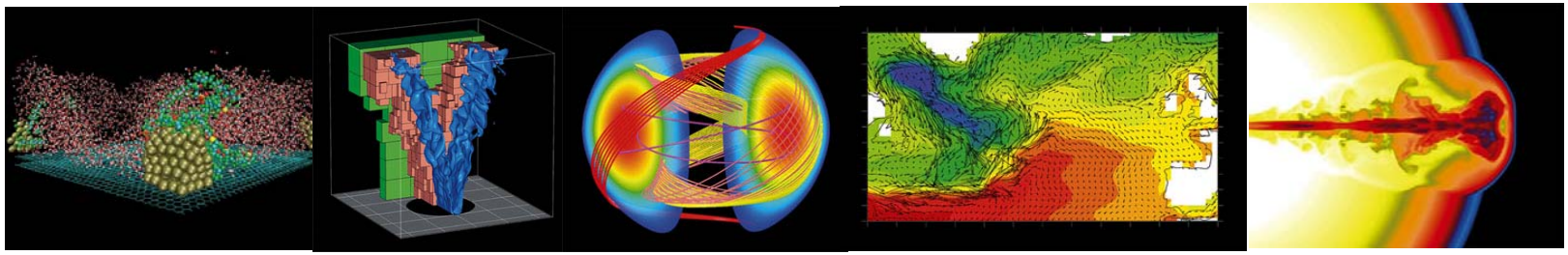


Franklin: User Experiences

Helen He, William Kramer,
Jonathan Carter, Nicholas Cardo

Cray User Group Meeting
May 5-8, 2008

BERKELEY LAB

NERSC

Office of Science
U.S. DEPARTMENT OF ENERGY

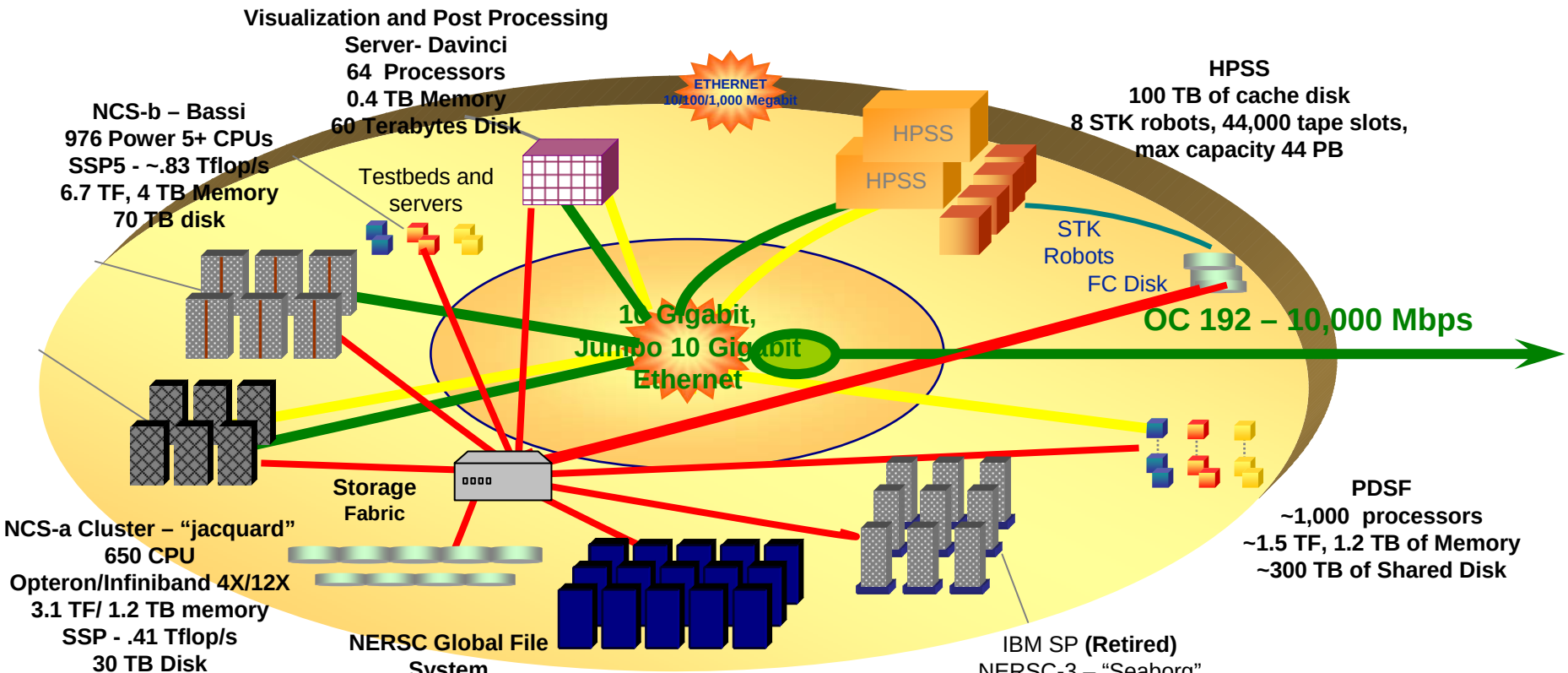
Outline

- Introduction
- Franklin Early User Program
- CVN *vs.* CLE
- Franklin Into Production
- Selected Successful User Stories
- Top Issues Affecting User Experiences
- Other Topics
- Summary

Franklin

Benjamin Franklin, one of America's first scientists, performed ground breaking work in energy efficiency, electricity, materials, climate, ocean currents, transportation, health, medicine, acoustics and heat transfer.

