec/proc
Processors
0
5
10
15
20
25
30
0
500
1000
1500
2000
2500


Scaling of Computational Cost
with Number of Levels in HOMME
5 elements/processor
Time (seconds)
800
700
600
500
400
300
200
100
0
0
20
40
60
80
100
120
140
Levels

Summary of remarks…

- Blue Gene/L seems to us to be an important architecture for many reasons.
  - Power/space & “fuel” efficiency
  - $/Tflops sustained
  - Fast reduction network
- 80 km, 20 level explicit HOMME model
  - We see 587 Gflops on 1944 nodes of BG/L.
  - 13.7 simulated years/day (useful climate rate)
- 1/10 degree “eddy permitting” POP ocean model would be the next interesting thing to look at. (barotropic CG needs fast global sums)

NCAR

BlueGene/L Collaboration

NCAR

CU Boulder

CU Denver

Blue Gene/L

NCAR

Questions?

NCAR

CRCP (super-parameterized) physics…

CRCP or super-parameterization: What is "super" about it?

- The basic idea is to represent sub grid scales of the 3D large-scale model (with horizontal resolution ~ 100 km) by embedding a 2D cloud resolving model in each column of the large-scale model.
- Involves thousands of 2D cloud resolving models interacting in a way consistent with large scale dynamics.
- Embarrassingly parallel but extremely expensive. (150x over traditional physics)

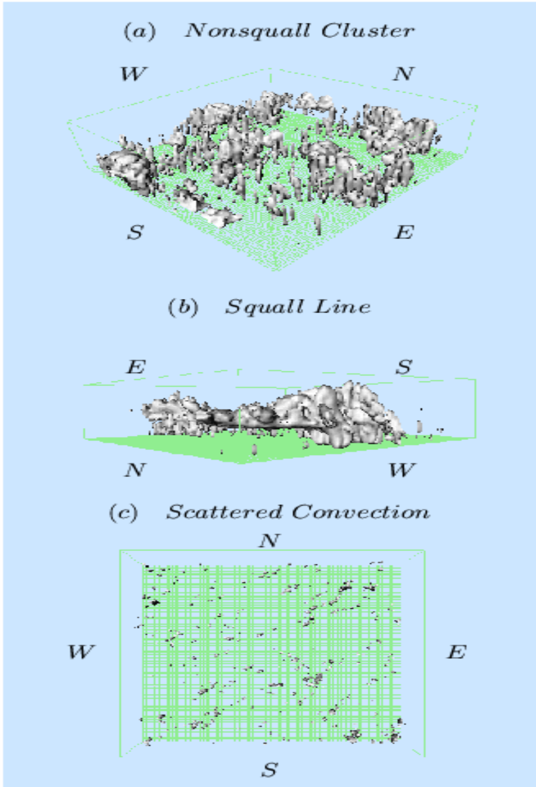
NCAR

CRCP is the next step in the quest for a cloud-system- resolving AGCM

- The computational cost of the following simulations are approximately the same:
  - A millennium-long simulation using a traditional climate model.
  - A few years-long simulation using a traditional climate model with CRCP
  - A day-long simulation of a cloud-system-resolving AGCM O(few km)
- Cost of each separated by 3 orders of magnitude!

