

---

## ***SDSC HPC Resources***

***DataStar: A 15.6 TF Power4 + Federation Switch System***

***Blue Gene: A 5.7 TF 2048 processor PowerPC***

**Mahidhar Tatineni**

**([mahidhar@sdsc.edu](mailto:mahidhar@sdsc.edu))**

**and**

**Amit Majumdar**

**([majumdar@sdsc.edu](mailto:majumdar@sdsc.edu))**

**September 2005**



SAN DIEGO SUPERCOMPUTER CENTER

at the UNIVERSITY OF CALIFORNIA, SAN DIEGO



---

# SDSC

- A NSF center with compute and data resources allocated *freely* through peer review process
- One of the emphasis of SDSC for national Cyberinfrastructure initiative is data intensive computing
- In the process of expanding DataStar with 96 8-way P655 nodes; being added to existing system
- SDSC Blue Gene will go into production in October 2005. ASC users can transfer their allocation from DataStar to BlueGene. One SU on DataStar will be transferred as 2.5 SUs on the Blue Gene.
- Applications benchmarked on BG with idea of moving appropriate ones to DataStar to reduce load on DataStar



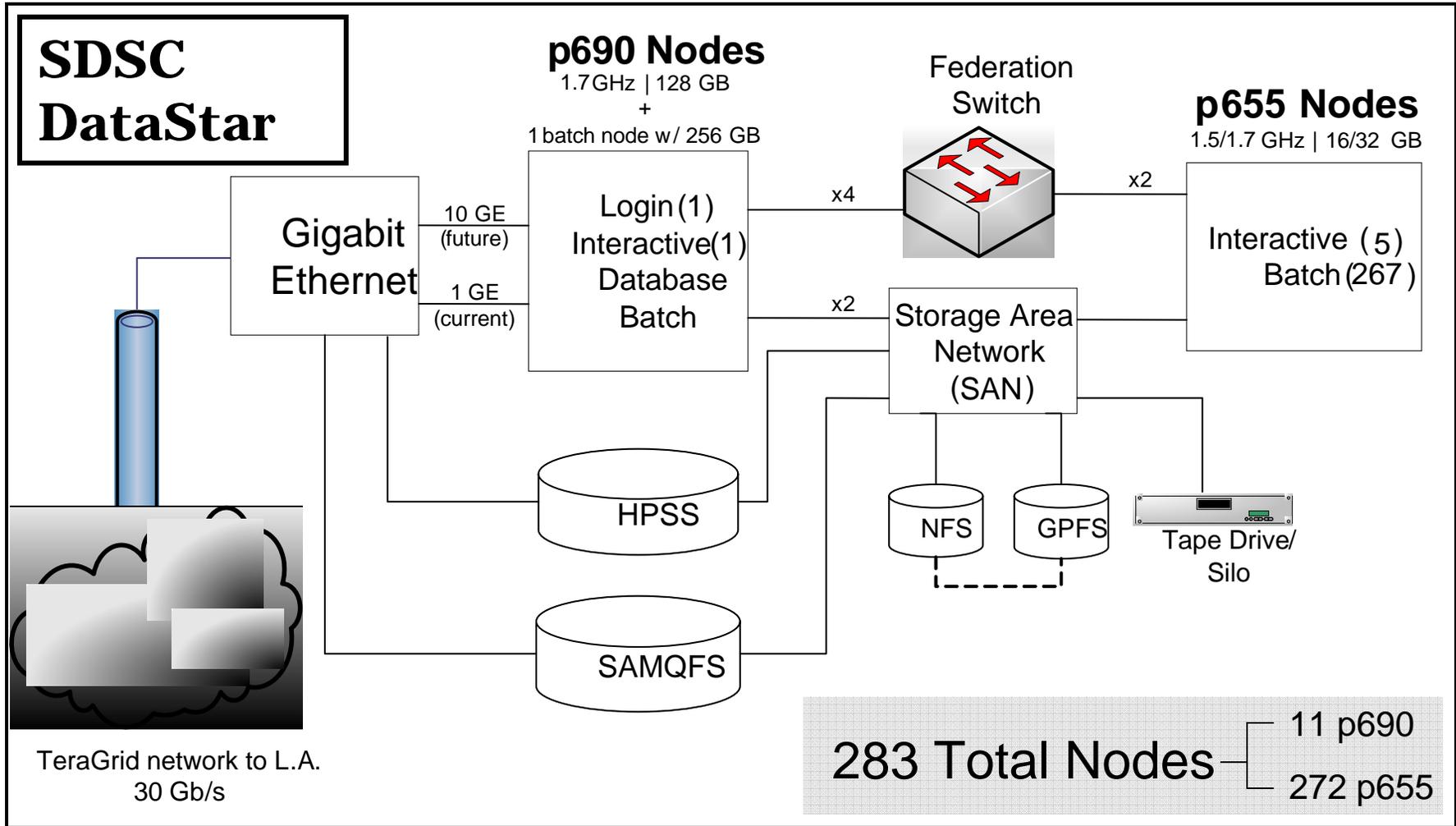
SAN DIEGO SUPERCOMPUTER CENTER

