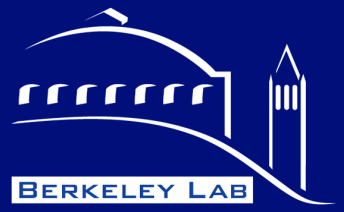
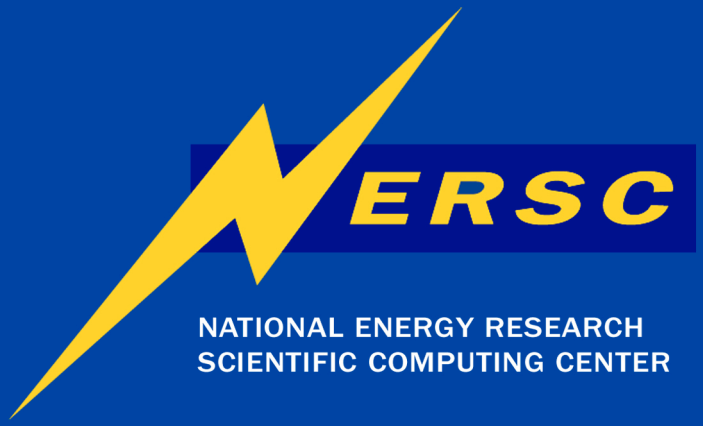




Experiences Configuring, Validating and Monitoring Bassi

Richard Gerber
NERSC User Services Group
RAGerber@lbl.gov

June 13, NUG @ Princeton Plasma Physics Lab

Bassi Delivery and Acceptance

- System delivery started 7/11/2005; system was integrated on-site.
- Because of power limitations, software was installed frame by frame, with switch integration after facility power upgrade completed
- Acceptance period began 10/14/2005; system was accepted on 12/15/2005.
- System availability ended with 99% + availability and 86% + utilization.
- Bassi went into production 01/09/2006.

Initial Configuration

- Although similar to Seaborg, Bassi is more complex.
- Benchmarks were initially run on the default system configuration.
- Under IBM's guidance, we started experimenting with various compiler and runtime settings.
- With NERSC & IBM playing about equal roles, most of the benchmark requirements were easy exceeded in what became the default configuration.

Significant Configuration Parameters

- There has been a learning curve for both NERSC and IBM
  - MEMORY_AFFINITY=MCM
  - MP_TASK_AFFINITY=MCM
  - OBJECT_MODE=64
  - MP_USE_BULK_XFER=yes
  - MP_SINGLE_THREAD=yes
  - -blpdata flag to compilers
  - …

SSP Results

Code	Contractual Commitment (Mflops/s/task)	Measured after tuning (as delivered)	Performance Ratio vs. Seaborg
NPB FT	670	822 (673)	8.95
NPB MG	800	1345 (889)	8.86
NPB SP	480	572 (492)	9.56
CAM	493	554 (517)	4.85
GTC	650	753 (658)	5.19
PARATEC	4400	4794 (4304)	5.65

Bassi Availability Period

- **NERSC invited selected users as Early Users:**
  - **INCITE5, J. Chen, Sandia, Chemical Sciences**
  - **m349, K. Ko, SLAC, Accelerator Physics**
  - **mp13, D. Toussaint, U. Arizona, HEP Theory**
  - **mp193, G. Potter, LLNL, Climate**
  - **mp19, W. Lee, PPPL, Fusion**
  - **Soon afterward: 2nd round of early users**
- **Got a lot of good feedback, which fed into how we set up the default user environment.**
- **Contract required that Bassi exhibit 96% availability over 30 days and meet benchmark variability requirements.**

User Reports

- **Completely non-scientific survey of some of the biggest Bassi users:**
  - **Sinclair: 6.15 x Seaborg @ 48 CPUs**
  - **Lie-Quan Lee: 11X (64 Bassi ~.5 time of 384 Seaborg)**
  - **Pieper: 4.68X**
  - **Lijewski: 4.12X**
  - **Toussaint: 5.11,5.27,7.11 (MILC CG, FF, LL)**
  - **Breslau: 3.37X**
  - **Colgan: 6.32X**
  - **Craig: CCSM 5.46X**
  - **Vary: 3.46X**
  - **Grabow: 5X**
  - **Chen, Hawke: 8-10X**
  - **Swesty, Myra: 3.80X**
  - **Mikkelssen: GS2 8.4X**

Performance Expectations

- **As part of the contractual commitment from IBM the system must meet benchmark performance and variation requirements.**
  - **Sustained System Performance (SSP): "Real World" performance**
  - **Synthetic benchmarks that probe the performance of individual system components (CPU, network, memory, etc.)**
  - **Full configuration, network, I/O, and reboot tests.**

Benchmark Suite Has Served Well

- **We've been very happy with our choice of benchmark for the procurement and subsequent performance monitoring.**
- **They helped identify a number of problems during acceptance and continue to do so.**
  - **problems with attempted Parallel Environment upgrades.**
  - **Uncovered an odd/even node HPS software bug.**
  - **Identified HPS switch firmware upgrade error.**
  - **Guided disk configuration for best performance.**

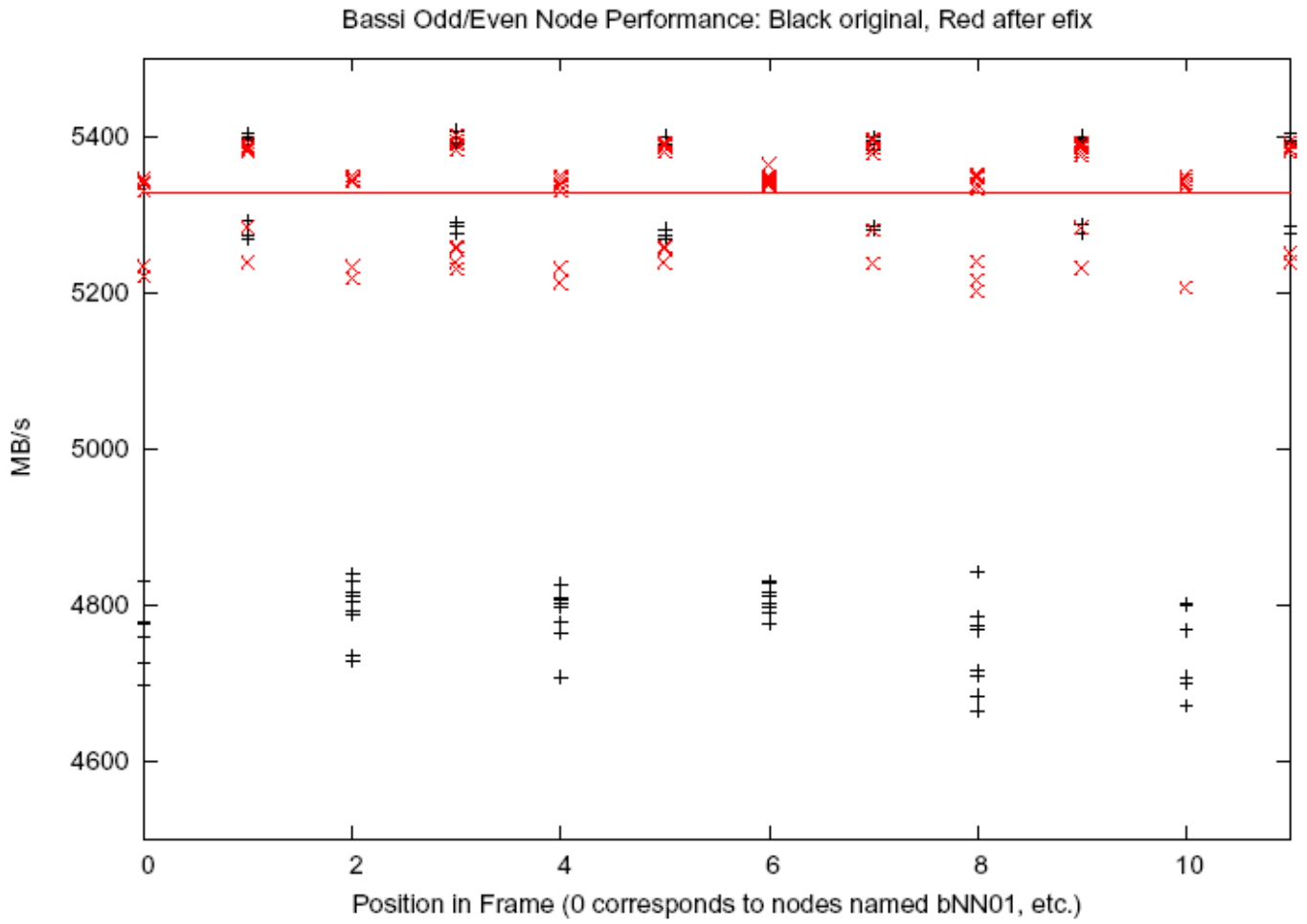
Full Benchmark Suite

- **With the SSP and additional benchmarks, we have a suite for performance evaluation and monitoring.**
  - **SSP suite of 6 codes (+ 2 OpenMP tests using CAM)**
  - **MEMRATE for memory bandwidth**
  - **MPITEST for network latency and bandwidth**
  - **NPB Serial codes: SP, MG, FT for single-node performance**
  - **PIORAW and METABENCH for I/O BW and metadata performance.**
  - **FFTW-driven full configuration test.**