NCAR

CRCP: 2D or not 2D: that is the question…



400 km box of clouds with dx = dy = 2 km

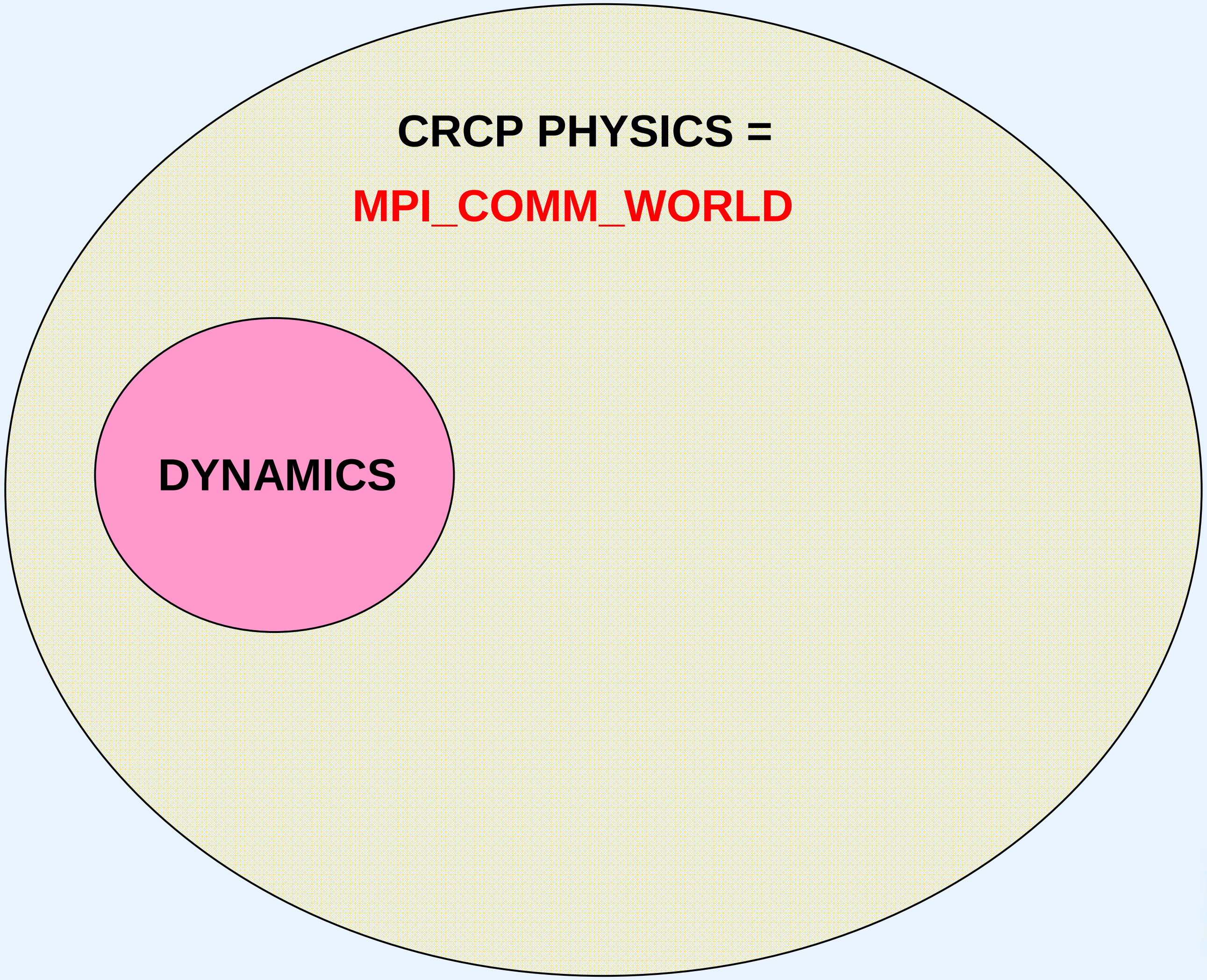
Advantages of CRCP

- Allows for better representation of convection, clouds and radiative transfer and surface exchange.
- Allows for a dynamic response to changes in cloud parameters (particle sizes, aerosol characteristics, precipitation mechanisms at approximately correct length scales.
- Example: cloud top longwave cooling in response to particle size change.