at the UNIVERSITY OF CALIFORNIA, SAN DIEGO



# DataStar Expansion

<b>Current Configuration</b>	<b>Expanded Configuration</b>
<b>10.1 TF, 1760 processors total</b>	<b>15.6 TF, 2528 processors total</b>
<b>11 32-way 1.7 GHz IBM p690s</b>	<b>11 32-way 1.7 GHz IBM p690s</b>
2 nodes 64 GB memory for login and interactive use	2 nodes 64 GB memory for login and interactive use
6 nodes 128 GB memory for scientific computation	6 nodes 128 GB memory for scientific computation
2 nodes 128 GB memory for database, DiscoveryLink	2 nodes 128 GB memory for database, DiscoveryLink
1 node 256 GB memory for batch scientific computation	1 node 256 GB memory for batch scientific computation
<b>176 8-way 1.5 GHz IBM p655</b>	<b>176 8-way 1.5 GHz IBM p655</b>
16 GB memory	16 GB memory
Batch scientific computation	Batch scientific computation
<b>All nodes Federation switch attached</b>	<b>All nodes Federation switch attached</b>
<b>All nodes SAN attached</b>	<b>All nodes SAN attached</b>
<b>Currently 66 TB GPFS</b>	<b>Estimated 110 TB GPFS</b>
	<b>96 8-way 1.7 GHz IBM p655</b>
	32 GB memory
	Batch scientific computation