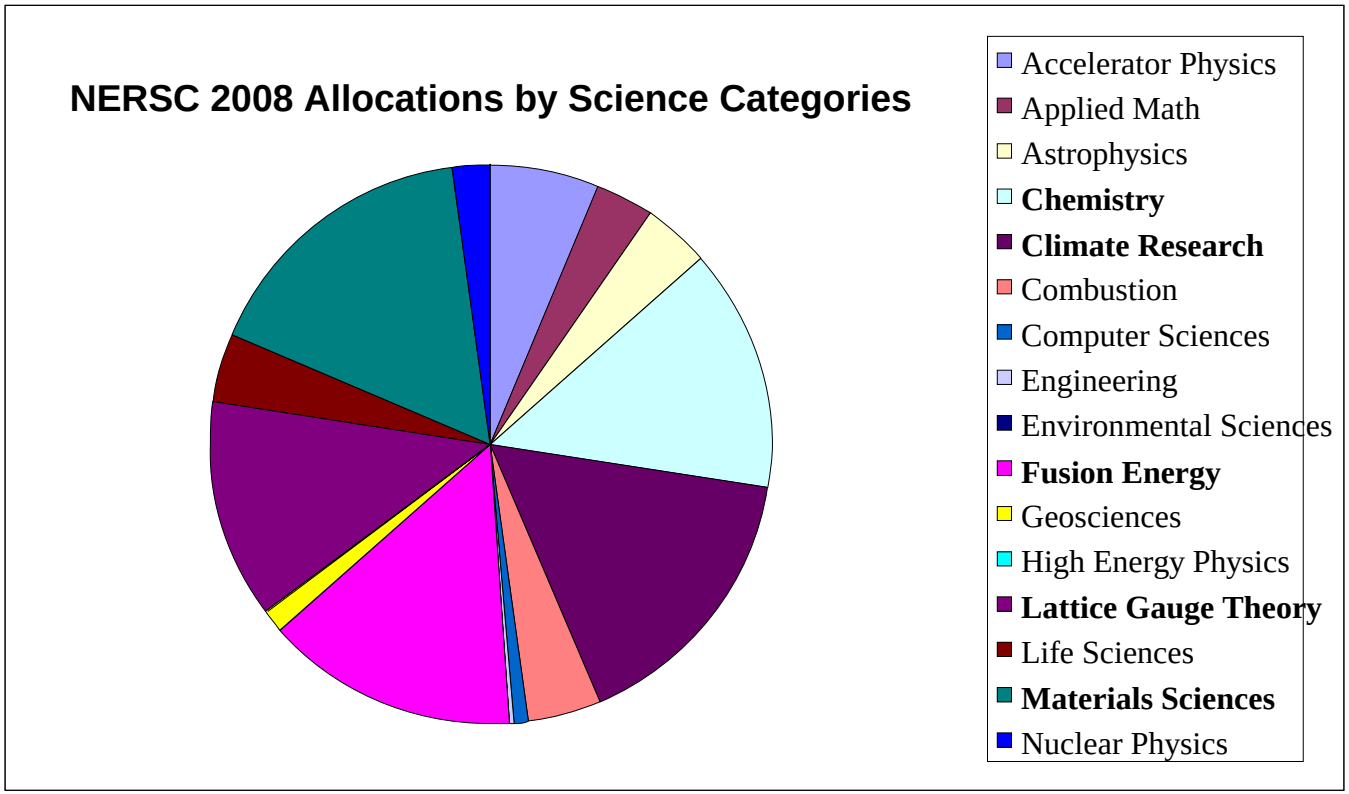
NERSC Systems



Franklin's Role at NERSC

- NERSC is US DOE's keystone high performance computing center.
- Franklin is the "flagship" system at NERSC after Seaborg (IBM SP3) retired after 7-years in January 2008.
- Increased available computing time by a factor of 9 for our ~3,100 scientific users.
- Serves the needs for most NERSC users from modest to extreme concurrencies.
- Expects significant percentage of time to be used for capability jobs on Franklin.

Allocation by Science Categories



- Large variety of applications.
- Different performance requirements in CPU, memory, network and IO.

Number of Awarded Projects

Allocation Year	Production	INCITE & Big Splash	SciDAC	Startup
2008	275	11	47	40
2007	291	7	45	44
2006	286	3	36	70
2005	277	3	31	60
2004	257	3	29	83
2003	235	3	21	76

NERSC was the first DOE site to support INCITE and is in its 6th year.

About Franklin

- 9,736 nodes with 19,472 CPU (cores)
- dual-core AMD Opteron 2.6 GHz, 5.2 GFlops/sec peak
- 102 node cabinets
- 101.5 Tflop/s theoretical system peak performance
- 16 KWs per cabinet (~1.7 MWs total)
- 39 TBs aggregate memory
- 18.5+ Tflop/s Sustained System Performance (SSP)
  (Seaborg - ~0.98, Bassi - ~0.83)
- Cray SeaStar2 / 3D Torus interconnect (17x24x24)
  - 7.6 GB/s peak bi-directional bandwidth per link
  - 52 nanosecond per link latency
  - 6.3 TB/s bi-section bandwidth
  - MPI latency ~ 8 us
- ~350 TBs of usable shared disk

Software Configuration

- SuSE SLES 9.2 Linux with a SLES 10 kernel on service nodes
- Cray Linux Environment (CLE) for all compute nodes
  - Cray's light weight Linux kernel
- Portals communication layer
  - MPI, Shmem, OpenMP
- Lustre Parallel File System
- Torque resource management system with the Moab scheduler
- ALPS utility to launch compute node applications

Programming Environment

- PGI compilers: assembler, Fortran, C, and C++
- Pathscale compilers: Fortran, C, and C++
- GNU compilers: C, C++, and Fortran F77
- Parallel Programming Models: Cray MPICH2 MPI, Cray SHMEM, and OpenMP
- AMD Core Math Library (ACML): BLAS, LAPACK, FFT, Math transcendental libraries, Random Number generators, GNU Fortran libraries
- LibSci scientific library: ScaLAPACK, BLACS, SuperLU
- A special port of the glibc GNU C library routines for compute node applications
- Craypat and Cray Apprentice2
- Performance API (PAPI)
- Modules
- Distributed Debugging Tool (DDT)

NERSC User Services

- Problem management and consulting.
- Help with user code debugging, optimization and scaling.
- Benchmarking and system performance monitoring.
- Strategic projects support.
- Documentation, user education and training.
- Third-party applications and library support.
- Involvement in NERSC system procurements.