NCAR

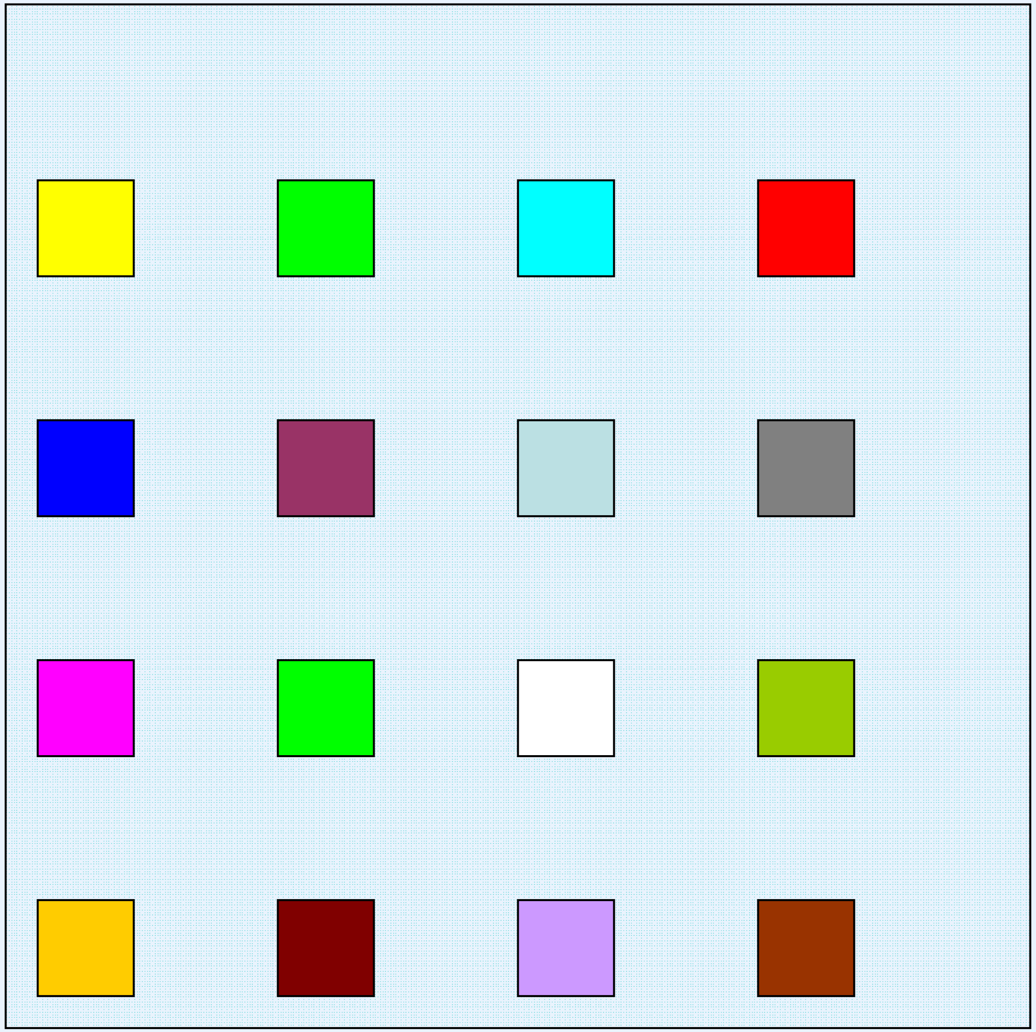
Coupling CRCP physics and dynamics together…

