CISL Update
Operations and Services

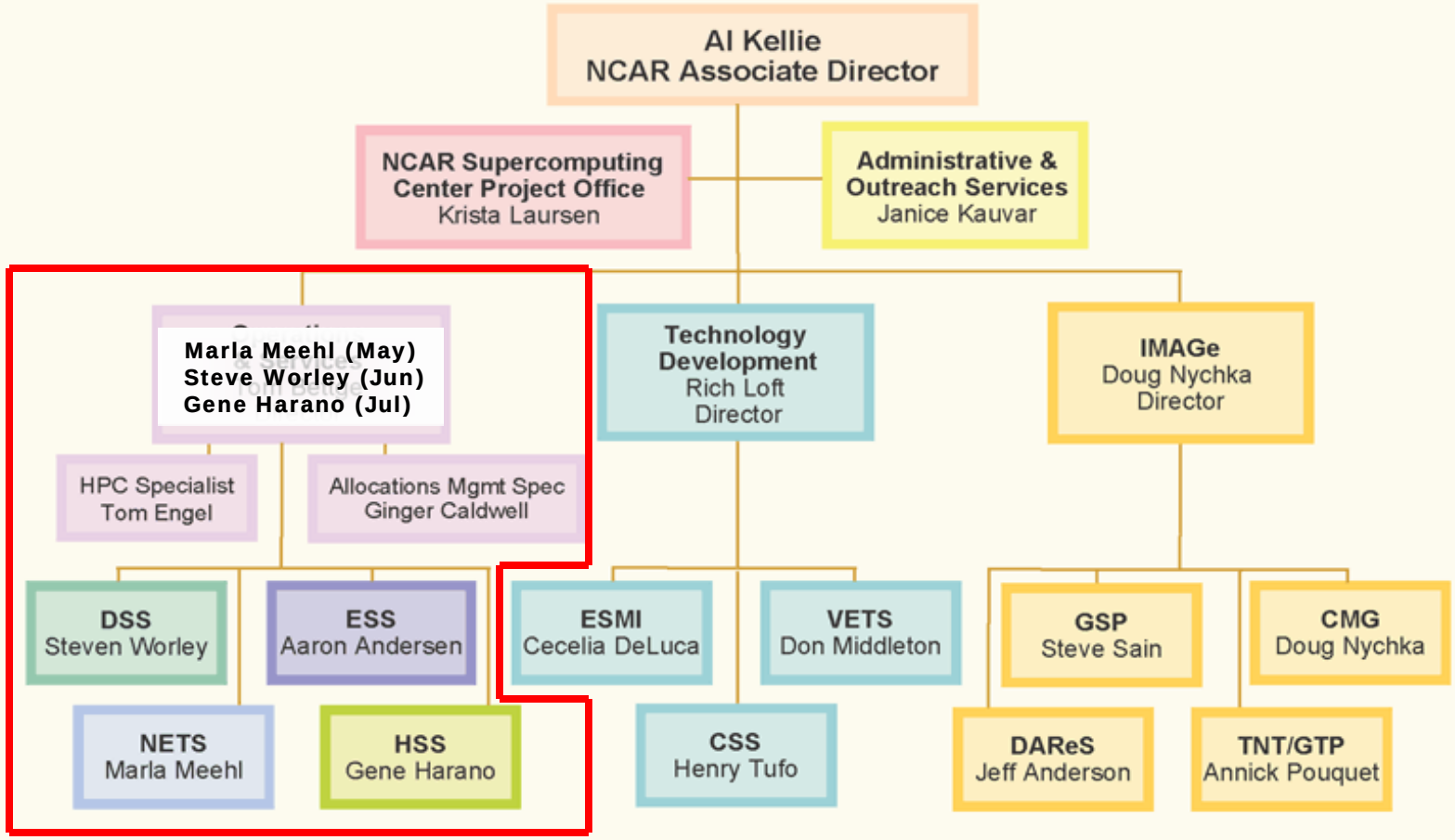
CISL HPC Advisory Panel Meeting
14 May 2009

Gene Harano & Tom Engel
Operations and Services Division
Computational and Information Systems Laboratory

CISL Computational & Information Systems Laboratory

NCAR

Computational and Information Systems Laboratory (CISL)