Early User Program

- NERSC has a diverse user base compared to most other computing centers.
- Early users could help us to mimic production work load, identify system problems.
- Early user program is designed to bring users in batches.
- Gradually increase user base as system is more stable.

Enabling Early Users

- **Pre-early users (~100 users)**
  - **Batch 1, enabled first week in March 2007**
    - **Core NERSC staff**
  - **Batch 2, enabled second week in March 2007**
    - **Additional NERSC staff**
    - **A few invited Petascale projects.**
- **Early users (~150 users)**
  - **Solicitation email sent in end of Feb 2007**
  - **Reviewed, approved, or deferred each application.**
    - **Criteria: User codes easily ported to and ready to run on Franklin.**
    - **Successful requests formed Batch 3 users.**
    - **Further categorized into sub-batches for the balance of science category, scale range and IO need, etc. Each sub-batch has about 30 users.**
  - **Batch 3a, enabled early July 2007.**
  - **Batch 3b, enabled mid July 2007.**
  - **Batch 3c, enabled early Aug 2007.**
  - **Batch 3d, enabled late Aug 2007.**
  - **Batch 3e, enabled early Sept 2007.**

Enabling Early Users (cont'd)

- Early users (cont'd)
  - Batch 4, enabled mid Sept 2007.
    - Requested early access, but dropped or deferred.
  - Batch 5, enabled Sept 17-20, 2007.
    - Registered NERSC User Group meeting and user training.
  - Batch 6, enabled Sept 20-23, 2007.
    - A few other users requested access.
  - Batch 7, enabled Sept 24-27, 2007.
    - All remaining NERSC users.

Pre-Early User Period

- Lasted from early March to early July.
- Created franklin-early-users email list. Written web pages for compiling and running jobs, and quick start guide.
- Issues in this period (all fixed):
  - Defective memory replacement, March 22 – April 3.
  - File loss problem, April 10-25.
  - File system reconfiguration, May 18-June 6.
  - Applications with heavy IO crashed the system. Reproduced and fixed the problem with “simple IO” test using full machine.
- NERSC and Cray collaboration “Scout Effort” brought in total of 8 new applications and/or new inputs.
- Installed CLE in the first week of June, 2007.
- Decision made to forward with CLE for additional evaluation and entering Franklin acceptance with CLE.

CVN *vs.* CLE

- CLE was installed on Franklin the week it was released from Cray development, which was ahead of its original schedule.
- CLE is the path forward eventually, so better for our users not have to go through additional step of CVN.
- More CLE advantages over CVN
  - Easier to port from other platforms with more OS functionalities and a richer set of GNU C libraries.
  - Quicker compiles (at least in some cases)
  - A path to other needed functions:
    - OpenMP, pthreads, Lustre failover, and Checkpoing/Restart.
  - Requirement for quad-core upgrade
  - More options for debugging tools
  - Potential for Franklin to be on NGF sooner

CVN *vs.* CLE (cont'd)

- CLE disadvantages
  - More OS footprint, ~extra 170 MB from our measurement.
  - Slightly higher MPI latencies for farthest intra-node.
- Holistic evaluation between CVN and CLE after several months on Franklin for each OS concluded:
  - CLE showed benefits over CVN in performance, scalability, reliability and usability.
  - CLE showed slightly, acceptable decreases in consistency.
- Mitigated risks, benefited DOE and other sites for their system upgrade plans.

Early User Period

- Lasted from early July to late Sept 2007.
- Franklin compute nodes running CLE.
- User feedback collected from Aug 9 to Sept 5, 2007.
- Top projects used over 3M CPU hours.
- Franklin user training from Sept 17-20, 2007.
- Issues in this period (all fixed):
  - NWCHEM and GAMESS crashed system
    - Both use SHMEM for message passing
    - Cray provided first patch to trap the shmem portals usage, exit user code.
    - Second patch solved the problem by throttling messages traffic.
  - Compute nodes lose connection after application started
  - Jobs intermittently run over the wallclock limit.
  - A problem related to a difficulty in allocating large contiguous memory in the portals level.
  - Specifying the node list option for aprun did not work.
  - aprun MPMD mode did not work in batch mode.
- User quota enabled Oct. 14, 2007.
  - Quota bug of not being able to set over 3.78 TB (fixed).
- Queue structure simplified to have only 3 instead of original 10+ buckets for the “regular” queue.

