

Reasons Cited for Using the IBM POWER 5 System Bassi at NERSC

Richard A. Gerber
NERSC User Services
June 2009

Overview

In March 2009, NERSC asked the top users of Bassi "Why do you use Bassi?" Representatives from 12 projects responded. Bassi is a 111-node IBM POWER 5 (1.9 GHz) system with 8 processors per SMP node for a total of 888 compute cores. Each node has 32 GB of memory. The largest system at NERSC in March 2009 is a Cray XT4 with 19,000 cores. Each Franklin node has a dual-core Opteron processor running at 2.6 GHz with 4 GB of memory.

Summary of User Responses to "Why do you use Bassi?"

Reason Cited	Number
Effort Needed to Port Code	7
System is more stable than Franklin	7
Computational speed is better than Franklin	6
Problems running or porting code	4
Large SMP memory (32 GB)	3
Good programming environment/compilers	1
Faster high-speed network than Franklin	1
Direct access to NGF from compute jobs	1
Job throughput (when combined with Franklin runs)	1
Long (36-hour) queues	1
NAG libraries (not on Franklin)	1

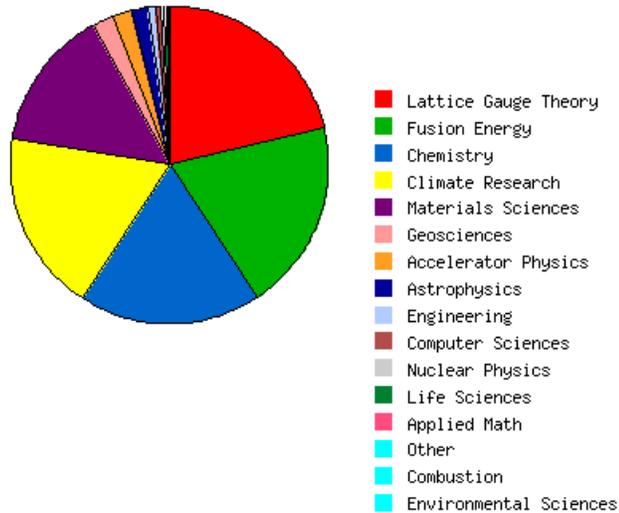
Bassi's Architectural Features

- 32 GB of memory on a single node
- 8-way OpenMP/threads/shared memory model available
- Extremely stable system
- High memory bandwidth (>7 GB/sec per processor on STREAMS TRIAD concurrently using all 8 processors/node)
- High Interconnect (MPI) point-to-point internode bandwidth (>3 GB/sec)
- MPI communication on-node uses shared-memory: extremely fast
- Low Latency Internode Interconnect (~4.6 us)
- High single-core performance (7.6 GFlops/sec theoretical max)
- Mature software environment; good programming environment

Top Projects Using Bassi

Repositories with the most usage on Bassi through the first quarter of the 2009 Allocation Year.

Raw Hours By Science Field



Repo	Project Title	PI	Science Category	Bassi Raw Hours	Bassi Avg. Cores	Franklin Raw Hours	Franklin Avg. Cores
mp27	Lattice Gauge Theory Simulations	Don Sinclair, Argonne National Laboratory	Lattice Gauge Theory	178,190	96	501,395	192
m249	Computational chemistry search of efficient catalysts	Perla Balbuena, Texas A&M University	Chemistry	109,701	32	1,658	32
m542	First principles simulations of nanostructures	Giulia Galli, UC-Davis	Materials Sciences	69,368	64	578,019	256
m328	Global cloud modeling	David Randall, Colorado State University	Climate Research	69,121	80	315,507	10,240
m172	Helicity Injected Torus Current Drive and Compact Toroid Studies	Brian Nelson, University of Washington	Fusion Energy	50,291	112	1,576	64
mp7	Lattice QCD Monte Carlo Calculation of Hadron Structure and Spectroscopy	Keh-Fei Liu, University of Kentucky	Lattice Gauge Theory	45,036	256	459,751	256
m411	Interaction of Atmospheric Chemistry and Aerosols with Climate	Philip Cameron-Smith, Lawrence Livermore National Laboratory	Climate Research	43,172	64	135	416
m102	Hadron Physics from Lattice Quantum Chromodynamics	Frank Lee, George Washington	Lattice Gauge Theory	41,434	256	21,936	128