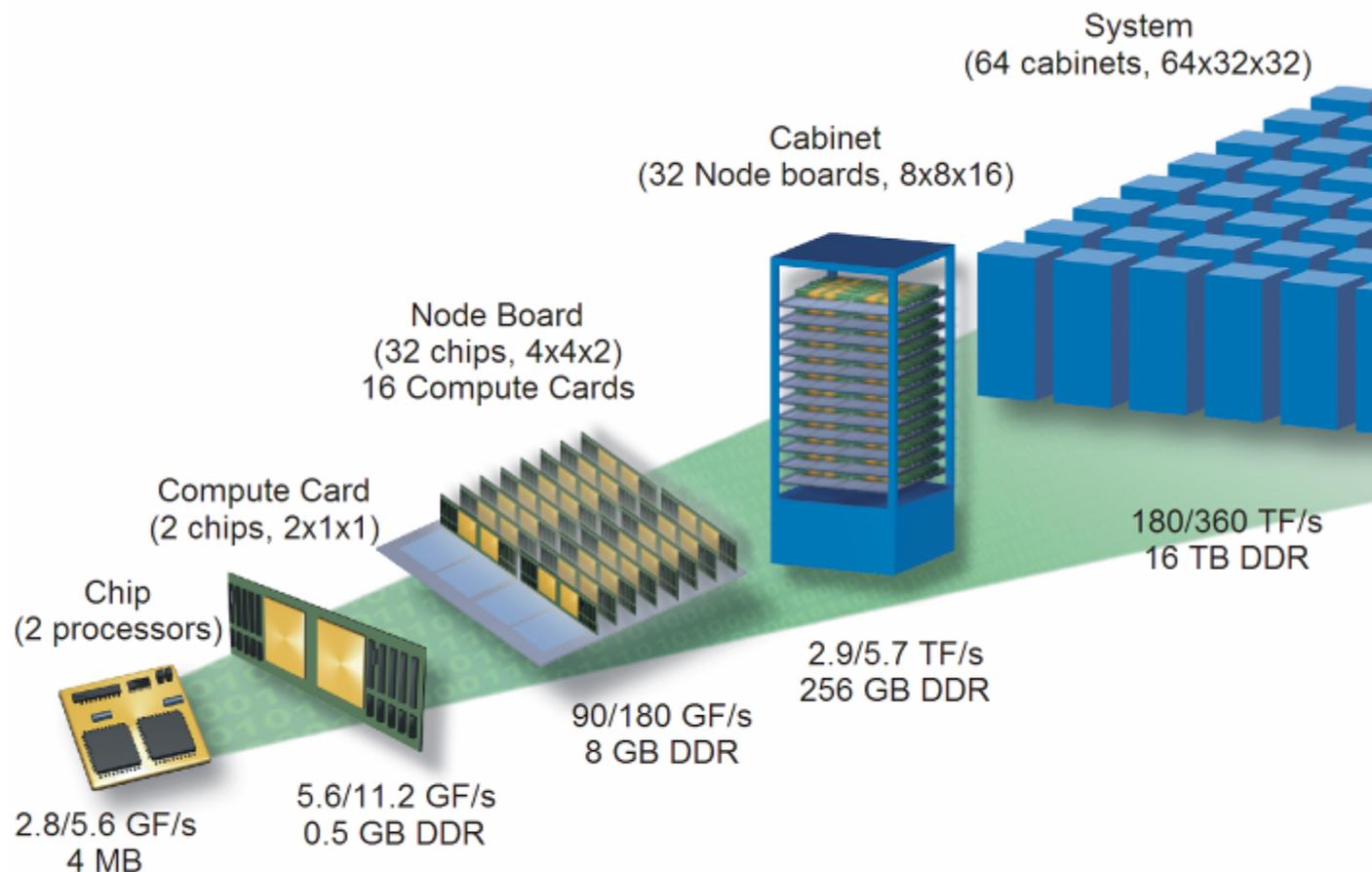


---

## ***SDSC DataStar Expansion***

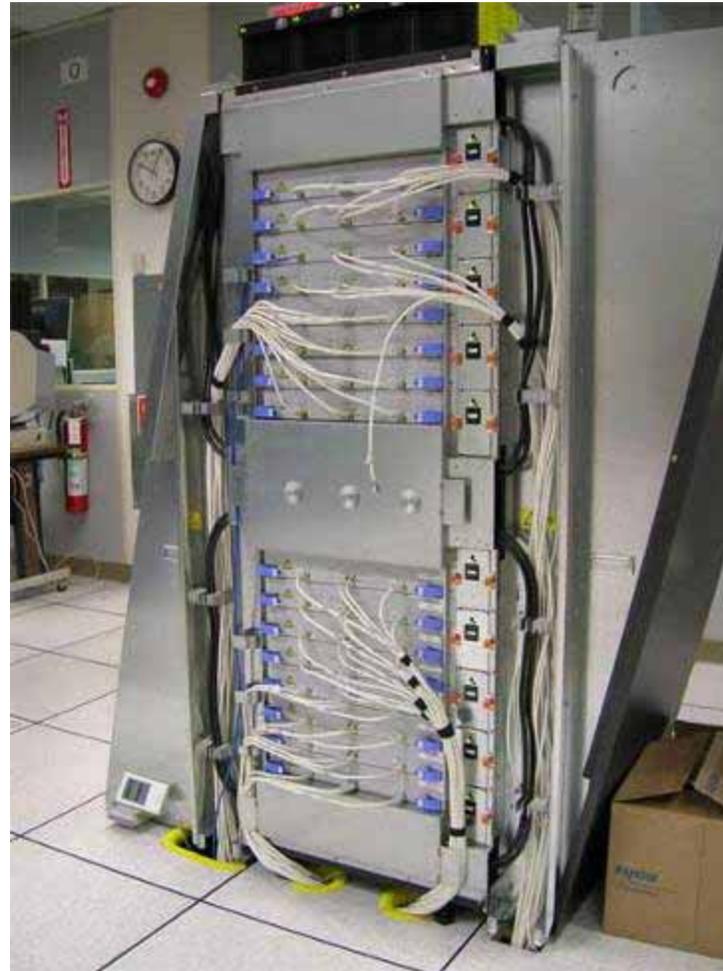
- **The job submission and queuing process will be the same as before using Loadleveler. Catalina will be the scheduler as before.**
- **There will be three sets of queues: 1) queues including the old p655 nodes (~171), 2) queues including the new p655 nodes (96), and 3) queues including all the p655 nodes (~267). Users will be able to runs of 2048 (or larger upto 2136) processors on the largest queue. ASC jobs get automatic boost in the queue.**
- **The new /gpfs filesystem (~110 TB) will be larger than before.**
- **The theoretical peak I/O performance for the new filesystem is around 6GB/s. We are currently setting up and testing the new filesystem and will update users with achieved performance numbers.**

# Blue Gene System Overview: Chips to Racks



---

# *Blue Gene System Overview: SDSC's single-rack system*



---

# ***BG System Overview: SDSC's single-rack system***

- **1024 compute nodes & 128 I/O nodes (each with 2p)**
  - Most I/O-rich configuration possible (8:1 compute:I/O node ratio)