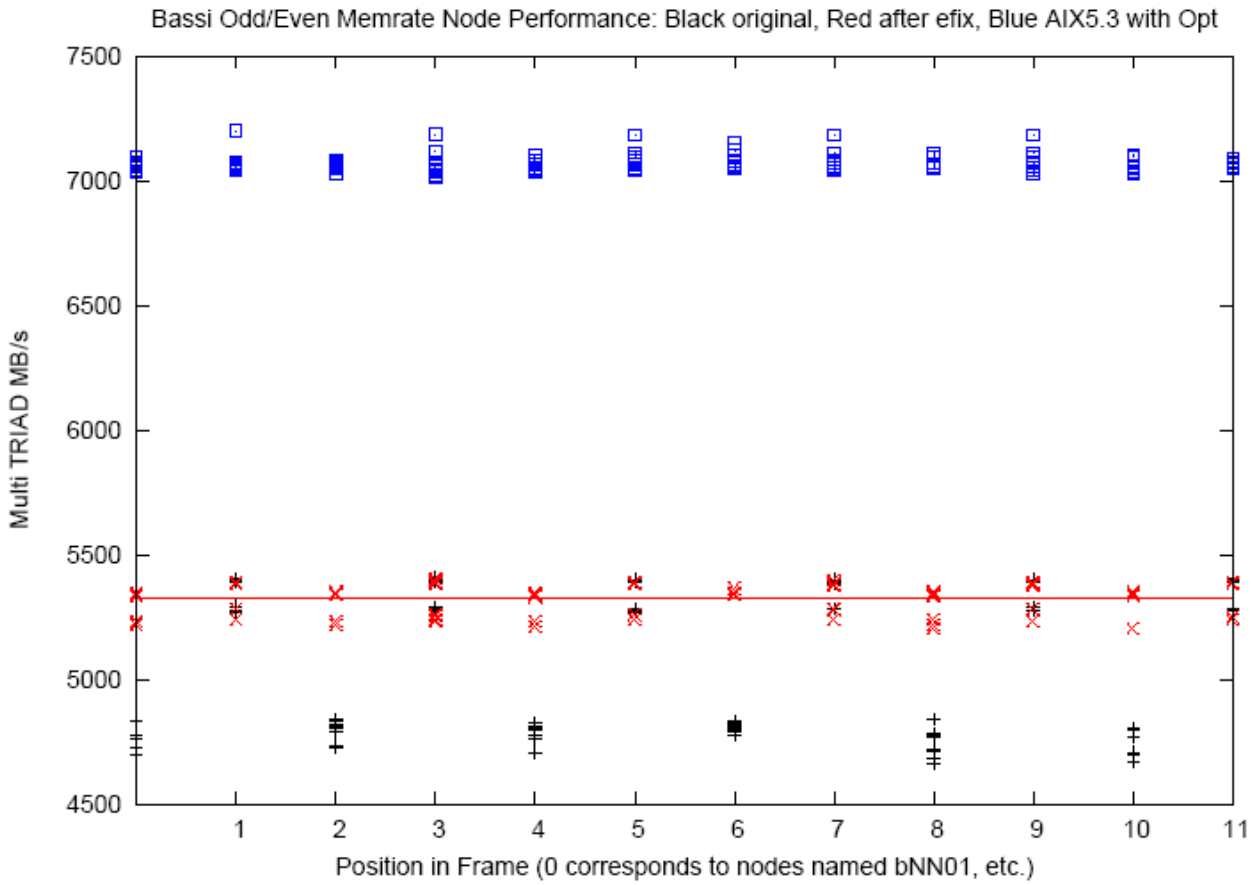
Debugging with Benchmarks

- **A memory bandwidth benchmark revealed a identified a problem with alternating nodes in the frame (rack).**
  - **IBM investigated and supplied a software fix.**
  - **The next slide shows a plot of memory bandwidth performance before (black) and after (red) the fix.**
  - **Subsequent work uncovered a compiler option that gained another 25% in the memory bandwith value; now 7200 MB/sec per CPU (packed node) as measured by HPC MEMRATE TRIAD code.**

Memory Bandwidth



Memrate 5.3 Optimized



Bassi from Then to Now

- We've encountered many problems while trying to get Bassi to it current state.
- The following slides track some of those problems and their resolution.
- Subsequent slides show the performance of the benchmarks over time and how they have been used to monitor and diagnose problems.

Bassi Upgrades and Fixes

- **October 7, 2005**
  - **After NERSC discovers that alternating nodes in a frame have memory bandwidth that differs by about 10%, IBM installs a software efix that boosts performance on the poorly performing nodes.**
- **November 30 to December 7, 2005**
  - **System software is updated to LoadLeveler 3.3.1.1 and Parallel Environment (PE) 4.2.2.1 on Nov. 30. Performance for some codes decreases by up to a factor of 4. The software levels are reinstated at 3.3.0.4 and 4.2.0.3 on Dec. 7.**
- **February 10, 2006**
  - **During a site-wide outage, NERSC attempts to migrate the system from AIX 5.2 to AIX 5.3 and to upgrade LoadLeveler and the IBM Parallel Environment. The migration scripts fail and the system remains at AIX 5.2.**