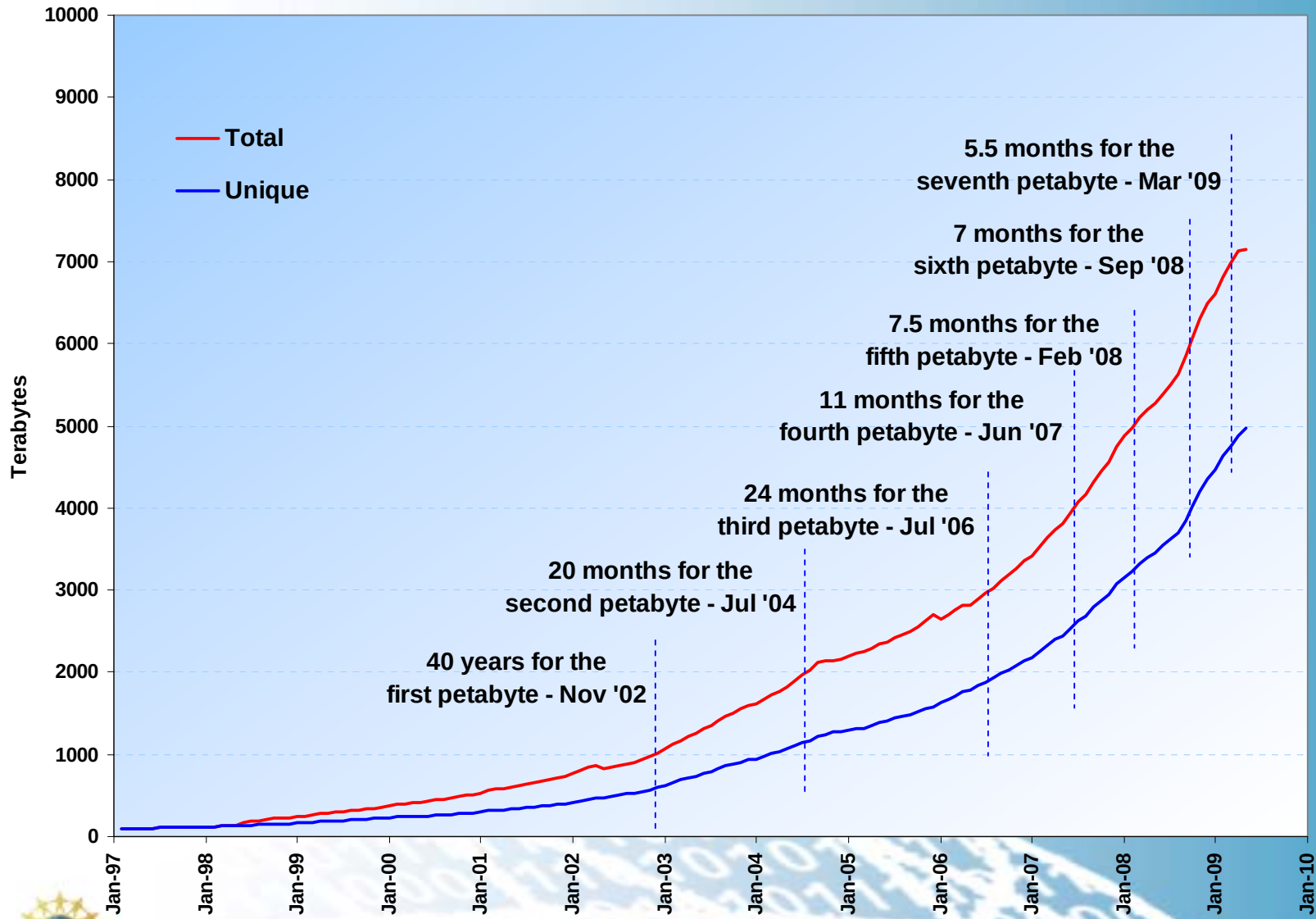
OSD Recent Personnel Changes

- **Departures/Changes**
  - *Tom Bettge* — *OSD Director*
    - » Remains as CISL Casual (CAS2K9 Program Chairman)
  - *George Fuentes* — *OSD/HSS/SSG Group Head*
    - » SSG Group Head - Irfan Elahi (acting)
  - *Mark Love* — *OSD/HSS/MSSG*
  - *Marc Genty* — *OSD/HSS/SSG*
    - » Moved to OSD/HSS/MSSG
  - *Wei Huang* — *OSD/HSS/CSG*
    - » Moved to CISL/TDD/VETS
  - *Ken Albertson* — *OSD/ESS/CPG*
    - » Moved to Research Applications Laboratory

- **Current Openings**
  - *OSD Director*
  - *OSD/HSS/SSG Software Engineer (2)*
  - *OSD/HSS/CSG Software Engineer*
  - *OSD/ESS/CPG Operator*

