Experiences Profiling and Characterizing DoD HPC Applications

Bill Ward
Computational Science and Engineering Group
ERDC MSRC

Goals of Gathering Profile Data

- Characterize application codes so as to be able to construct a representative set of benchmarks
- Provide relevant data for construction of synthetic benchmarks
- Optimize performance of benchmark codes
- Provide data to enable performance modeling and prediction

Data to Be Collected

- Three levels
  - Increasing levels involve more effort but allow for more detailed analysis.
- Gather data for each benchmark code for
  - at least two compiler optimization levels
  - three or four different CPU counts
  - two or more sets of input data
  - two or more different machine architectures

Level 1

- The following data for the entire run of a program
  - Wallclock time
  - Cycle count
  - Floating-point operation counts by type
  - Memory operation counts by type
  - Branch counts
  - Cache (L1, L2, etc.) and TLB miss rates
  - Communication time broken down by MPI message type

Level 1 (continued)

- Easy to acquire
- Would give picture of workload with respect to broad characteristics but would not enable optimization and modeling

Level 2

- Same data as Level 1
  - Level 2a: broken down by routine
  - Level 2b: broken down by basic block and loop
- Would provide a time-dependent profile of the application that would enable optimization and modeling to some degree
  - Could be used to construct “machine-dependent” performance models

Level 3

- Data required for “machine-independent” performance models
  - Memory access patterns
  - I/O profiles
  - Scaling profiles
  - Branch profiles
- Difficult to collect
- Collection will slow down execution significantly.

Benchmark Application Codes

Aero – Aeroelasticity CFD code

Fortran, serial vector, 15,000 lines of code (LOC)

Cobalt-60 – Turbulent flow CFD code

Fortran, MPI, 19,000 LOC

GAMESS – Quantum chemistry code

Fortran, MPI, 330,000 LOC

HYCOM – Ocean circulation modeling code

Fortran, MPI, 31,000 LOC

NAMD – Molecular dynamics code

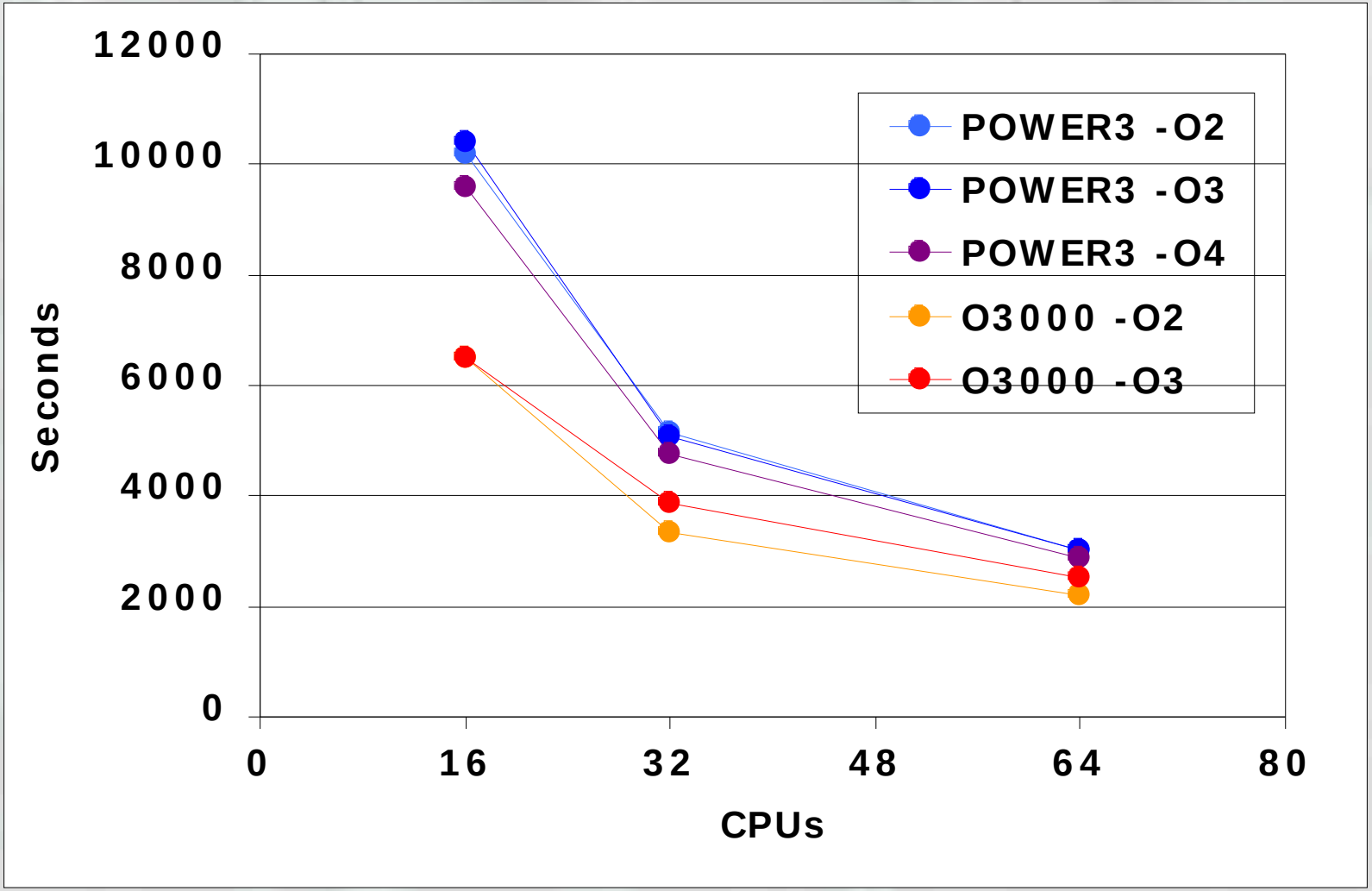
C, MPI, 57,000 LOC

OOCore – Out-of-core solver

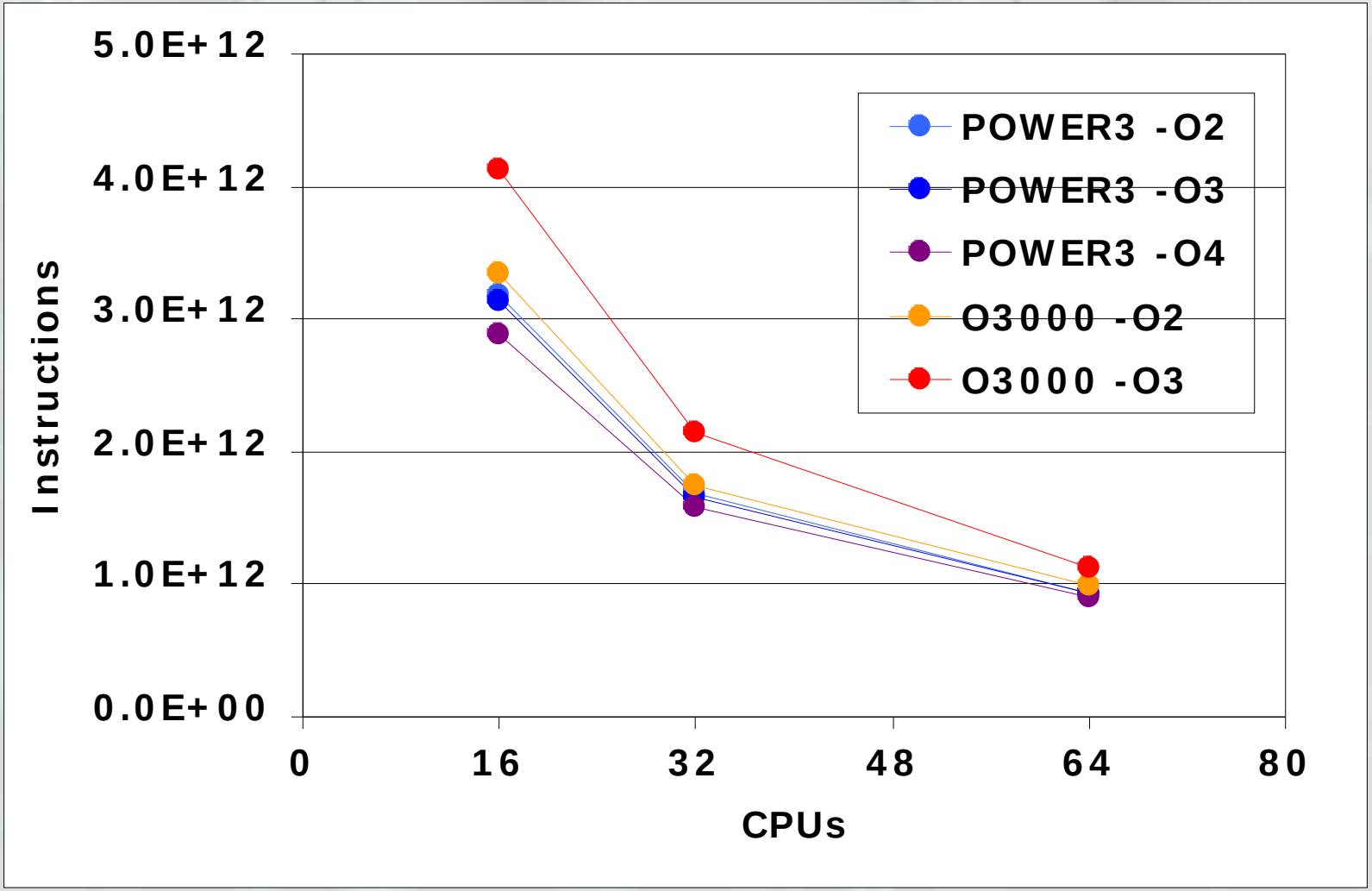
HPCMPO Systems Profiled

Vendor	System	MHz	CPUs
IBM	POWER3	375	1336
IBM	POWER4	1300	1184
SGI	O3900	700	512

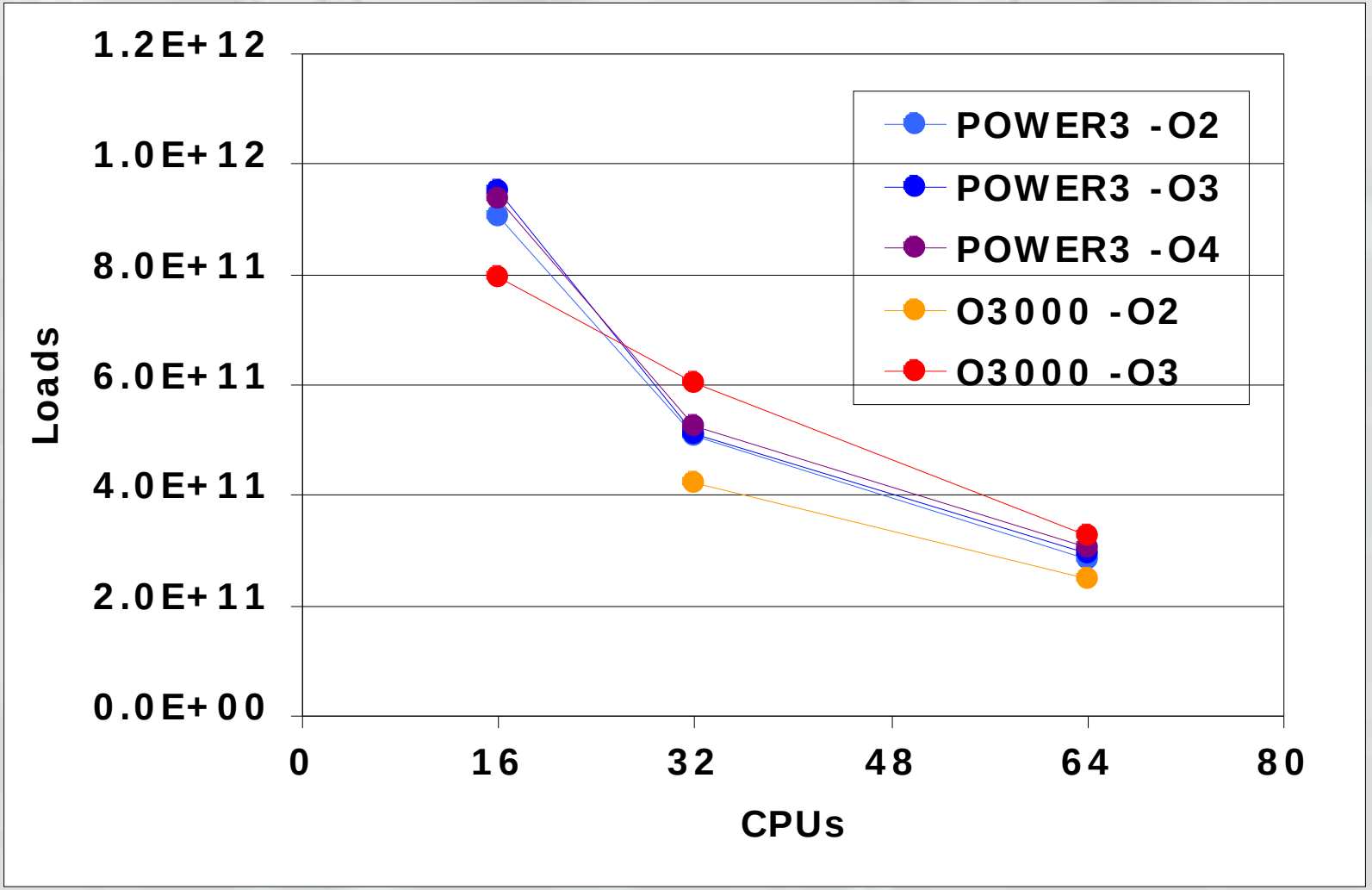
RFCTH Wall Time (Average per Process)



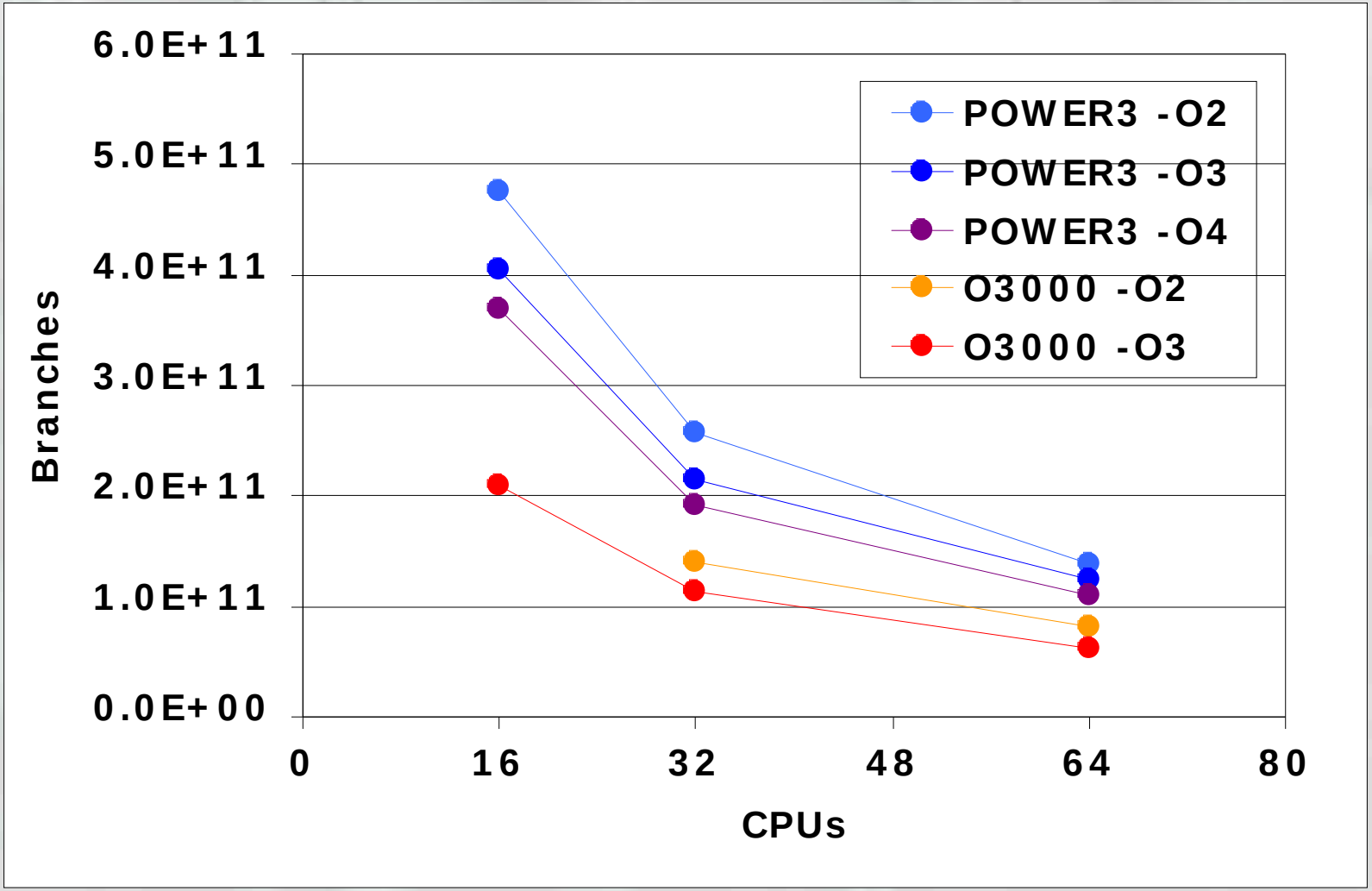
RFCTH Instruction Count (Average per Process)



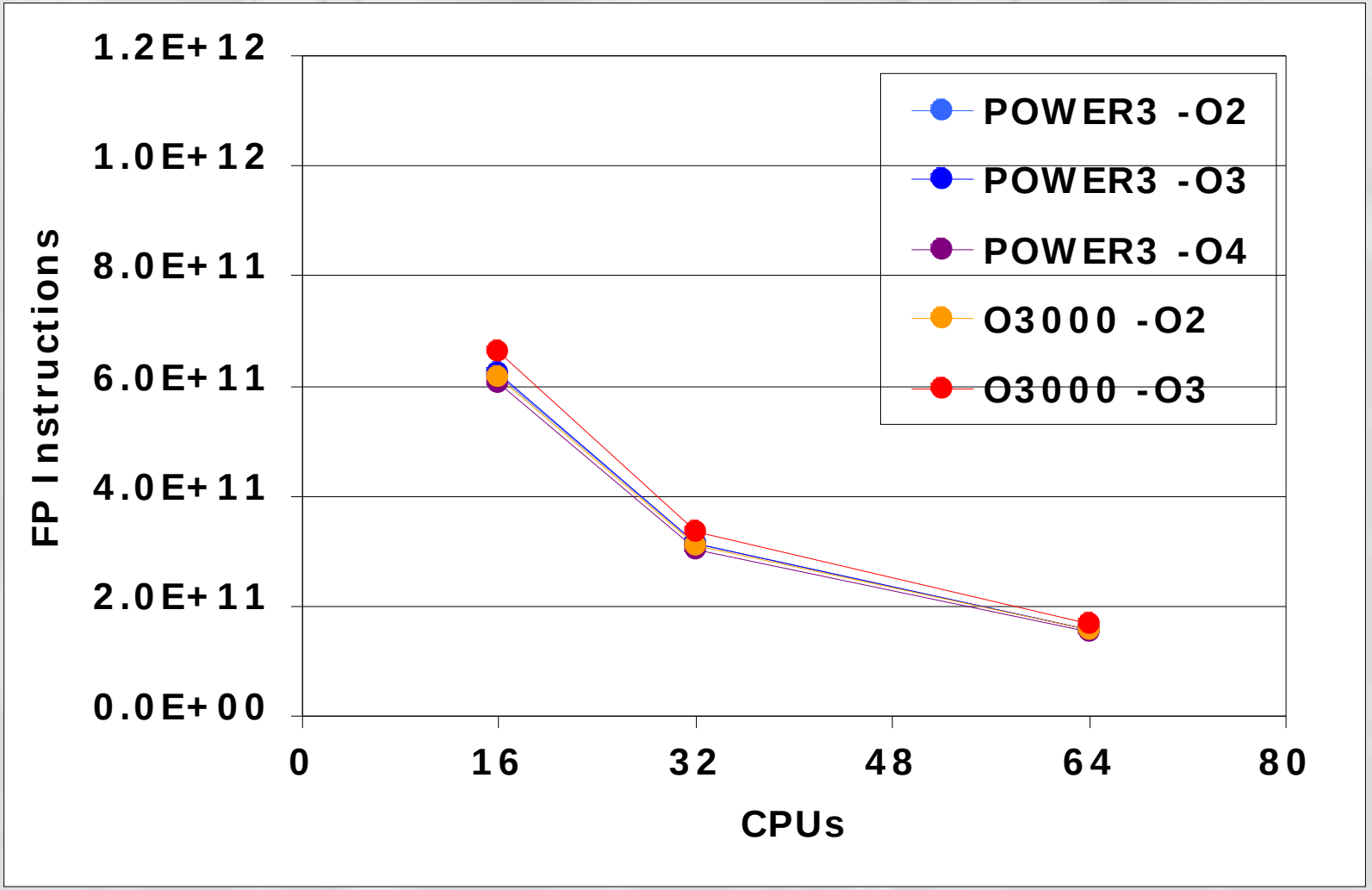
RFCTH Loads Performed (Average per Process)



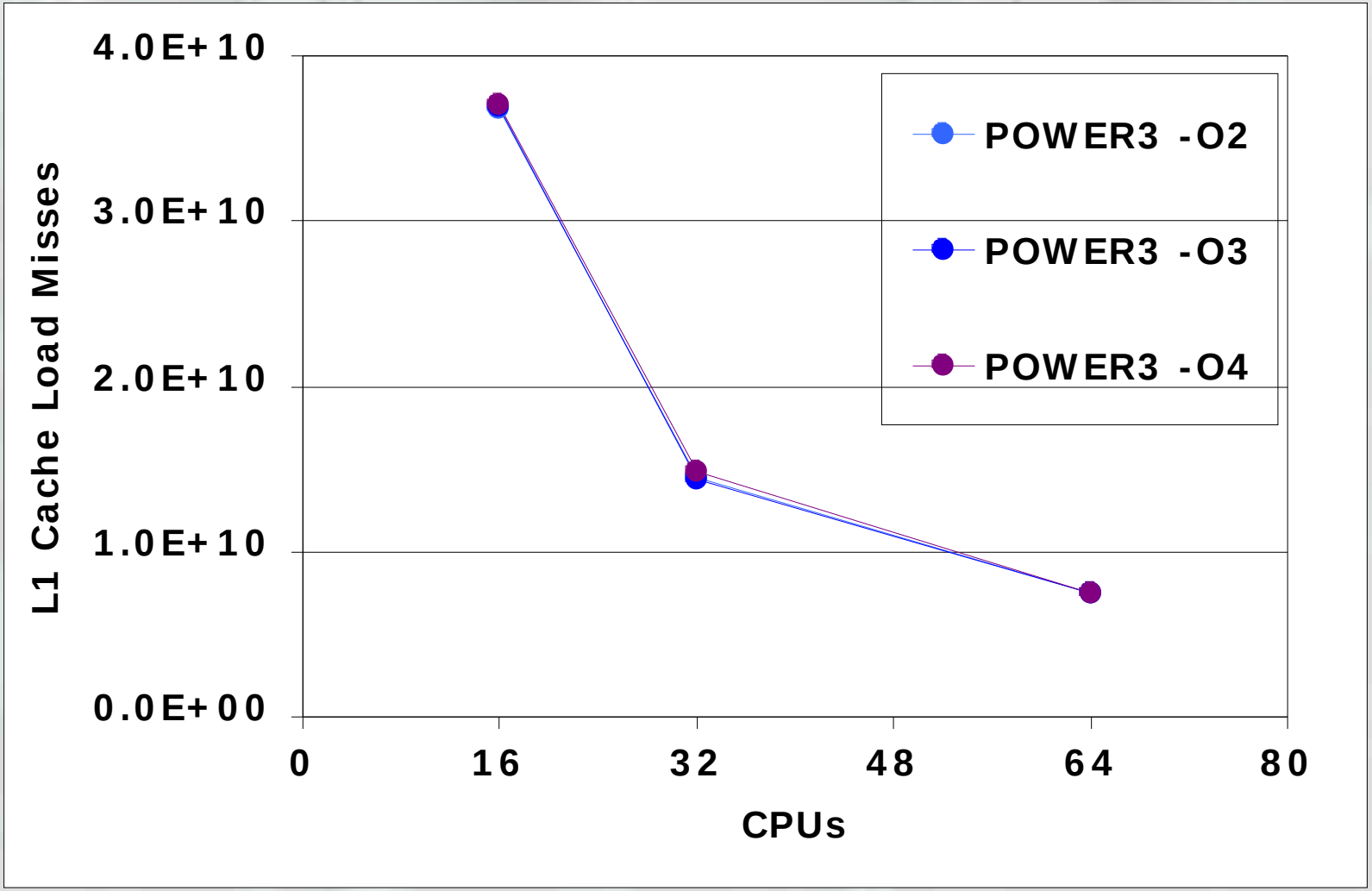
RFCTH Branches Performed (Average per Process)