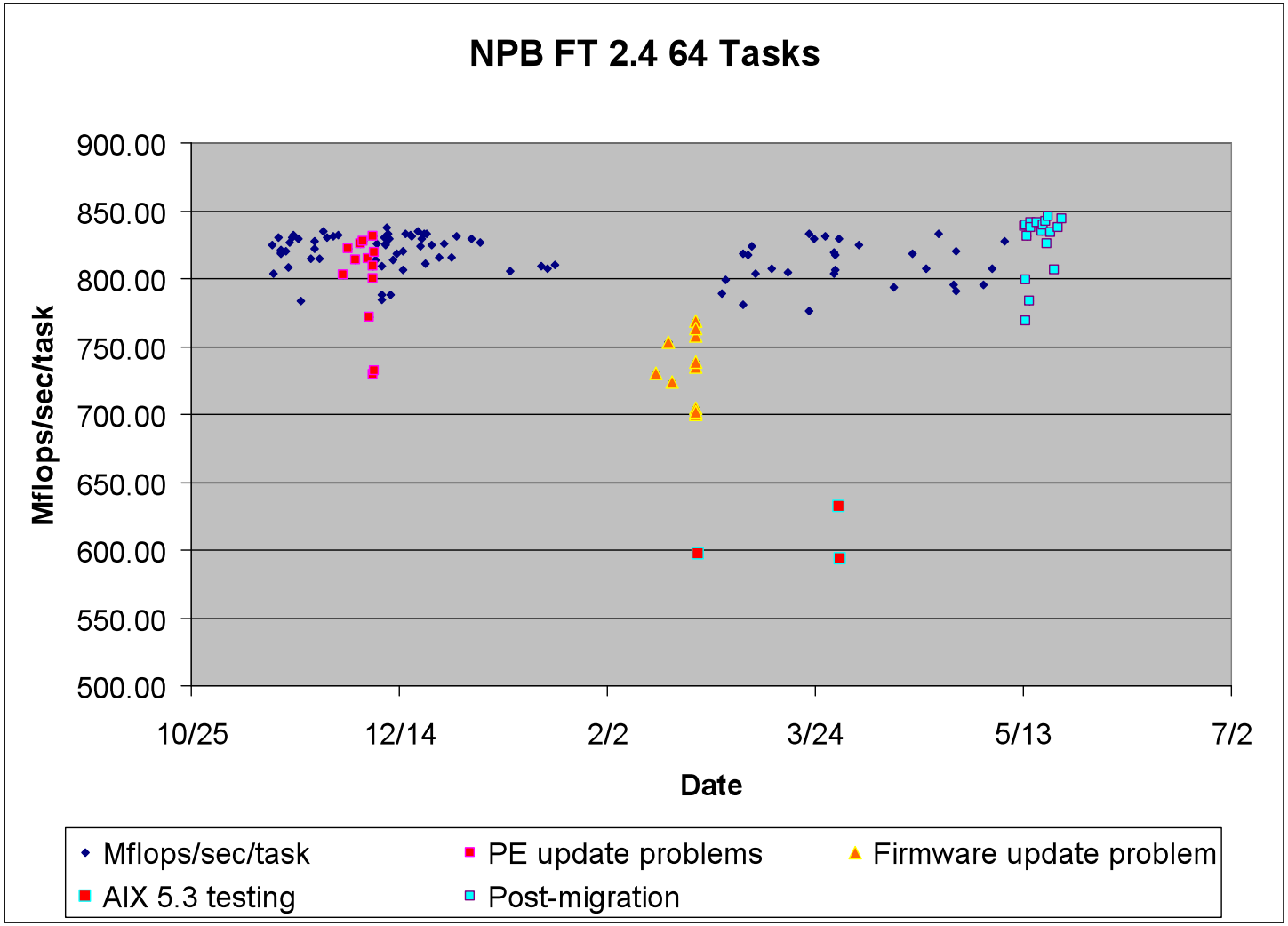
Bassi Upgrades and Fixes

- **February 17, 2006**
  - **NERSC's two-node test/development p575 system, which had already been successfully migrated to AIX 5.3, is upgraded to PE 4.2.2.2. The upgrade is successful and benchmark performance is acceptable.**
- **February 23, 2006**
  - **NERSC again attempts to update AIX and PE. One frame of 12 nodes is migrated to AIX 5.3 and PE 4.2.2.2. Benchmark performance on the migrated nodes is unacceptably poor and the attempt is aborted**
- **March 1, 2006**
  - **The entire system is rebooted after network performance degradation is confirmed as a result of a firmware upgrade that was installed on Feb. 10. The reboot restores network performance. In an attempt to boot 12 nodes to AIX 5.3/PE 4.2.2.2 for testing, GPFS becomes unavailable and AIX 5.3 testing is aborted.**
- **March 29, 2006**
  - **Downtime was taken to load and test AIX 5.3 and Parallel Environment 4.2.2.2 on 10 Bassi nodes. The performance of NERSC benchmarks on the 10 5.3 nodes was unacceptable and the machine was returned to service with all nodes running AIX 5.2 and POE 4.2.0.3.**

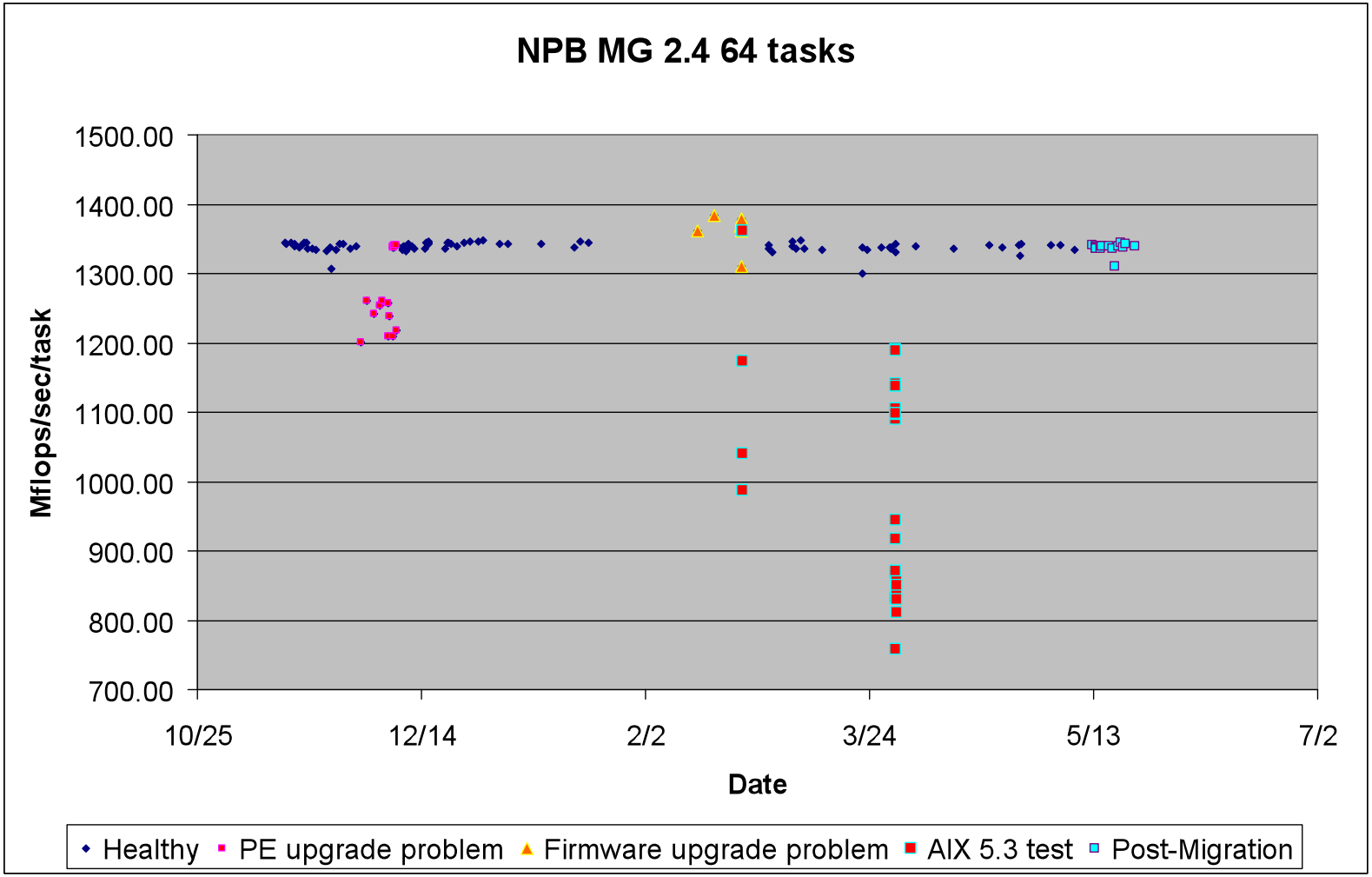
Bassi Upgrades and Fixes

- **April 26, 2006**
  - Dedicated system time was taken to evaluate AIX 5.3 on 12 nodes. During the outage a number of security patches were applied to the production system. Acceptable benchmark performance was attained on the 12 AIX 5.3 nodes, but a problem with indexing authentication database files was found.
- **May 10, 2006**
  - Bassi's operating system was migrated from AIX 5.2 to AIX 5.3.
- **May 24, 2006**
  - Bassi is running AIX 5.3 and performance across the entire system is believed to be comparable to that under AIX 5.3. Following the May 10 AIX 5.3 upgrade some nodes had to be remigrated, others had incorrect large-page memory configurations, GPFS was misconfigured, and a bad node was identified and removed