		University					
m41	Computational Atomic Physics for Fusion Energy	Mitch Pindzola, Auburn University	Fusion Energy	38,888	64	427,125	2,000
m881	High dimensional quantum dynamics studies of molecular spectroscopy	Hua-Gen Yu, Brookhaven National Laboratory	Chemistry	37,336	120	0	-
mp2	LLNL MFE Supercomputing	Bruce Cohen, Lawrence Livermore National Laboratory	Fusion Energy	35,612	24	0	-
m189	Carbon Data Assimilation with a coupled Ensemble Kalman Filter	Inez Fung, UC-Berkeley	Climate Research	32,459	64	314	64
mp47	Clay Mineral Surface Geochemistry	Garrison Sposito, UC-Berkeley	Geosciences	27,168	24	1	-
m428	Aerosol Forcings and Consequences on Climate Changes	Catherine Chuang, Lawrence Livermore National Laboratory	Climate Research	23,337	16	0	-
incite15	Three-Dimensional Particle-in-Cell Simulations of Fast Ignition	Chuang Ren, University of Rochester	Fusion Energy	22,757	384	293,098	4,096
m936	Computational Studies of Lithium Carbenoid Reaction Mechanisms and Structure, Bonding, and Reactivity of Other Organolithium Compounds in CPME	Larry Pratt, Fisk University	Chemistry	21,775	8	0	-
mp110	Computational Materials Science	Daryl Chrzan, UC-Berkeley	Materials Sciences	17,965	96	0	-
m783	Computational Studies at BNL of the Chemistry of Energy Production and Use	James Muckerman, Brookhaven National Laboratory	Chemistry	15,303	8	674	32
incite11	Surface Input Reanalysis for Climate Applications (SIRCA) 1850-2011 Surface Input Reanalysis for Climate Applications (SIRCA) 1850-2011	Gil Compo, University of Colorado	Climate Research	15,220	8	2,372,535	448
m633	Simulations of the Formation, Compression and Merging of Compact Tori	Simon Woodruff, Woodruff Scientific	Fusion Energy	15,142	128	6,067	80
m527	Decadal Climate Studies with Enhanced Variable and Uniform Resolution GCMs Using Advanced Numerical	Michael Fox-Rabinovitz, University of Maryland	Climate Research	13,538	16	0	-

	Techniques						
--	------------	--	--	--	--	--	--

Case Studies

Climate Research

The followings are the particular features we like about Bassi:

1. Very stable. Compared to Franklin, it almost does not have unscheduled down time. This is particular important for us because we need to do a lot of short-time testing of the code on daily basis. 2. The friendly IBM fortran compiler installed on Bassi. The "xlf95" fortran compiler is widely used in atmospheric science community, so it does not need to do much extra work when importing the code to Bassi. On the other hand, I particular had some difficulty importing some of my codes which works on Bassi to Franklin. 3. The computational speed is faster (maybe this is true for our applications particularly).

With the same amount of nodes, the computational speed is faster in Bassi than in Franklin.

These are the characteristics we like about Bassi. However, Bassi machine relatively has longer queue time, which may be because many people find it more friendly to use.

The main reason that our group is using Bassi is because of the similarity between Bassi and Seaborg. We were so used to run our CODENAME model on Seaborg and it will require extra work to modify the code to run on other NERSC's platform. We realize that Bassi may be gone in a year or two, so our computer scientist has started the preparation for other machines on NERSC. He encountered some library problems but thought they can be solved if he has more time on these issues.

We also run CODENAME on Bassi, though CODENAME is more platform flexible. It would be more straightforward to switch CODENAME on Franklin than CODENAME. However, to keep the consistency of simulations, we are still using Bassi. Before the time when a switch is inevitable, we will find the right point to make a smooth transition to other machines.

The answer is multi-faceted. There is nothing fundamentally stopping us moving to Franklin, but my own cost-benefit analysis is not particularly favorable.

The main costs in moving to Franklin are: the effort to port codes to Franklin, and the effort to become familiar with a new computer, its operating system, and its utilities and scheduling software.

Rumors about stability and bugs with the XT4s doesn't help.