Mass Storage
Usage and Performance

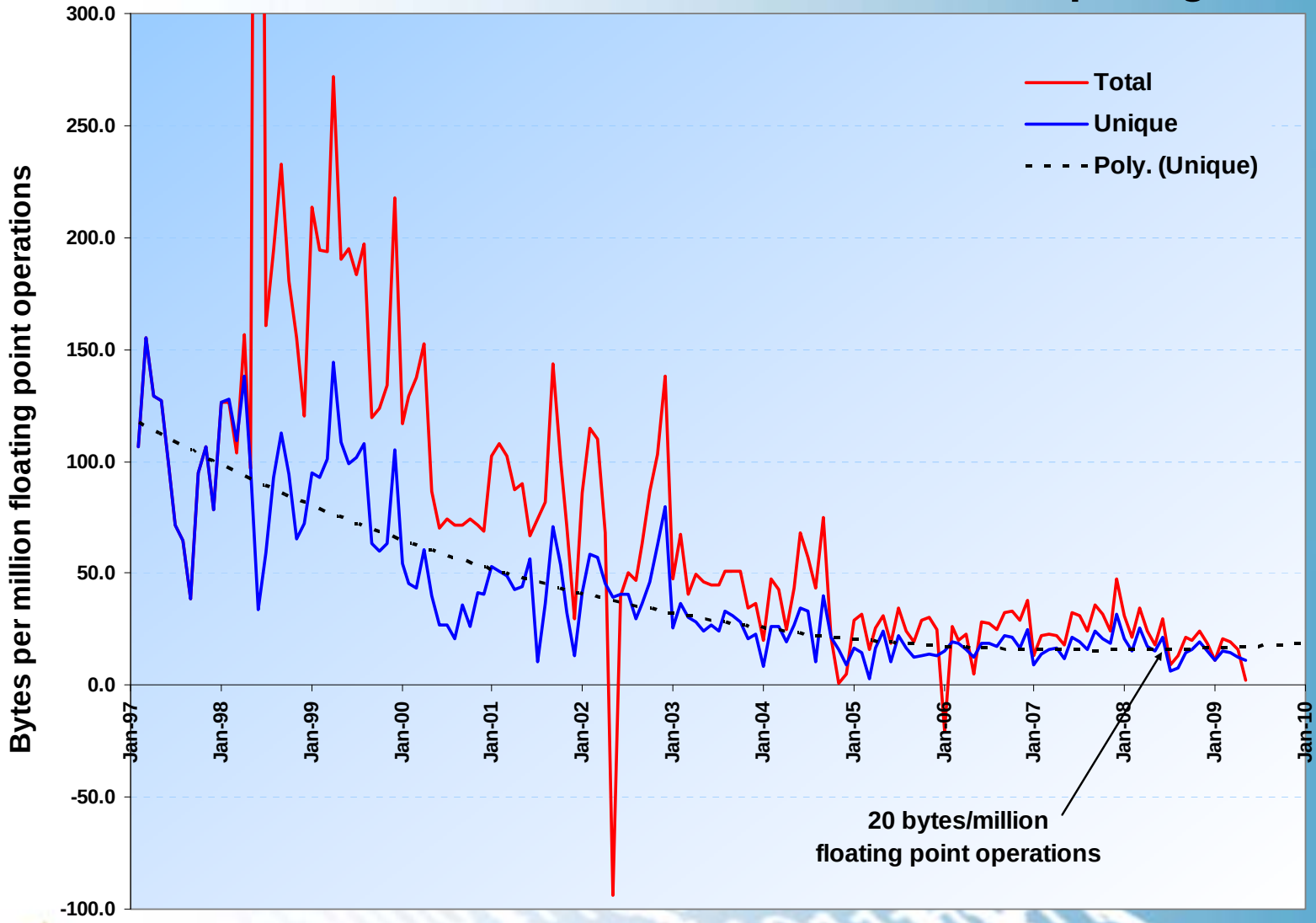
NCAR MSS - Total Data in Archive



CISL Computational & Information Systems Laboratory

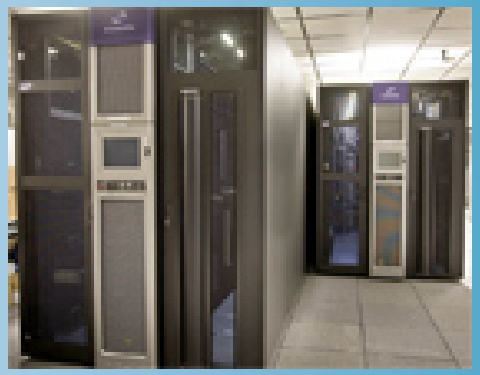
NCAR

NCAR MSS - Net Growth vs. Sustained Computing



MSS Status

May 2009

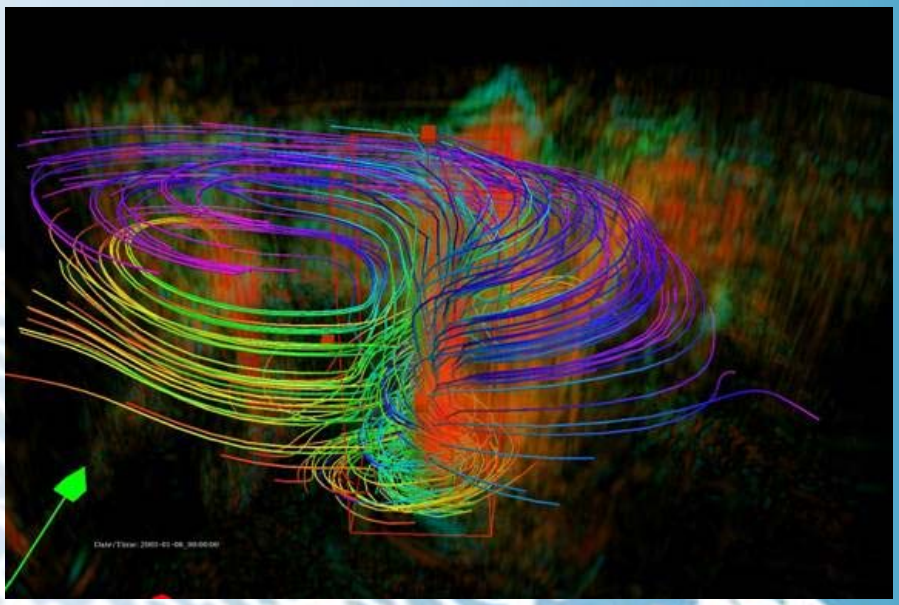


- **AMSTAR Libraries in Production**
  - *MSS Capacity now 14 PB, with plans to upgrade to ~30 PB by 2012*
  - *Maximum file size increased to 100 GB*
- **MSS Metadata Database & Server**
  - *Significant upgrade – performance increased by a factor of 3x (one user reported 10x!)*
- **HPSS Deployed as Production TeraGrid Resource on March 1**
  - *Within six months there will be direct access to HPSS from the TeraGrid*

CAVES Summary

Computational Analysis Visualization Enabled Storage (CAVES)

- Increased total capacity by 100 TB; now totals 324 TB
- 1 GigE GridFTP and HPN-ssh access deployed
- 10 GigE connection to be deployed soon
- GPFS/HPSS HSM interface prototype roll-out in the next 6 months



Evolution of storms near the equator: Julie Caron (CGD) and Alan Norton (CISL), data from WRF NRCM ASD project.