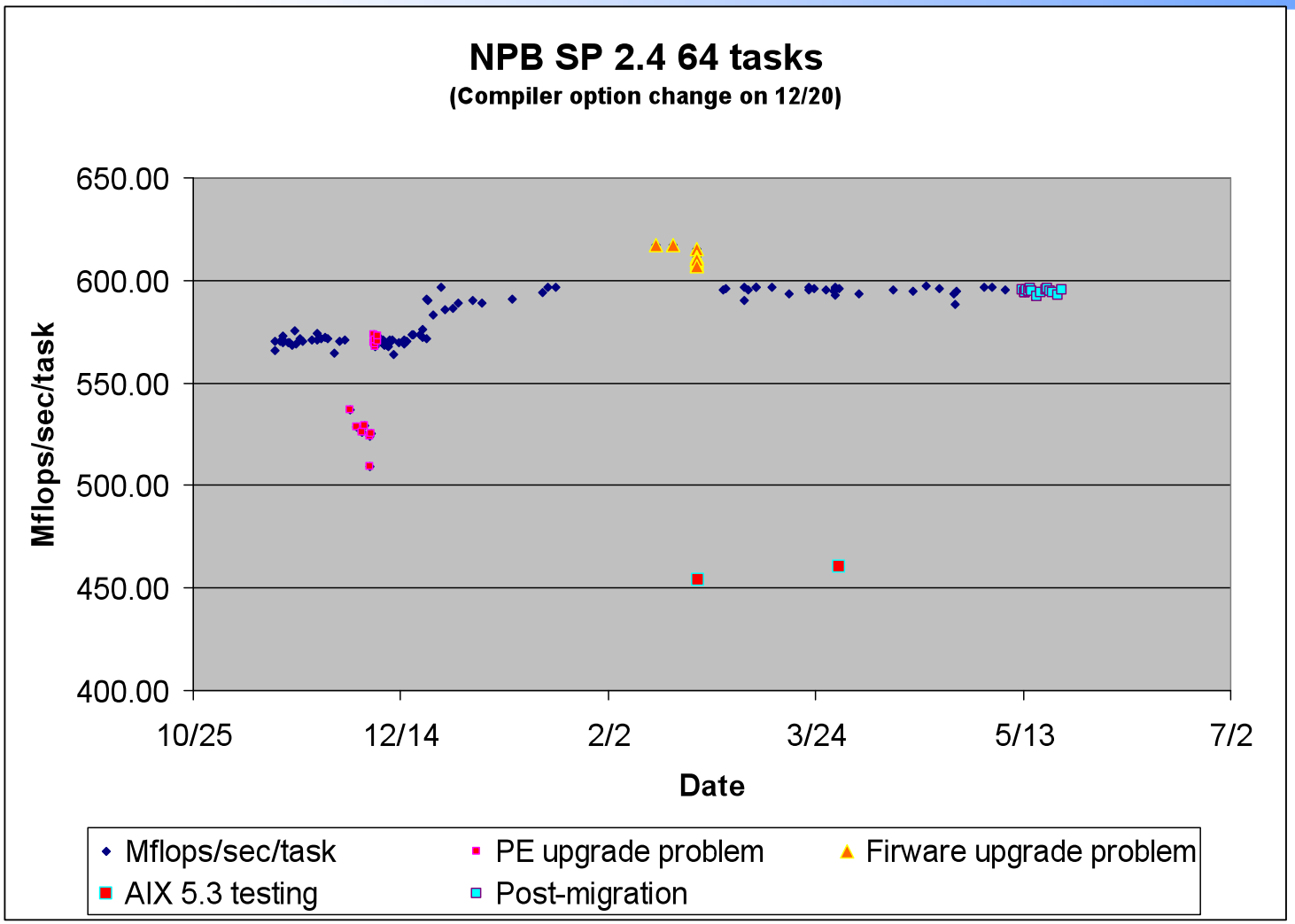
NPB FT Parallel



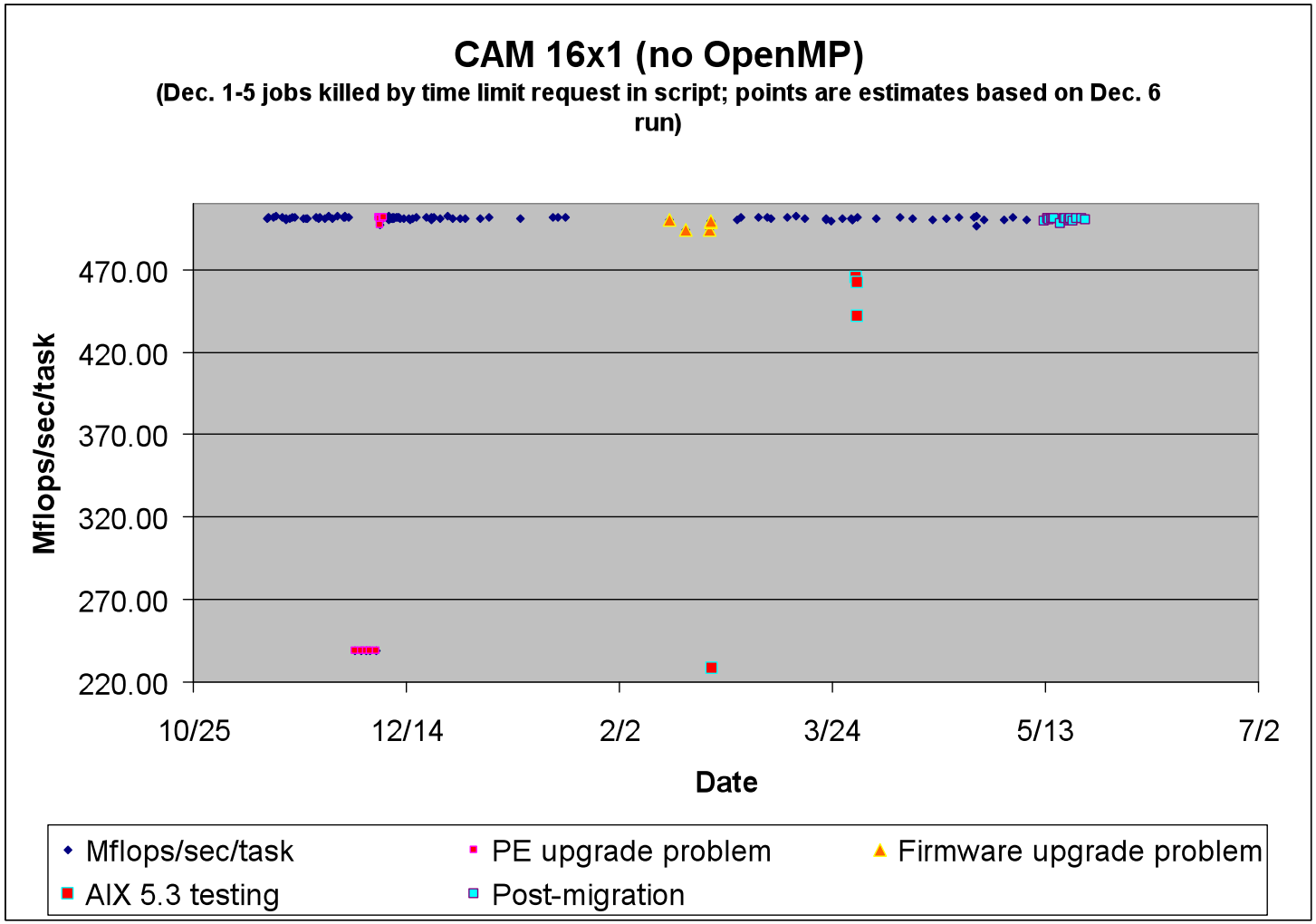
NBP MG


NPB MG 2.4 64 tasks
1500.00
1400.00
1300.00
1200.00
1100.00
1000.00
900.00
800.00
700.00
Mflops/sec/task
10/25
12/14
2/2
3/24
5/13
7/2
Date
Healthy
PE upgrade problem
Firmware upgrade problem
AIX 5.3 test
Post-Migration