  - Identical hardware in each node type with different networks wired
  - Compute nodes connected to: torus, tree, global interrupt, & JTAG
  - I/O nodes connected to: tree, global interrupt, Gigabit Ethernet, & JTAG
- **Two half racks (also confusingly called midplanes)**
  - Connected via link chips
- **Front-end nodes (4 B80s, each with 4p)**
- **Service node (p275 with 2p)**

---

# ***BG System Overview: Processor Chip (= System-on-a-chip)***

- **Two 700-MHz PowerPC 440 processors**
  - Each with two floating-point units
  - Each with 32-kB L1 data caches that are noncoherent
  - 4 flops/proc-clock peak (=2.8 Gflops/proc)
  - 2 8-B loads or stores / proc-clock peak in L1 (=11.2 GBps/proc)
- **Shared 2-kB L2 cache (or prefetch buffer)**
- **Shared 4-MB L3 cache**
- **Five network controllers (though not all wired to each node)**
  - 3-D torus (for point-to-point MPI operations: 175 MBps nom x 6 links x 2 ways)
  - Tree (for most collective MPI operations: 350 MBps nom x 3 links x 2 ways)
  - Global interrupt (for MPI\_Barrier: low latency)
  - Gigabit Ethernet (for I/O)
  - JTAG (for machine control)
- **Memory controller for 512 MB of off-chip, shared memory**