After Acceptance Early User Period

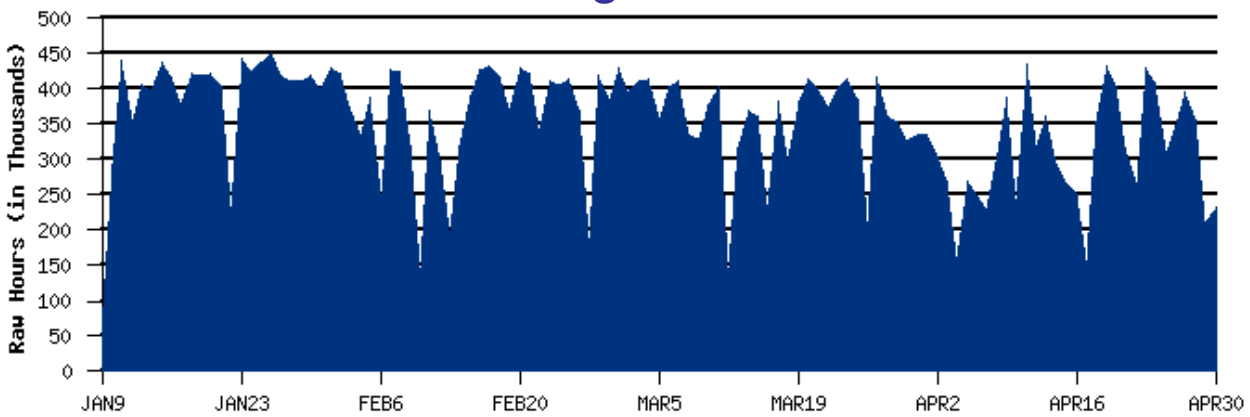
- Lasted from late Sept 2007 to Jan 8, 2008.
- Franklin accepted Oct 26, 2007.
- User feedback collected from Nov 1-26, 2007.
- Accommodated some very large applications and massive amounts of time.
- Batch queue backlogs started to grow. Small to medium jobs showed large delays. Idle and global run limits were modified to address above issues.
- Issues in this period (3 of 4 fixed):
  - Inode quota bug occasionally. Could not cross over certain inode bucket boundaries.
  - "Exit codes: 13" problem related resulted from implicit barrier from MPI_allreduce not functioning properly.
  - Jobs intermittently over wall clock limit due to bad memory nodes left by previously over-subscribing memory jobs.
  - Multiple apruns simultaneously did not work in batch.

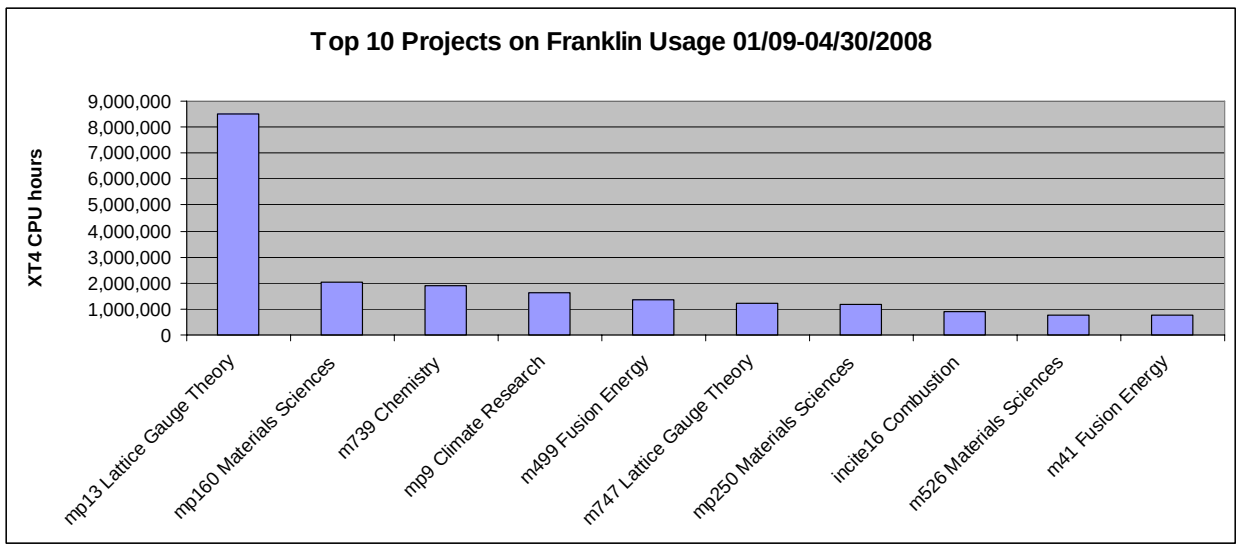
Selected Early Users Feedback

- Overall user feedback was quite positive.
- Most applications were relatively easy to port to Franklin.
- Familiar user environment, batch system working well.
- 51 science projects participated. Free of charge from user allocations. Many able to run high concurrency jobs for large problems that were impossible before.
- Broader range of user applications helped to identify problems, and providing fixes.
- Selected User feedbacks:
  - "Franklin has been easy to use in both programming (porting) and running codes. I am very pleased and impressed with the quality of this machine. I believe it is an exceptional asset to the computational physics community in the US."
  - "The friendly user period on Franklin has significantly impacted our science by allowing us to test the capabilities of our code and to establish that such high resolution simulations will be useful and constructive in understanding within-canopy turbulent transport of carbon dioxide."
  - "I have been able to compile and run large scaling studies with very few problems. The queuing system has worked well and I have not had any problems with libraries, etc."
  - "Overall, I am impressed with the performance and reliability of Franklin during the testing stage."