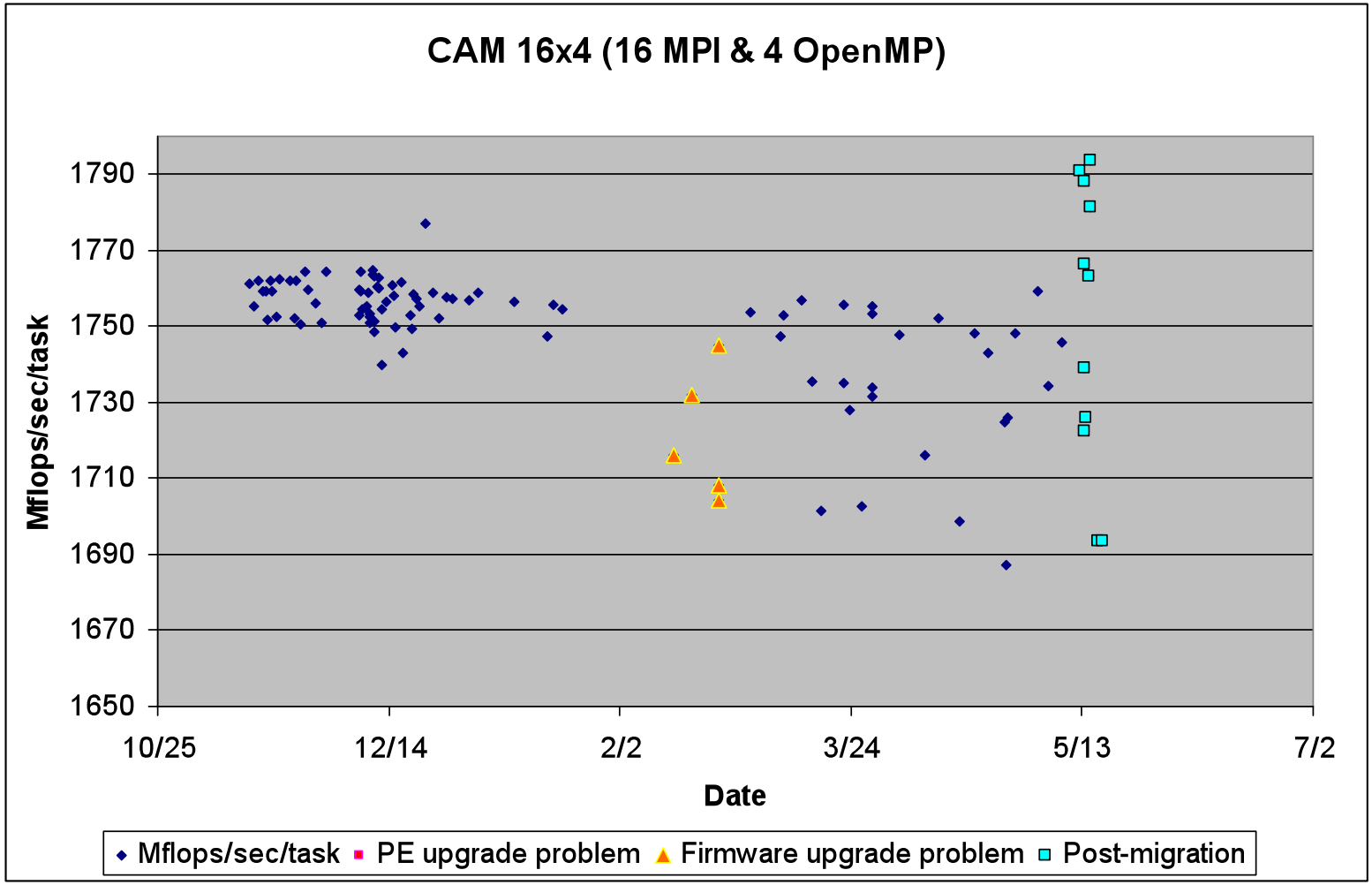
NPB SP


NPB SP 2.4 64 tasks
(Compiler option change on 12/20)
650.00
600.00
550.00
500.00
450.00
400.00
Mflops/sec/task
10/25
12/14
2/2
3/24
5/13
7/2
Date
Mflops/sec/task
PE upgrade problem
Firware upgrade problem
AIX 5.3 testing
Post-migration

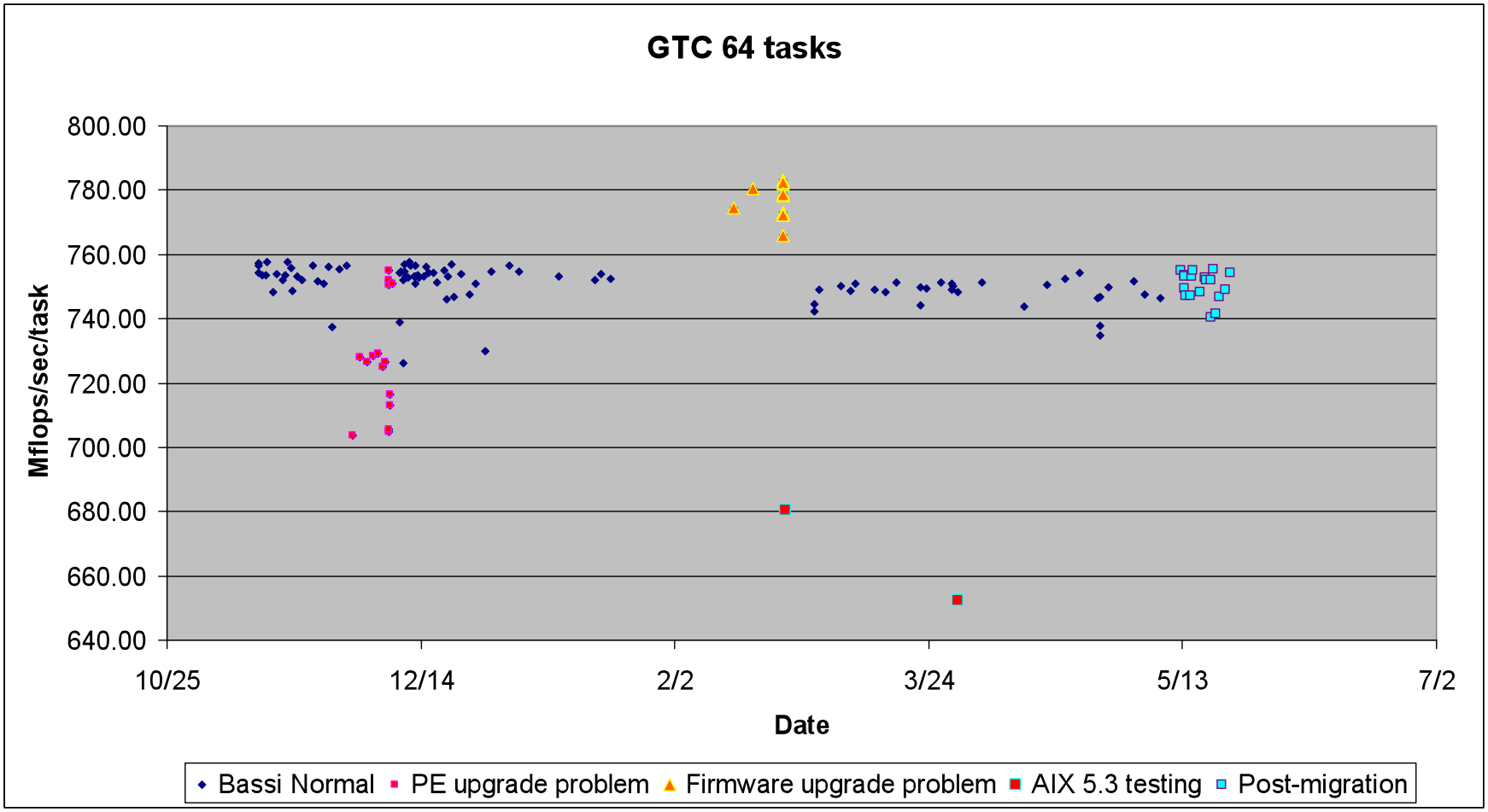
CAM 3.0


CAM 16x1 (no OpenMP)
(Dec. 1-5 jobs killed by time limit request in script; points are estimates based on Dec. 6 run)
Mflops/sec/task
470.00
420.00
370.00
320.00
270.00
220.00
10/25
12/14
2/2
3/24
5/13
7/2
Date
Mflops/sec/task
PE upgrade problem
Firmware upgrade problem
AIX 5.3 testing
Post-migration