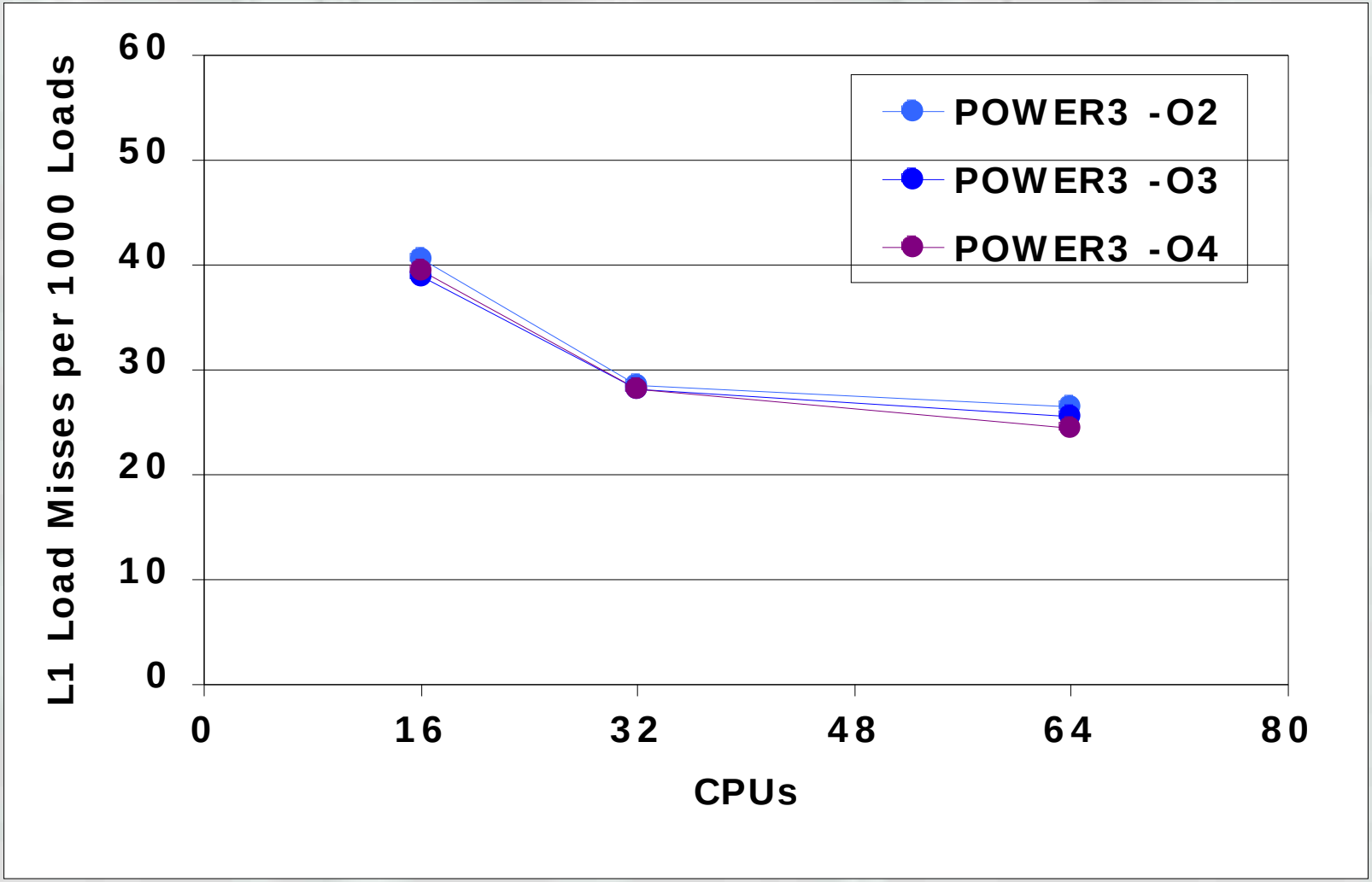
RFCTH FP Instructions (Average per Process)



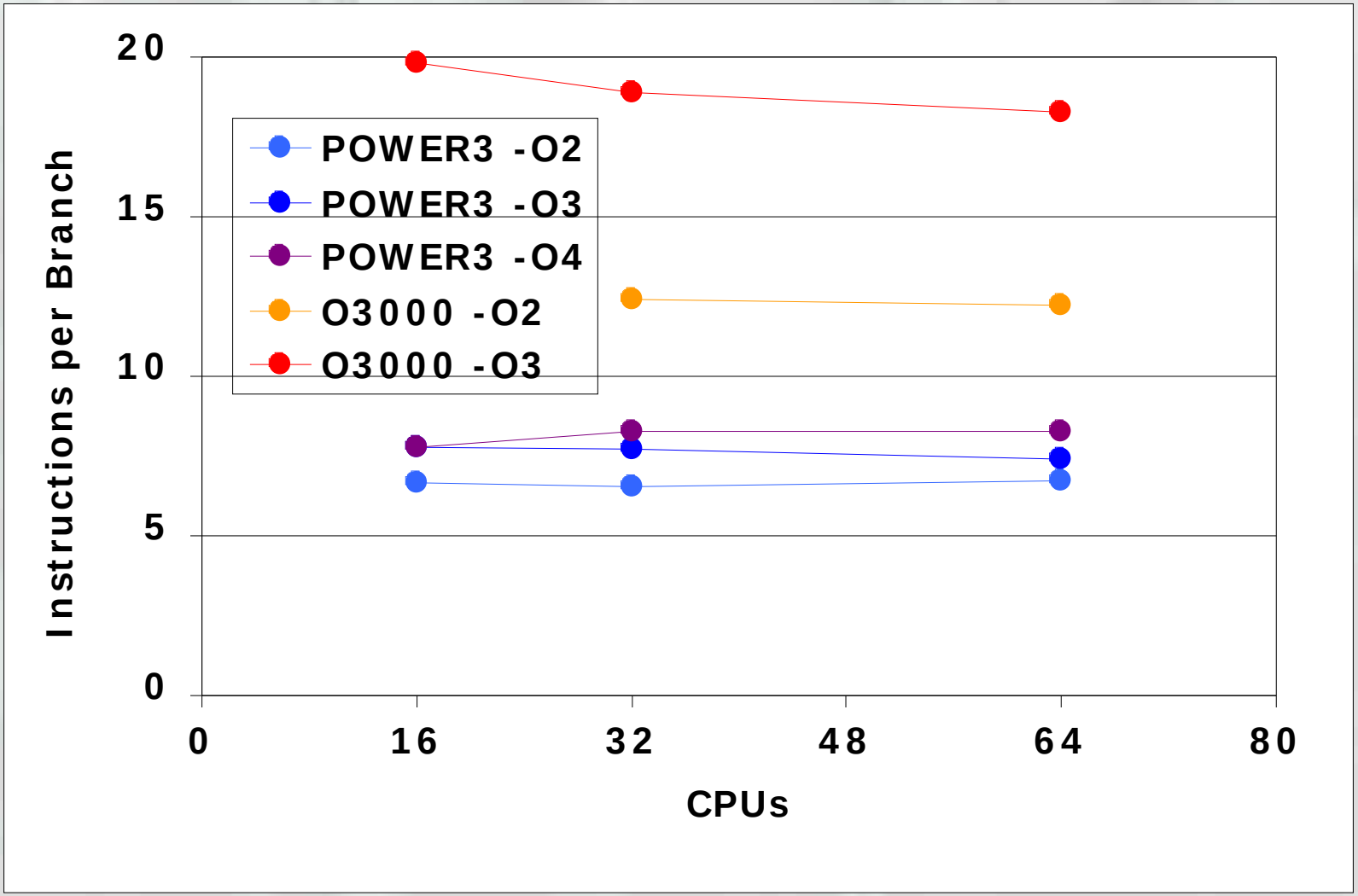
RFCTH L1 Load Misses (Average per Process)



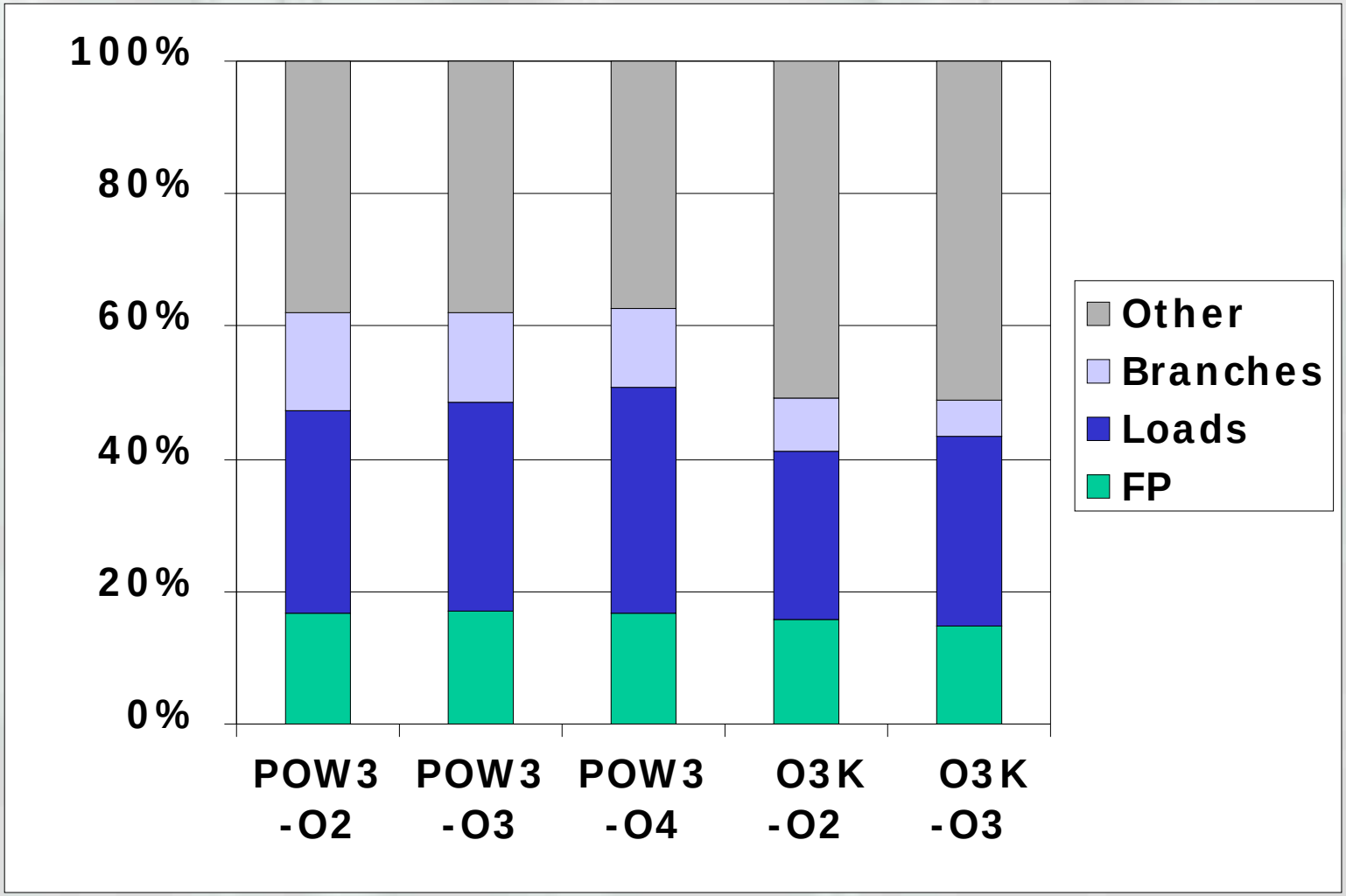
RFCTH L1 Load Misses per 1000 Loads (Average per Process)



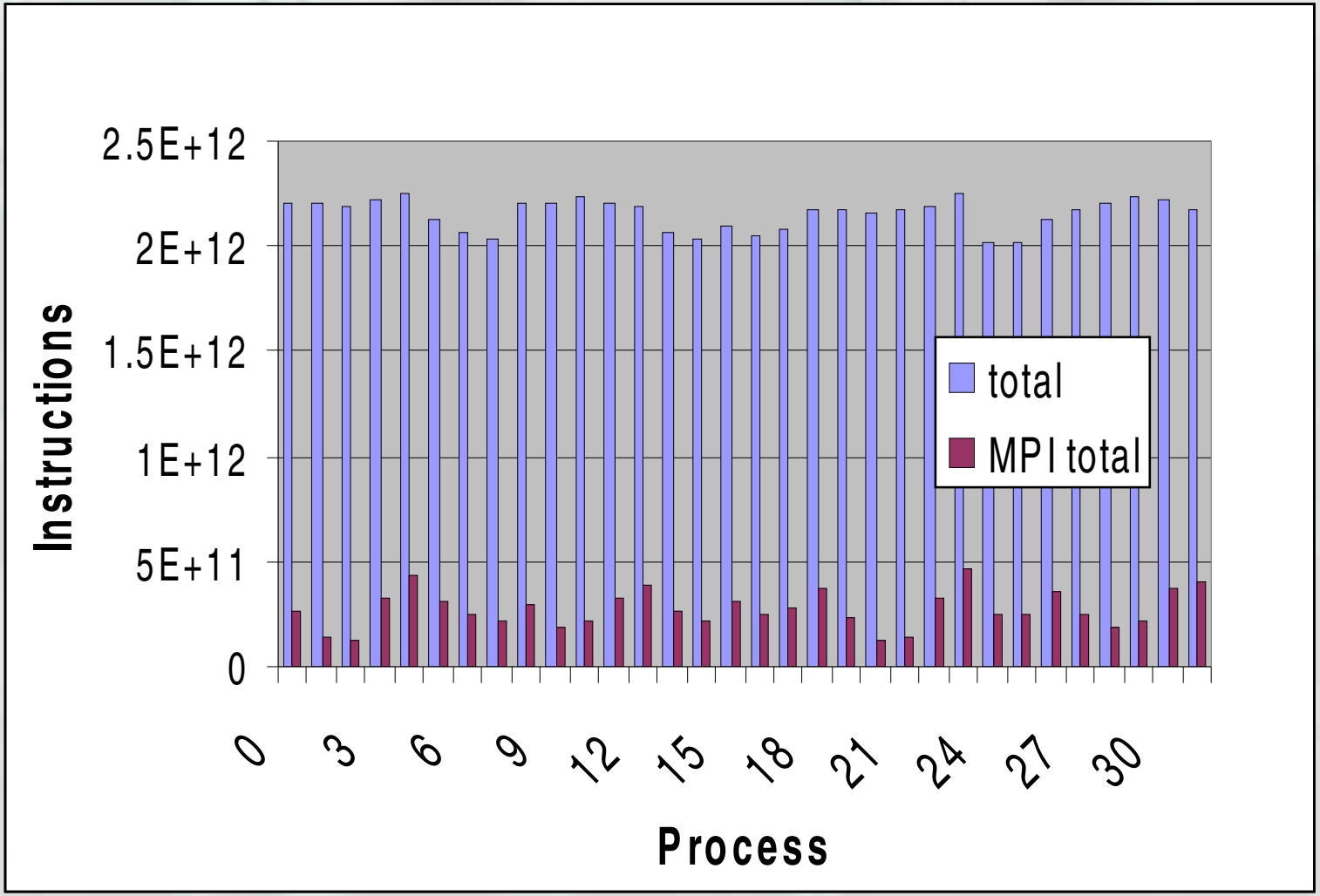
RFCTH Instructions per Branch (Average per Process)