Franklin Into Production

Franklin Usage 01/09-04/30/2008

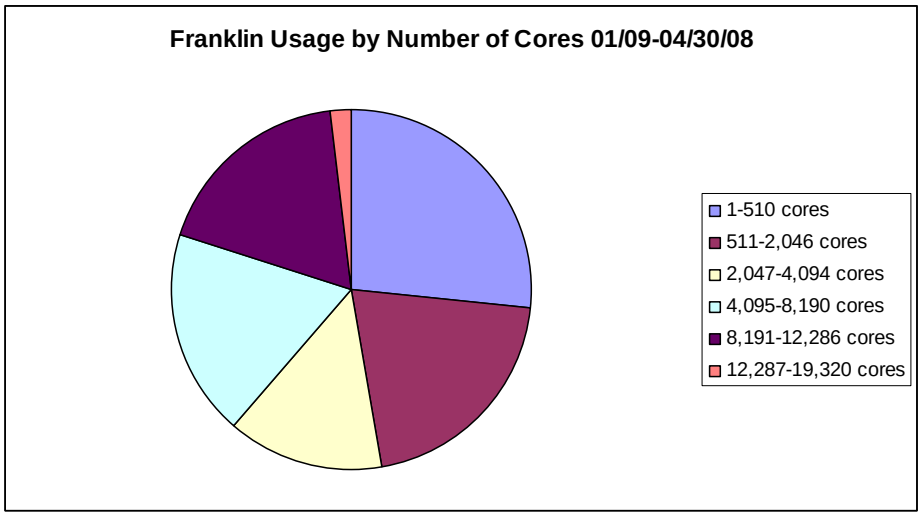


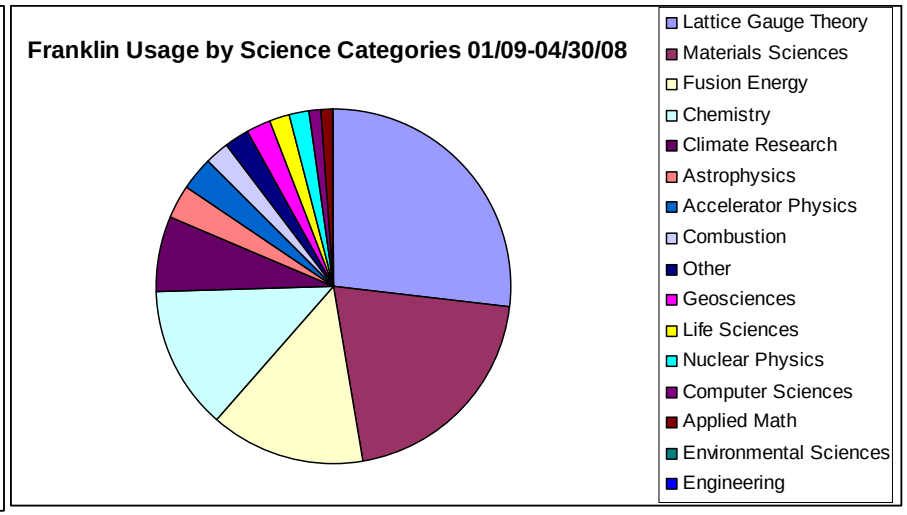


Franklin Into Production

- System utilization was high except during certain periods in April.
- Top 10 projects have used over 20M CPU hours.
- Scaling reimbursement program
  - DOE set aside 26M MPP hours (equivalent of 4M Franklin CPU hours).
  - Help projects to understand and improve the scaling characteristics of their codes, and to scale efficiently to 2,416+ processors (1,208+ nodes).
  - NERSC has to meet the DOE metric that at least 40% of the time used on Franklin is by jobs running on 1/8th or more of its processors.
  - 23 users now enrolled in this year's program.
  - Opened for all projects on May 6, 2008. (2M MPP hours cap per project)

Franklin Production Usage



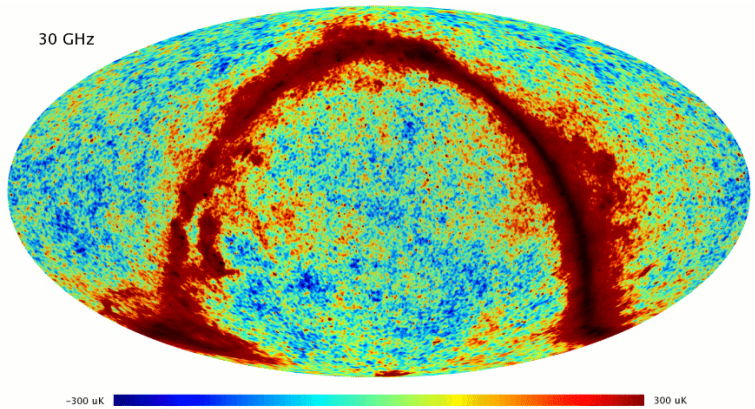


- **Over 50% jobs use >= 2,048 cores.**
- **Top 5 science categories used are Lattice Gauge Theory, Material Sciences, Fusion Energy, Chemistry, and Climate Research.**

Planck Cosmic Microwave Background Map Constructed Using Franklin

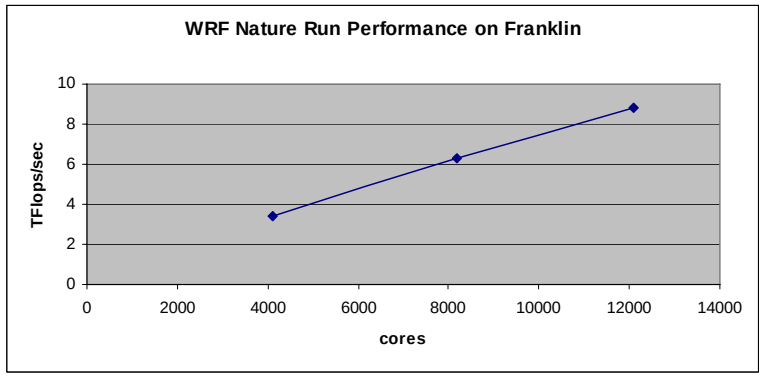
- PI Julian Borrill, Berkeley Lab
- Uses the massively parallel MADmap code
  - A PCG solver for the maximum likelihood map given the measured noise statistics.