---

# ***BG System Overview:***

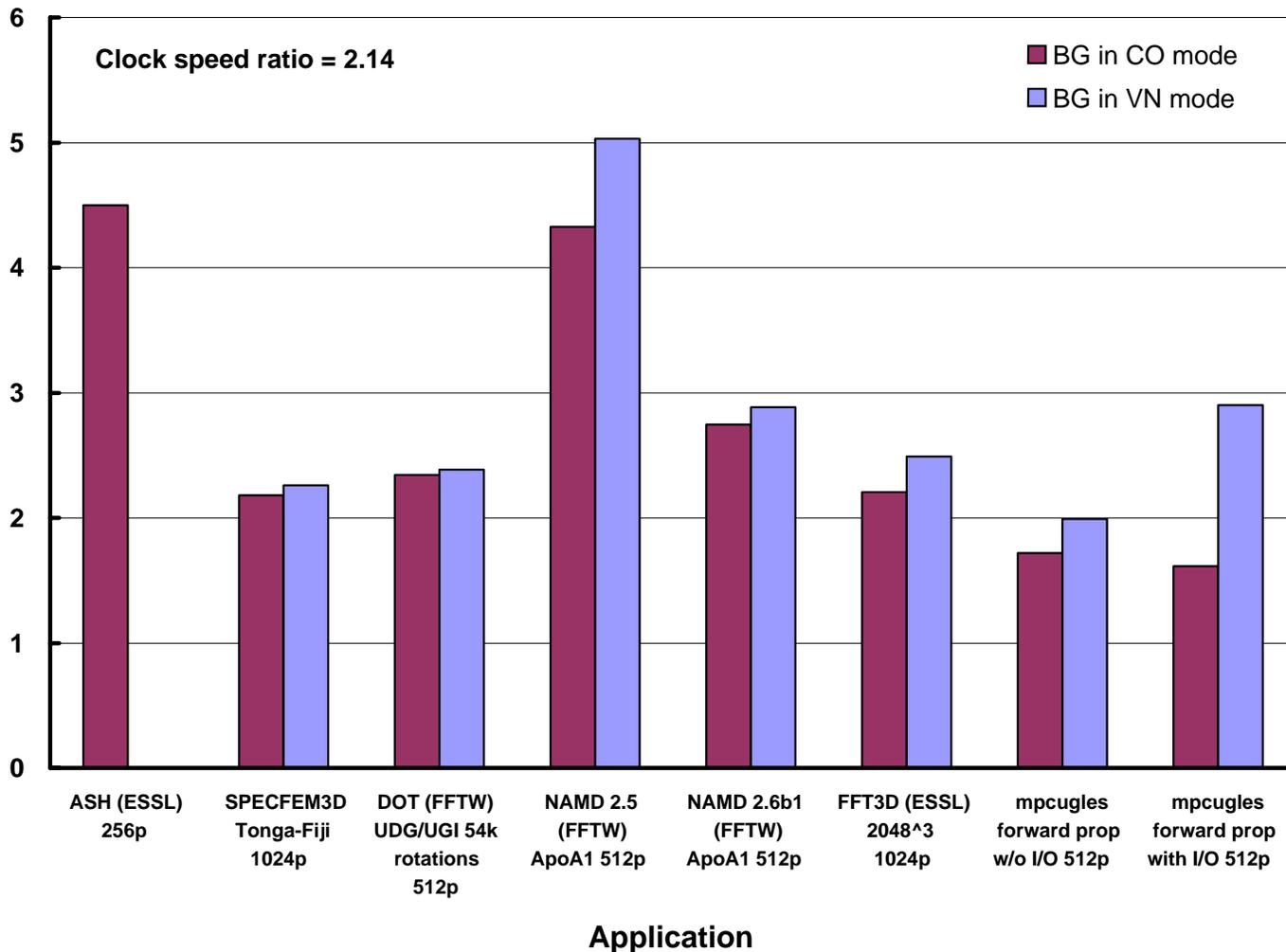
## ***Multiple operating systems & functions***