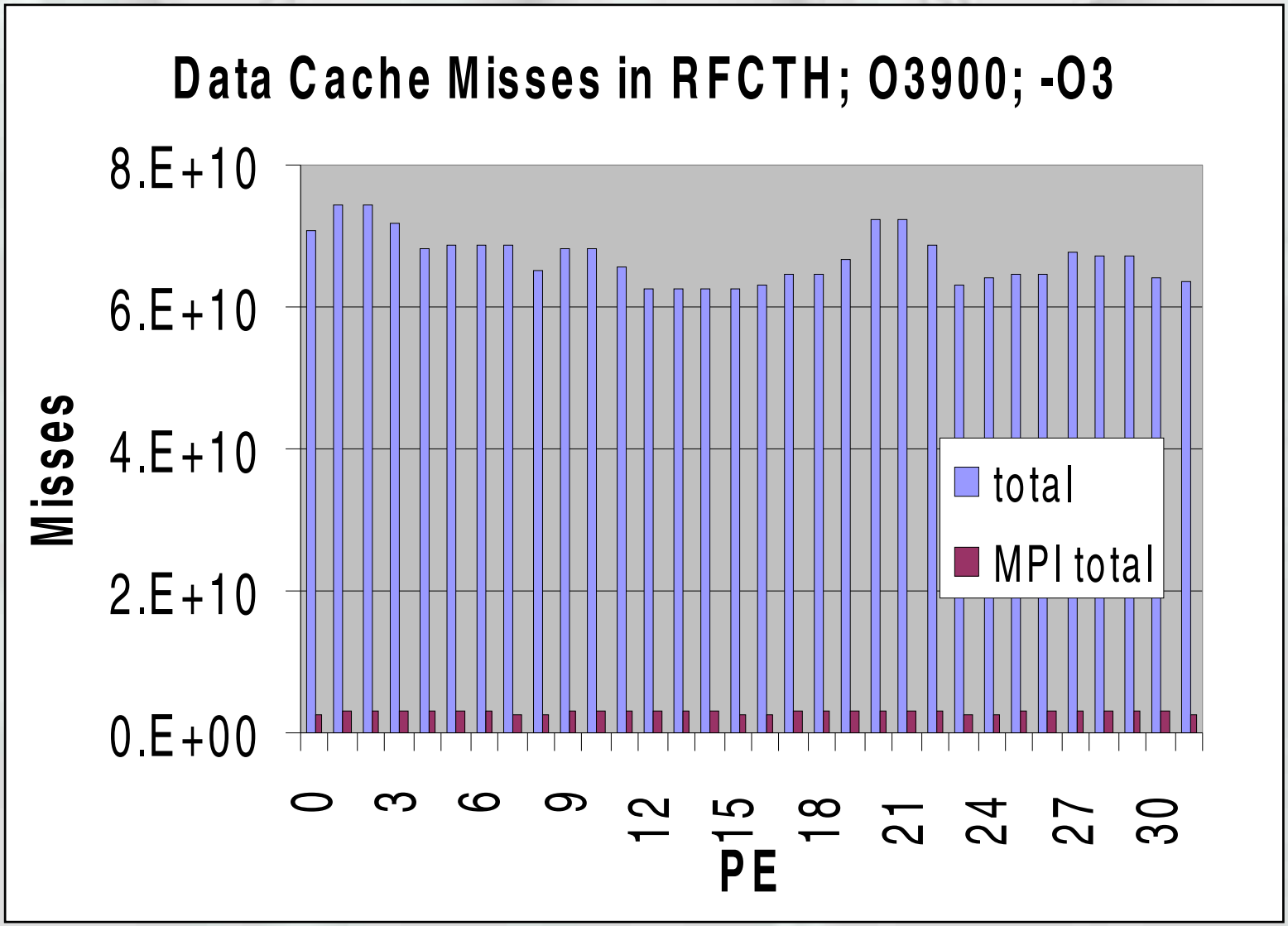
RFCTH Instruction Mix at 64 CPUs



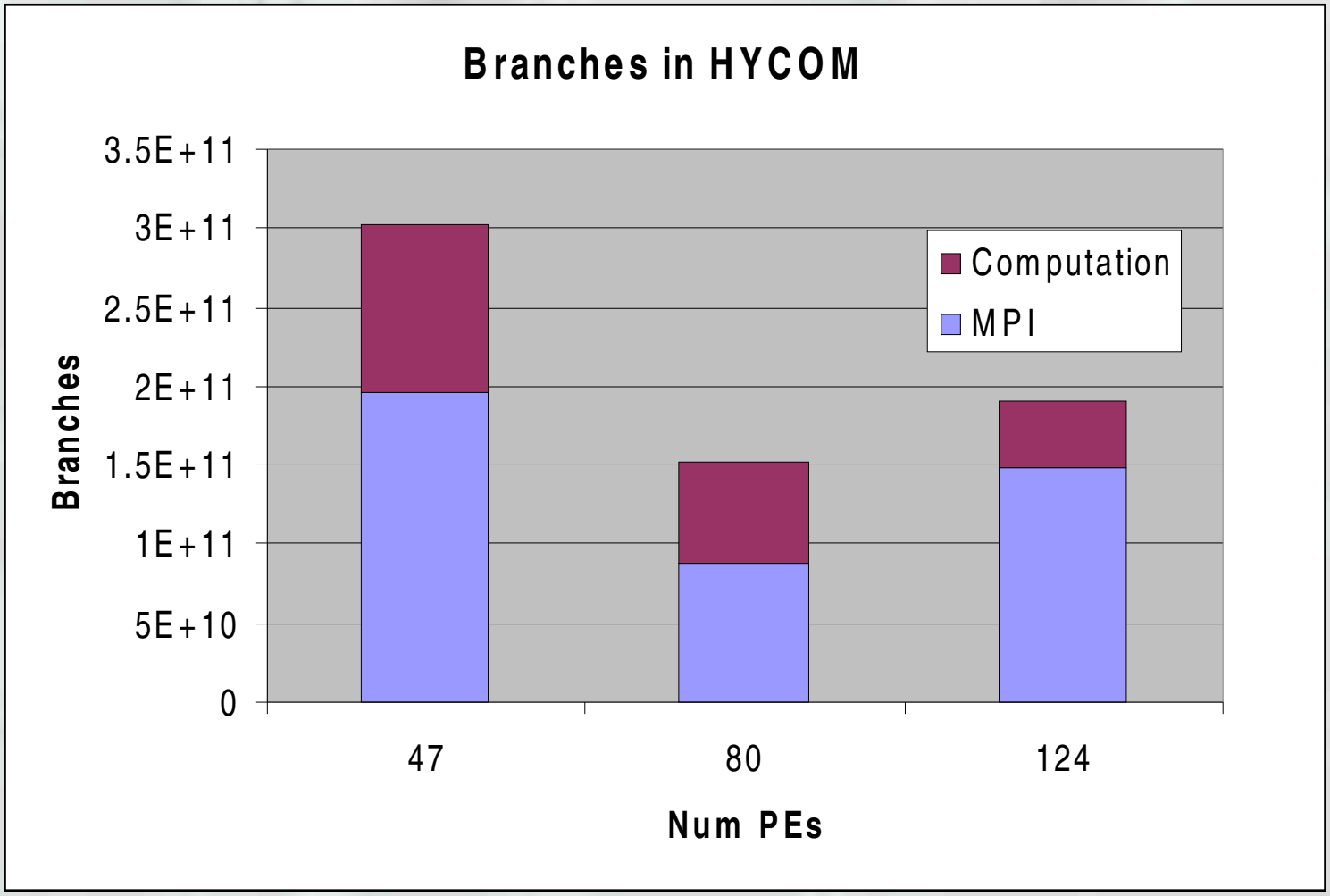
RFCTH Instruction Count by Process



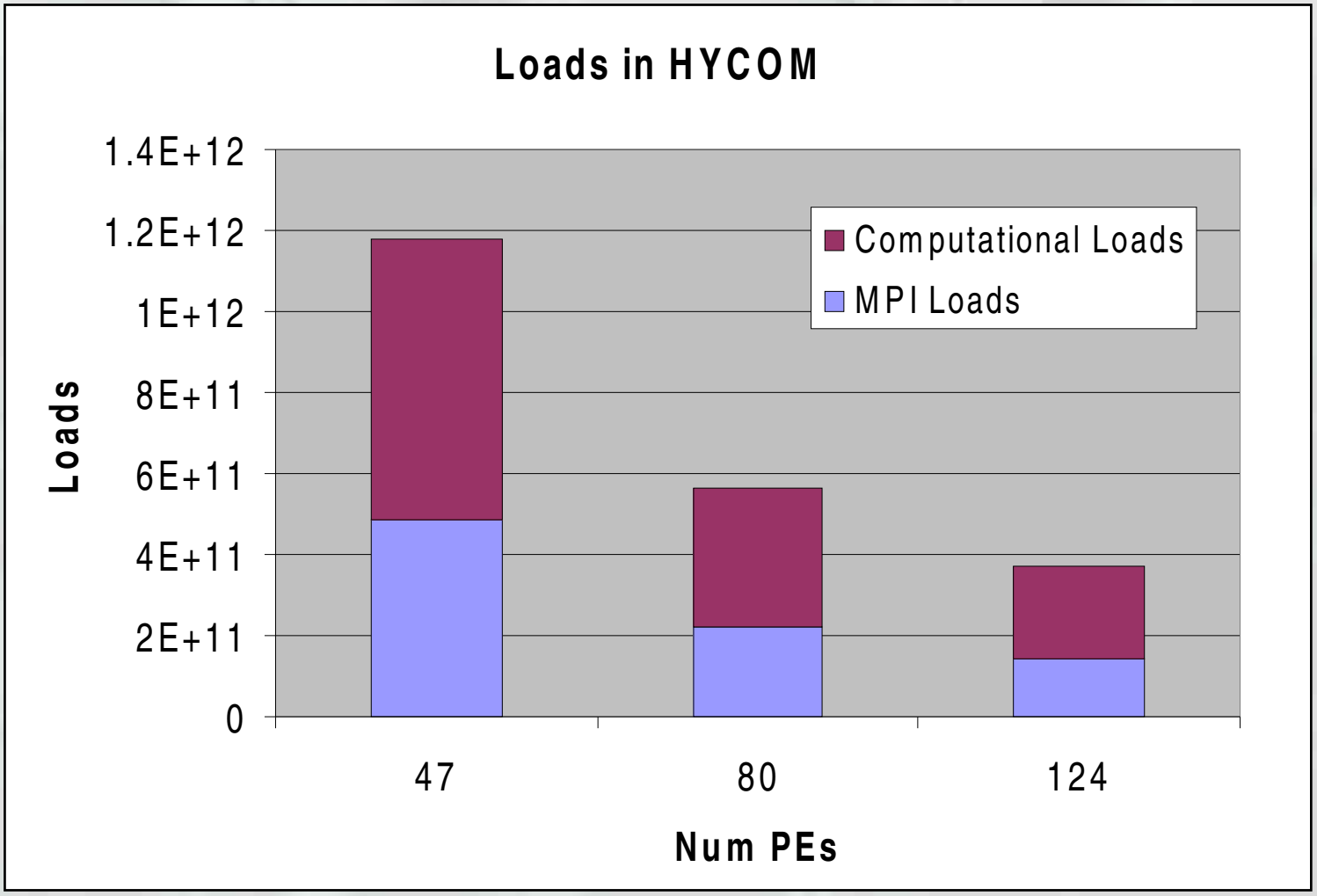
RFCTH Cache Misses



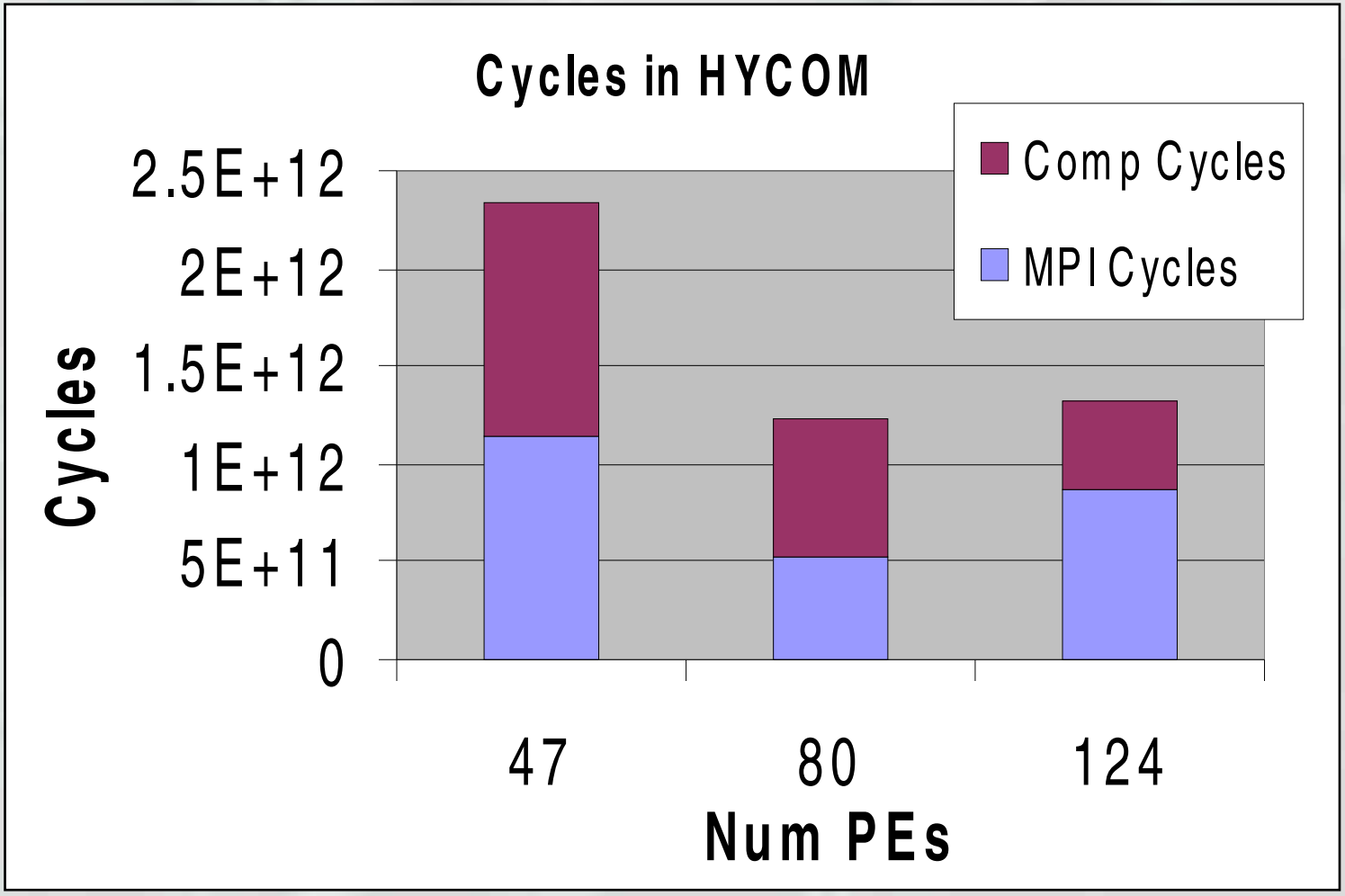
HYCOM Profiling



HYCOM Profiling



HYCOM Profiling



Profiling Issues

- Vendors occasionally unable to tell us what's being counted
- 10-100% increase in runtime on the O3K when MPI is instrumented
- Slight speedup on POWER3 when PAPI is used
- Different vendors count different events

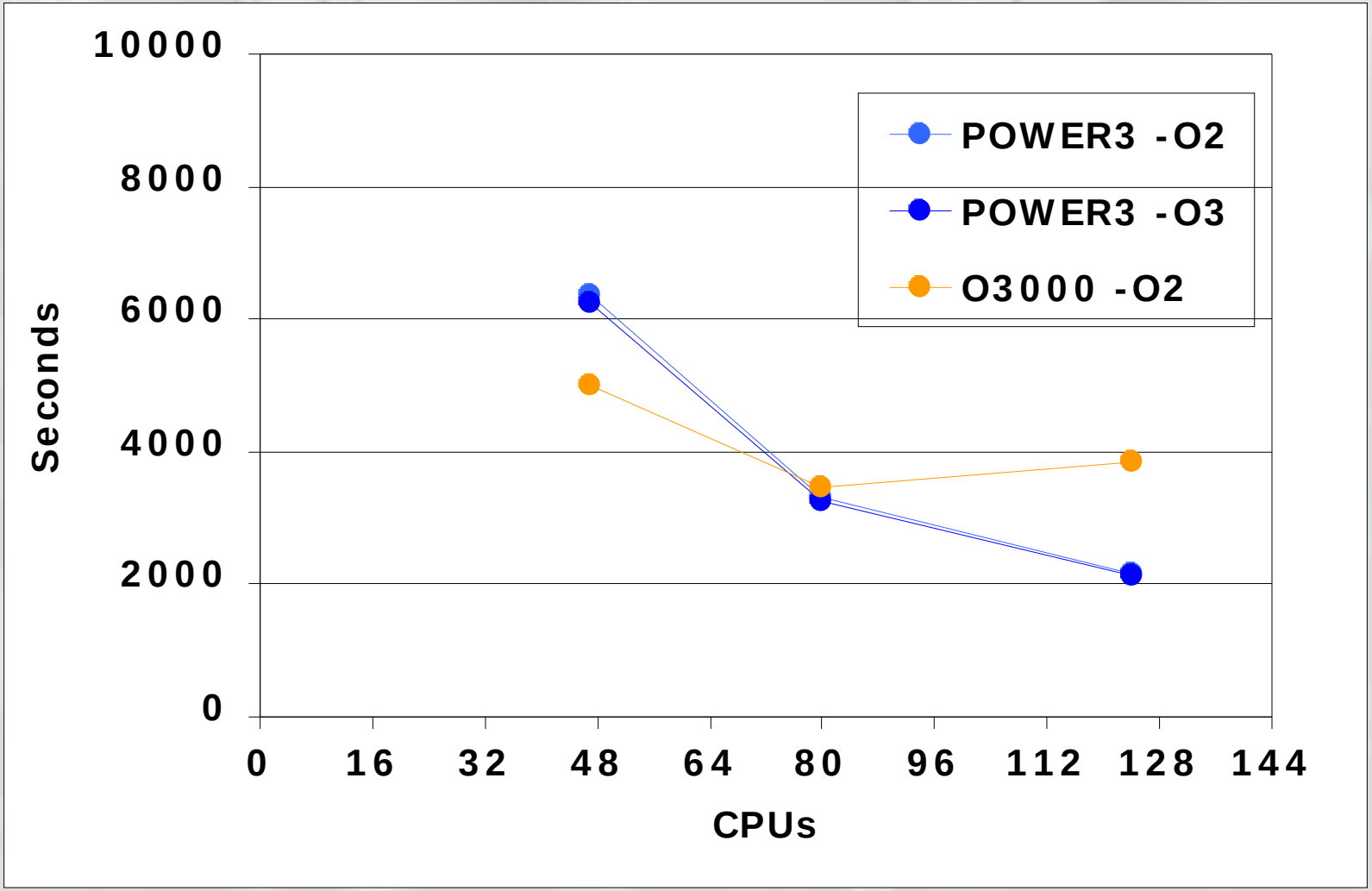
Profiling Issues (continued)

- Data management – amount of data is large and growing; had to create a data base to hold it
- Data gathering environment – production environment, not dedicated, so run-to-run variation
- Need to quantify effects of each source of variation

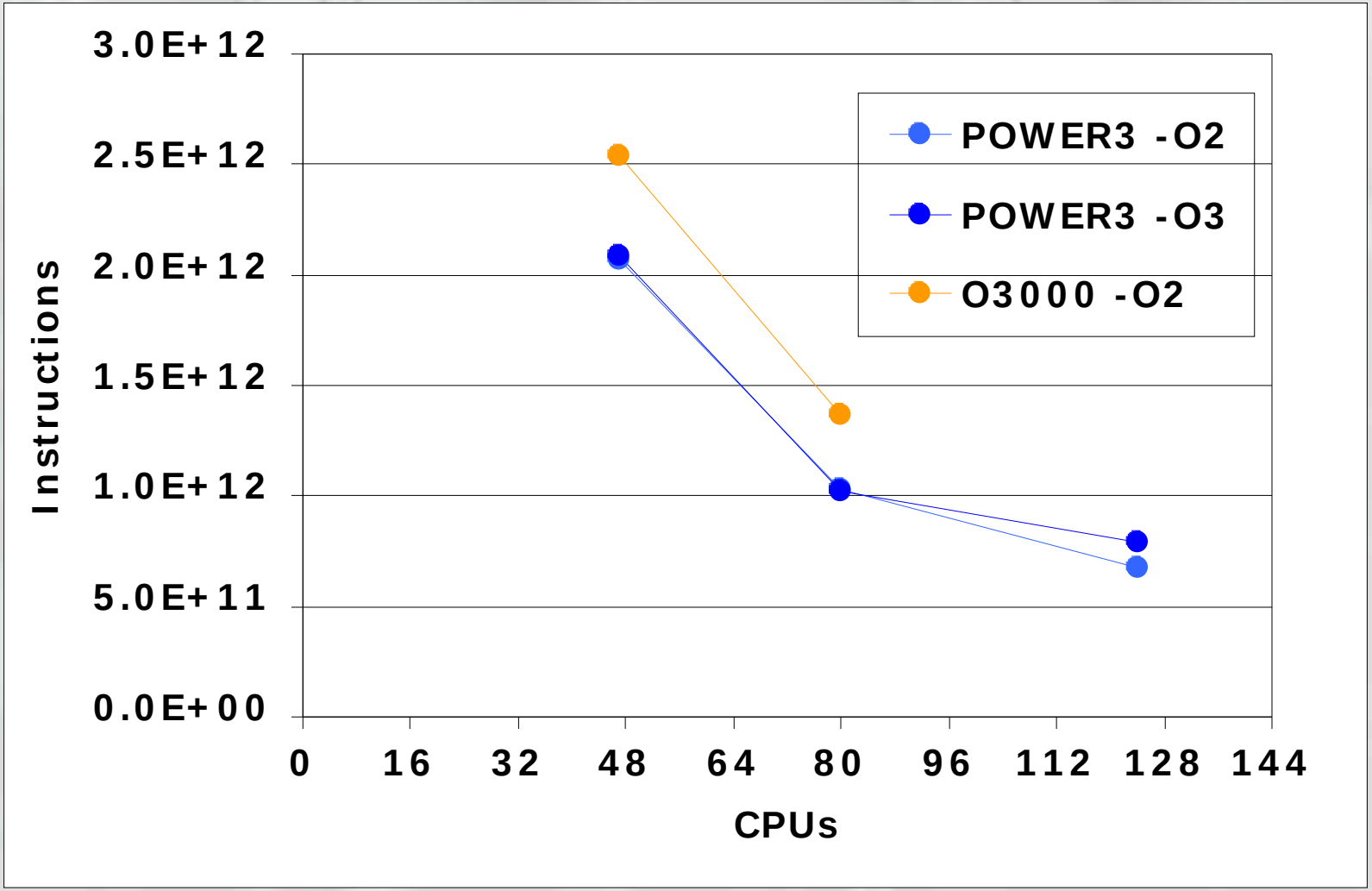
Acknowledgements

- Computational Science & Engineering Group gathered this data
  - Bob Alter
  - Paul Bennett
  - Sam Cable
  - Carrie Leach
  - Tom Oppe

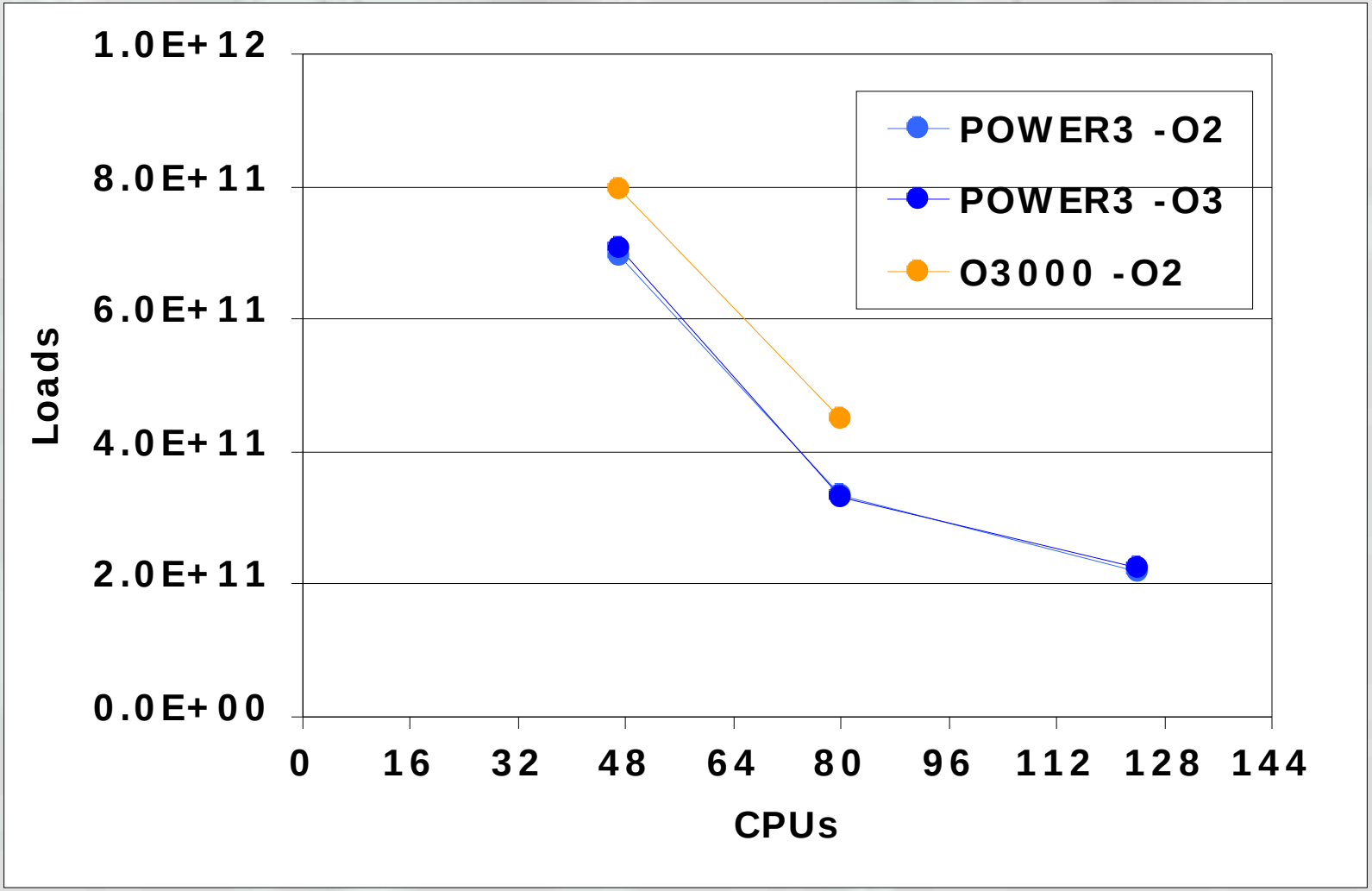
HYCOM Wall Time (Average per Process)



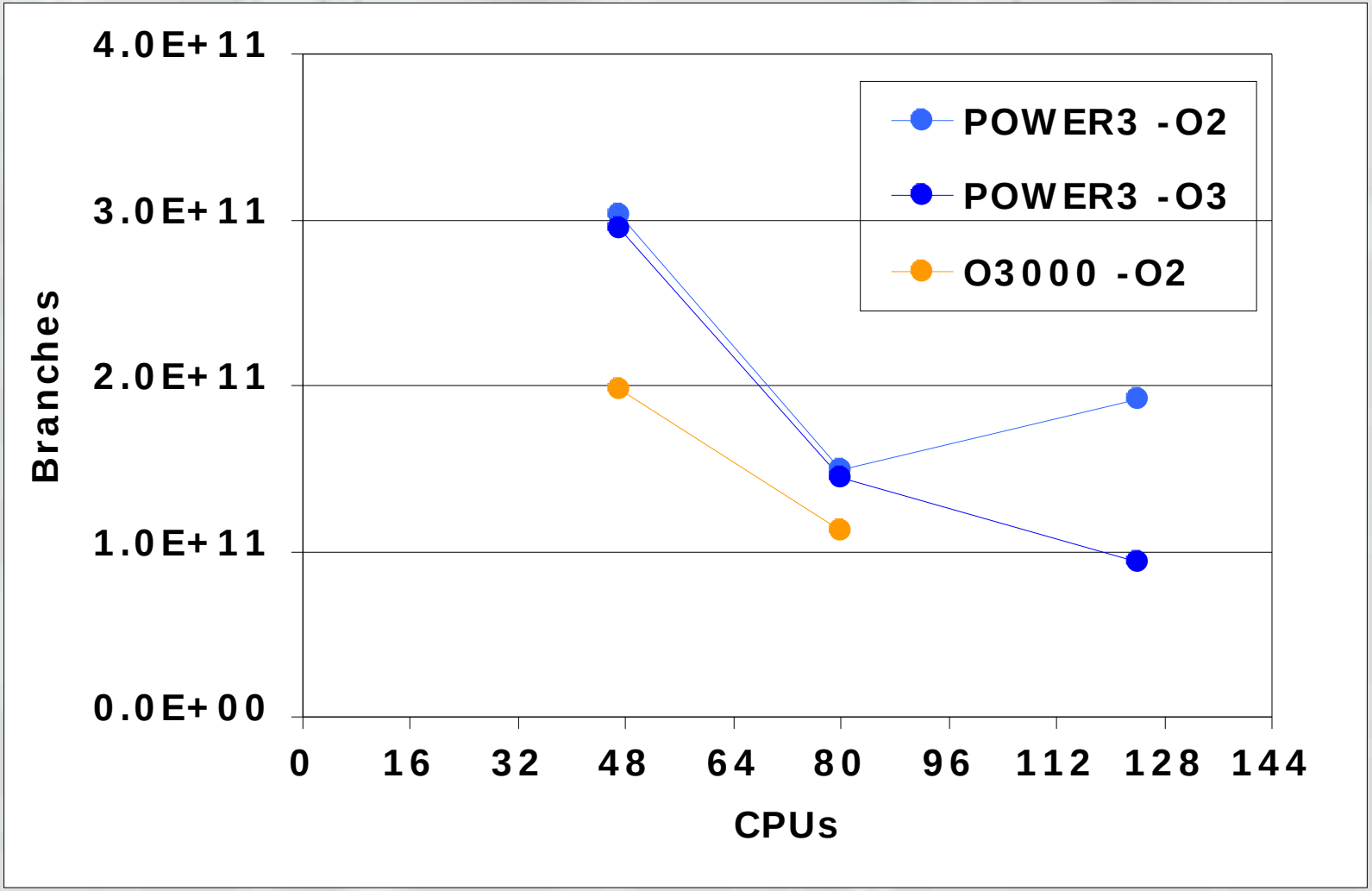
HYCOM Instruction Count (Average per Process)



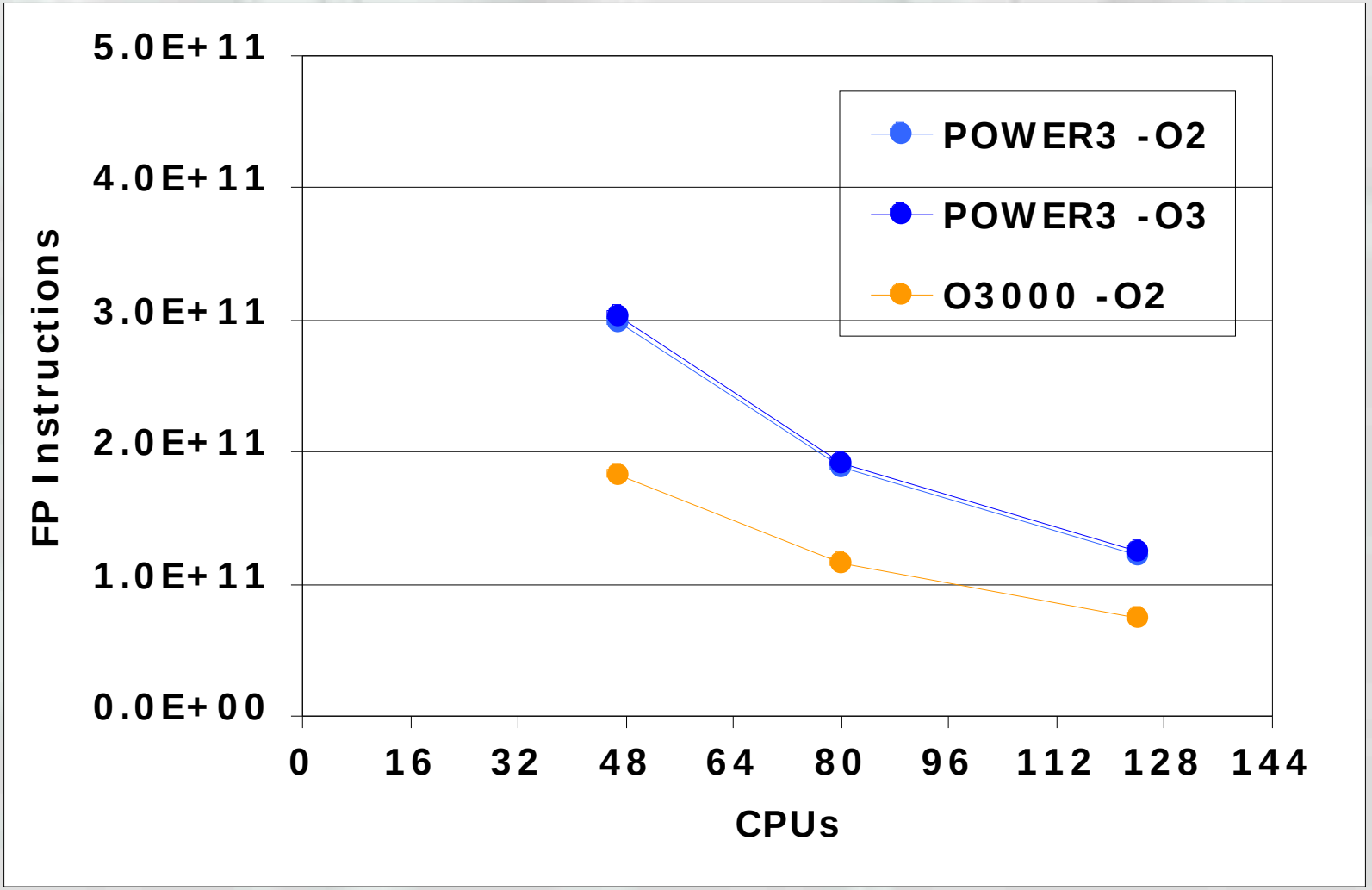
HYCOM Loads Performed (Average per Process)