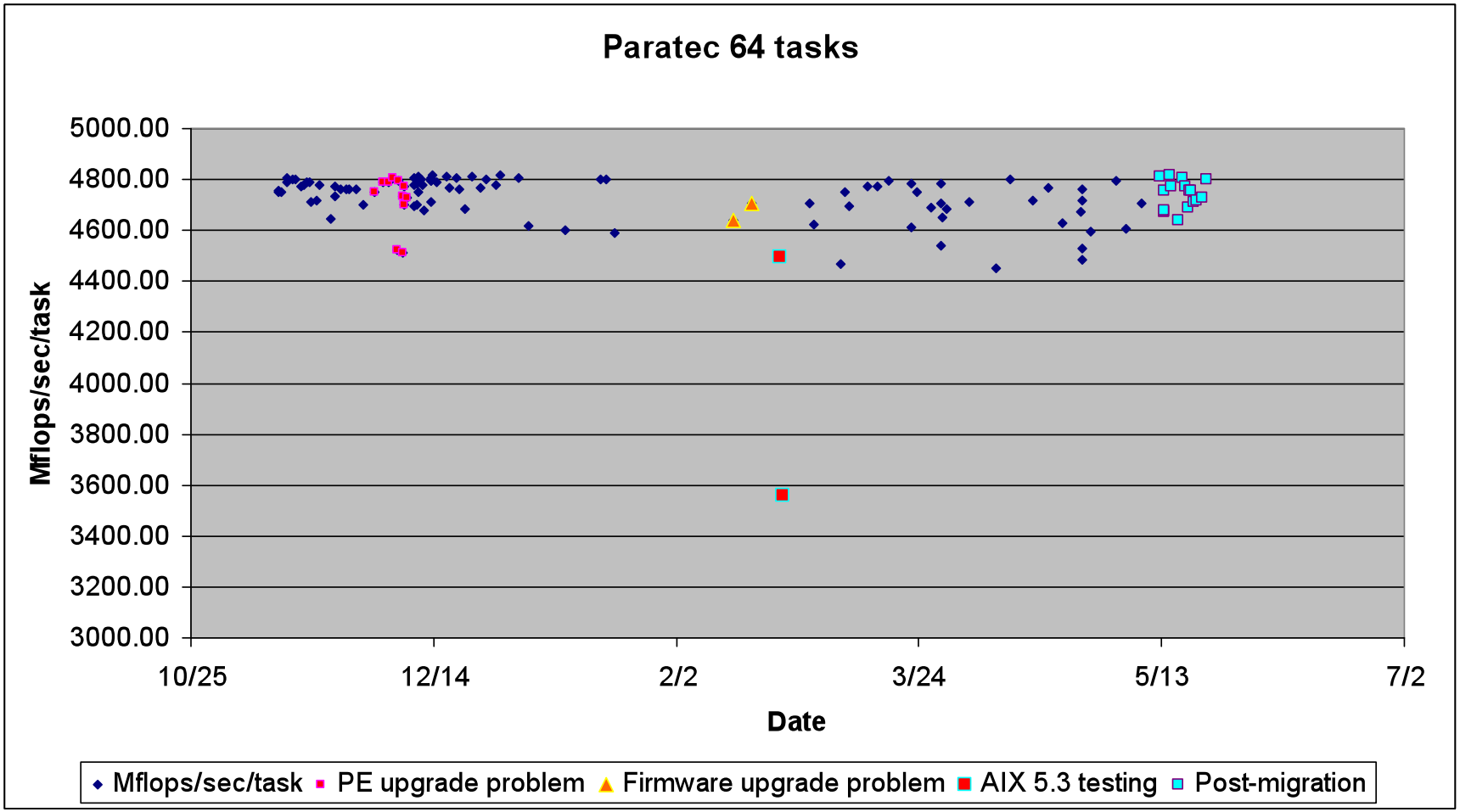
CAM 3.0 OMP


CAM 16x4 (16 MPI & 4 OpenMP)
Mflops/sec/task
1790
1770
1750
1730
1710
1690
1670
1650
10/25
12/14
2/2
3/24
5/13
7/2
Date
Mflops/sec/task
PE upgrade problem
Firmware upgrade problem
Post-migration

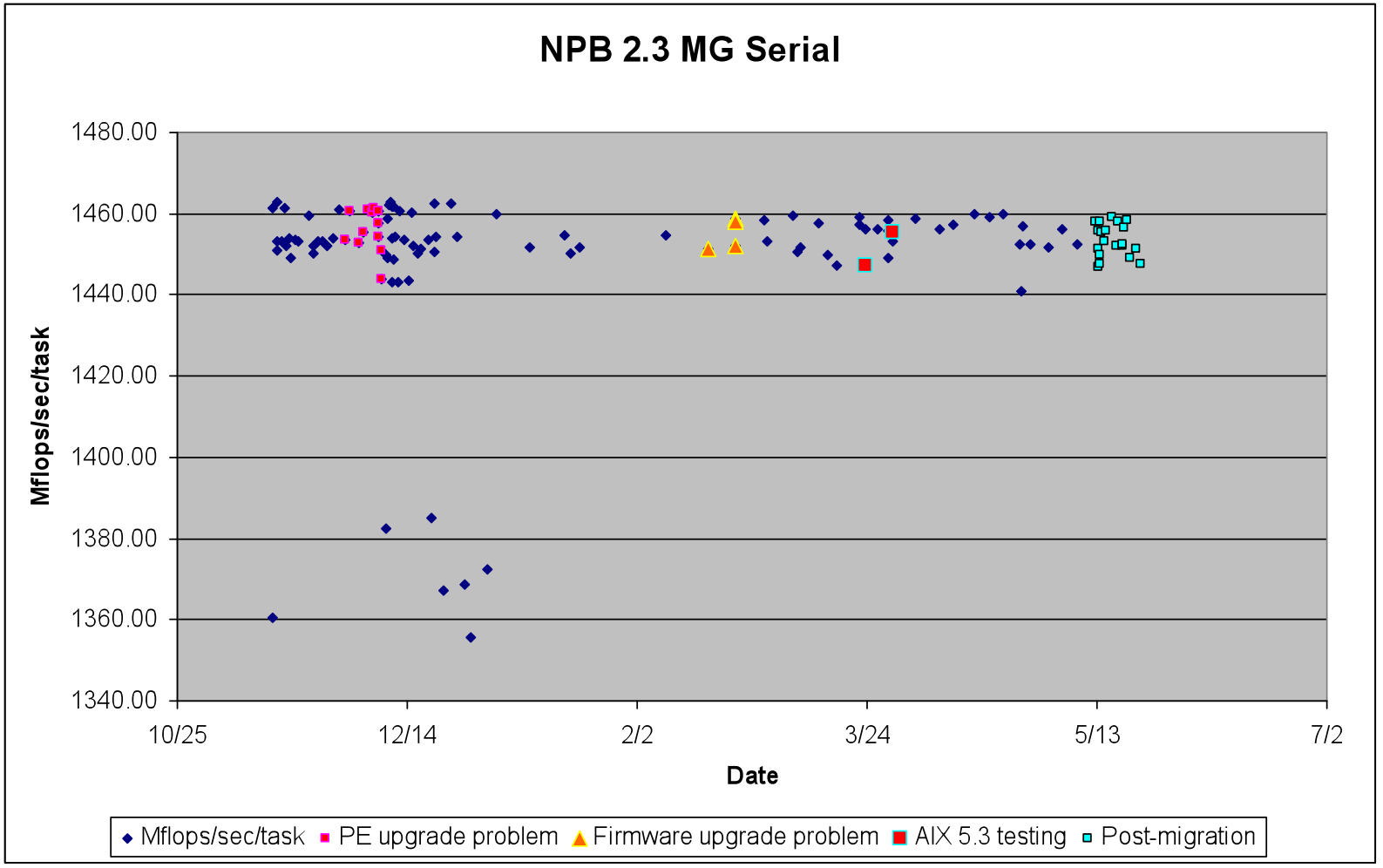
GTC


GTC 64 tasks
800.00
780.00
760.00
740.00
720.00
700.00
680.00
660.00
640.00
Mflops/sec/task
10/25
12/14
2/2
3/24
5/13
7/2
Date
Bassi Normal
PE upgrade problem
Firmware upgrade problem
AIX 5.3 testing
Post-migration