- **Compute nodes:** run **Compute Node Kernel (CNK = blrts)**
  - Each run only one job at a time
  - Each use very little memory for CNK
- **I/O nodes:** run **Embedded Linux**
  - Run CIOD to manage compute nodes
  - Perform file I/O
  - Run GPFS
- **Front-end nodes:** run **SuSE SLES9 Linux/PPC64**
  - Support user logins
  - Run cross compilers & linker
  - Run parts of mpirun to submit jobs & LoadLeveler to manage jobs
- **Service node:** runs **SuSE SLES9 Linux/PPC64**
  - Uses DB2 to manage four system databases
  - Runs control system software, including MMCS
  - Runs other parts of mpirun & LoadLeveler
- **(Software comes in drivers: currently running Driver 202)**

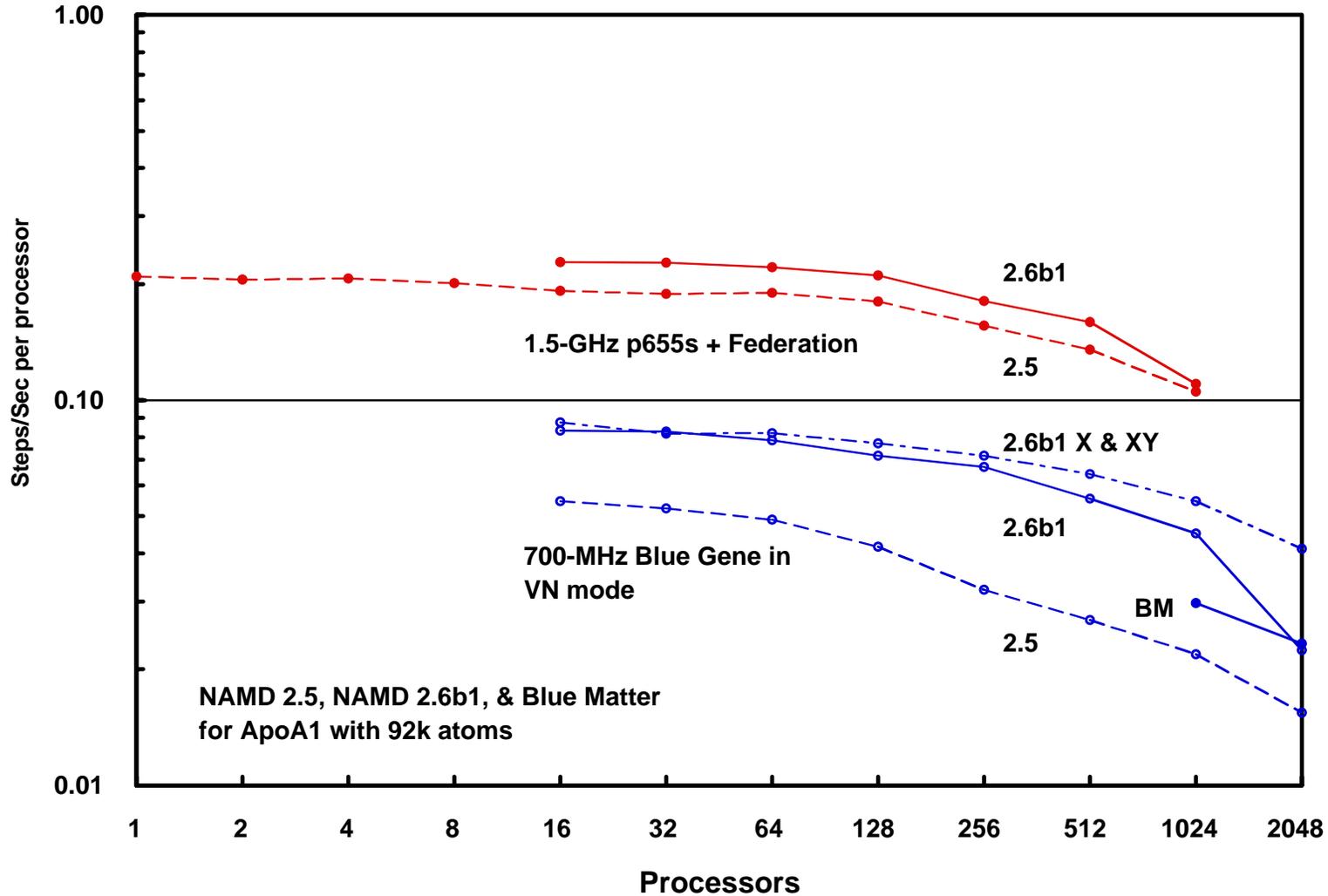
***Additional applications have been implemented at SDSC Blue Gene following the user workshop in July: current list follows***