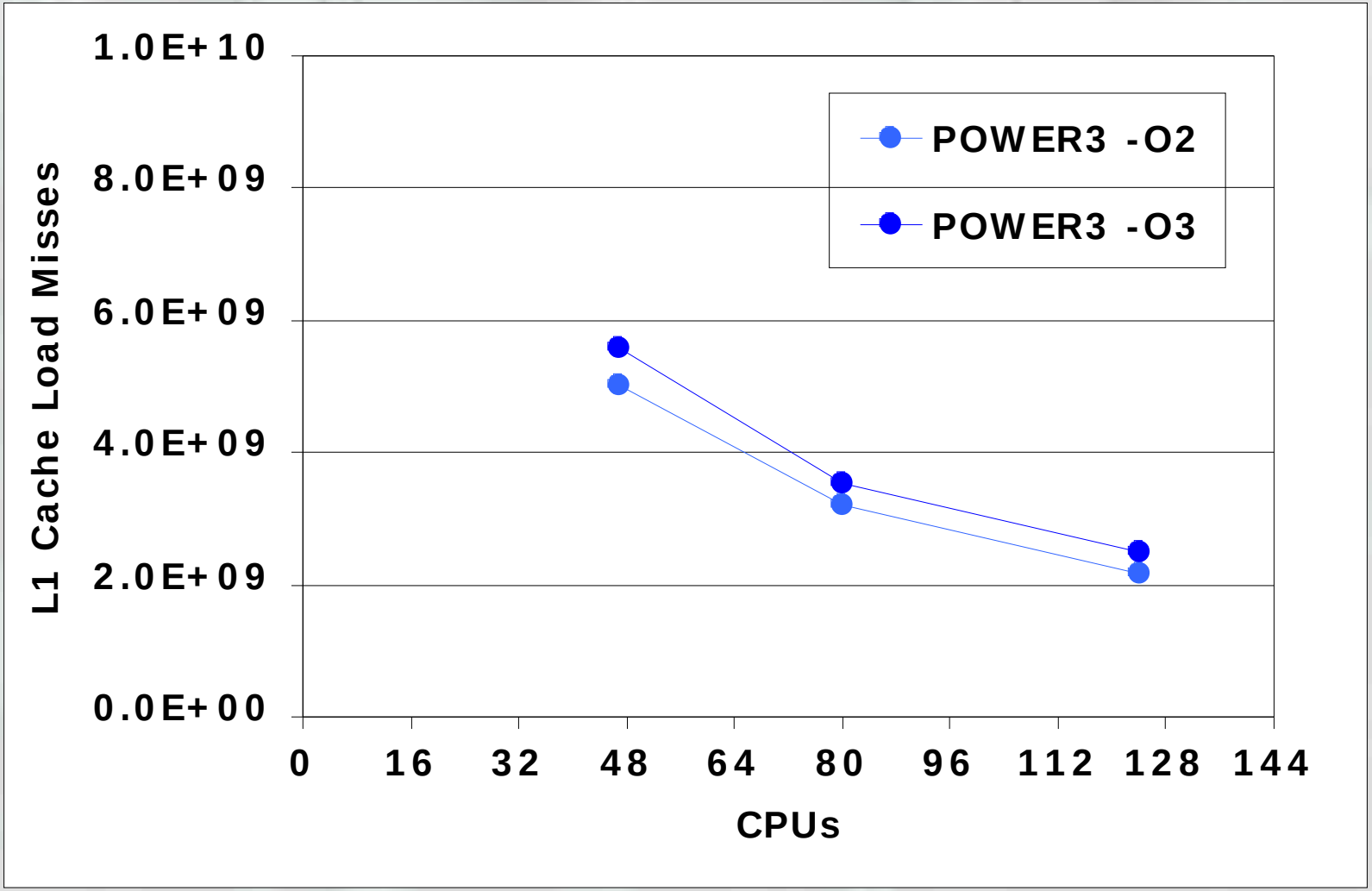
HYCOM Branches Performed (Average per Process)



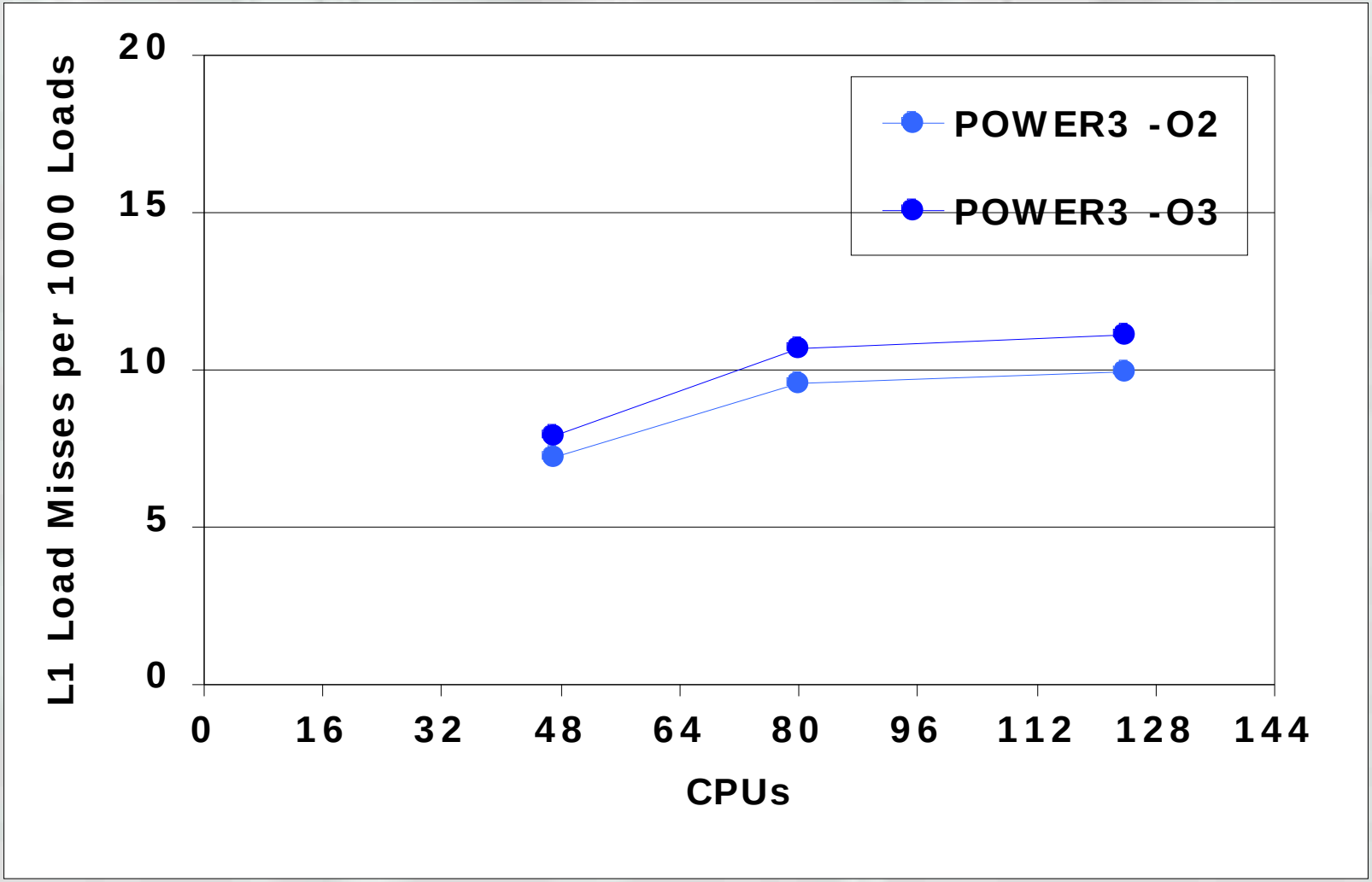
HYCOM FP Instructions (Average per Process)



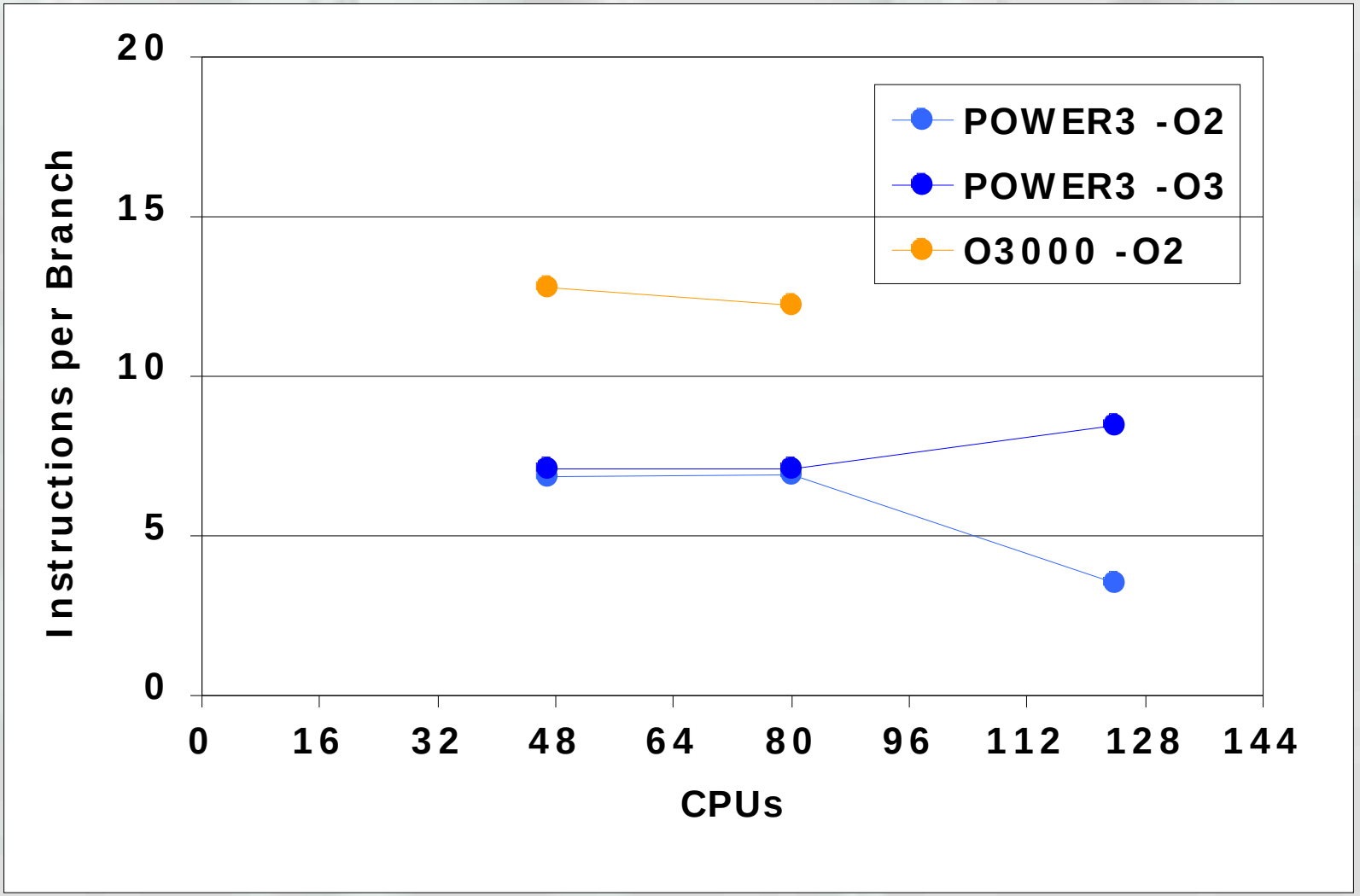
HYCOM L1 Load Misses (Average per Process)



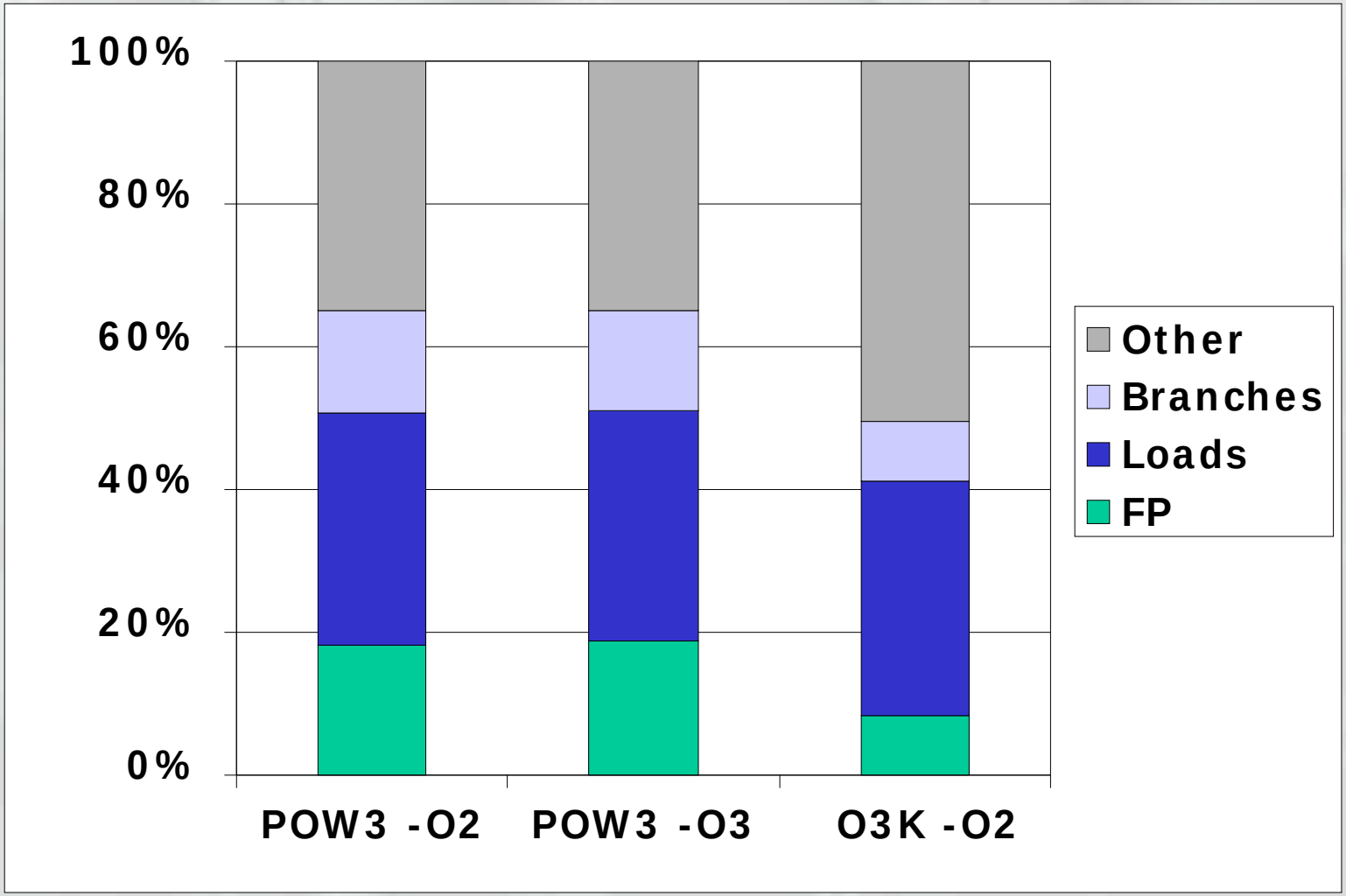
HYCOM L1 Load Misses per 1000 Loads (Average per Process)



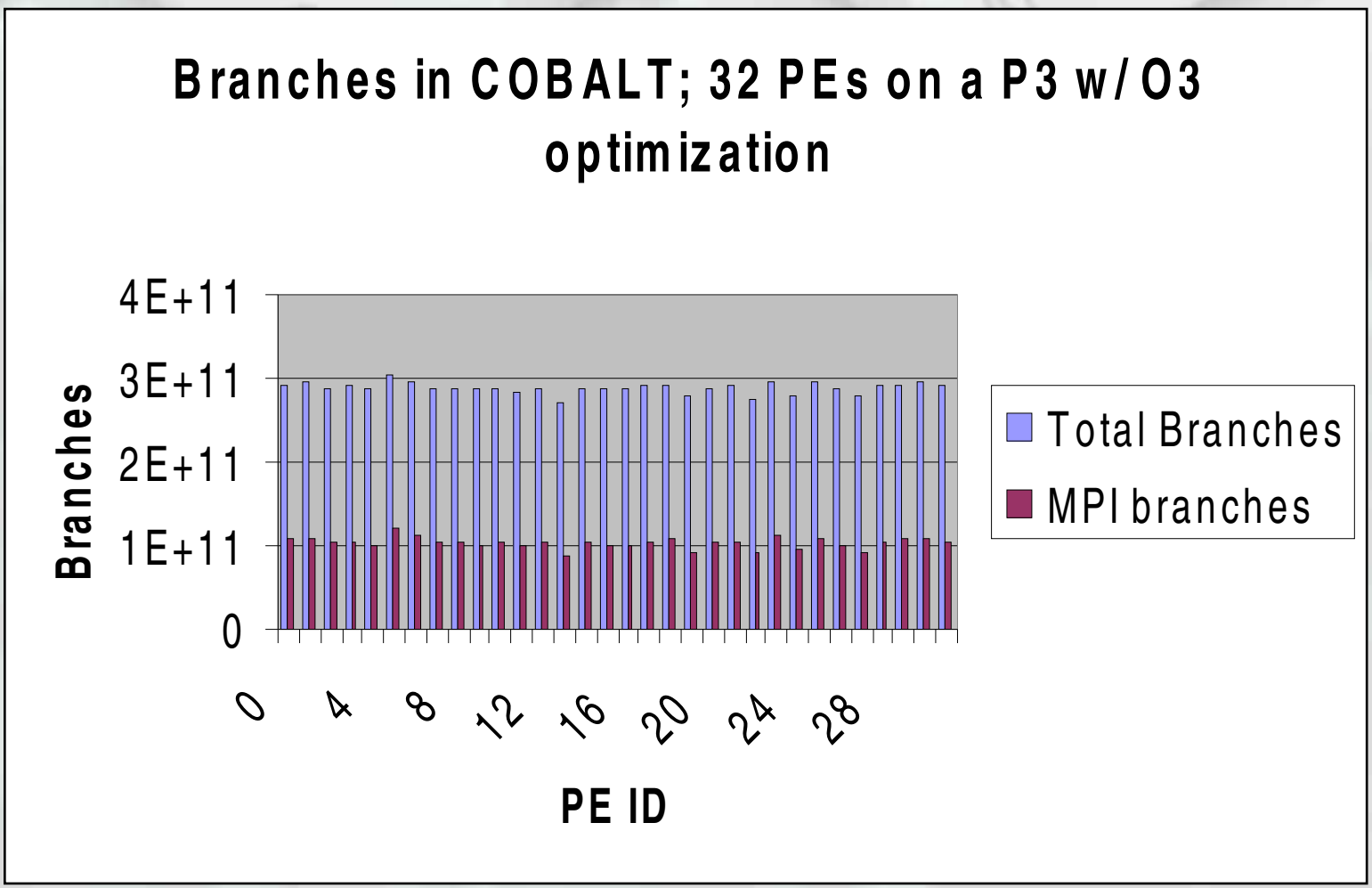
HYCOM Instructions per Branch (Average per Process)