The main expected benefit is: a less crowded machine, and a little better performance.

It also appears that the Power machines are well balanced for running atmosphere models, which offsets the speed and scale advantages of newer machines.

To be specific, we have tried porting CODENAME, one of our 2 main codes, to Franklin, but ran into problems. We got suggestions from the consultants, but haven't had the time to try them out.

We have also tried using CODENAME on Franklin, since it was ported by other people. We encountered some bugs when running a standard atmospheric chemistry configuration that only manifested themselves on the XT4s. These problems are probably fixable, but take time.

In my experience, such problems are common whenever moving to a new architecture. This is in stark contrast to the nearly seamless transition between Seaborg and Bassi.

In short, the costs of transitioning to new architectures is typically underestimated, IMHO, and one is rarely given funding, or relief from deadlines, in order to make the transition.

Our group uses Bassi because of its operating system. We run a special version of the CODENAME that supports variable-resolution stretched grids. It was originally developed on an IBM Power 3. We attempted to port it to a supercomputer running Linux, namely SGI Altix, but with no success. Since Franklin runs Linux as well we did not try porting our software to it.

In a word, comfort level. Our model runs well on bassi and bassi has proven reliable. We have ported to franklin but the effort was not quite seamless. Our production has been on bassi because of franklin reliability. These runs have been on a relatively low number of procs (40 - 80) and bassi actually runs the code some 25-30% faster than franklin.

We are preparing production with higher resolution versions of the code that will scale to more procs (160+). We anticipate franklin turnaround with this configuration will be much better than bassi, and trust that franklin is now much more stable. We are now satisfied with our port and we expect higher resolution production of this code to be on franklin.

Fusion Energy

For small jobs using 256 processors, bassi runs our codes about 3 times faster than franklin. We also have access to the NAG and NAG-parallel libraries on bassi. Also, the IBM machine appears to be much more stable than the Cray machine.

We're mostly using Bassi for "legacy" reasons, rather than switching to a new platform. I don't believe there are any capability issues.

Reliability of Franklin is perhaps another issue, for now.

We do plan to use Franklin more in the future.

Chemistry

Most of my students prefer Bassi because running VASP they find this platform to be efficient. Most of them have not tried Franklin though, but the general opinion is that VASP runs better in IBM platforms.

Some comments I have received are these:

Bassi is very reliable, powerful (get results fast) and the allowed simulation time vs. number of used nodes is reasonable. The only thing I don't like about Bassi is the waiting time, it's quite long 3-5 days.

Bassi has nodes with large memory (4*8 GB) that are idea for quantum dynamics (not as quantum chemistry) studies. In dynamic calculations, shared-memory parallelizations do not work well. Thus I also like to use davinci. However, the time and node number limitations on davinci make my calculations less possible.

I run Gaussian 03. This is one of a very few systems that I have used that can handle large calculations that have become routine in physical organic chemistry. Bassi can handle about 95% of the calculations that I need to run. The other 5% include jobs that time out after 48 hours and can't be restarted (frequency and NMR calculations) and jobs that use more than the maximum allotted memory, such as CCSD(T).

Lattice Gauge Theory

The per processor performance on Bassi is superior to that of Franklin, due in part to faster interprocessor communication.

The ability to run directly from adequate PERMANENT disk space (/project) is an important advantage of running on Franklin where I am forced to use volatile disk space (\$SCRATCH) as permanent disk space.

Bassi does not penalize me for running on small numbers of processors. The largest jobs I could conceivably run in the near future would be 432 cores. Currently my largest jobs use 192 cores. My highest priority jobs use 48 and 72 cores.

With the 7 job limit in the regular-small batch queue, it is impossible for me to use my allocation fast enough, running solely on Franklin. Note that I am currently also running premium jobs on Franklin to improve throughput.

Bassi is necessary because it increases the number of jobs I can run concurrently.

Bassi has a 36 hour queue. Franklin queues now have a 24 hour limit.

Last but not least, Bassi is a STABLE platform where as Franklin is not.

Bassi's unique feature is the large memory per node (32GB compared to Franklin's 8GB), so I ran some jobs which required a lot of memory (but not an enormous amount of cpu power) on Bassi in the past (I haven't used it for quite some time, though).