<b>Code name</b>	<b>Discipline</b>	<b>Description</b>	<b>Implementors</b>
ASH (ESSL)	Astrophysics	3-D turbulent convection	Ben Brown &
	U.Colorado	of the Sun	Robert Harkness
SPECFEM3D	Geophysics	3-D seismic wave	Brian Savage
	Caltech	propagation	
DOT (FFTW)	Biophysics	Protein docking	Susan Lindsey &
	SDSC		Wayne Pfeiffer
NAMD 2.6b1 (FFTW)	Biophysics	Molecular dynamics	Sameer Kumar
	UIUC		
FFT3D (ESSL: kernel of DNS)	Engineering	Direct numerical simulatio	Dmitry Pekurovsky &
	GeorgiaTech	of 3-D turbulence	Giri Chukkapalli
mpcugles	Engineering	3-D fluid dynamics	Giri Chukkapalli
	U.Minn.		

**Speed ratio of p655s to BG is now between 2 & 3  
for 5 of 6 applications on 512p or 1024p,  
given realistic conditions of VN mode with I/O**



***NAMD 2.6b1 is much faster than NAMD 2.5 & even faster than Blue Matter; BG is less than 3x slower than p655s now***



---

## *Performance of Blue Gene relative to p655s has improved for several apps recently*

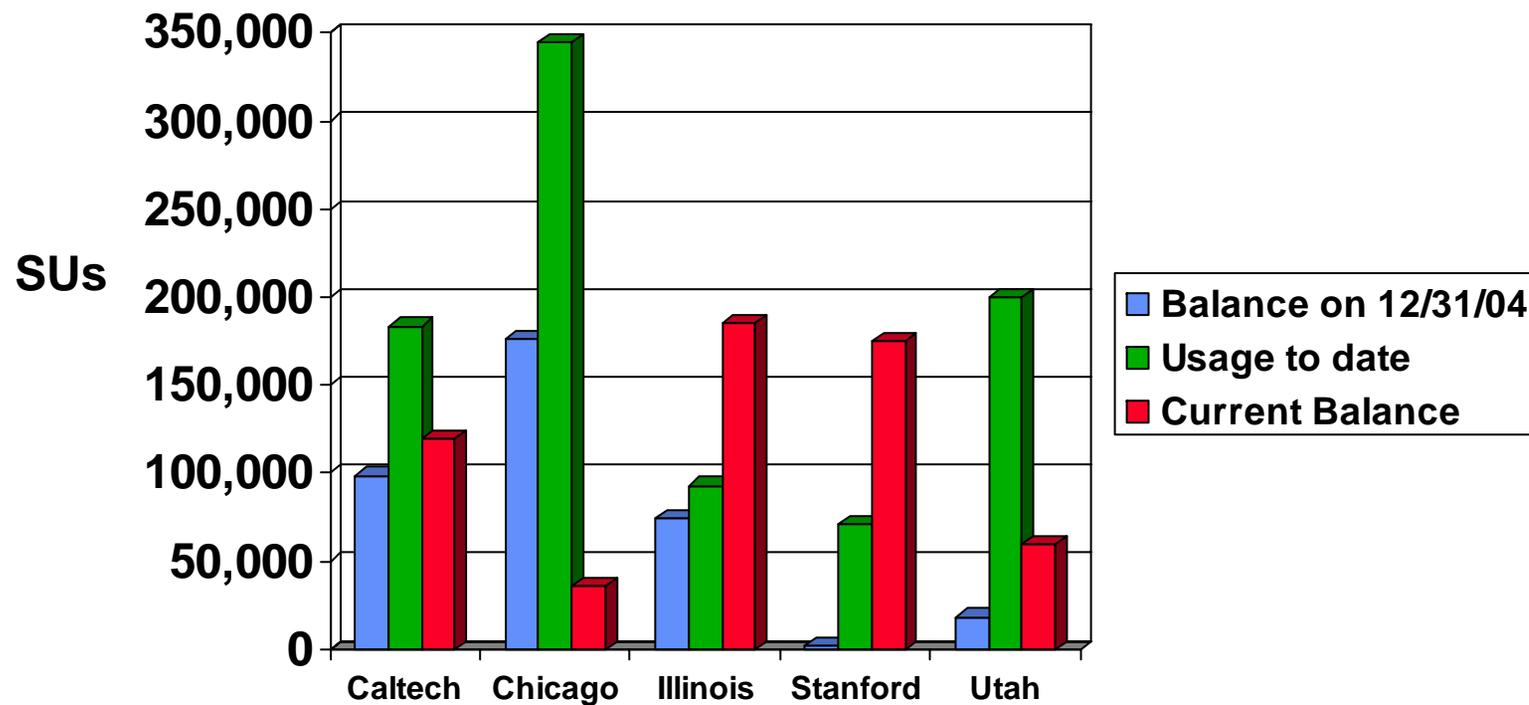
- **Two new apps – SPECFEM3D & DOT – run well**
  - p655/BG VN speed ratios of 2.3 to 2.4
- **NAMD 2.6b1 is much faster than NAMD 2.5**
  - due to efforts at UIUC & IBM
  - p655/BG VN speed ratio about 2.9 now vs 5.0 before on ApoA1 problem
  - still needs work for larger problems
- **FFT3D in VN mode has improved**
  - due to new compiler and/or driver
  - p655/BG VN speed ratio about 2.5
- **mpcugles results are unchanged**
  - p655/BG VN speed ratios between 2.0 & 2.9 depending upon I/O
- **ASH is still relatively slow on BG**
  - tuning underway

---