Planck Full Focal Plane Map

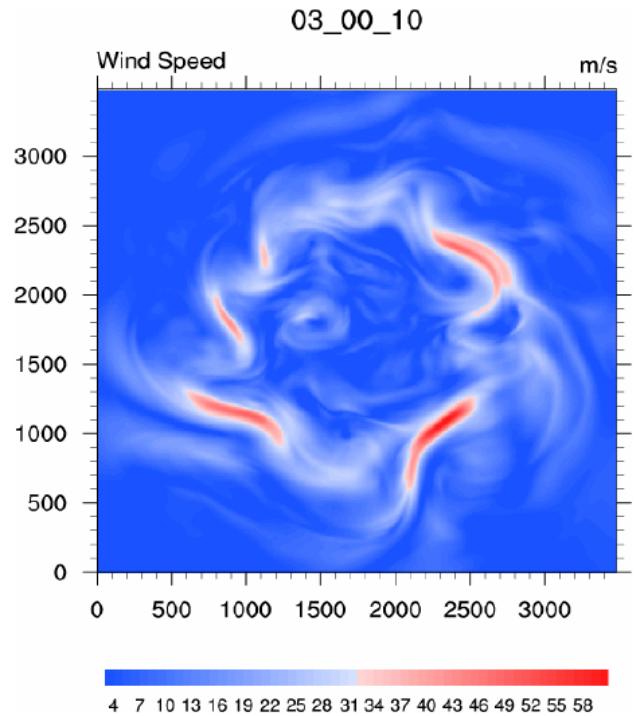


- 2007: First map-making of one mission year of Planck data from all detectors at all frequencies (100% data)
  - 750 billion observations mapped to 1.5 billion pixels (74 detectors, 3TB, 50K files)
  - Early user access to 10,000 cores of Franklin.
  - Previously intractable calculation.
  - "This is the first time that so many data samples have been analyzed simultaneously, and doing so has been the primary goal of our group's early Franklin efforts. "
- The team also developed MADbench2, a stripped down MADcap code
  - Retains full computational complexity (calculation, communication, and IO).
  - Removed scientific complexity using self-generated pseudo data.
  - Used in procurement for Franklin.
  - First user application crashed the system.

WRF Nature Run on Franklin



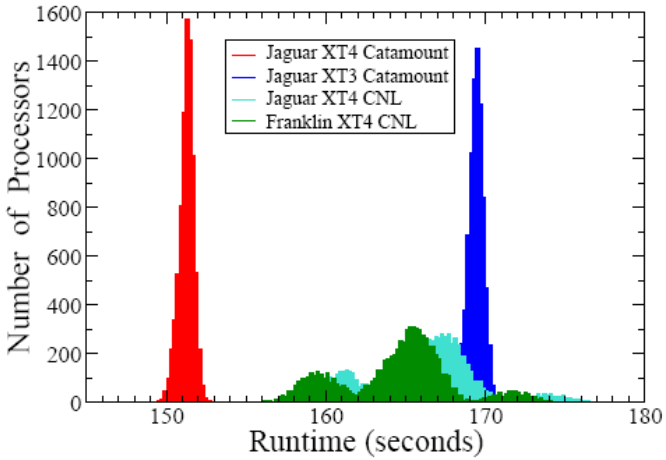
- A team from NCAR, SDSC, LLNL and IBM.
- WRF model is an atmosphere model for mesoscale research and operational numerical weather prediction.
- The nature run involves an idealized high resolution rotating fluid on the hemisphere; at a size and resolution never before attempted – 2 billion cells @ 5km resolution
- Huge volume of data! – 200GB input and 40GB per simulated hour output.
- 8.8 TF @ 12,090 cores, 6.27 TF @ 8192 cores.
- The fastest performance of any weather model on a US computer. SC07 Gordon Bell finalist.



WRF Nature Run with 5km (idealized) resolution captures large scale structure such as Rossby Waves. (Courtesy Wright)

Science Goal: To provide very high-resolution "truth" against which more coarse simulations or perturbation runs may be compared for purposes of studying predictability, stochastic parameterization, and fundamental dynamics.

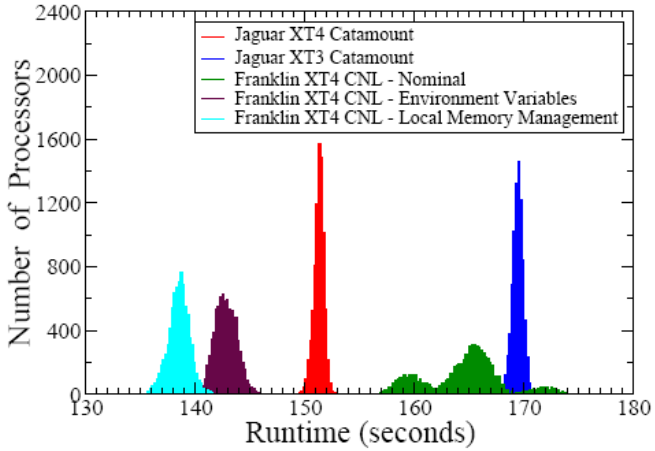
OS Jitter or Something Else?



Histogram of run time on Jaguar and Franklin.
(Courtesy Van Straalen *et al.*)

- **This story is presented here to illustrate the positive and healthy vendor interaction with NERSC users.**
- **Chombo AMR team created an embarrassingly parallel benchmark to study the potential of strong scaling for AMR benchmarks**
  - **Extracting a Fortran kernel from the AMR gas dynamic code**
  - **Assigning same workload for each proc**
  - **No IO, no MPI, no communication barriers, no system calls.**
  - **Expect to see almost perfect load balancing.**
- **Initial results showed 3-peak distribution on Franklin and Jaguar CLE, but not on Jaguar CVN.**

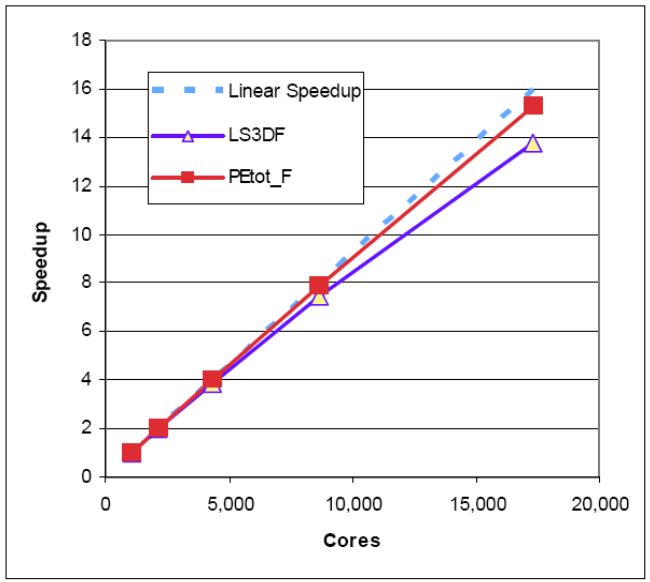
OS Jitter or Something Else?