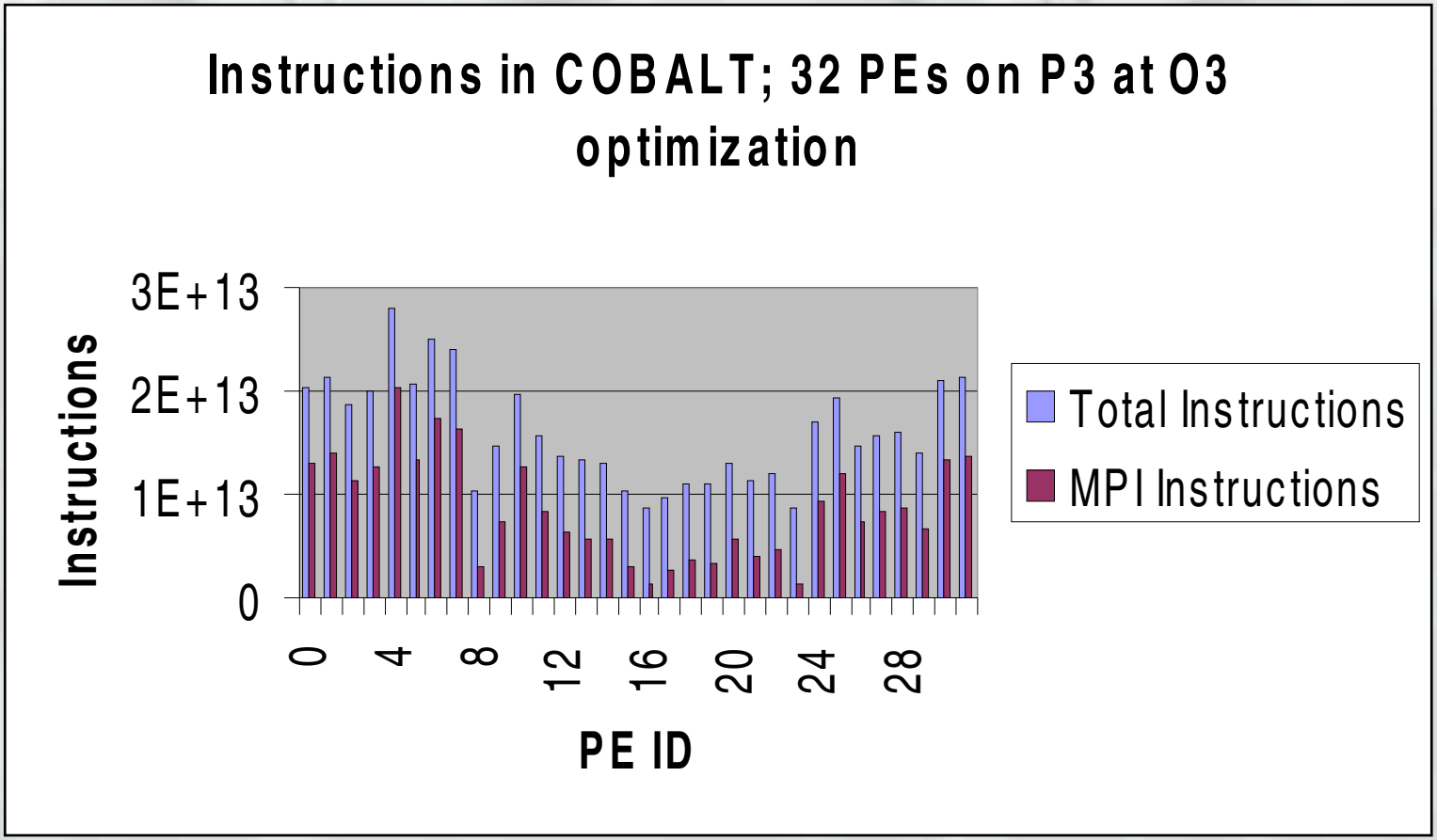


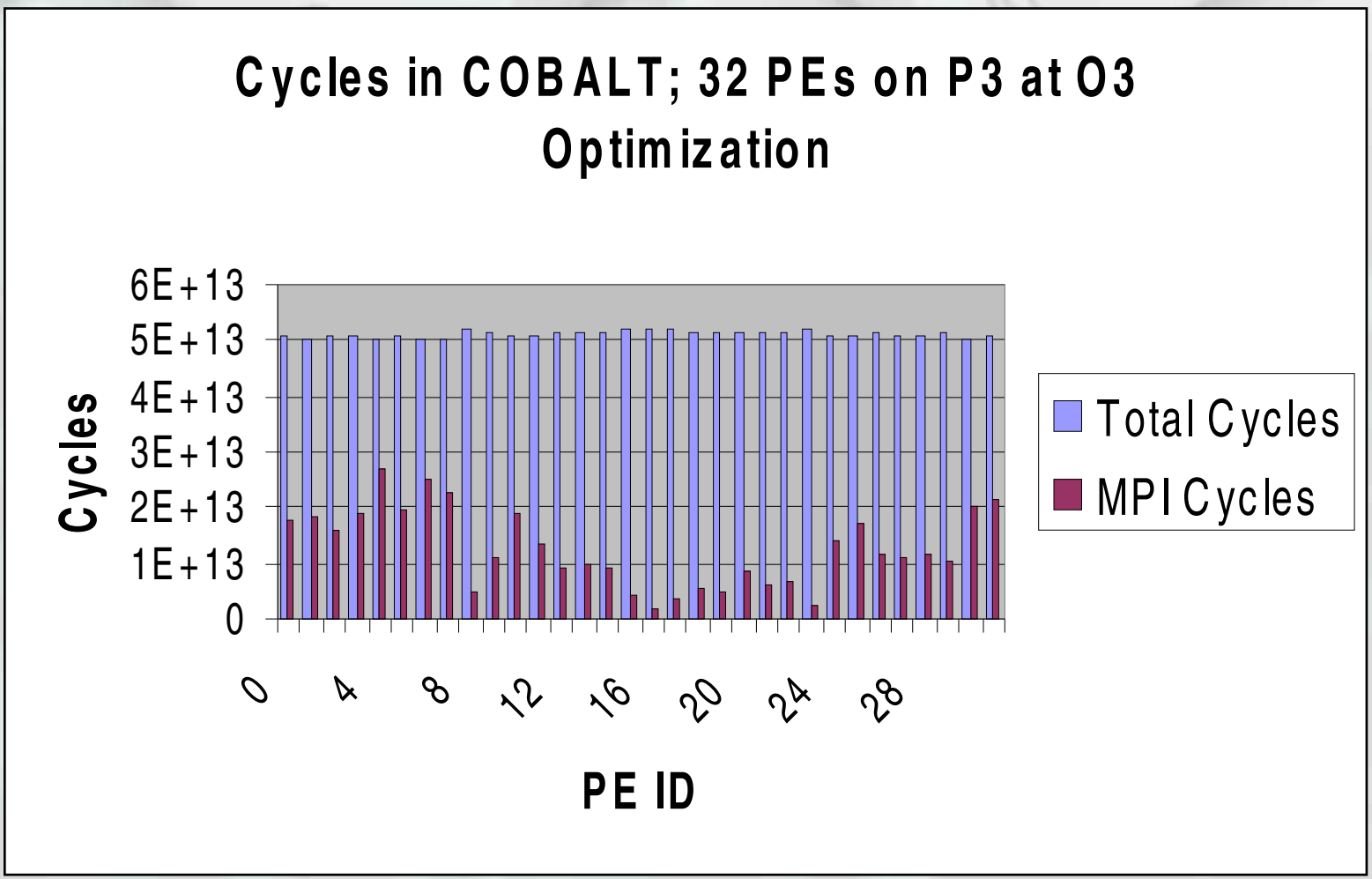
HYCOM Instruction Mix at 80 CPUs

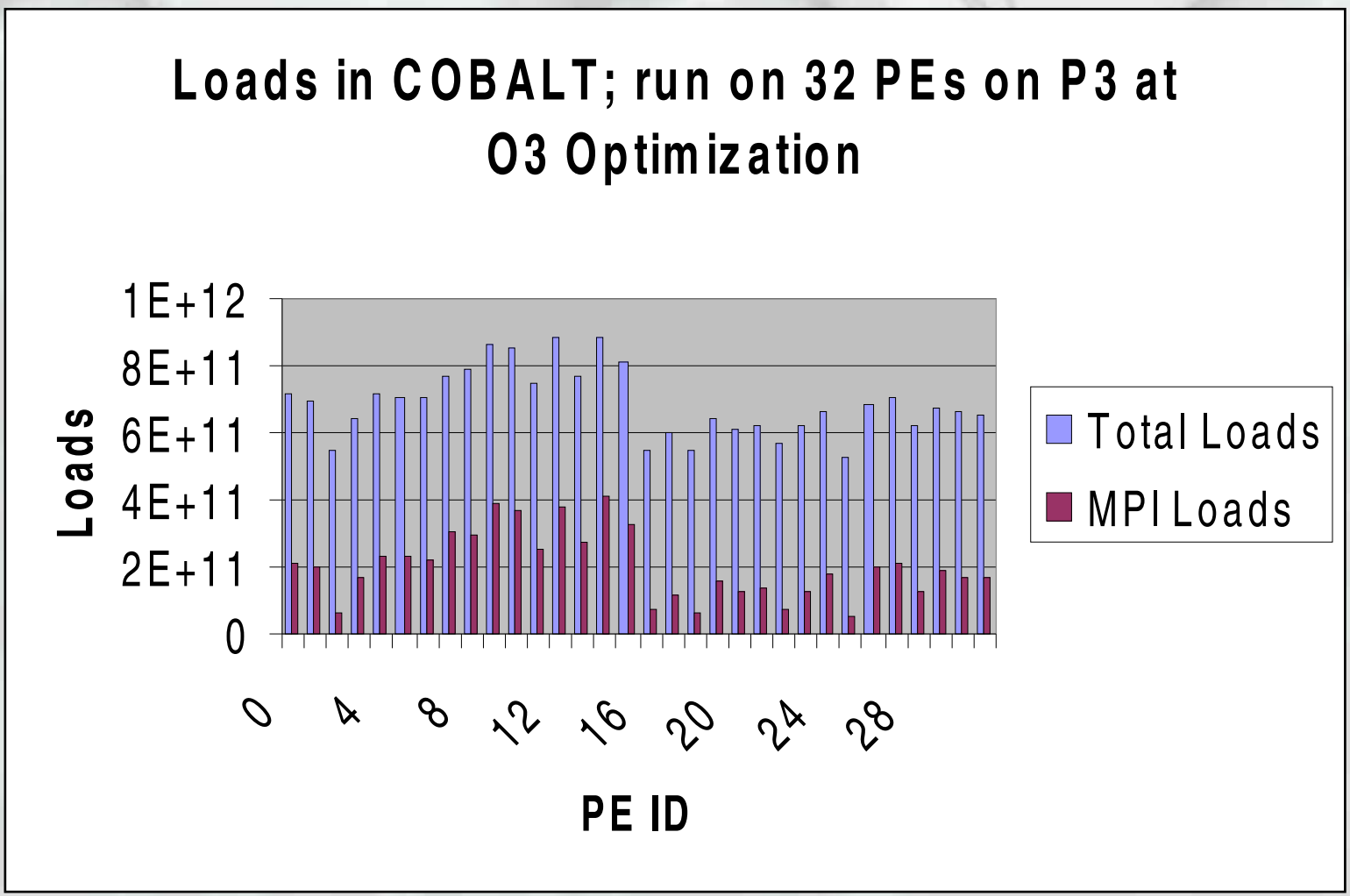


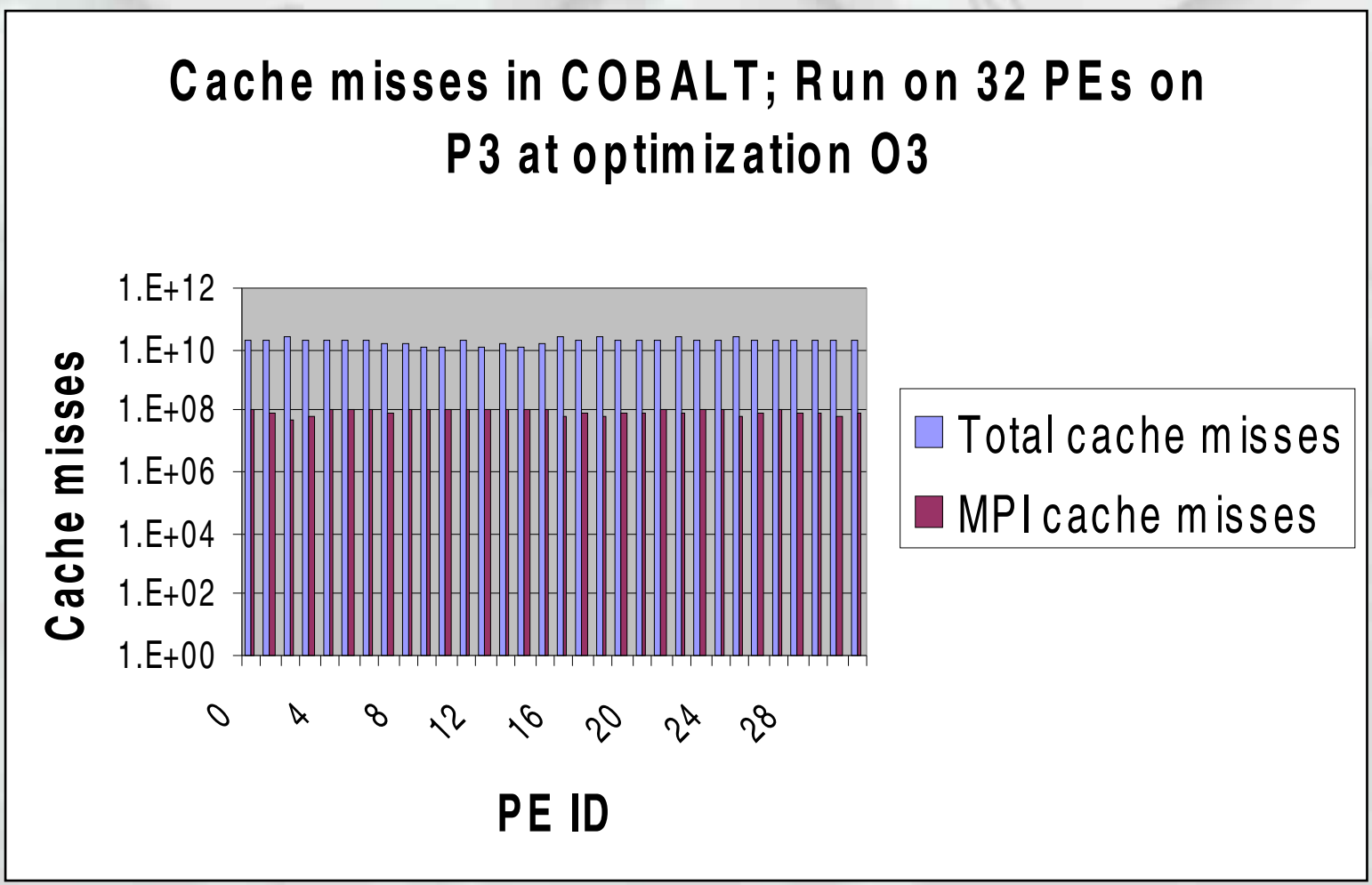
Extras

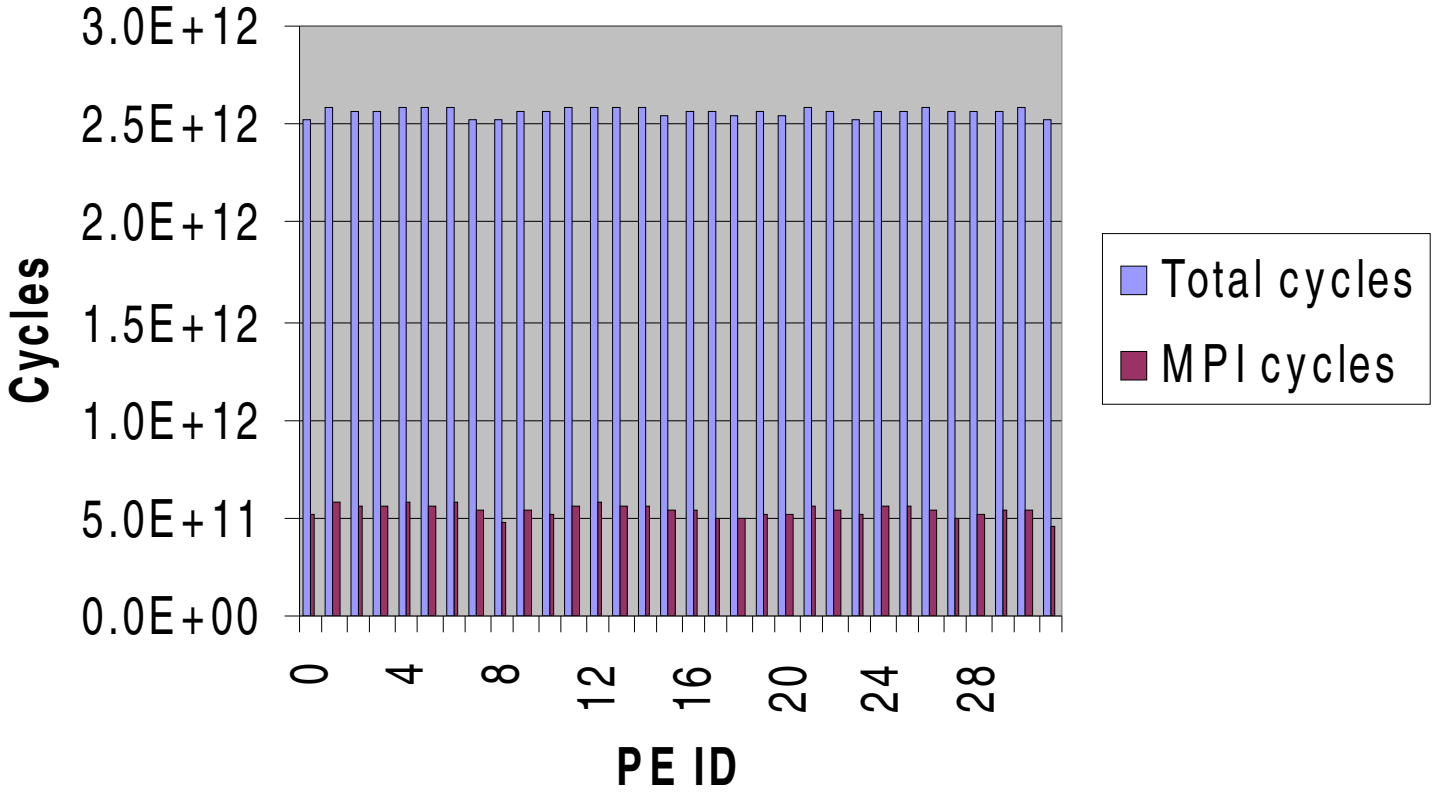

Branches in COBALT; 32 PEs on a P3 w/ O3 optimization
Branches
4E+11
3E+11
2E+11
1E+11
0
0
4
8
12
16
20
24
28
PE ID
Total Branches
MPI branches

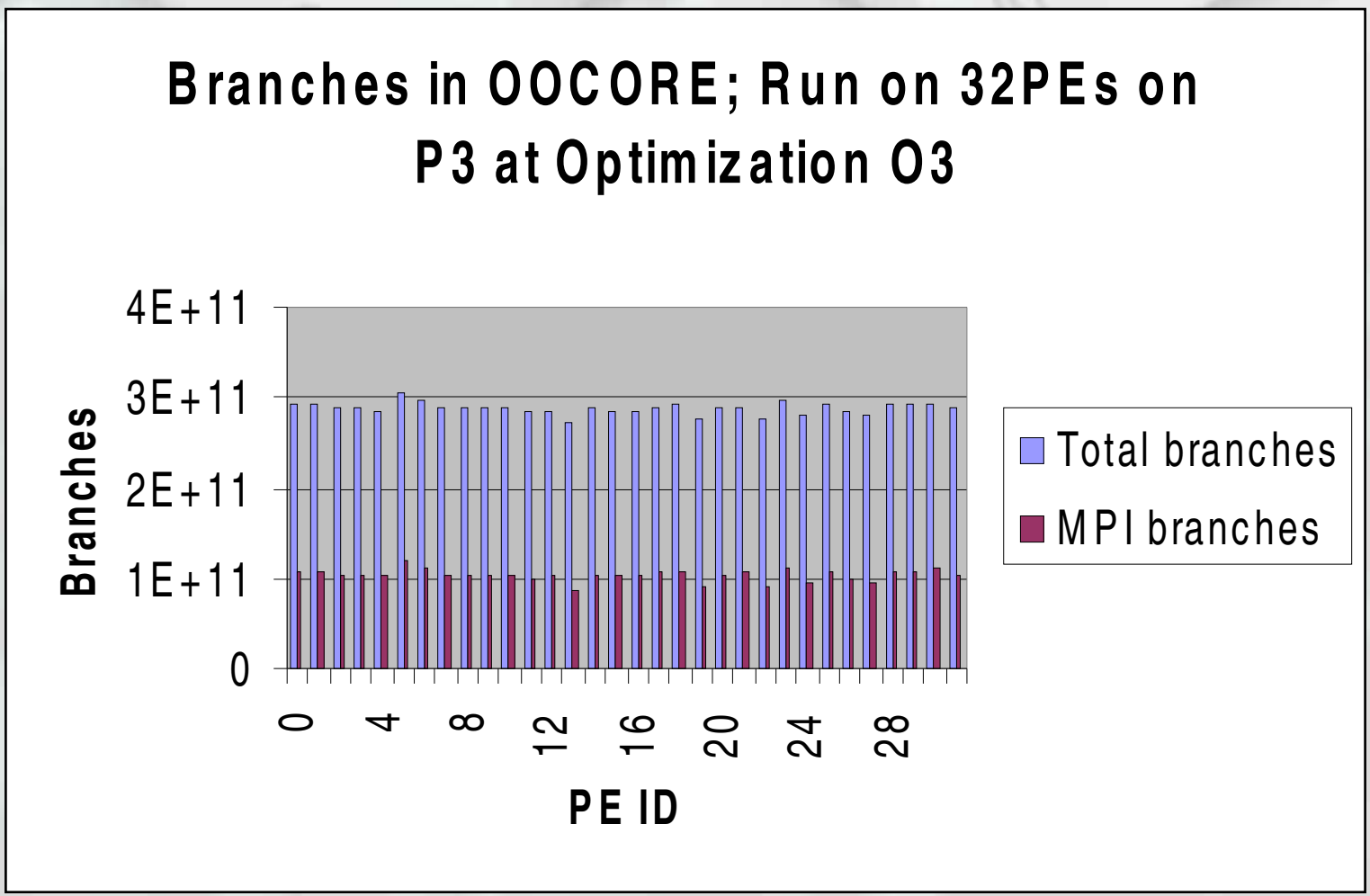

Instructions in COBALT; 32 PEs on P3 at O3 optimization
Instructions
3E+13
2E+13
1E+13
0
0
4
8
12
16
20
24
28
PE ID
Total Instructions
MPI Instructions

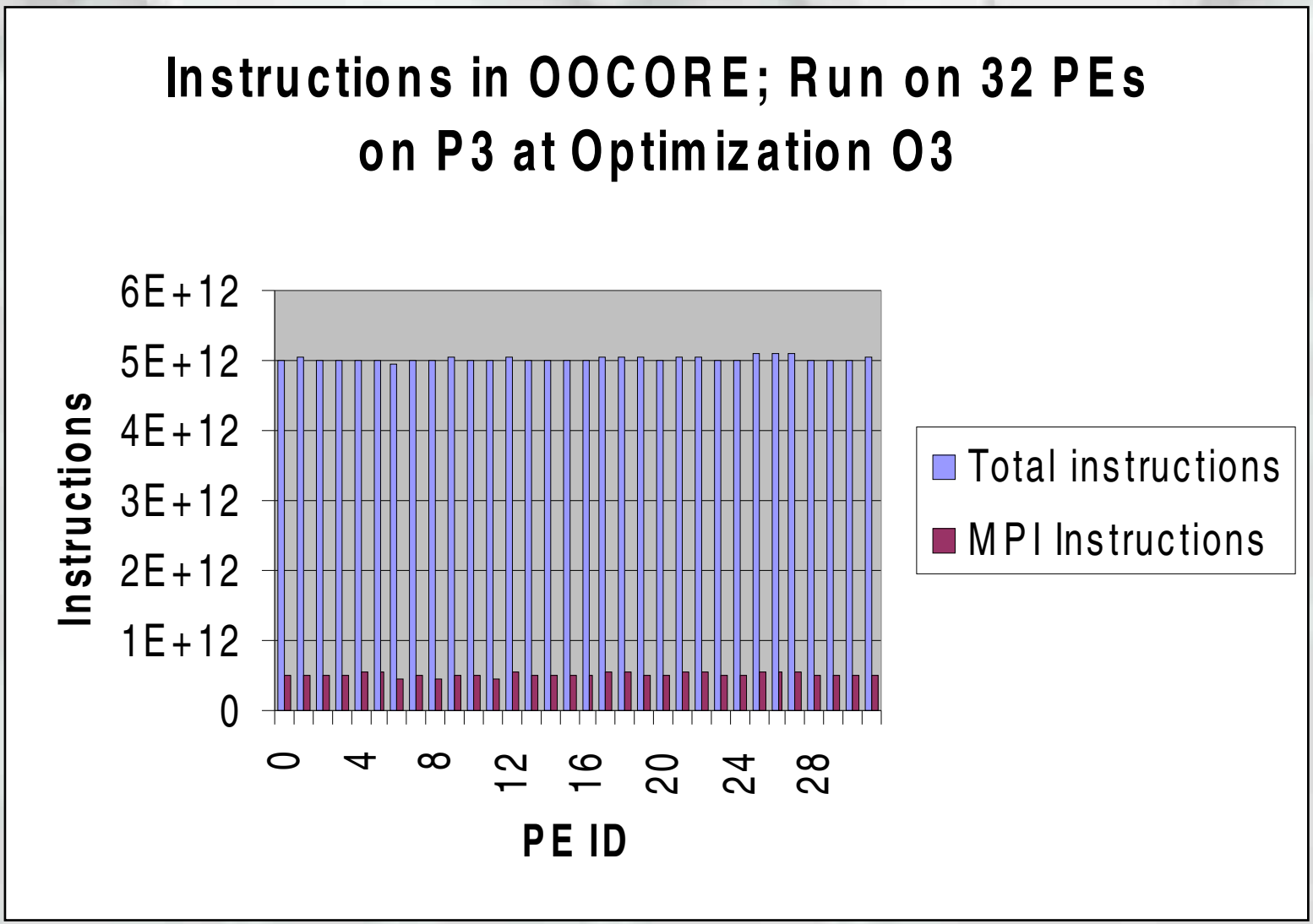

Cycles in COBALT; 32 PEs on P3 at O3 Optimization
6E+13
5E+13
4E+13
3E+13
2E+13
1E+13
0
Cycles
0
4
8
12
16
20
24
28
PE ID
Total Cycles
MPI Cycles

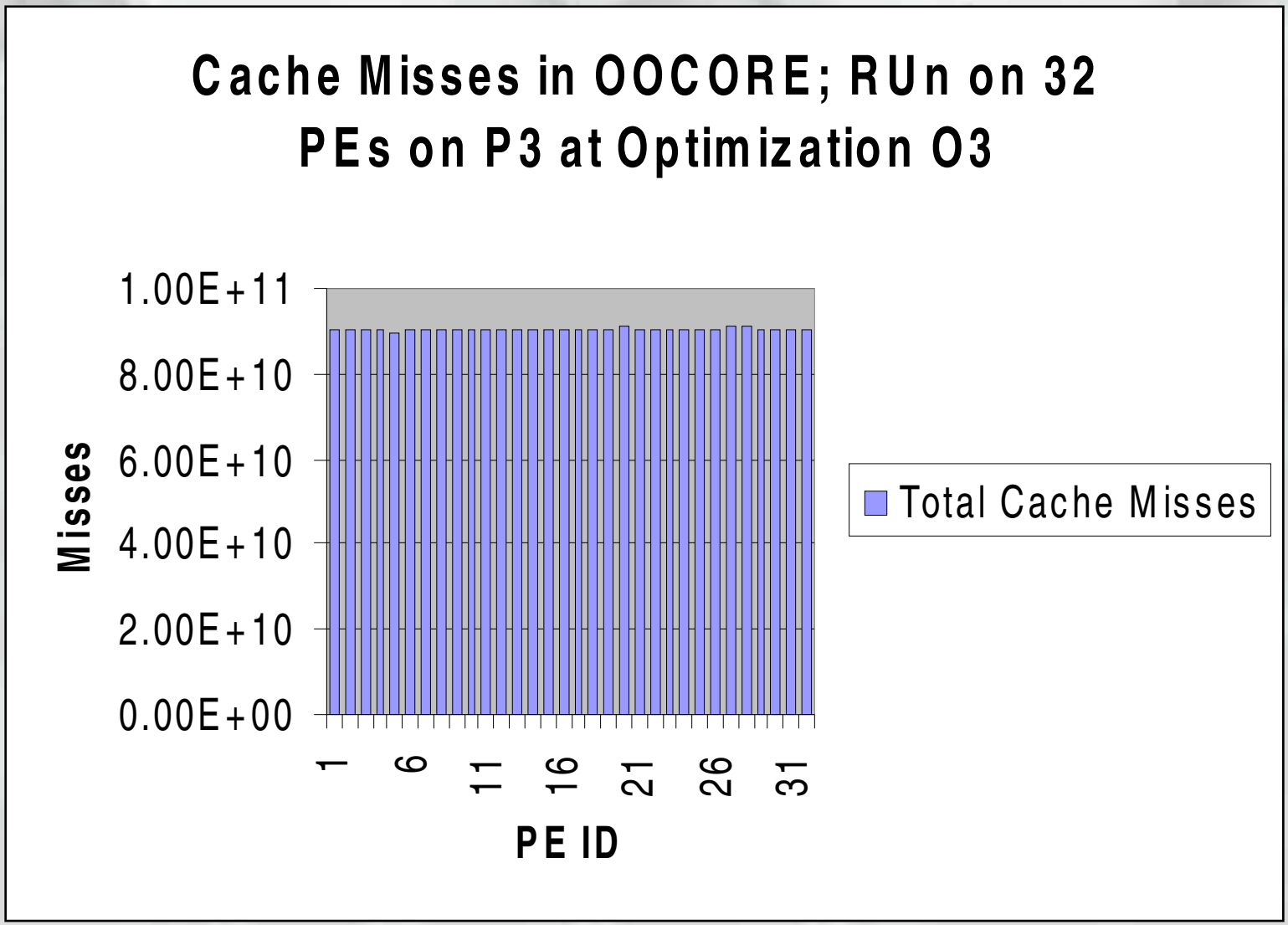

Loads in COBALT; run on 32 PEs on P3 at O3 Optimization
Loads
1E+12
8E+11
6E+11
4E+11
2E+11
0
0
4
8
12
16
20
24
28
PE ID
Total Loads
MPI Loads