## *Prospects for the future are encouraging*

- **SPECFEM3D should scale well based on Earth Sim runs**
  - Tests will begin shortly on 4 racks of BG/W
- **NAMD 2.6b1 is much more attractive than 2.5 on BG**
  - Code needs to run on more problems
- **DNS, the full code based on FFT3D, is nearly ready to run**
  - Memory will be tight for  $2048^3$  problem on 2048p
- **Use of second FPU should boost performance**
- **Four codes use FFTs in ESSL or FFTW**
  - Improvements in these libraries could further boost performance

# ASC: Current Usage at SDSC



---

## ***SDSC Resources and Links***

- **SDSC's User Services page provides online user guides, frequently updated user news, consultant support and training:**  
[http://www.sdsc.edu/user\\_services/](http://www.sdsc.edu/user_services/)
- **The DataStar userguide is located online at:**  
[http://www.sdsc.edu/user\\_services/datastar/](http://www.sdsc.edu/user_services/datastar/)
- **The Blue Gene userguide is located online at:**  
[http://www.sdsc.edu/user\\_services/bluegene/](http://www.sdsc.edu/user_services/bluegene/)
- **For SDSC consulting information please visit:**  
[http://www.sdsc.edu/user\\_services/consulting/](http://www.sdsc.edu/user_services/consulting/)
- **[Submit a Ticket](#) or E-mail [consult@sdsc.edu](mailto:consult@sdsc.edu) (M - F 5:00am - 5:00pm PST).**
- **For technical support issues that cannot be submitted electronically, call 1-866-336-2357 (M - F 9:00am - 5:00pm PST).**



SAN DIEGO SUPERCOMPUTER CENTER

---

at the UNIVERSITY OF CALIFORNIA, SAN DIEGO