Data Support Section (DSS)

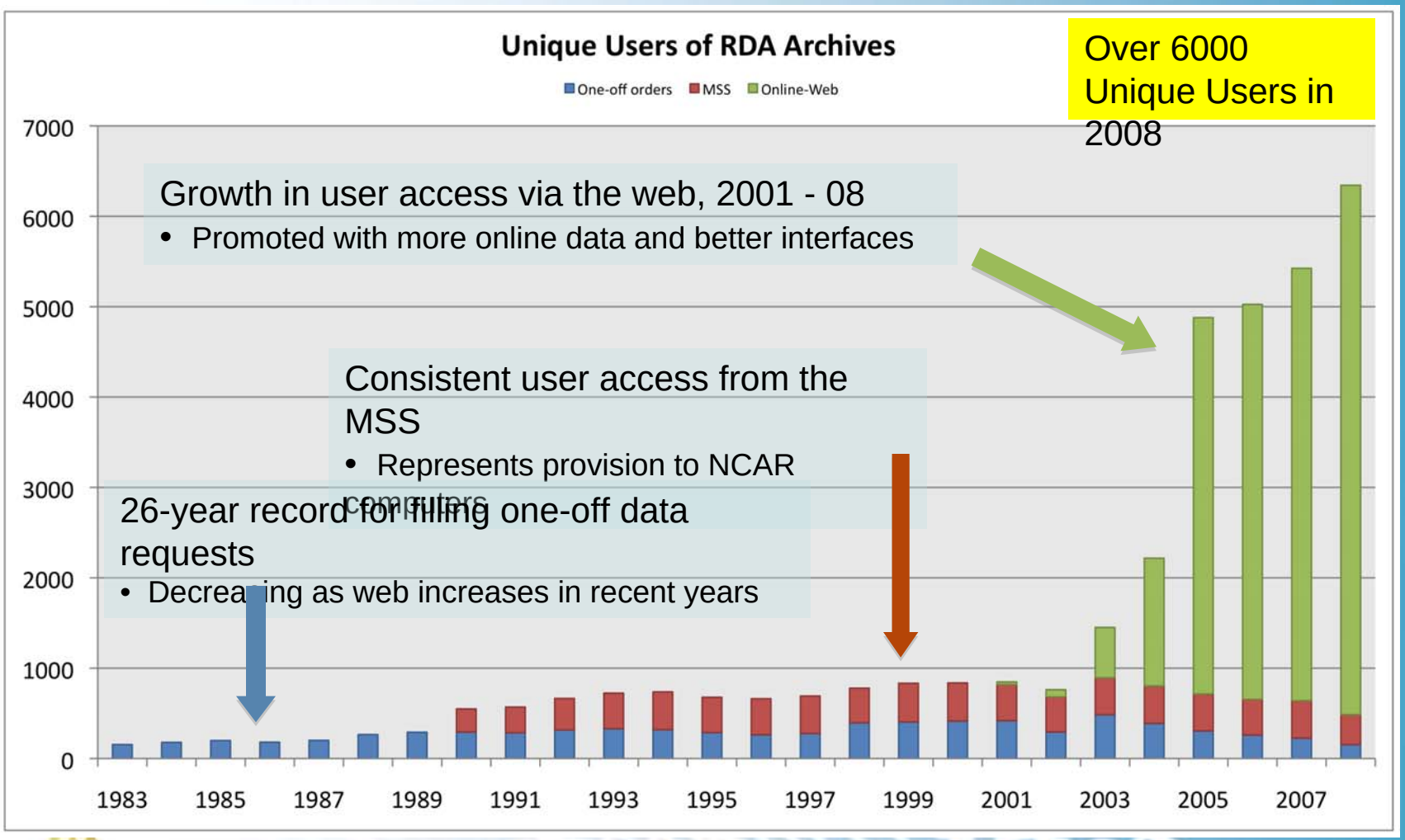
New Features in the Research Data Archive

- **NOAA 20th Century Reanalysis Ver. 1, 1908-1958**
  - *Low resolution (180x90), model T62L28*
  - *Ver. 2, 1891-2000, is in archive preparation @ NCAR now*

- **ERA-Interim reanalysis, 1989-onward**
  - *Arriving @ NCAR now*
  - *CISL is computing a high resolution (512x256) regular gridded product from the model (T255L60 4DVAR) output*

- **New system to offer RDA MSS holdings**
  - *Semi-automatic access promoted first on key datasets (reanalyses, TIGGE) and to the university community*