Histogram of run time on Jaguar and Franklin with CLE malloc environment variable setting and with AMR local memory management show a reduced single peak distribution. (Courtesy Van Straalen *et al.)*

- **Chombo team met with Cray on-site support**
  - **Hypothesis: CLE has a more sophisticated, but stochastic, heap manager than CVN.**
  - **Test 1: simplify heap manager by setting two malloc env variables:**
    - **MALLOC_MMAP_MAX_ and MALLOC_TRIM_THRESHOLD_.**
  - **Test 2: Change the order of memory allocation and free operations. Chombo was tested with its own memory allocation routine "CArena".**
  - **Both worked. Only 1-peak and much faster!**
  - **However, why under CLE with heap manager removed, the run time variation is still twice larger than under CVN?**
- **User wrote: "This Franklin research is some of the best vendor interaction I have had in my time using a supercomputer", and thanks for "taking us seriously, and being careful and open, and honest vendor collaborators".**

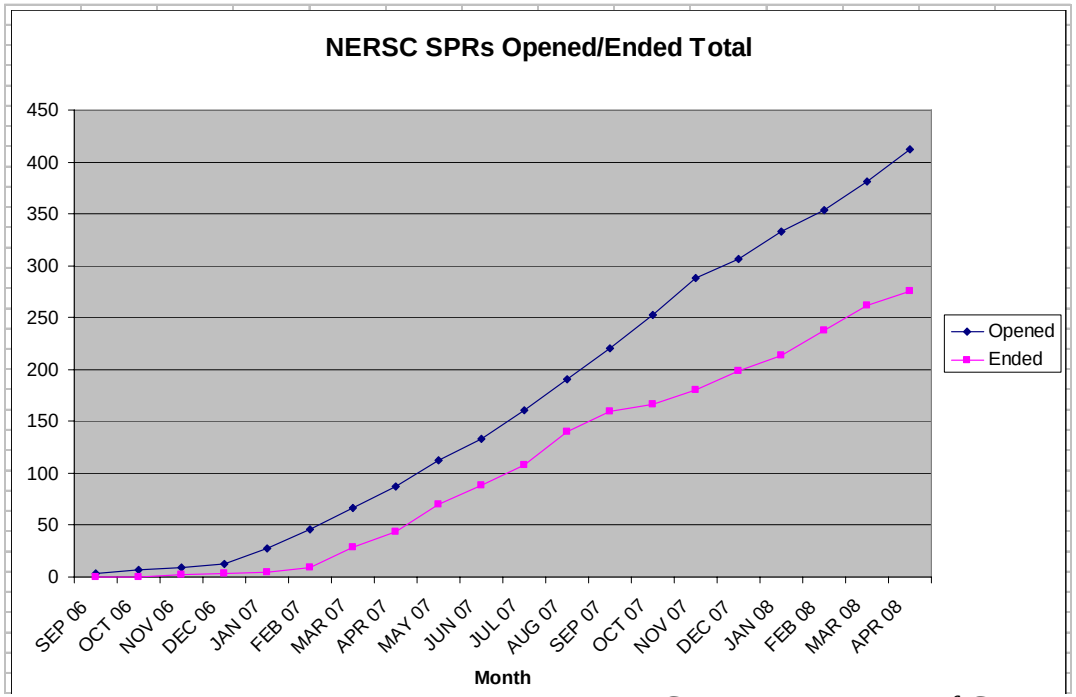
Large Scale Electronic Calculations (LS3DF)



LS3DF and PEtot_F speedup. The speedup (and parallel efficiency) for 17,280 cores (from the base 1,080-core run) for PEtot_F and LS3DF are 15.3 (and 95.8%) and 13.8 (and 86.3%), respectively. (Courtesy Wang *et al.*)

- **PI: Lin-Wang Wang, LBNL.**
- **LS3DF model is an O(N) method (compared to conventional O(N3) methods) for large scale *ab initio* electronic structure calculations**
- **It uses a divide-and-conquer approach to calculate the total energy self-consistently on each subdivision of a physical system.**
- **Almost perfect scaling for higher numbers of processors. Achieved 35.1 TFlop/sec, 39% of the peak speed, using 17,280 cores on Franklin.**
- **Submitted to SC08 Gordon Bell competition.**
- **LS3DF is capable of simulating tens of thousands of atoms, and is a candidate for petascale computing when the computing hardware is ready.**