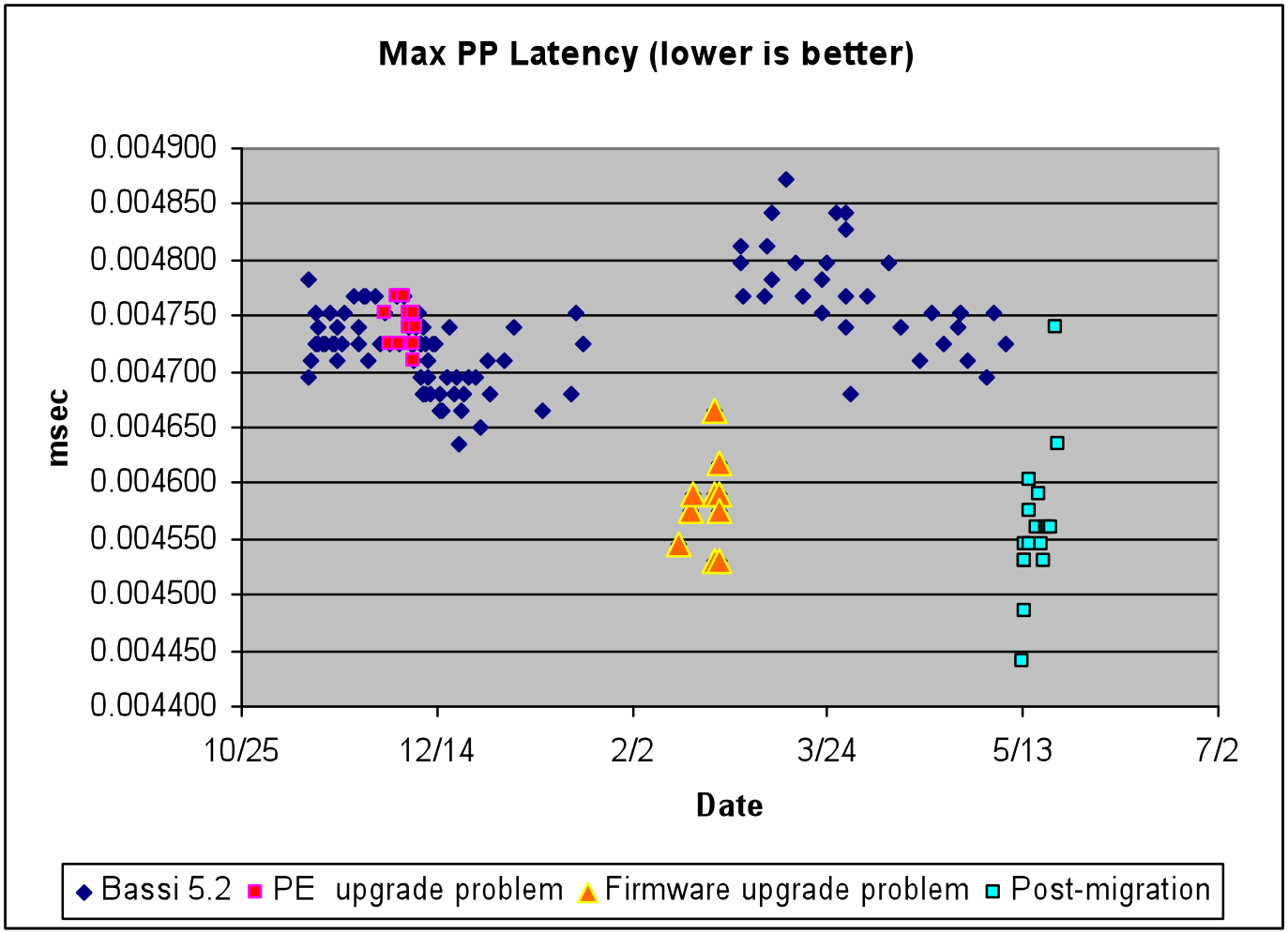
PARATEC


Paratec 64 tasks
5000.00
4800.00
4600.00
4400.00
4200.00
4000.00
3800.00
3600.00
3400.00
3200.00
3000.00
Mflops/sec/task
10/25
12/14
2/2
3/24
5/13
7/2
Date
Mflops/sec/task
PE upgrade problem
Firmware upgrade problem
AIX 5.3 testing
Post-migration

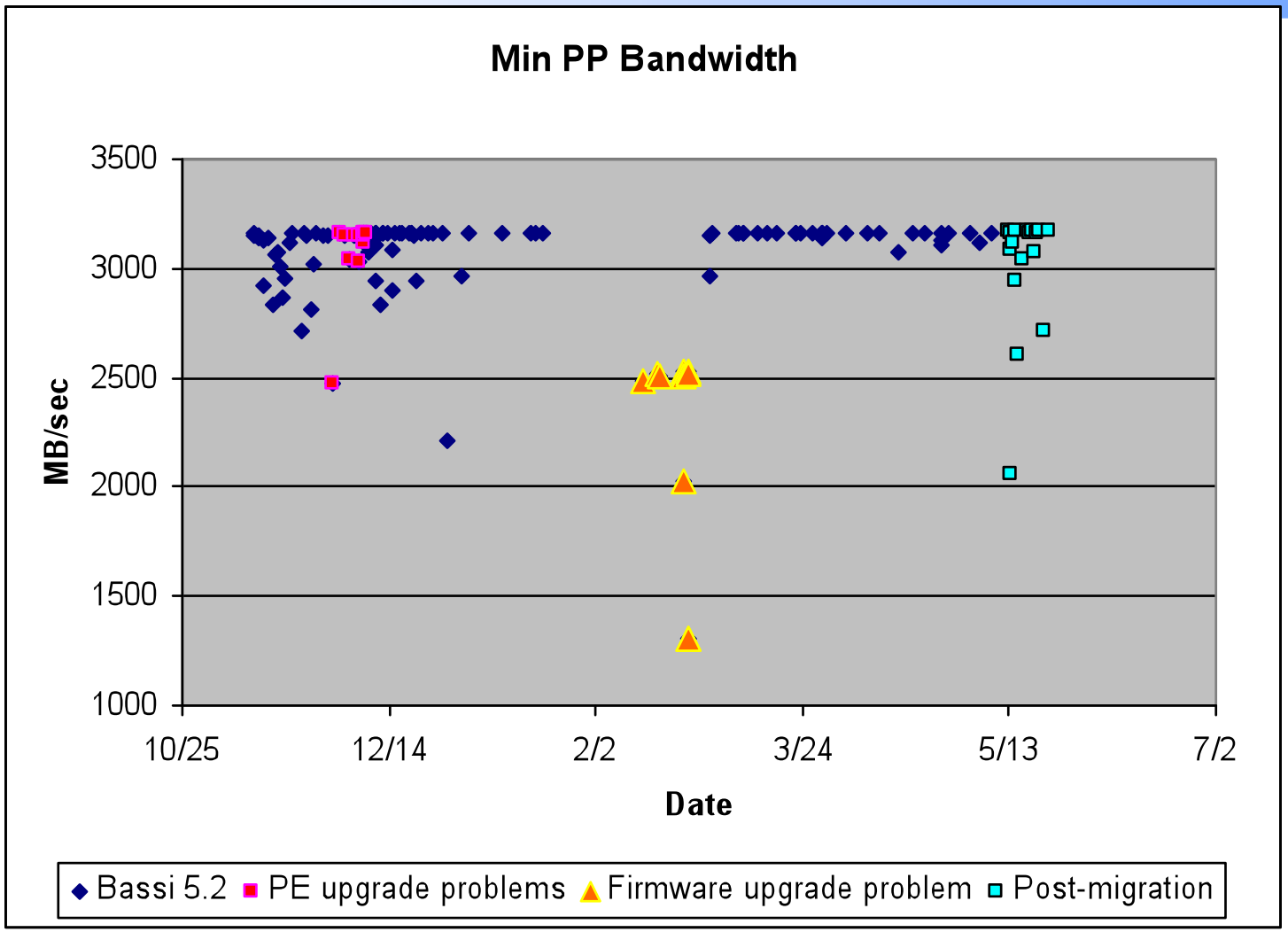
NPB MG SERIAL



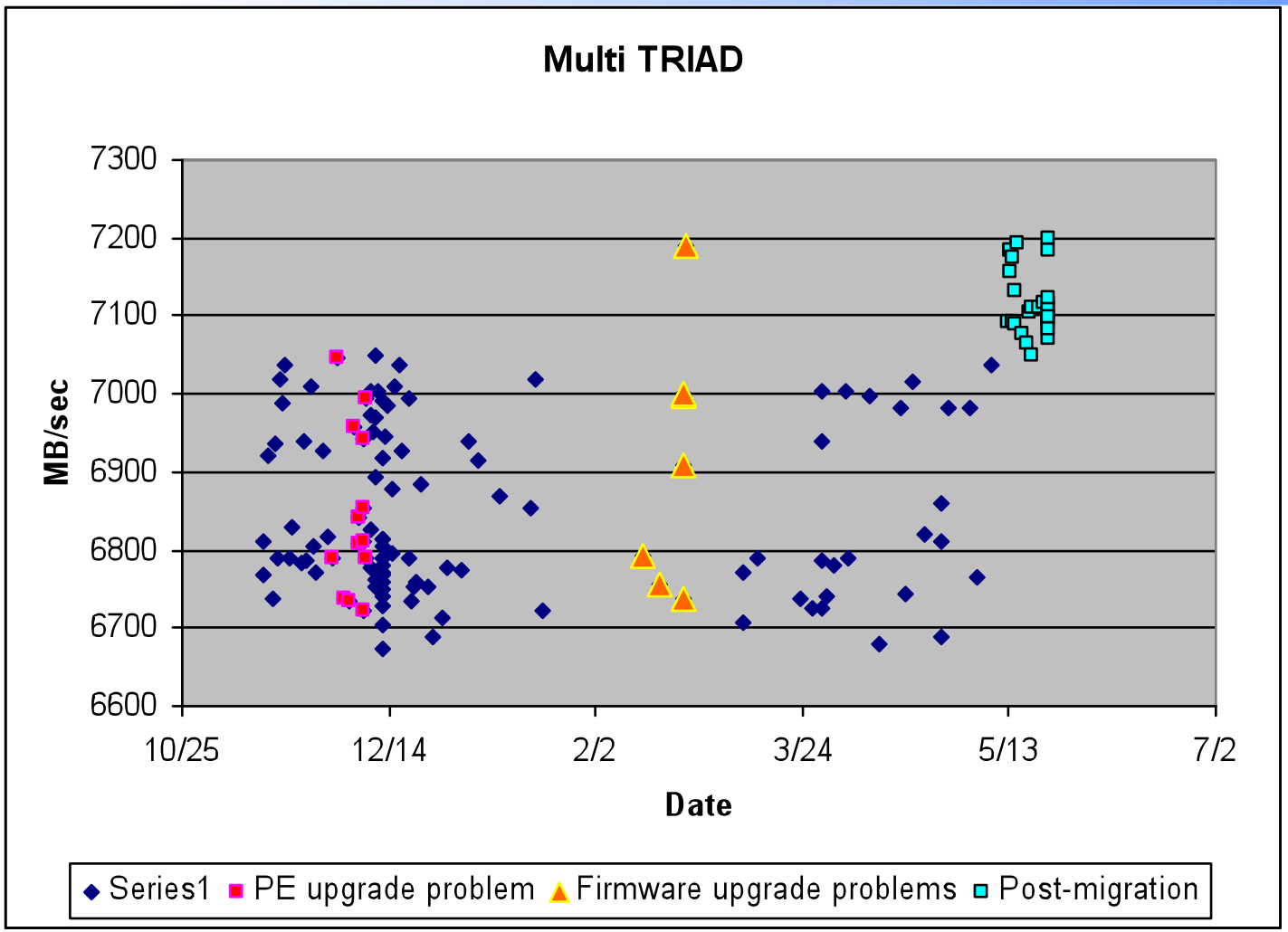
MPI Latency


Max PP Latency (lower is better)
msec
0.004900
0.004850
0.004800
0.004750
0.004700
0.004650
0.004600
0.004550
0.004500
0.004450
0.004400
10/25
12/14
2/2
3/24
5/13
7/2
Date
Bassi 5.2
PE upgrade problem
Firmware upgrade problem
Post-migration