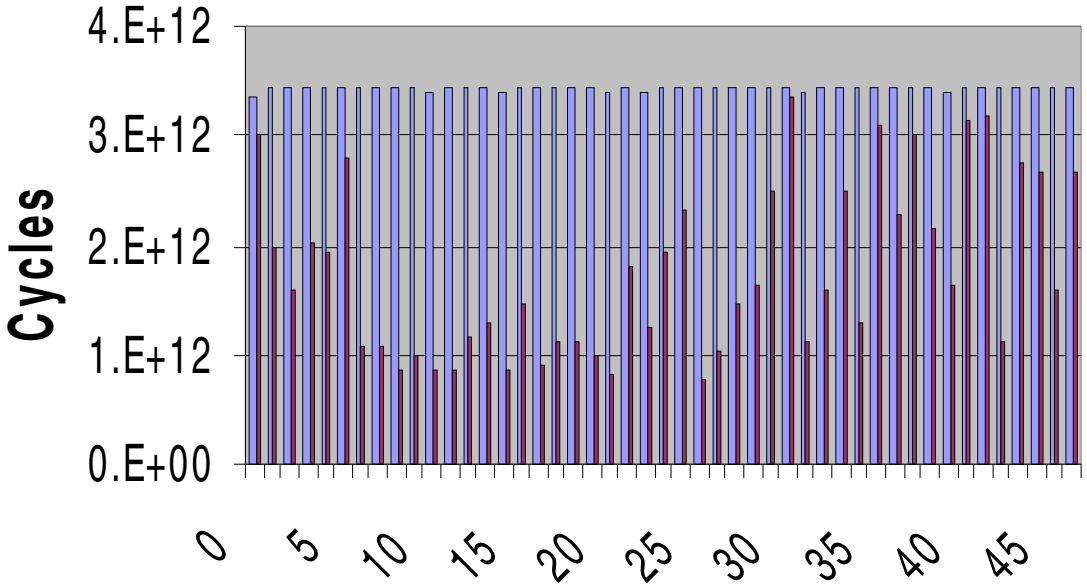

Cache misses in COBALT; Run on 32 PEs on P3 at optimization O3
Cache misses
1.E+12
1.E+10
1.E+08
1.E+06
1.E+04
1.E+02
1.E+00
0
4
8
12
16
20
24
28
PE ID
Total cache misses
MPI cache misses

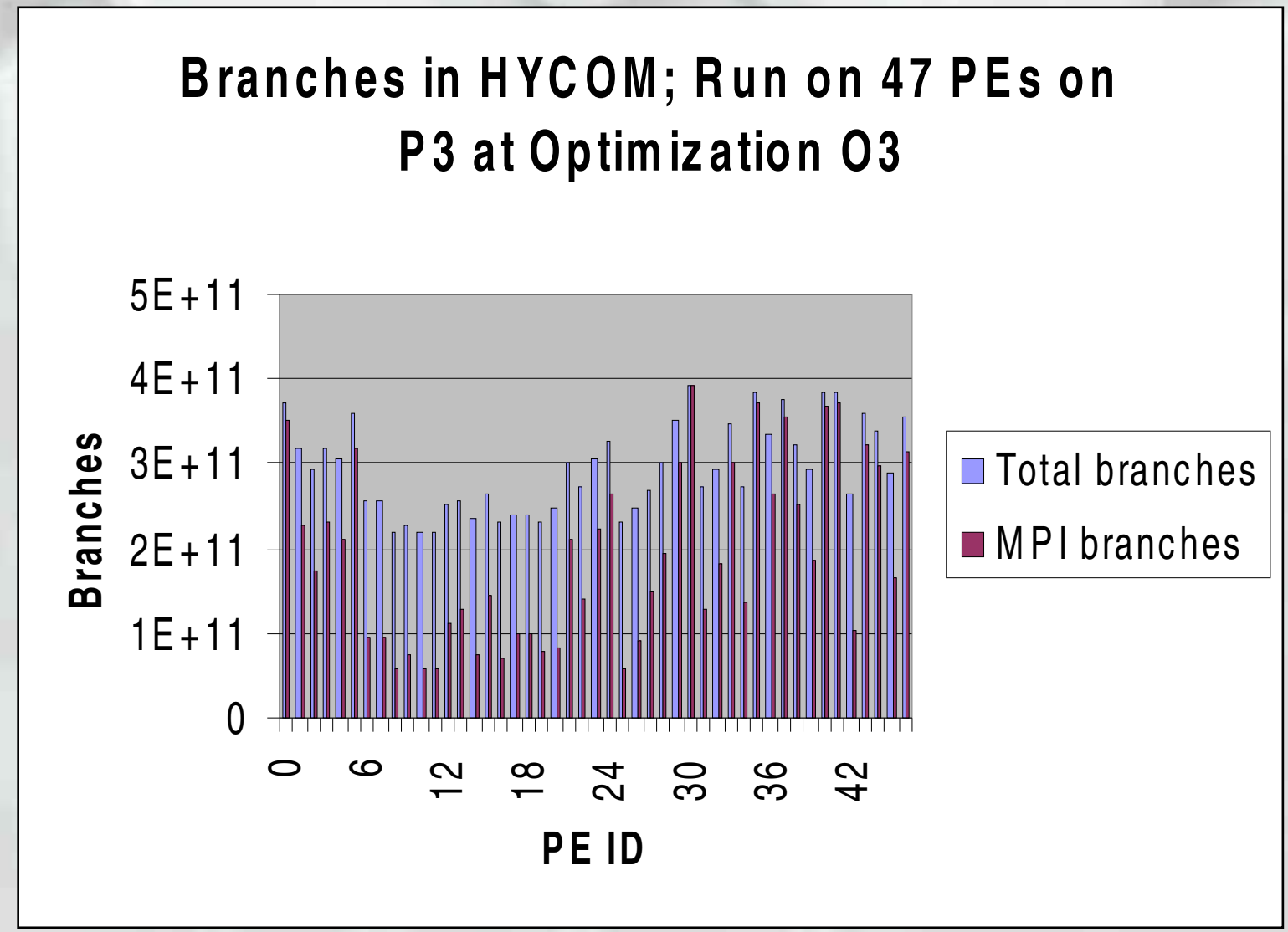

Cycles in OOCORE; RUn on 32 OEs on P3 at Optimization O3
3.0E+12
2.5E+12
2.0E+12
1.5E+12
1.0E+12
5.0E+11
0.0E+00
Cycles
0
4
8
12
16
20
24
28
PE ID
Total cycles
MPI cycles

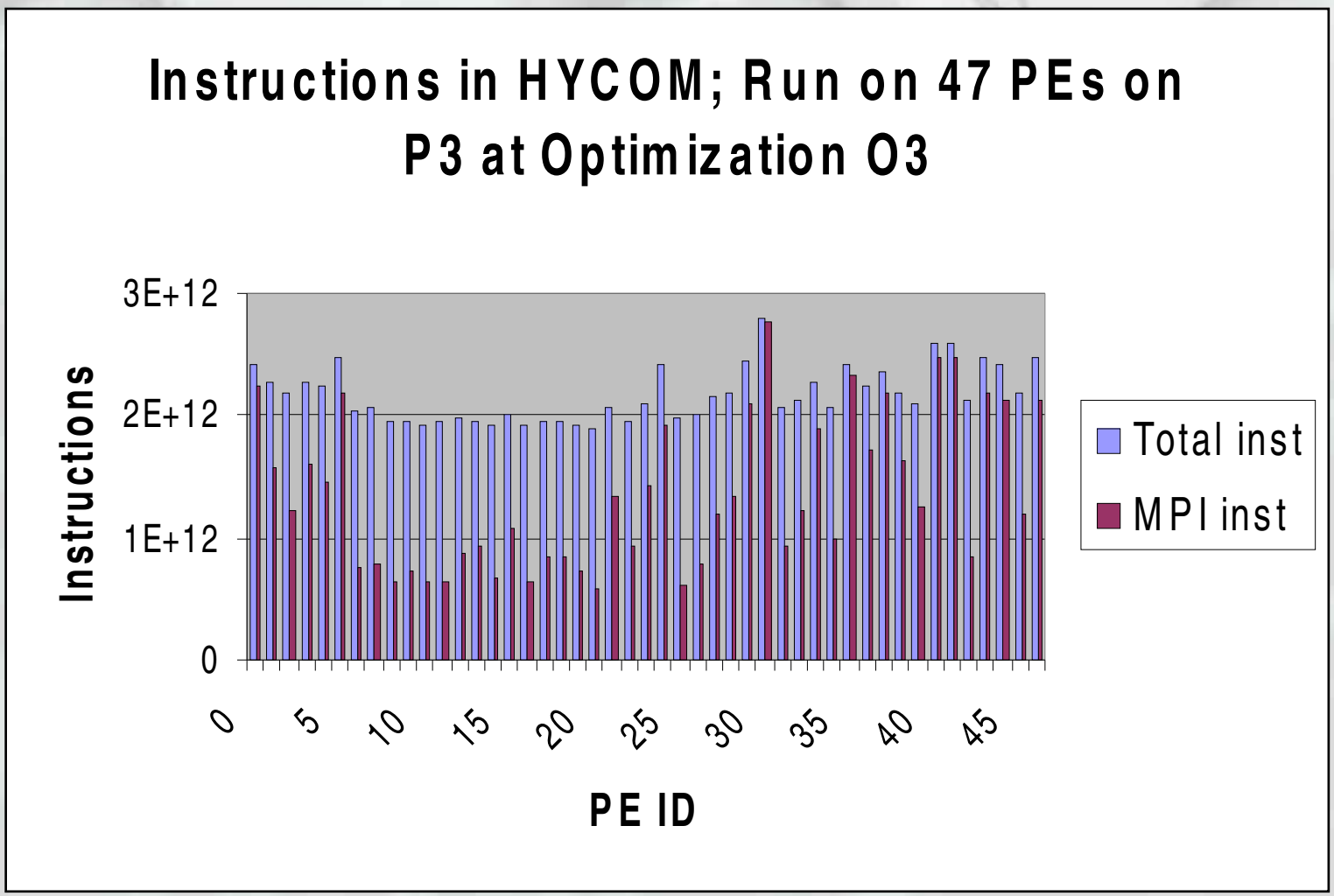

Branches in OOCORE; Run on 32PEs on P3 at Optimization O3
Branches
4E+11
3E+11
2E+11
1E+11
0
0
4
8
12
16
20
24
28
PE ID
Total branches
MPI branches

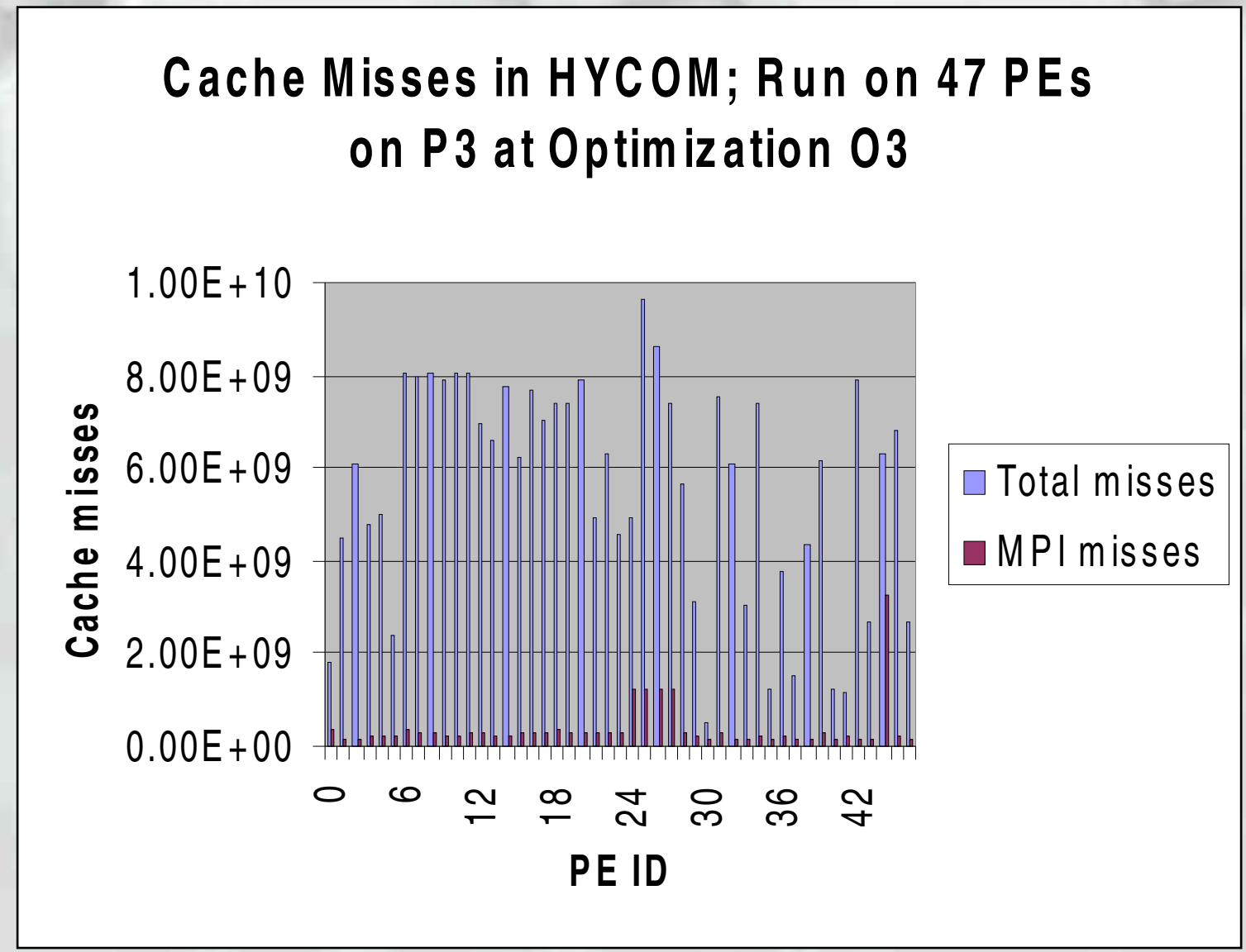

Instructions in OOCORE; Run on 32 PEs on P3 at Optimization O3
Instructions
6E+12
5E+12
4E+12
3E+12
2E+12
1E+12
0
0
4
8
12
16
20
24
28
PE ID
Total instructions
MPI Instructions

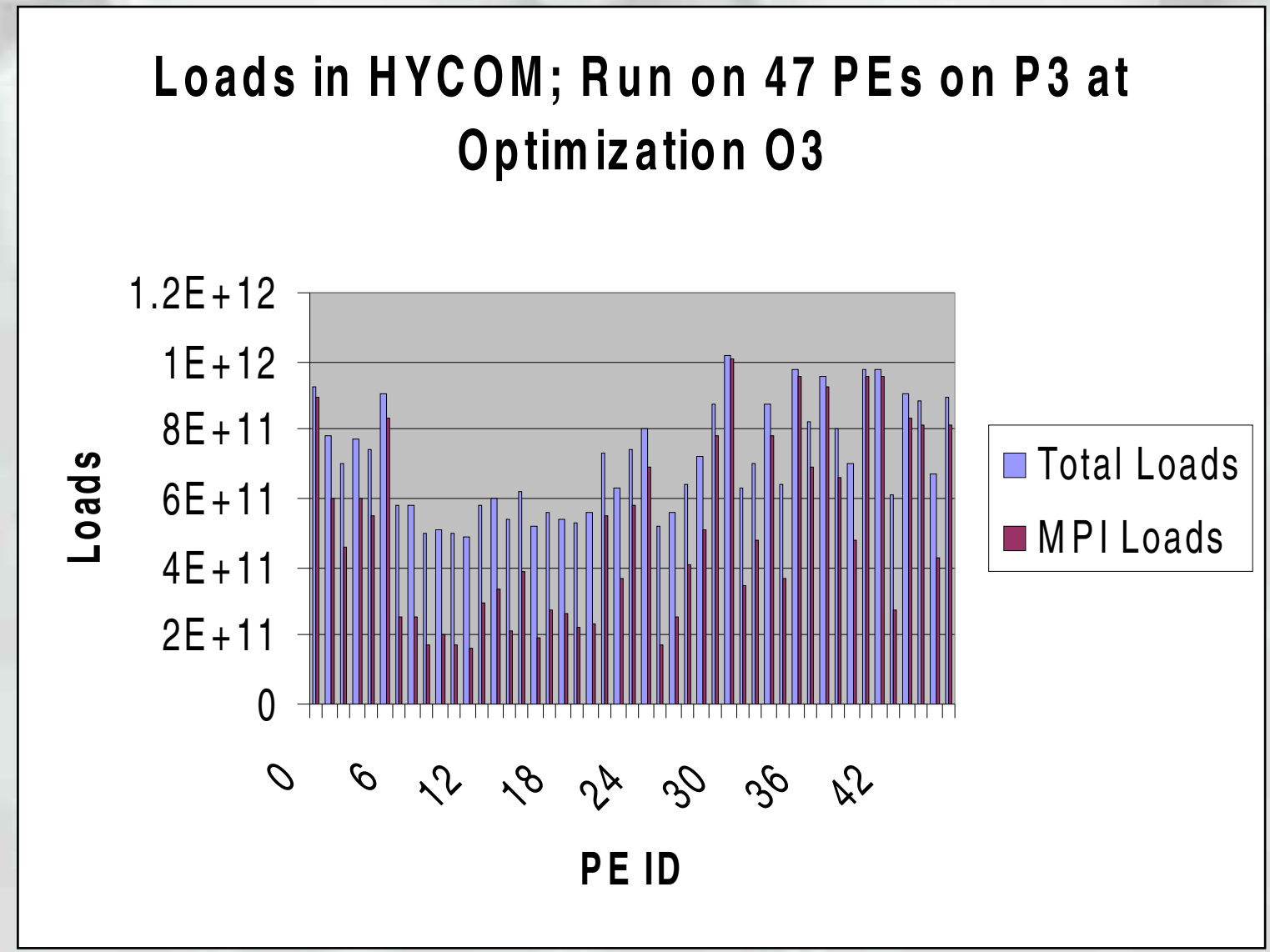

Cache Misses in OOCORE; RUn on 32 PEs on P3 at Optimization O3
1.00E+11
8.00E+10
6.00E+10
4.00E+10
2.00E+10
0.00E+00
Misses
1
6
11
16
21
26
31
PE ID
Total Cache Misses


Cycles in HYCOM; Run on 47 PEs on P3 at Optimization O3
4.E+12
3.E+12
2.E+12
1.E+12
0.E+00
Cycles
0
5
10
15
20
25
30
35
40
45
PE ID
Total cycles
MPI cycles

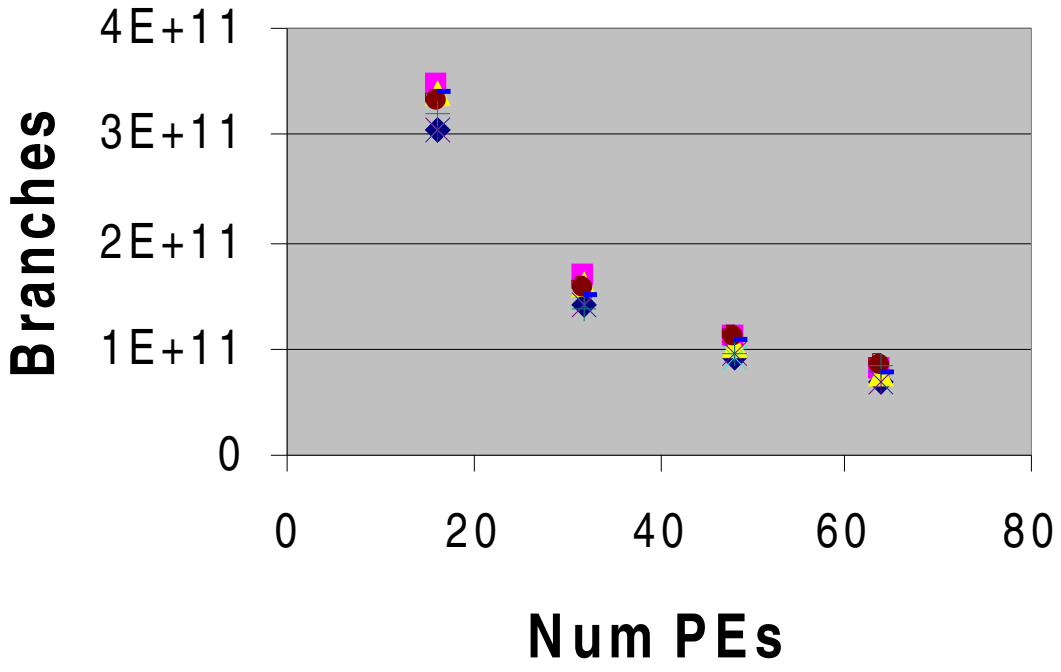

Branches in HYCOM; Run on 47 PEs on P3 at Optimization O3
5E+11
4E+11
3E+11
2E+11
1E+11
0
Branches
0
6
12
18
24
30
36
42
PE ID
Total branches
MPI branches

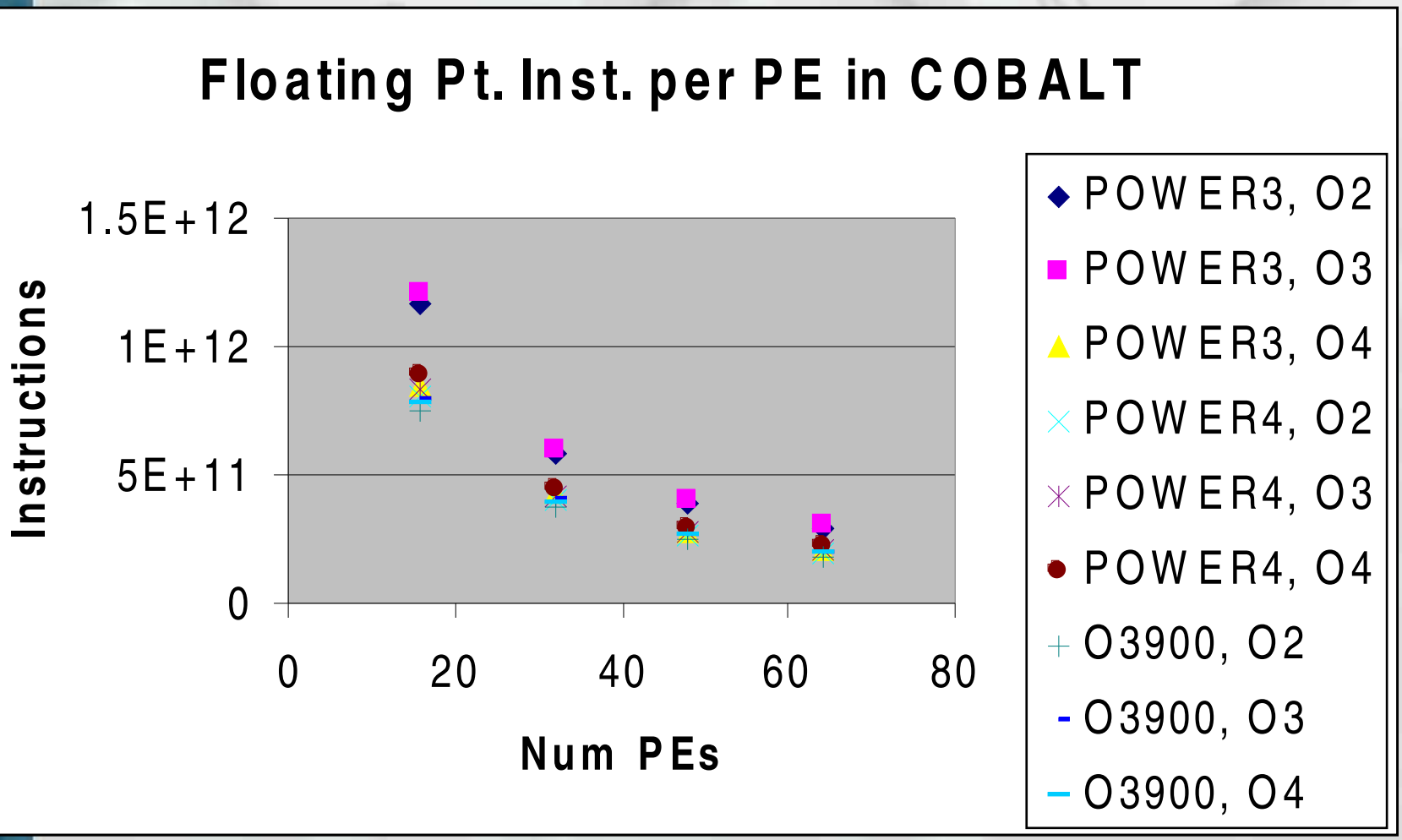

Instructions in HYCOM; Run on 47 PEs on P3 at Optimization O3
3E+12
2E+12
1E+12
0
Instructions
0
5
10
15
20
25
30
35
40
45
PE ID
Total inst
MPI inst