NERSC SPRs Filed and Fixed

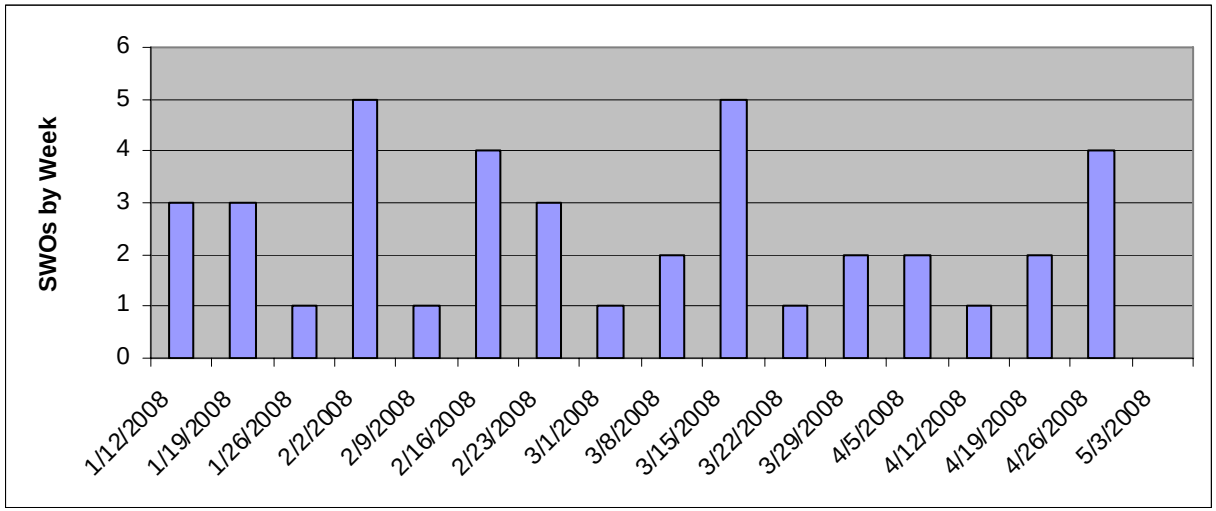


Courtesy Dan Unger of Cray

- **Increased gap since October 2007 when all NERSC users were enabled.**
- **The large number of problems being solved is a great credit to the efforts of Cray development and support teams.**

Top Issues Affecting User Experiences

- **System Stability**



Has been up for >8 days!

- When a system crashes, all jobs die.
- Longer user jobs become unrealistic with short MTBF.
- One heavy user reported 27% of job failure rate due to the combination of system failures, compute node failures, and job hung.
- Frequent system crashes cause light system usages.
  - Running jobs killed.
  - Users not submitting new jobs.

Top Issues Affecting User Experiences (cont'd)

- Shared Login Nodes
  - Prone to crash, with single or combinations of
    - User jobs launched without aprun
    - Large-scale parallel makes
    - Resource intensive scripts such as python or visualization packages
  - Educate users
  - Set 60 min of CPU limit for process.
- Hung Jobs and "Bad" Nodes
  - Jobs hung without aprun started, wall clock limit exceeded.
  - Hypothesis is this could be related to "bad" nodes in the system.
  - Needs better node health checking.
- Job Error Messages
  - Users not getting enough details of why their jobs failed.
  - For example, out-of-memory jobs are killed with no explicit messages in stderr file.

Top Issues Affecting User Experiences (cont'd)

- Quota Related Issues
  - Quota bugs severe enough to crash system prevented us to set user quotas.
  - Inode quota set to 0 on Jan 4, 2008.
  - Space quota set to 0 on Feb 4, 2008.
  - /scratch file system fill up quickly.
  - User Services had to contact users to clean up.
- Slow Interactive Response Time
  - Users report occasional slow interactive response.
  - Maybe associated with heavy IO load on the system.
  - "ls" default set to no coloring to avoid stats of files.
  - Issue to investigate for IO team.
- Run Time Variations
  - Some users reported large run time variations.
  - SPR tracking IOR variations.
  - Issue to investigate for IO team.

DDT *vs.* Totalview

- Totalview is the standard debugger for XT systems. However, it is very expensive.
- The launch of CLE allowed us to evaluate another debugger, Alliena's Distributed Debugging Tool (DDT).
- DDT has very similar interfaces from Totalview
  - Totalview has been on other major NERSC systems.
  - Learning curves for NERSC users are small.
- Major improvements for DDT over Totalview are:
  - Parallel stack view
  - Parallel data comparison
  - Easy group control of processes
- NERSC has DDT as production parallel debugger with a floating license for 1,024 cores.

ACTS PETSc *vs.* Cray PETSc

- DOE ACTS softwares are standard on selected DOE platforms, maintained by ACTS group.
- Cray PETSc has module name conflict with ACTS PETSc.
- ACTS PETSc advantages:
  - More varieties of PETSc modules for different versions: optimized, C++, debug
  - Software likely to be more up-to-date.
  - Has more complete support for ParMETIS, HYPRE, and SuperLU standalone packages.
- Cray PETSc advantages:
  - More official support for the software
  - Performance tuning for XT4 via CASK
  - Support on all three compilers
  - Support for ParMETIS, HYPRE, and SuperLU packages within PETSc
- We like to have both. Is it possible to rename Cray PETSc module?
  - xt-petsc would imply there is x1-petsc or others
  - Issues with existing customers
  - However, the library itself is named craypetsc, maybe we could rename the module name to cray-petsc?
  - Raised the issue at the Applications and Programming Environment SIG.

Summary

- Franklin has delivered large amount of high performance computing resources to NERSC users.
- Produced lots of scientific accomplishments.
- Cray is a company we enjoy to work with: hardworking and efficient.
- Supporting demand for Franklin is still very high during early production period.
- Two teams formed at NERSC in mid April 2008
  - Stability and quality of services issues
    - Stability tiger team
    - Franklin general issues team
  - IO issues: IO tiger team
  - Cray involves in two tiger teams activities.

- Looking forward to an improved Franklin user environment and more satisfied Franklin users.

Acknowledgement

- Cray support teams (on-site and remote) for Franklin.
- NERSC User Services and Systems groups.
- Chombo, LS3DF, WRF, and MADmap groups.
- Authors are supported by the Director, Office of Science, Advanced Scientific Computing Research, U.S. Department of Energy.