Logical View of CRCP-HOMME Coupling on Blue Gene/L



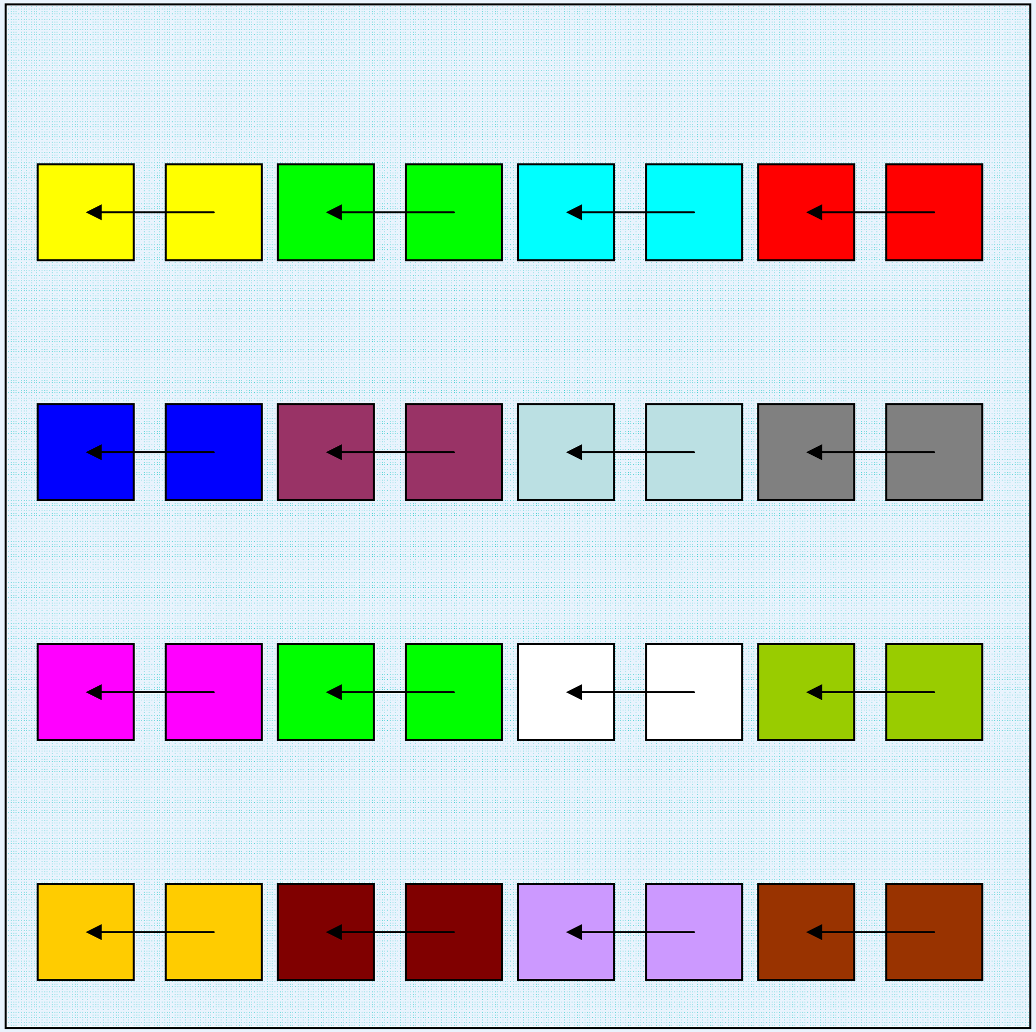
NCAR

Hardware View of CRCP-HOMME Coupling on Blue Gene/L



Begin with elements laid out for dynamics scheme

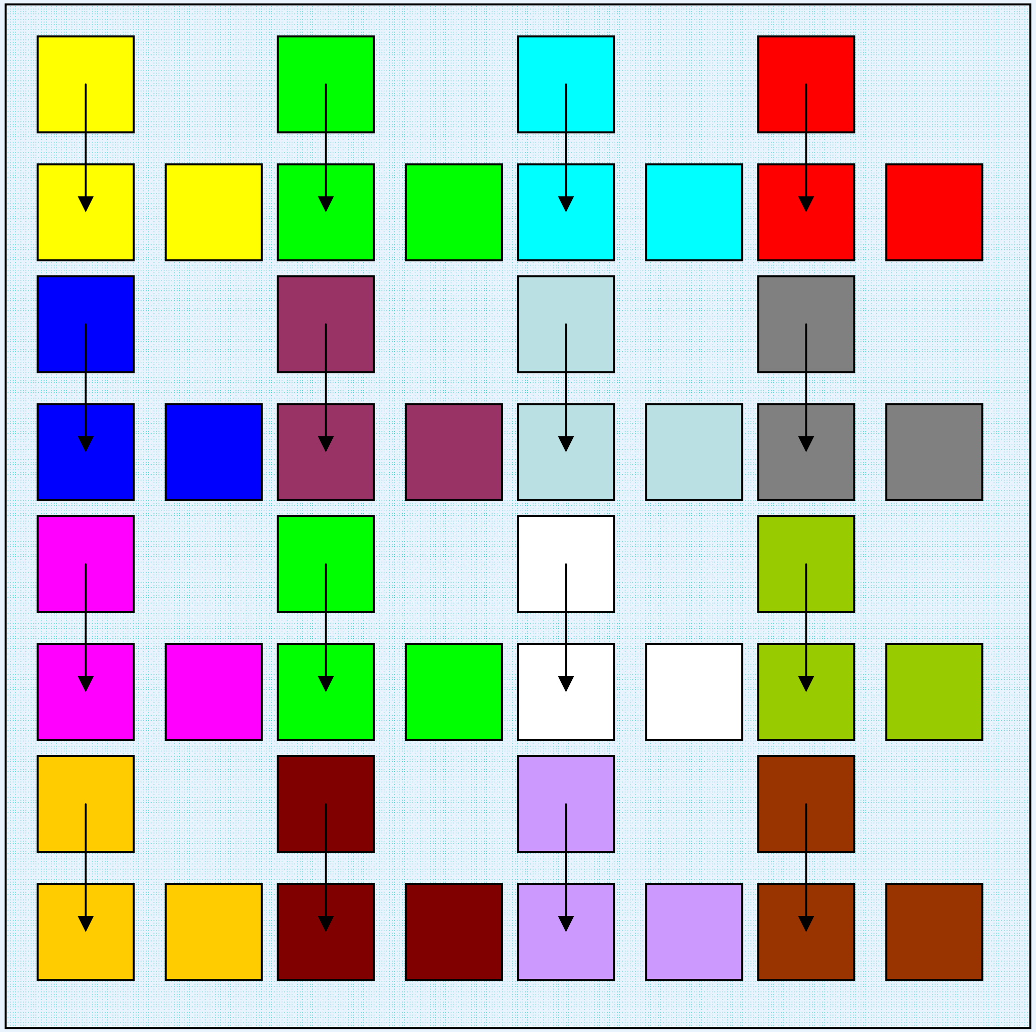
Hardware View of CRCP-HOMME Coupling on Blue Gene/L