Long-Term RDA User Metrics



bluefire
Usage and Performance

Bluefire facts & figures

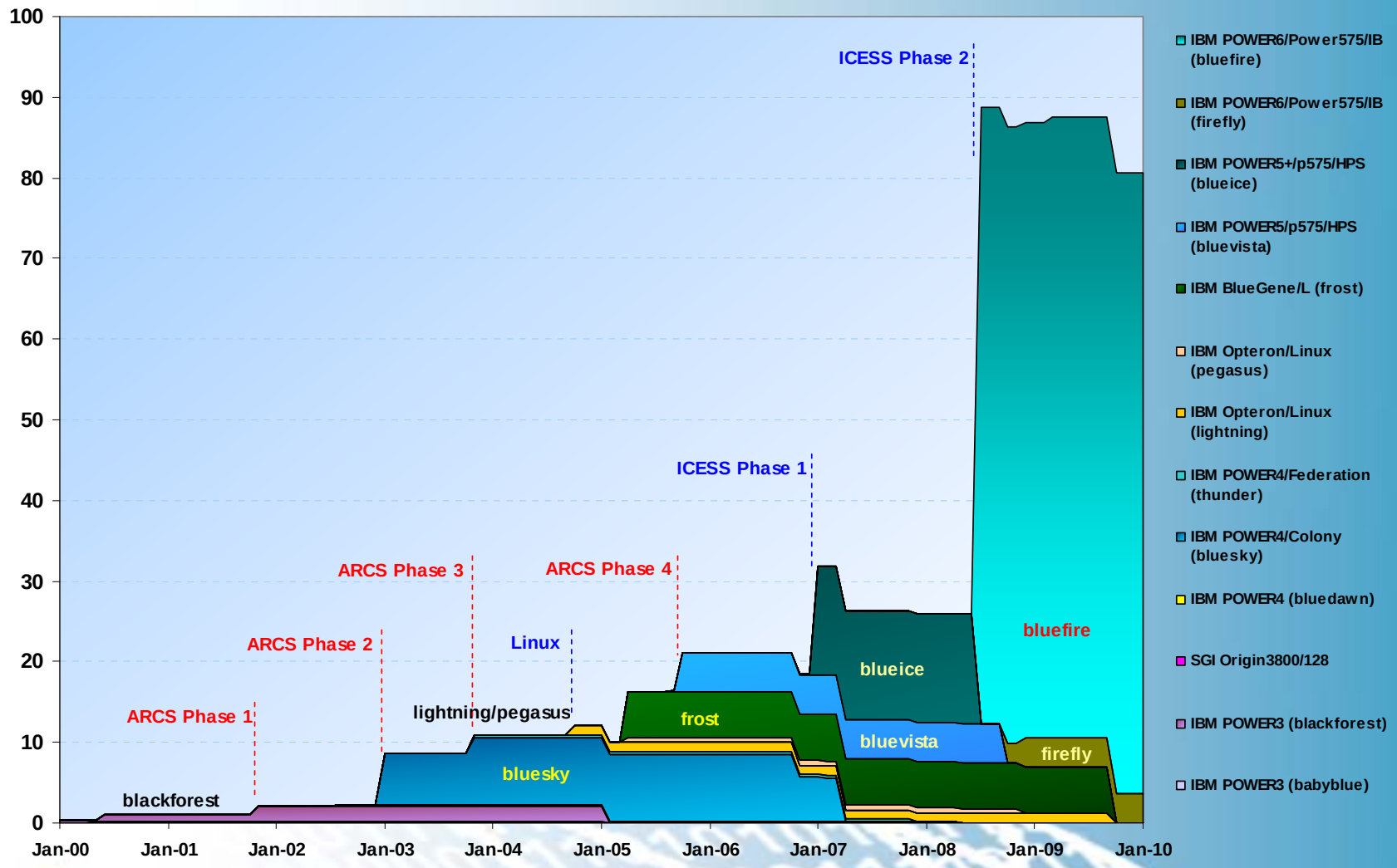
- **Undergoing installation one year ago**
- **Entered initial, early production July 2, 2008**
- **ASD: prep Aug 2008, prod Sep-Nov 2008**
- **Full University / NCAR / CSL production Dec 1, 2008**
- **Lifetime Availability: 95.0% (96.8% Vendor)**
- **Upgrades**
  - *AMPS*
    - 2 nodes added to firefly
    - one additional batch node added to bluefire (128 total)
  - *May 2-4 weekend upgrade*
    - AIX SP 6.2, p575 firmware, InfiniBand firmware, MPI libraries
    - Most complex (and well-planned) upgrade ever @ NCAR
    - improved stability & performance
  - *~1 June: 2 service nodes -> batch*
    - Batch capacity: 120 nodes, 3840 CPUs, 72 TFLOPs
- **Dec 12, 2008 - UCAR Outstanding Accomplishment Award for Scientific & Technical Advancement awarded to the Bluefire Facilities Install Team**
  - *... for designing, implementing and commissioning significant new Mesa Lab infrastructure to support bluefire*

Accelerated Scientific Discovery (ASD)

- **September – November 2008**
  - *3.9 M GAUs allocated, 4.0 M GAUs used*

PI	Institution	Title	Initial Allocation	GAUs Used
CHAP Proposals				
Tulich	CU	Convection-Wave Interactions	360,000	399,838
McClean	Scripps	Eddy-Induced Tracer Variability	489,309	620,812
Trapp	Purdue	CC Radiative Forced Convective Storms	399,000	405,808
NSF Proposals				
Ridley	Univ of Michigan	Small-scale Processes – Corona, Magnetosphere, and Ionosphere.	645,000	298,490
Shay	Univ of Delaware	Collisionless Magnetic Reconnection in Magnetosphere	336,000	345,358
NCAR Proposals				
Rasmussen	NCAR	Modeling Winter Precip Processes	500,000	510,579
Holland	NCAR	Nested Regional Climate Modeling	500,000	549,266
Mininni	NCAR	Rotation and Helicity in Turbulent Flows	700,000	884,079
		Totals	**3,929,309**	**4,014,230**