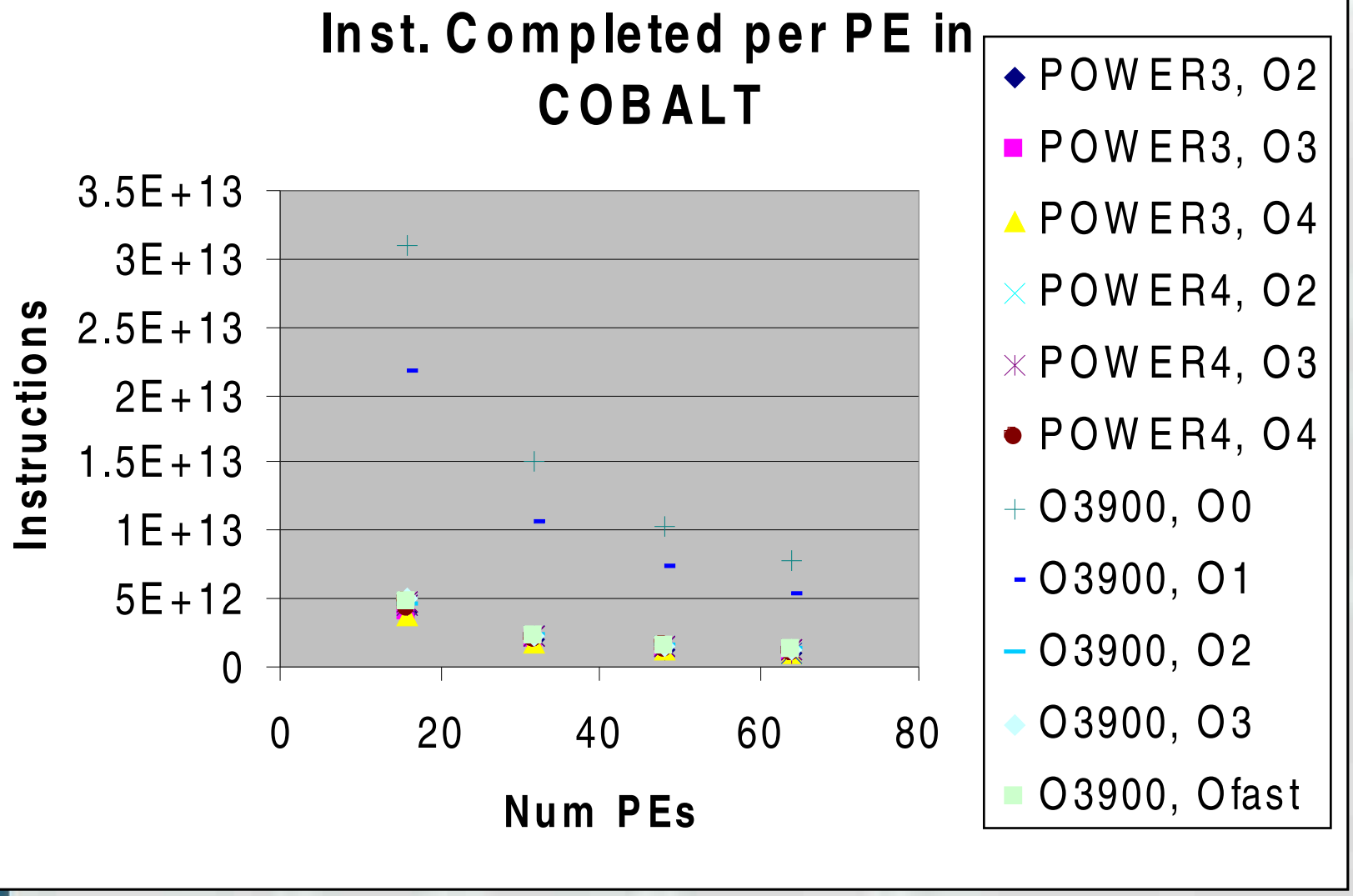

Cache Misses in HYCOM; Run on 47 PEs on P3 at Optimization O3
Cache misses
1.00E+10
8.00E+09
6.00E+09
4.00E+09
2.00E+09
0.00E+00
0
6
12
18
24
30
36
42
PE ID
Total misses
MPI misses

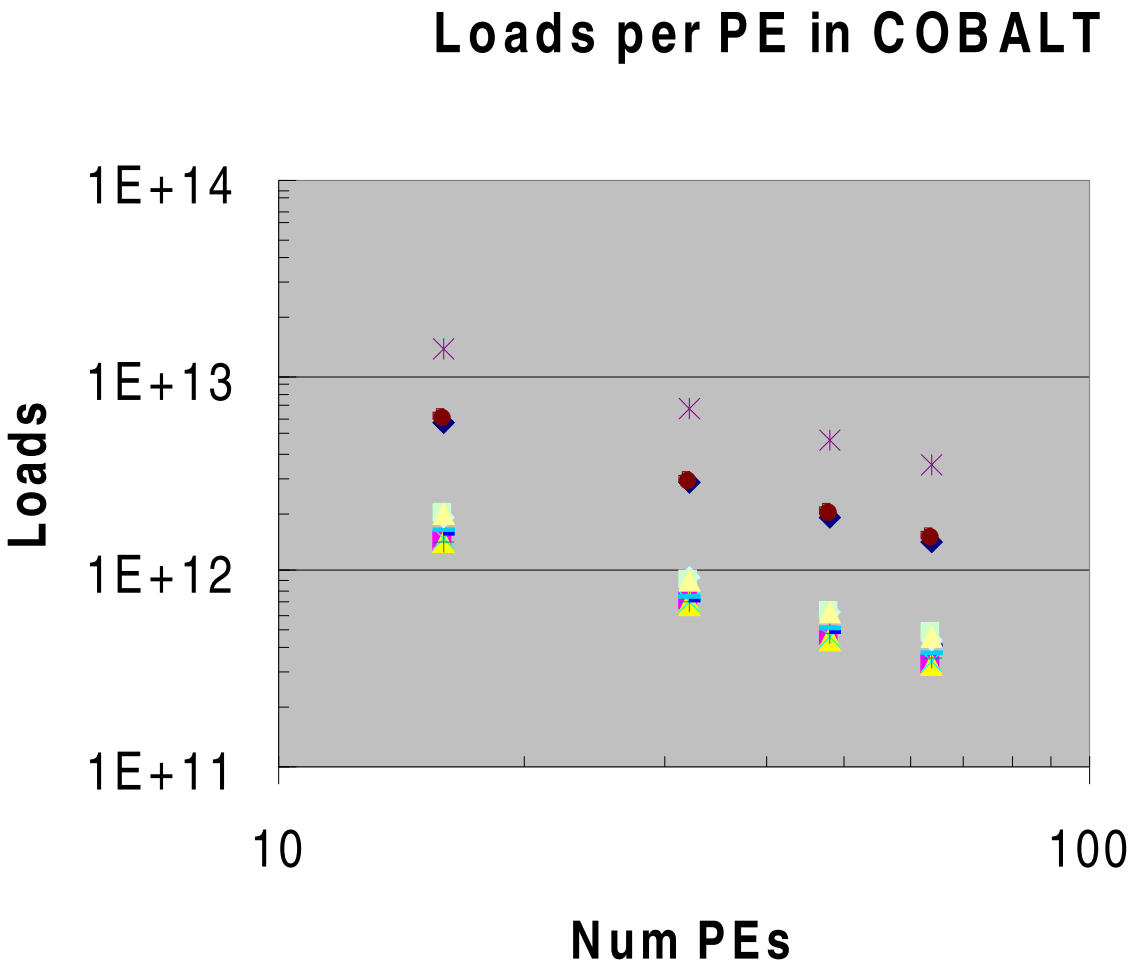

Loads in HYCOM; Run on 47 PEs on P3 at Optimization O3
1.2E+12
1E+12
8E+11
6E+11
4E+11
2E+11
0
Loads
0
6
12
18
24
30
36
42
PE ID
Total Loads
MPI Loads

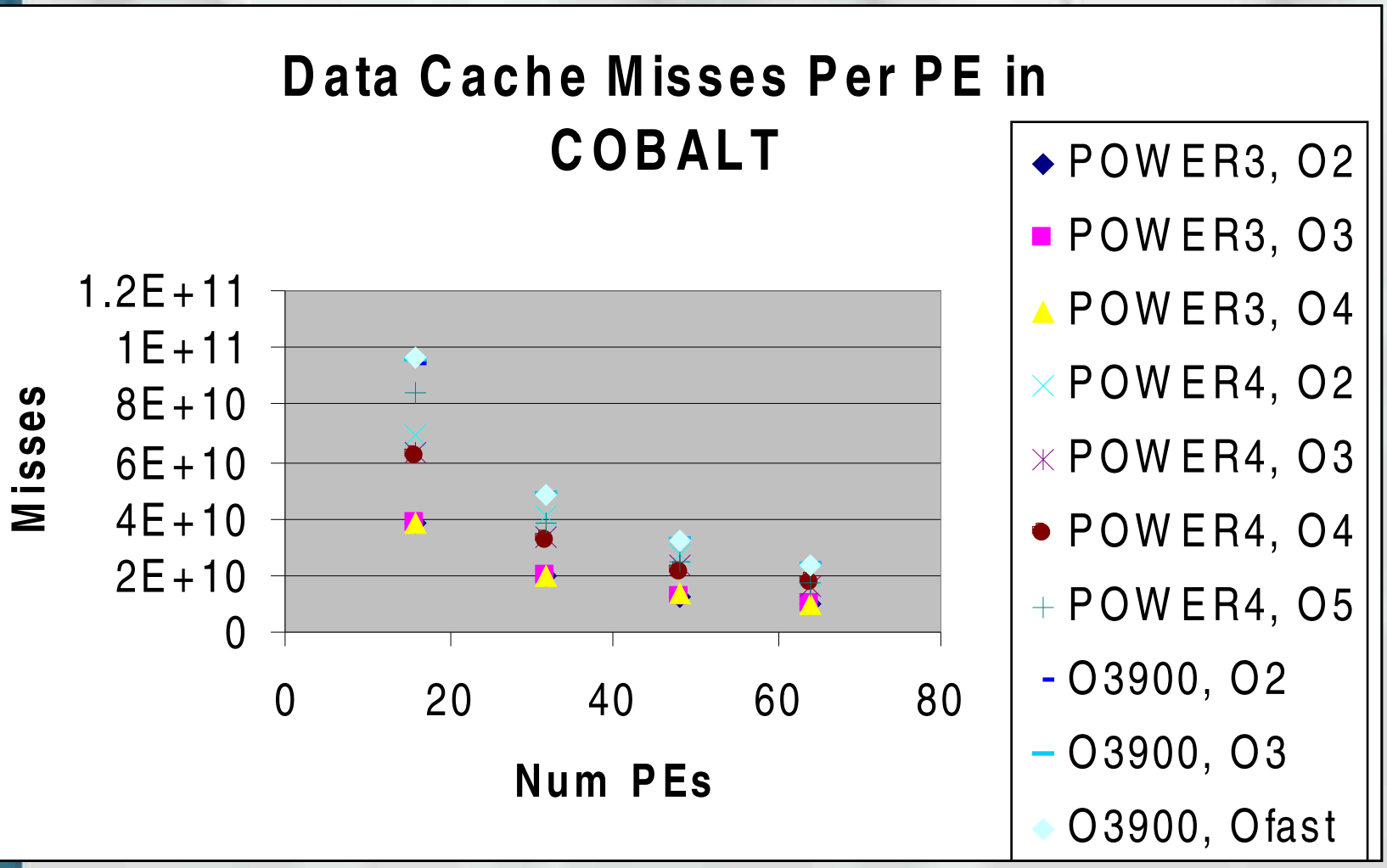
Extras

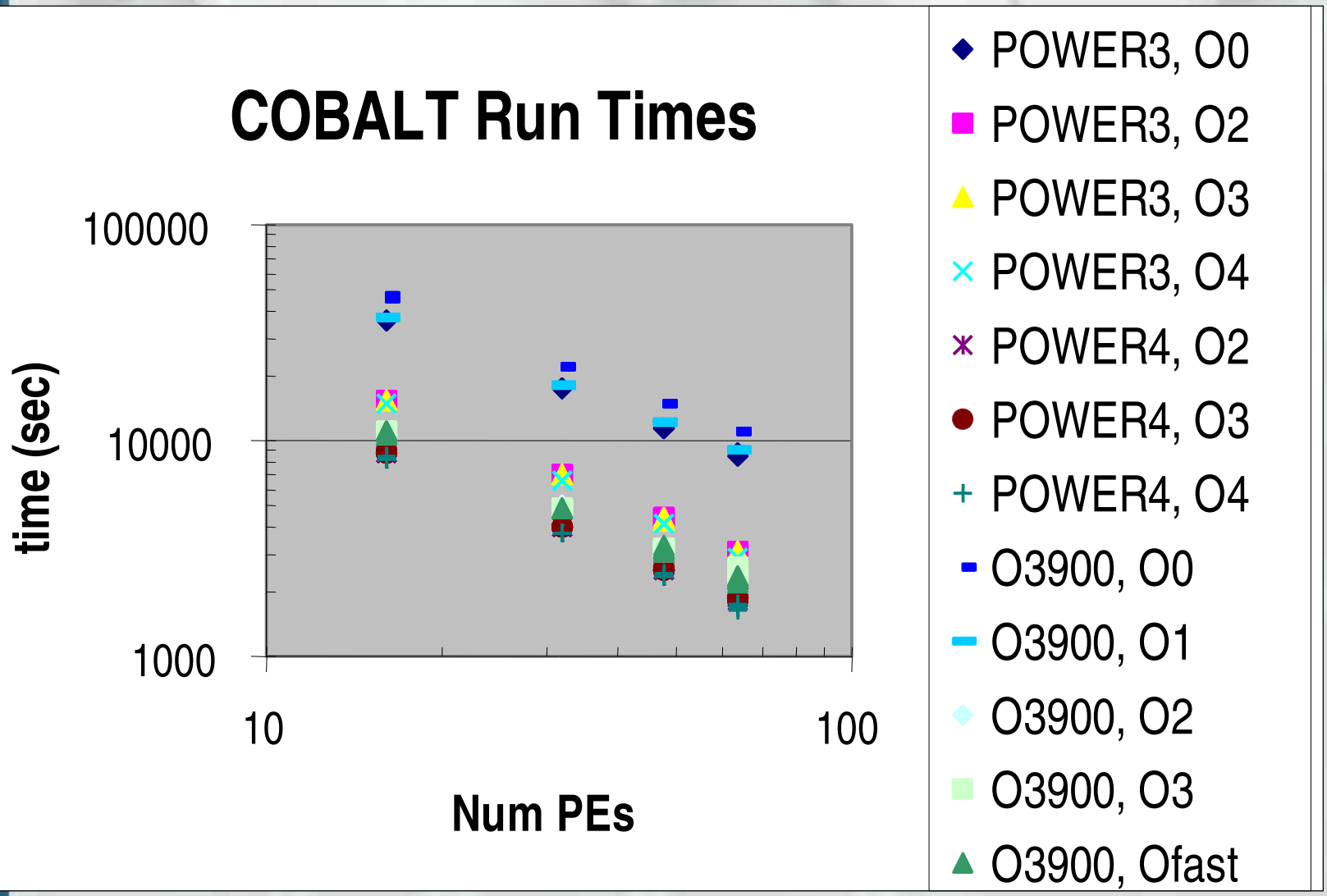

Branches per PE in COBALT
Branches
4E+11
3E+11
2E+11
1E+11
0
0
20
40
60
80
Num PEs
POWER3, O0
POWER3, O2
POWER3, O3
POWER3, O4
POWER3, O5
O3900, O2
O3900, O3
O3900, Ofast

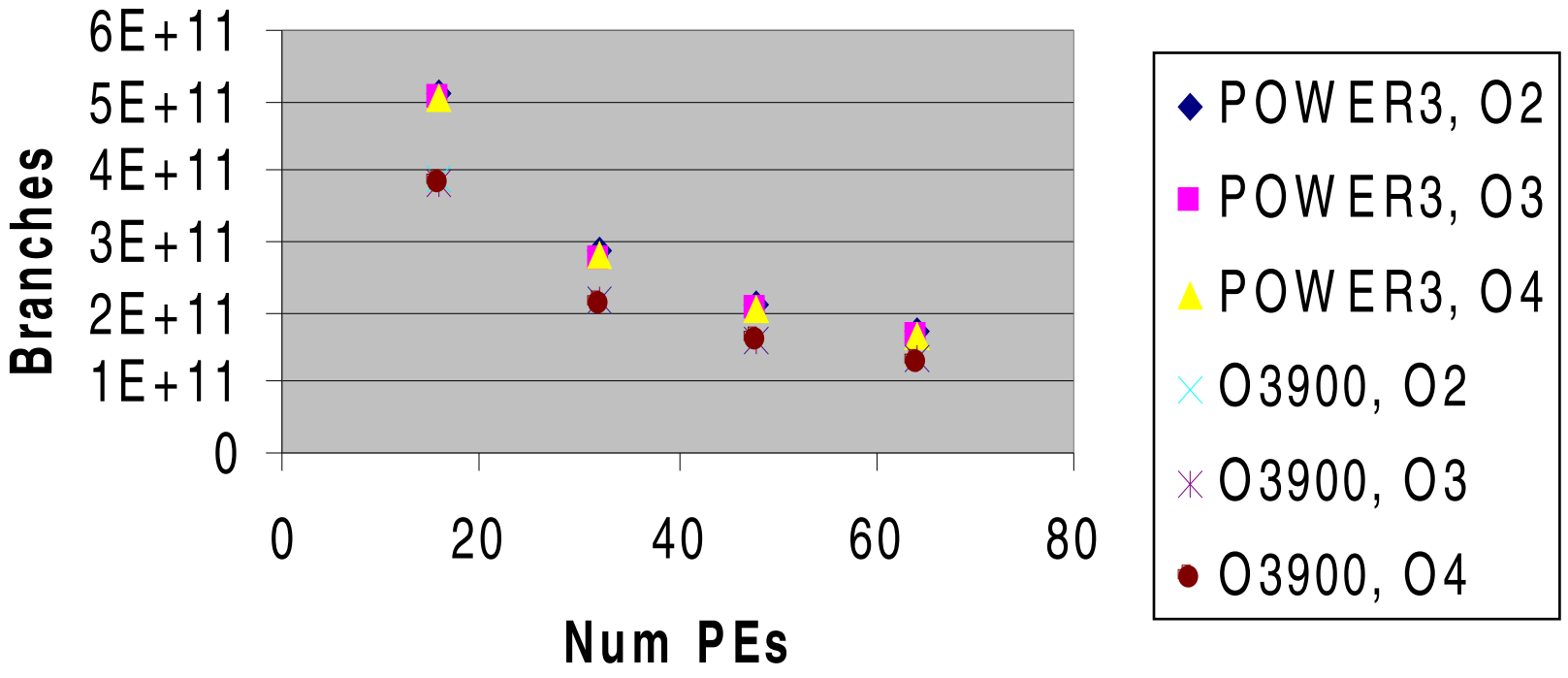

Floating Pt. Inst. per PE in COBALT
Instructions
1.5E+12
1E+12
5E+11
0
0
20
40
60
80
Num PEs
POWER3, O2
POWER3, O3
POWER3, O4
POWER4, O2
POWER4, O3
POWER4, O4
O3900, O2
O3900, O3
O3900, O4

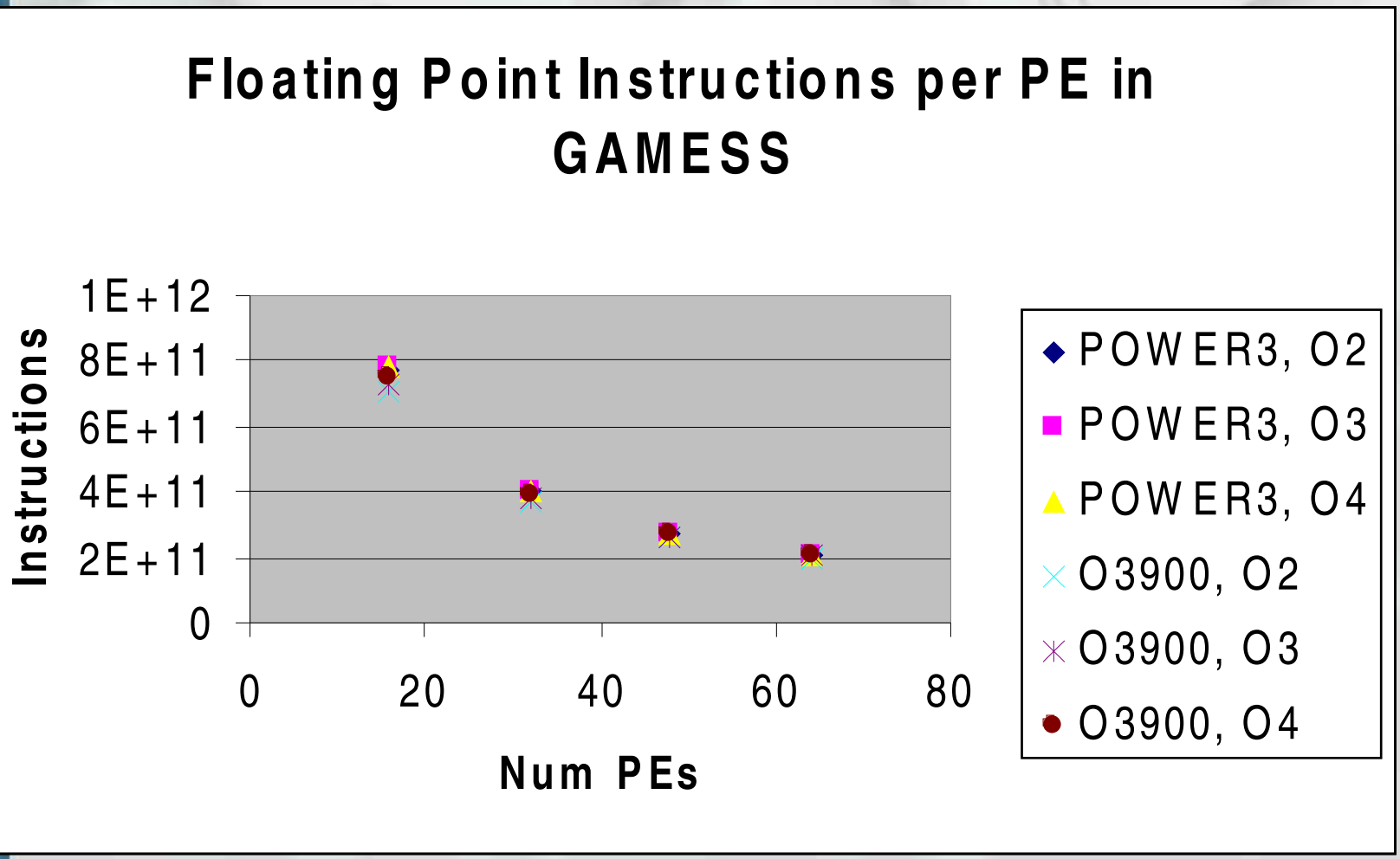

Inst. Completed per PE in COBALT
Instructions
3.5E+13
3E+13
2.5E+13
2E+13
1.5E+13
1E+13
5E+12
0
0
20
40
60
80
Num PEs
POWER3, O2
POWER3, O3
POWER3, O4
POWER4, O2
POWER4, O3
POWER4, O4
O3900, O0
O3900, O1
O3900, O2
O3900, O3
O3900, Ofast