Begin scattering the columns

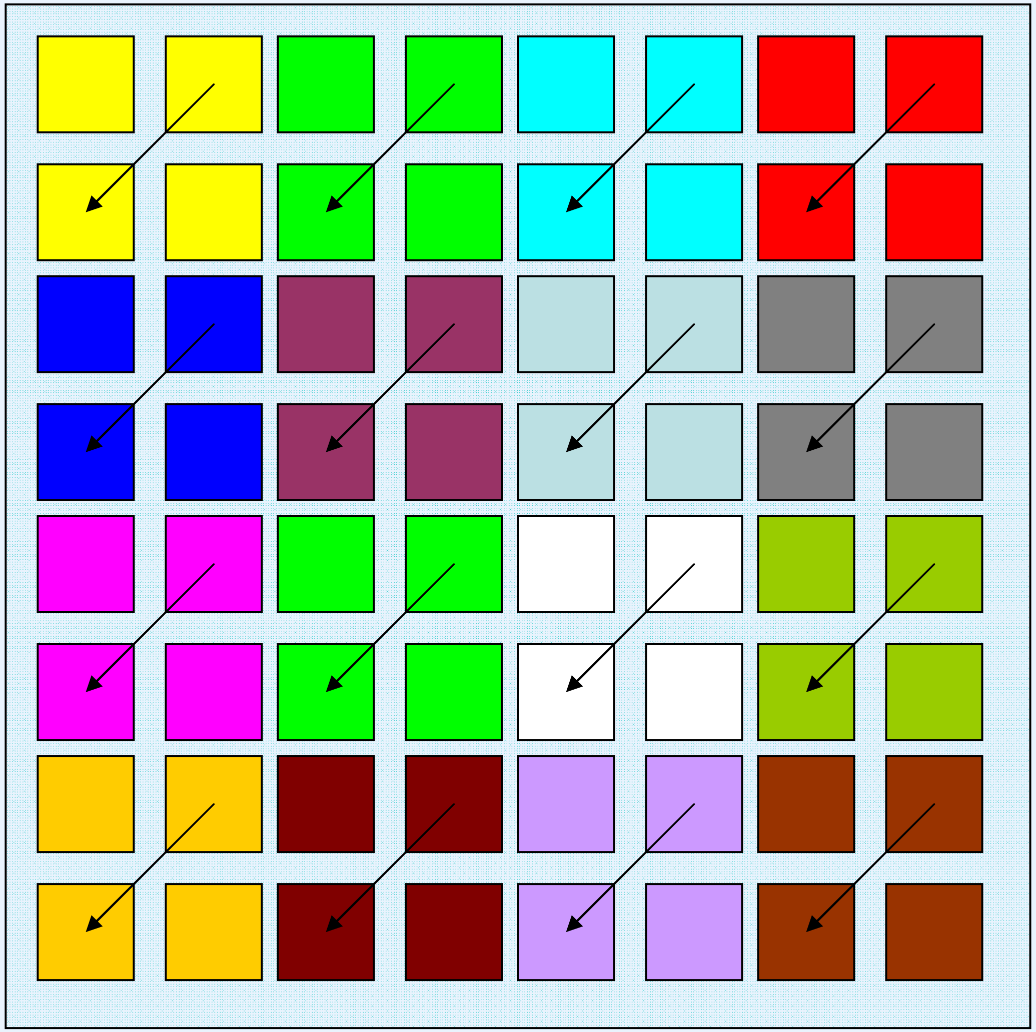
NCAR

Hardware View of CRCP-HOMME Coupling on Blue Gene/L



Continue scattering columns

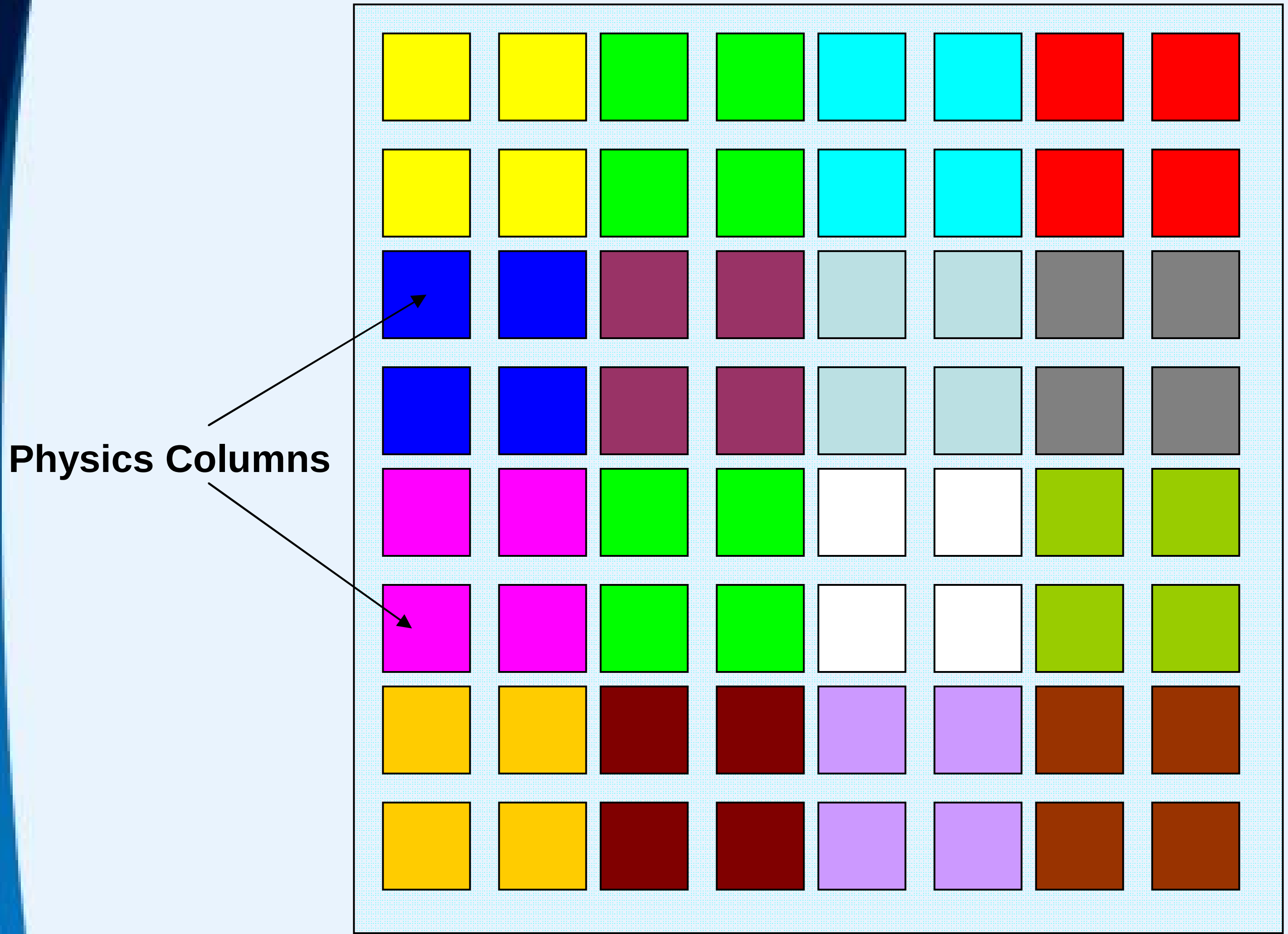
Hardware View of CRCP-HOMME Coupling on Blue Gene/L



Continue scattering columns

NCAR

Hardware View of CRCP Physics Layout on Blue Gene/L



Colors denote elements

Simplified 2-d BG/L topology