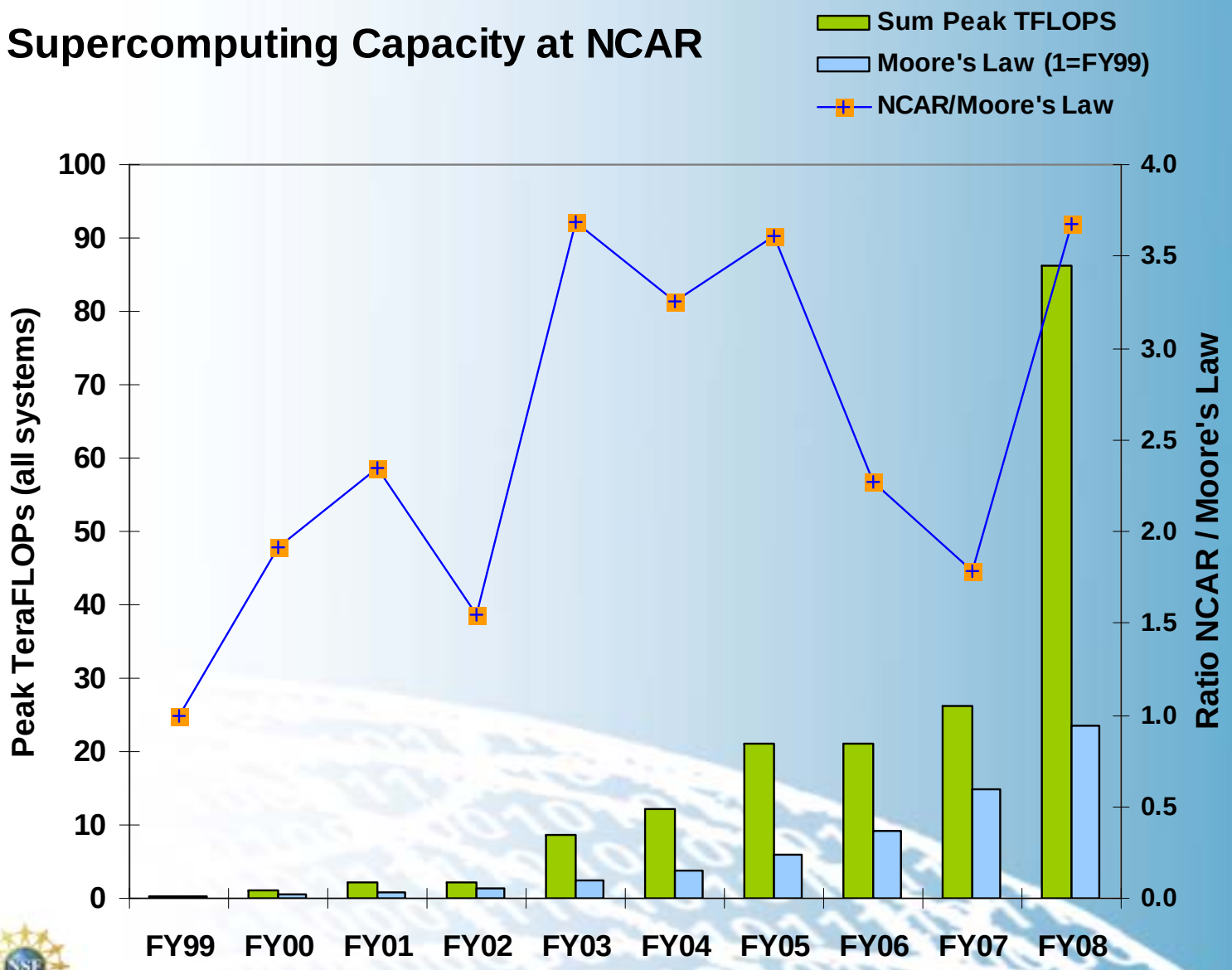
Peak TFLOPs at NCAR (All Systems)