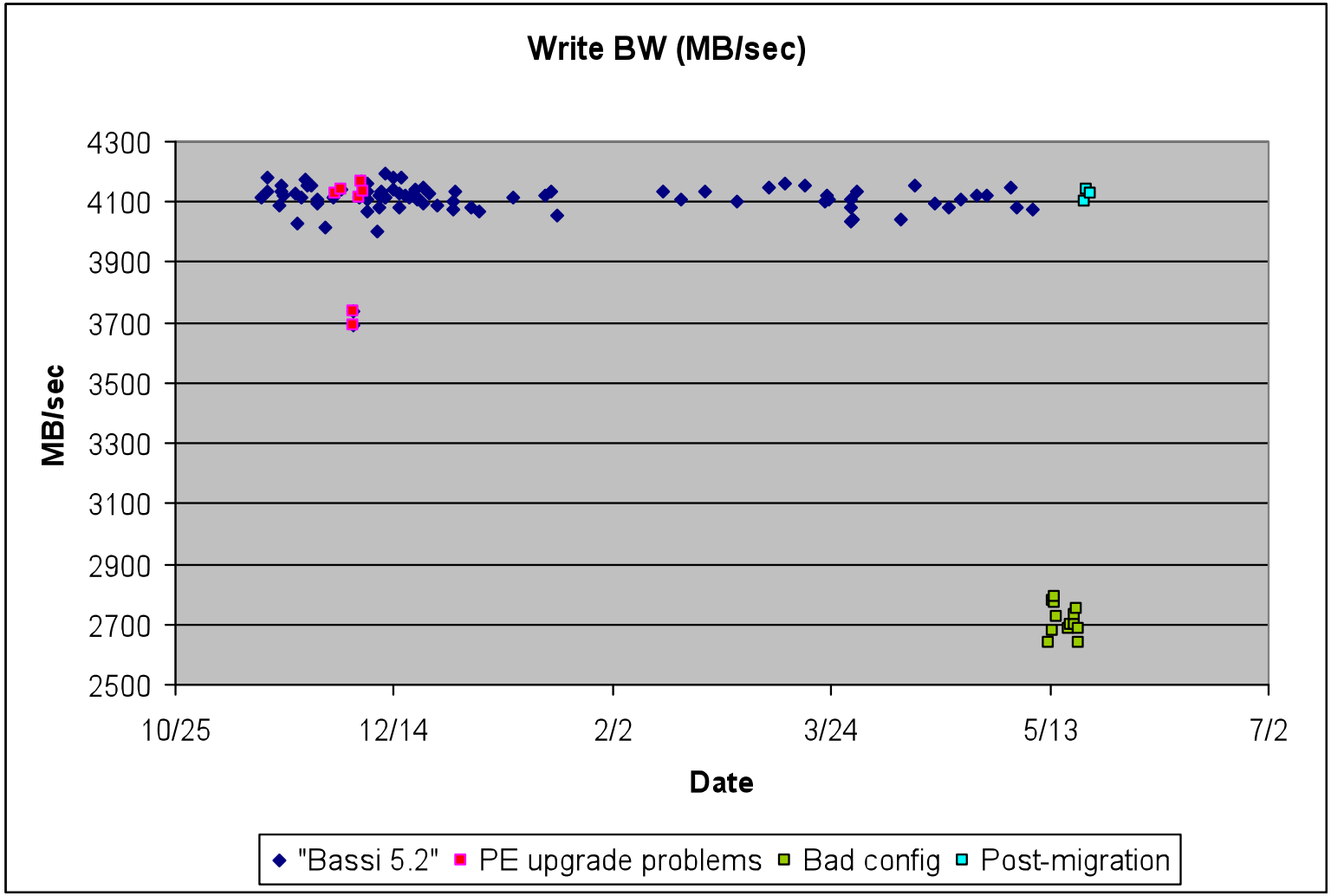
HPS Bandwidth



MEMRATE


Multi TRIAD
7300
7200
7100
7000
6900
6800
6700
6600
MB/sec
10/25
12/14
2/2
3/24
5/13
7/2
Date
Series1
PE upgrade problem
Firmware upgrade problems
Post-migration

PIORAW


Write BW (MB/sec)
MB/sec
4300
4100
3900
3700
3500
3300
3100
2900
2700
2500
10/25
12/14
2/2
3/24
5/13
7/2
Date
"Bassi 5.2"
PE upgrade problems
Bad config
Post-migration

LDAP Integration

- NERSC is using OpenLDAP for common authentication among NIM, Bassi, Jacquard, DaVinci, Web, NERSC 5.
- AIX 5.2 does not support OpenLDAP – no Linux-like PAM functionality. IBM LDAP "solution" only supports weak crypt() password hash.
- NERSC workaround was to script a pull of info from LDAP and create AIX password, group and security files.
- AIX 5.3 adds PAM functionality and has other security enhancements.

AIX 5.3 Migration

- IBM originally proposed building the cluster onsite in Oakland with AIX 5.2 and migrating to 5.3 during the acceptance period.
- NERSC thought the migration was risky and pushed the acceptance period too far beyond the fiscal year, so we negotiated to make 5.3 a deliverable for early 2006.
- AIX 5.3 promised PAM, improved security, SMT support, dynamic large page configuration, improved large-page memory allocation, lower HPS latency, a path forward for LoadLeveler and Parallel Environment support.

5.3 Migration Problems

- We suffered through many failed attempts to migrate from 5.2 to 5.3.
- It turned out that we were the first to perform this on a large system.
- LLNL Purple was already at 5.3E, HPCx was running 5.3 from a re-install. NCAR and PNL were waiting to see what happened with us.
- IBM migration scripts locked out root from the nodes on first attempt. We backed off and returned to production at 5.2.

5.3 Migration Continued

- After first failed attempt, NERSC and IBM agreed to break the disk mirroring on 12 nodes, install 5.3 on those nodes, and make them dual-boot so we could easily back off if there were problems.
- The installation failed on the first attempt.
- Second attempt was successful, but all parallel benchmark and application performance was degraded (~30%).
- No obvious reason because PP HPS bandwidth was good, latency was excellent, and single-node serial benchmarks performed very well.

5.3 Migration Continued

- IBM and NERSC spent many weeks experimenting and debugging.
- We found a number of bugs in PE and AIX, but none of them fixed the basic problem of poor benchmark performance.
- NERSC got IBM to build a small Bassi clone system in Poughkeepsie. But all runs on it were up to spec.
- Traded boot disks, but still no clues.

5.3 Migration Continued

- After exhausive discussion and system comparisons, we whittled possible differences down to the NERSC password and group files.
- We had not wanted to give IBM our password files, even if they didn't contain the hashes. But we finally agreed that there were no security implications.
- When IBM ran with our authentication files, the were able to reproduce the problem.
- It turned out the various system demons were inefficiently parsing password, group, and security files with 1000s of lines. The interrupts were stealing cycles, ejecting memory pages, etc., causing the poor performance. (This is not the first time we've seen OS interrupts causing big effects.)

AIX 5.3 continued

- We found that with small password files, performance was restored. This was also true if the files were indexed.
- So we indexed the files each time the files were created from the LDAP pull scripts.
- This was thought to be a very temporary solution, since were were going to use the full PAM functionality under 5.3 and get all info from LDAP directly – the password files would contain only minimal entries (e.g., root)

AIX 5.3 Continued

- We thought the issue was resolved with the indexing work-around and began regularly pulling from LDAP and indexing authentication files.
- Began getting reports of job launch failures. The reports accelerated over a few days and we were able to repeat them, seemingly randomly.
- After many days of intense work NERSC & IBM discovered a bug in the initial indexing and IBM group lookup, which resolved itself after a single launch failure, but reappeared every 2 hours with the LDAP pull.

Current Status

- We've turned off password file updates on the compute nodes.
- Hoping this will eliminate the "getpcred" job launch failures.
- Working hard to implement full PAM/ LDAP solution.