CISL Computational & Information Systems Laboratory

NCAR

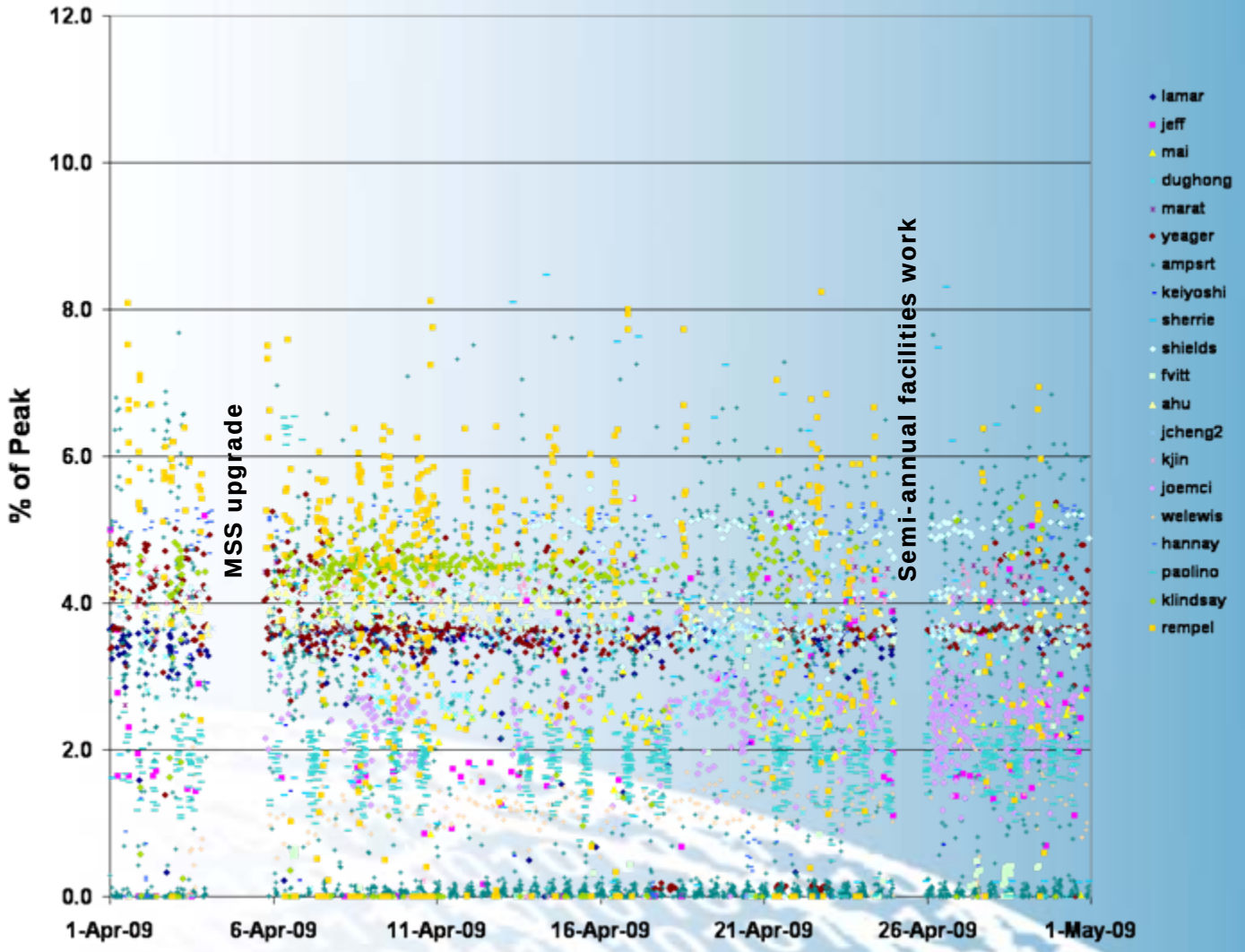
Supercomputing Capacity at NCAR



CISL Computational & Information Systems Laboratory

NCAR

Job Efficiency of Top 20 Users on bluefire during Apr 2009



Average bluefire job floating point efficiency: 4.66%
It's still early, but: we have observed ~0.5% improvement in fp efficiency since the "6.2/IB/firmware upgrade"

NCAR

Benchmarking University Codes

Soon after the POWER6 (bluefire) came online, CSG found that performance lagged below POWER5 (blueice) for many codes

- **Advertised the "Big 3" Performance Improvement Techniques to Users in July/August**
  - *Processor binding (most important for performance)*
  - *Simultaneous Multi-threading (SMT)*
  - *64K page sizes*
- **Benchmark was required for 15 projects receiving more than 120K GAUs from CHAP**
  - *First systematic effort by CISL to work with universities to improve code performance (ala BTS & ASD)*
  - *10K GAUS given initially for testing & benchmarking*
  - *CISL provided documentation on implementing the "Big 3" in the panel allocation letter*
- **All 15 projects supplied at least one benchmark**

Benchmarking University Codes (cont'd)

Goal: Improve "average sustained performance" on bluefire (floating-point "asp" currently 4.66%)

- Six projects submitted initial benchmark results demonstrating satisfactory code performance
  - Straus, Schneider, Paolino, Randall, Khairoutdinov, Fox-Kemper

Challenges

- Ensemble models difficult to run efficiently
  - LSF scheduler lacks support for multiple instances of the mpirun command
- Sustained floating-point performance a poor indicator for data assimilation models
  - CSG worked with Manganello (COLA) to profile phases and optimize (now achieving 3.1-3.3% of peak)
- Parallel I/O essential to some codes

Noteworthy Results

- 50% improvement on Montgomery project
- 42% improvement on Chang project
- 2 year/day to 9.8 year/day for Kirk-Davidoff project
  - allocation was based on 10.5 year/day

Questions and Discussion


CISL
Computational & Information Systems Laboratory
NSF
NCAR

Networking Engineering and Telecommunications Section (NETS)

- **NETS deployed approximately 2000 IPT phones over five months**
  - *The phones were at or beyond their estimated 5 year life*
  - *We had been having ongoing and increasing problems with hook-switch failures on the phones*
  - *Phone network port supports GigE to the desktop - part of NETS Strategic Plan to provide GigE to the desktop*
  - *Required a complimentary project to install new GigE Ethernet switch cards, power supplies, and cable management to support ports requiring increased power over Ethernet*
- **Expanding 10Gbps support in the ML Computer Room**
- **Upgrading the Bi-State Optical Network (BiSON) to 10Gbps**

bluefire Installation/Deployment

- **24 April 2008: *bluefire* Delivered**
  - ***First P6 Delivery/Installation Worldwide***
  - ***117 compute nodes delivered (127 total)***
    - **60+ TFLOPs peak planned 76 TFLOPs peak delivered**

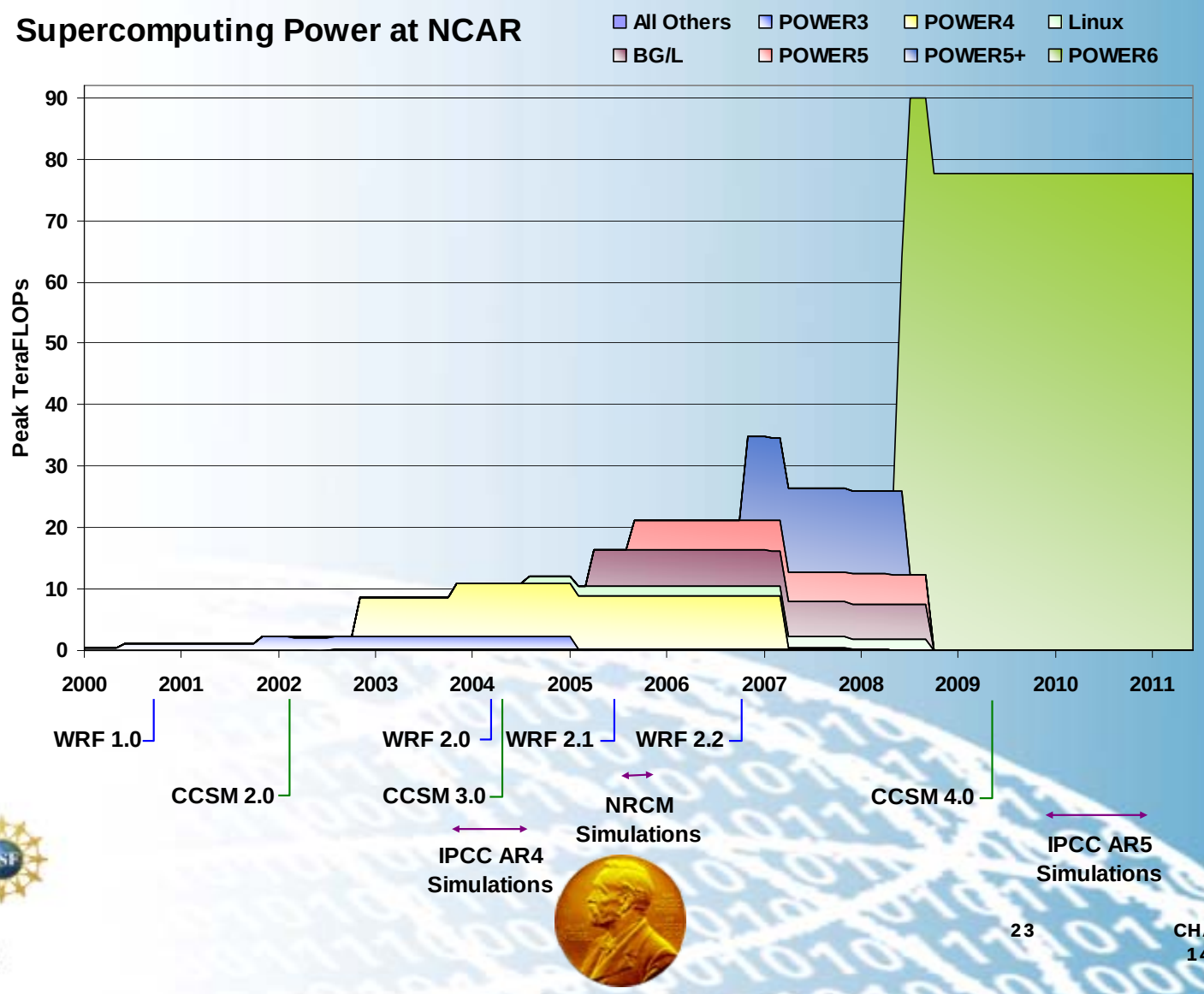
- **8 June 2008: *bluefire* Stage 1 Accepted**
- **14 June 2008: *blueice* Decommissioned**
- **16 June 2008: *bluefire* Stage 2 Power-up**
- **18 June 2008: *bluefire* Debuts at #30 on Top500**

- **30 June 2008: *bluefire* Accepted**
  - ***NCAR and IBM staff worked tirelessly to achieve this goal***

- **29 Sept 2008: *bluevista* Decommissioned**

- **12 December 2008: Bluefire Facilities Install Team wins UCAR Outstanding Accomplishment Award for Scientific & Technical Advancement**
  - *... for designing, implementing and commissioning significant new Mesa Lab infrastructure to support bluefire*

Peak TFLOPS at NCAR (Production Systems)



AMSTAR Contract

- **Facts and Figures**
  - ***New libraries to replace silos– SL8500***
  - ***Expansion to over 30 PB by end of 2012 in six phases into <u>three</u> 10,000 slot SL8500s***
  - ***Immediate expansion from 200 GB capacity tapes to 1 TB tapes***
  - ***Two-year move of data from silos to SL8500s***
- **Schedule**
  - ***1 Dec 2008 Production Deployment 8 PB***
  - ***1 Sep 2009 Capacity increased to 12.2 PB***
  - ***1 Sep 2010 Capacity increased to 18.2 PB***
  - ***1 Sep 2011 Capacity increased to 24.5 PB***
  - ***1 Sep 2012 Capacity increased to 30 PB***
    - *With compression, etc., could be > 37.5 PB*