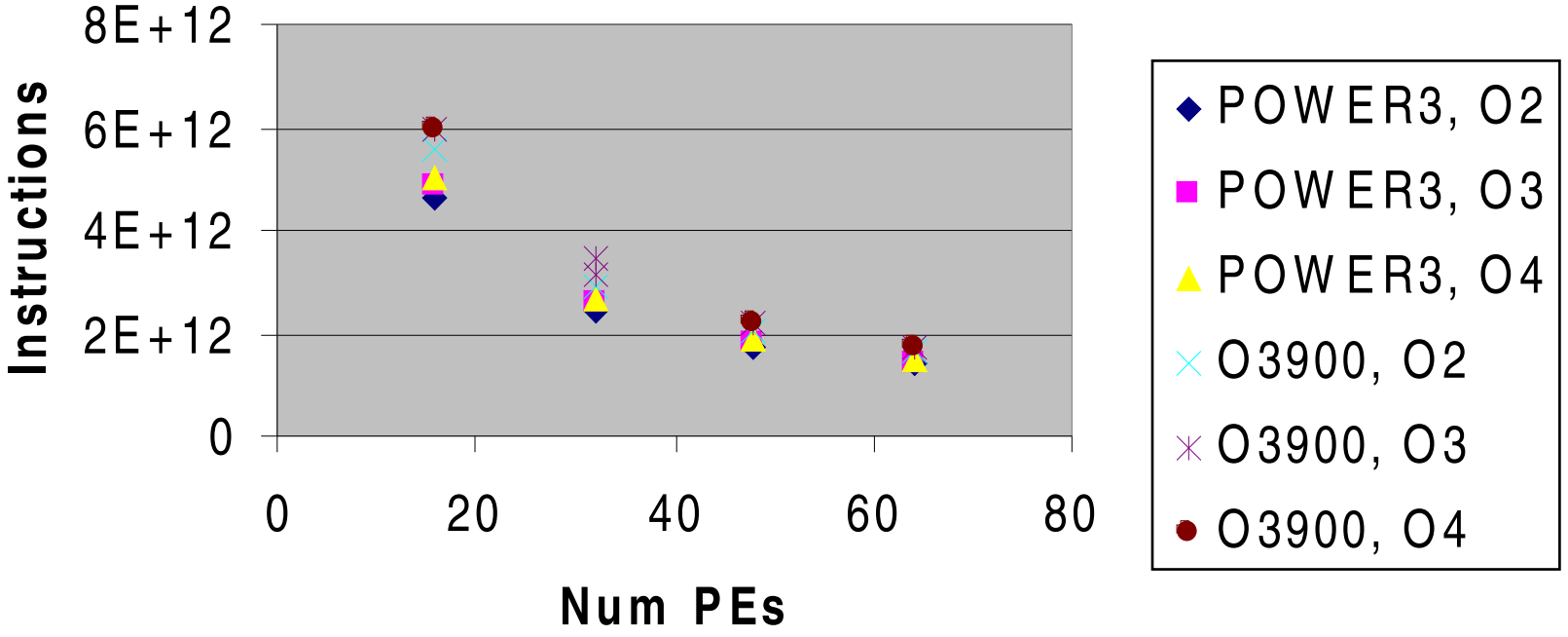

Loads per PE in COBALT
1E+14
1E+13
1E+12
1E+11
Loads
10
100
Num PEs
POWER3, O0
POWER3, O2
POWER3, O3
POWER3, O4
O3900, O0
O3900, O1
O3900, O2
O3900, O3
O3900, Ofast
POWER4, O2
POWER4, O3
POWER4, O4

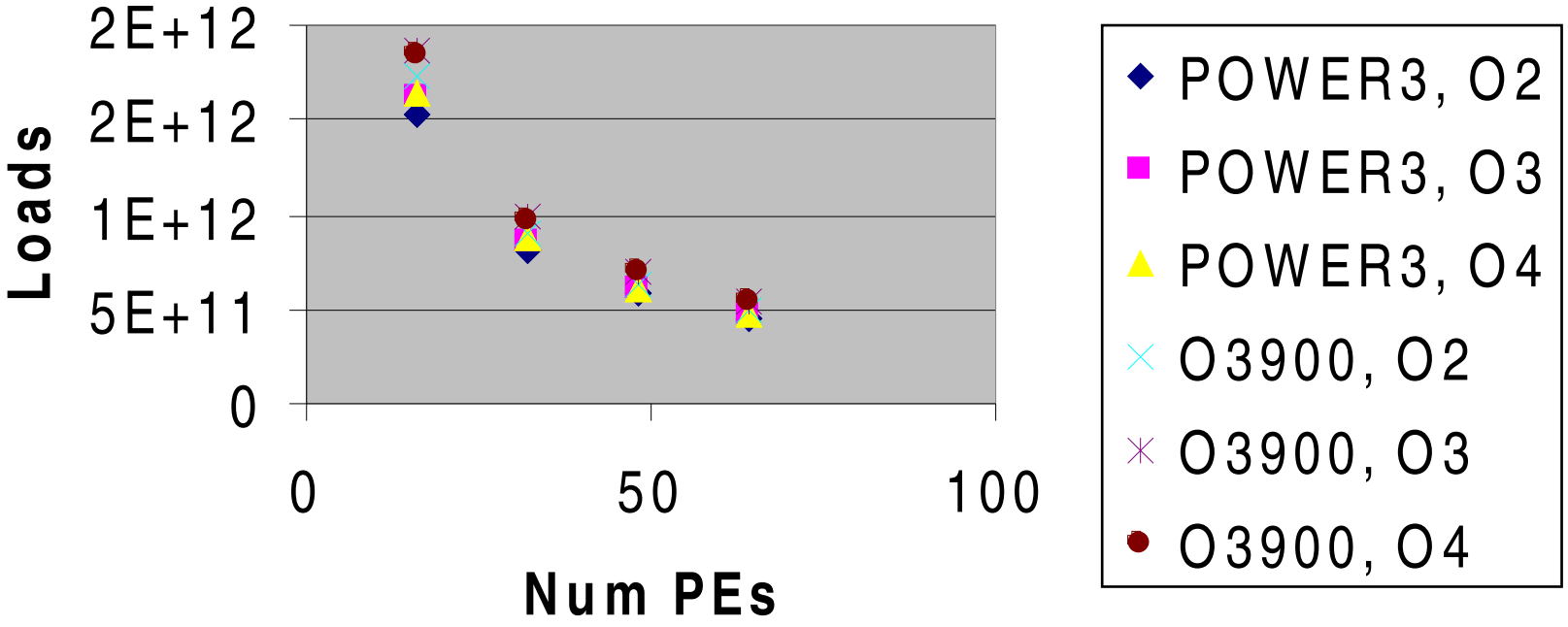

Data Cache Misses Per PE in COBALT
Misses
1.2E+11
1E+11
8E+10
6E+10
4E+10
2E+10
0
0
20
40
60
80
Num PEs
POWER3, O2
POWER3, O3
POWER3, O4
POWER4, O2
POWER4, O3
POWER4, O4
POWER4, O5
O3900, O2
O3900, O3
O3900, Ofast


COBALT Run Times
100000
10000
1000
time (sec)
10
100
Num PEs
POWER3, O0
POWER3, O2
POWER3, O3
POWER3, O4
POWER4, O2
POWER4, O3
POWER4, O4
O3900, O0
O3900, O1
O3900, O2
O3900, O3
O3900, Ofast

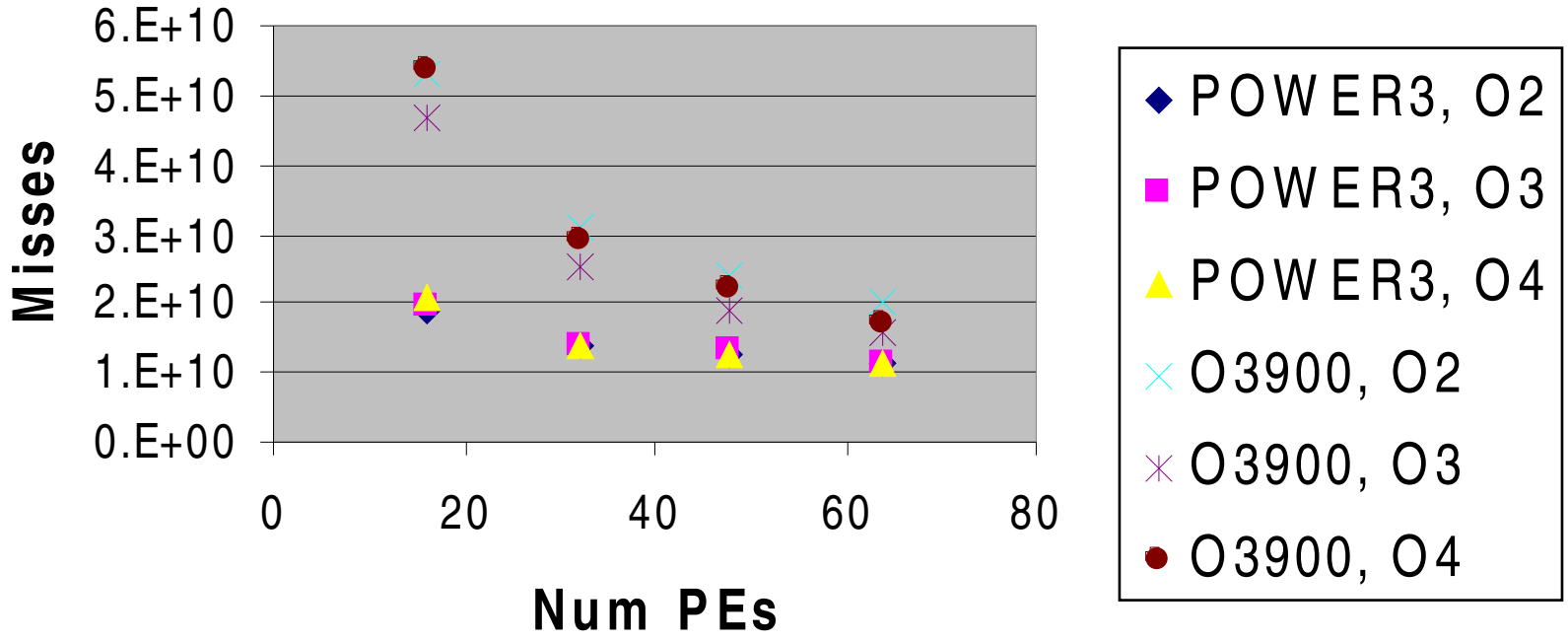

Branches per PE in GAMESS
Branches
6E+11
5E+11
4E+11
3E+11
2E+11
1E+11
0
0
20
40
60
80
Num PEs
POWER3, O2
POWER3, O3
POWER3, O4
O3900, O2
O3900, O3
O3900, O4

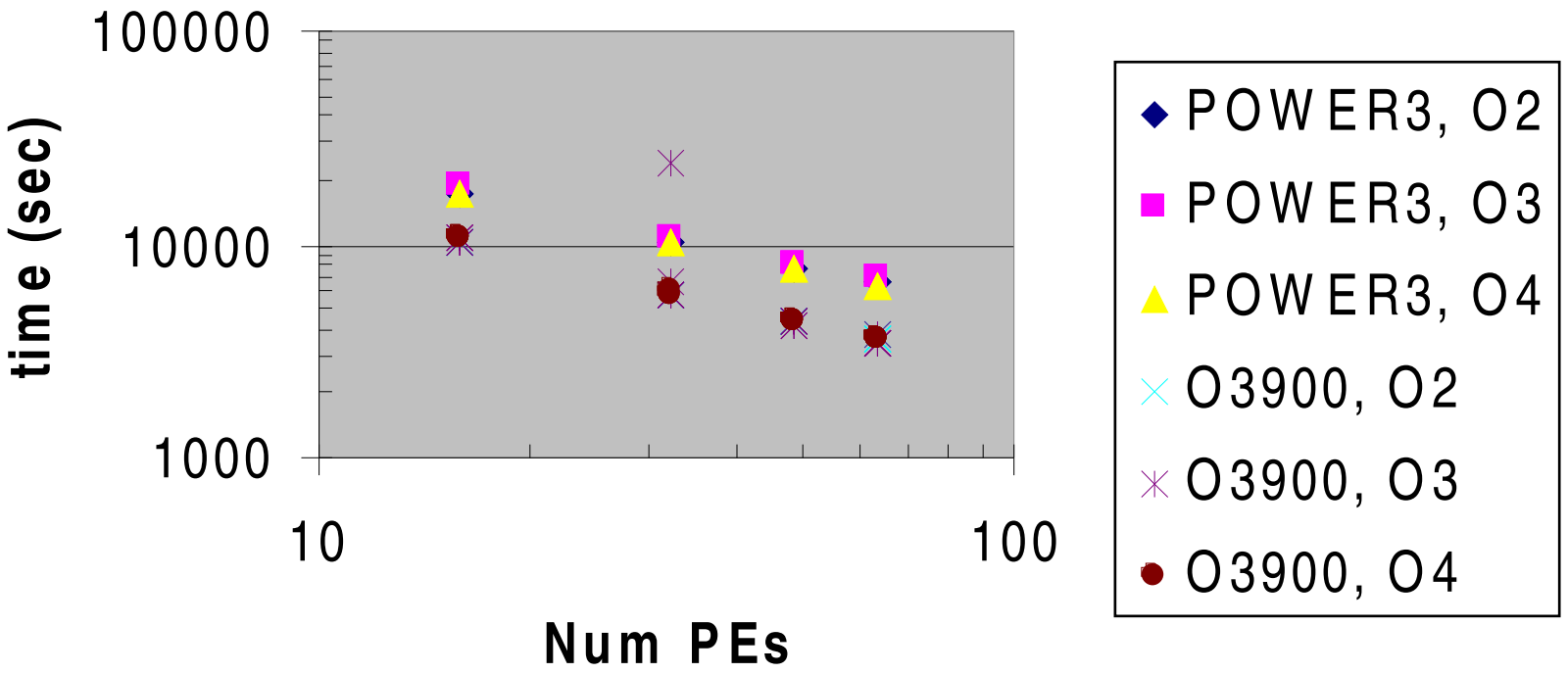

Floating Point Instructions per PE in GAMESS
Instructions
1E+12
8E+11
6E+11
4E+11
2E+11
0
0
20
40
60
80
Num PEs
POWER3, O2
POWER3, O3
POWER3, O4
O3900, O2
O3900, O3
O3900, O4

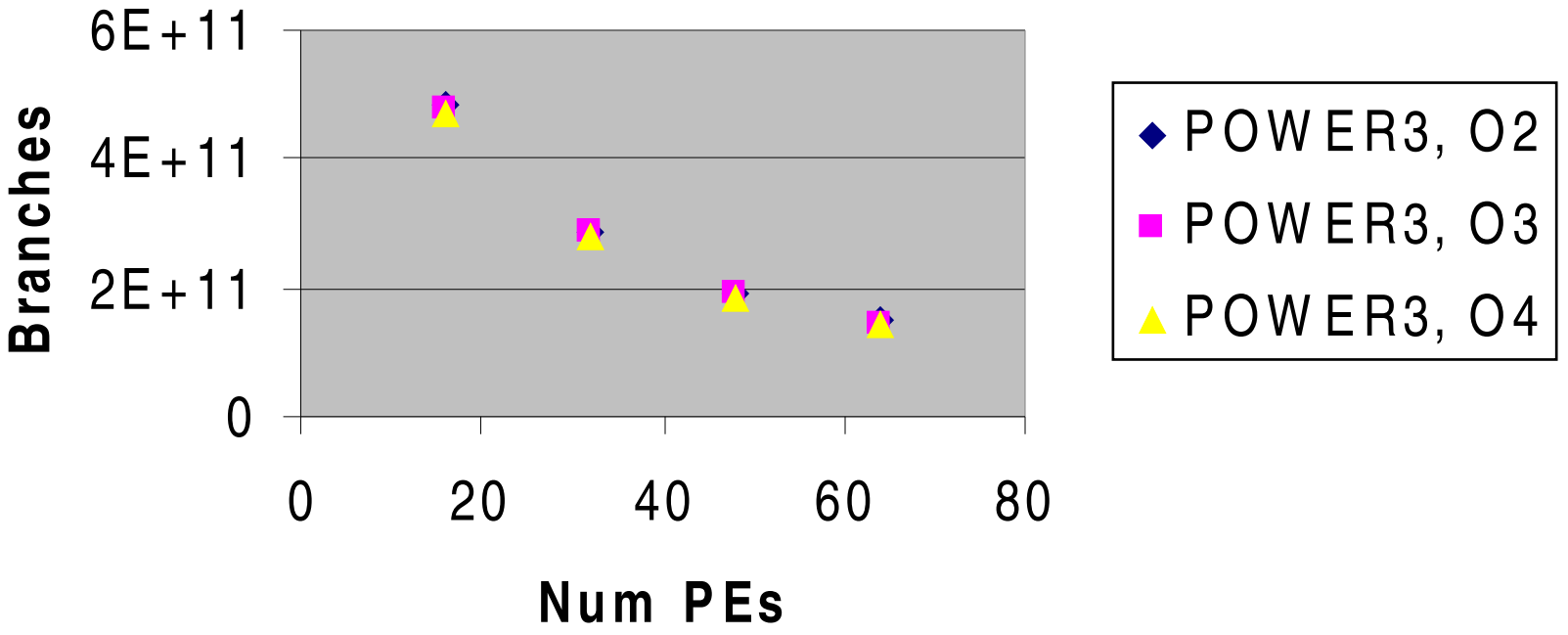

Instructions per PE in GAMESS
Instructions
8E+12
6E+12
4E+12
2E+12
0
0
20
40
60
80
Num PEs
POWER3, O2
POWER3, O3
POWER3, O4
O3900, O2
O3900, O3
O3900, O4


Loads per PE in GAMESS
Loads
2E+12
2E+12
1E+12
5E+11
0
0
50
100
Num PEs
POWER3, O2
POWER3, O3
POWER3, O4
O3900, O2
O3900, O3
O3900, O4

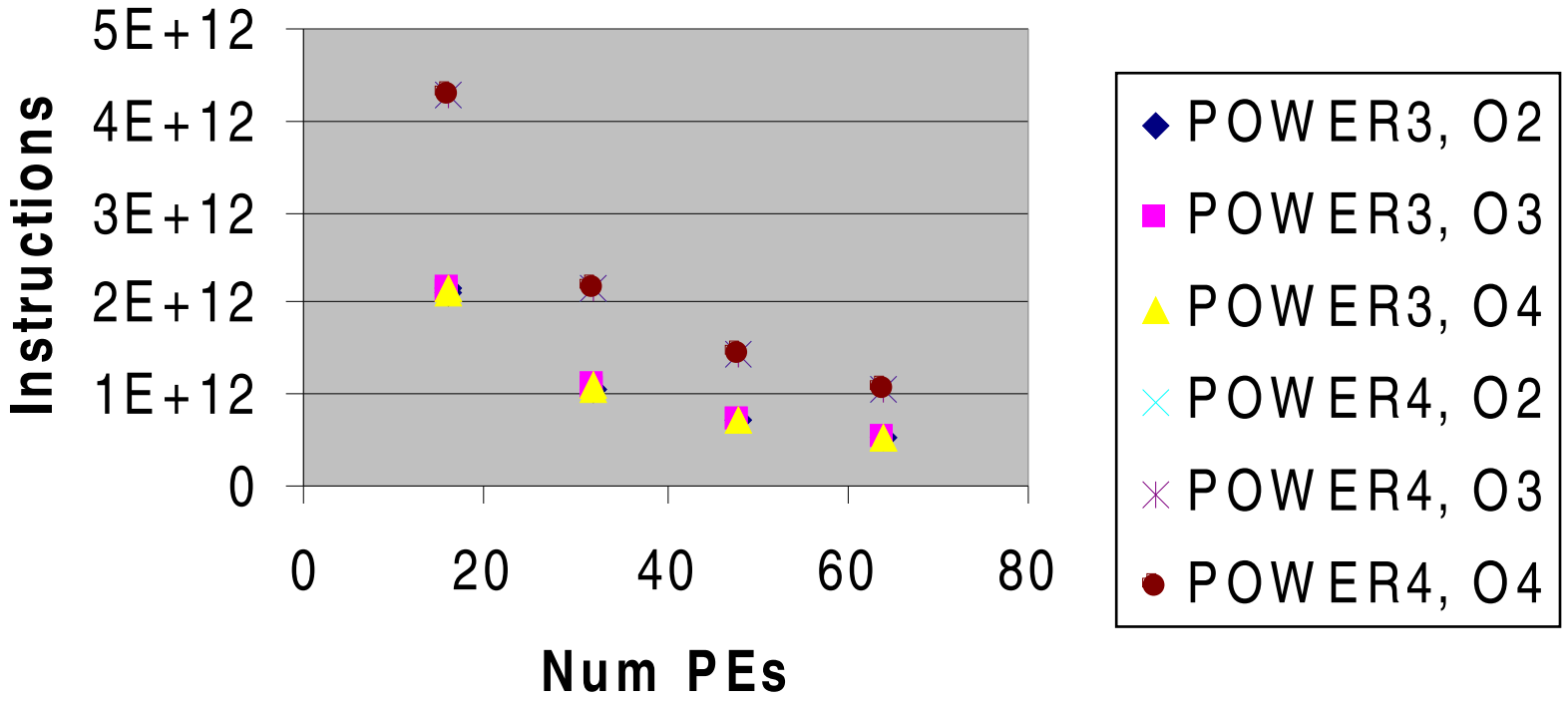
Data Cache Misses per PE in GAMESS

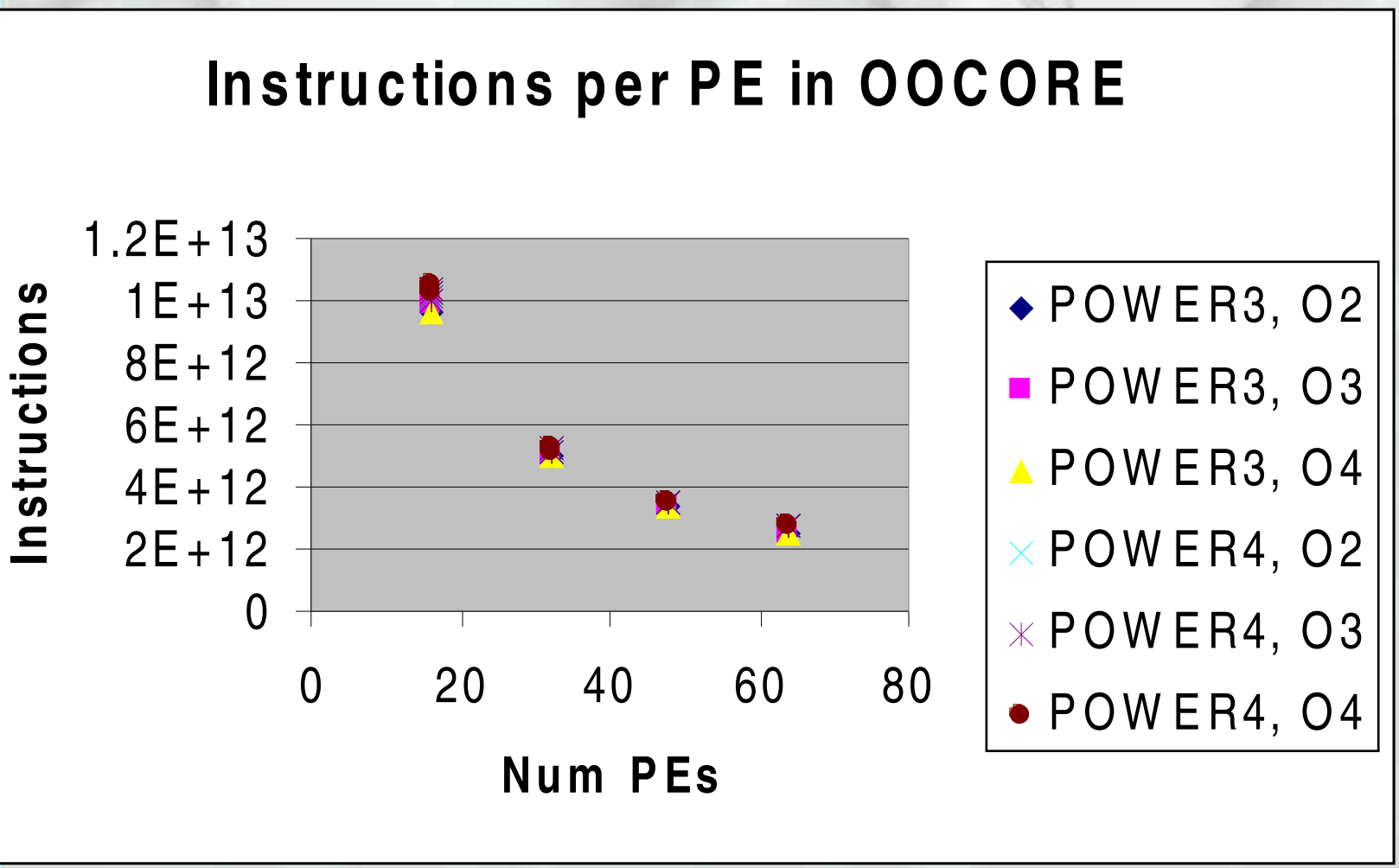


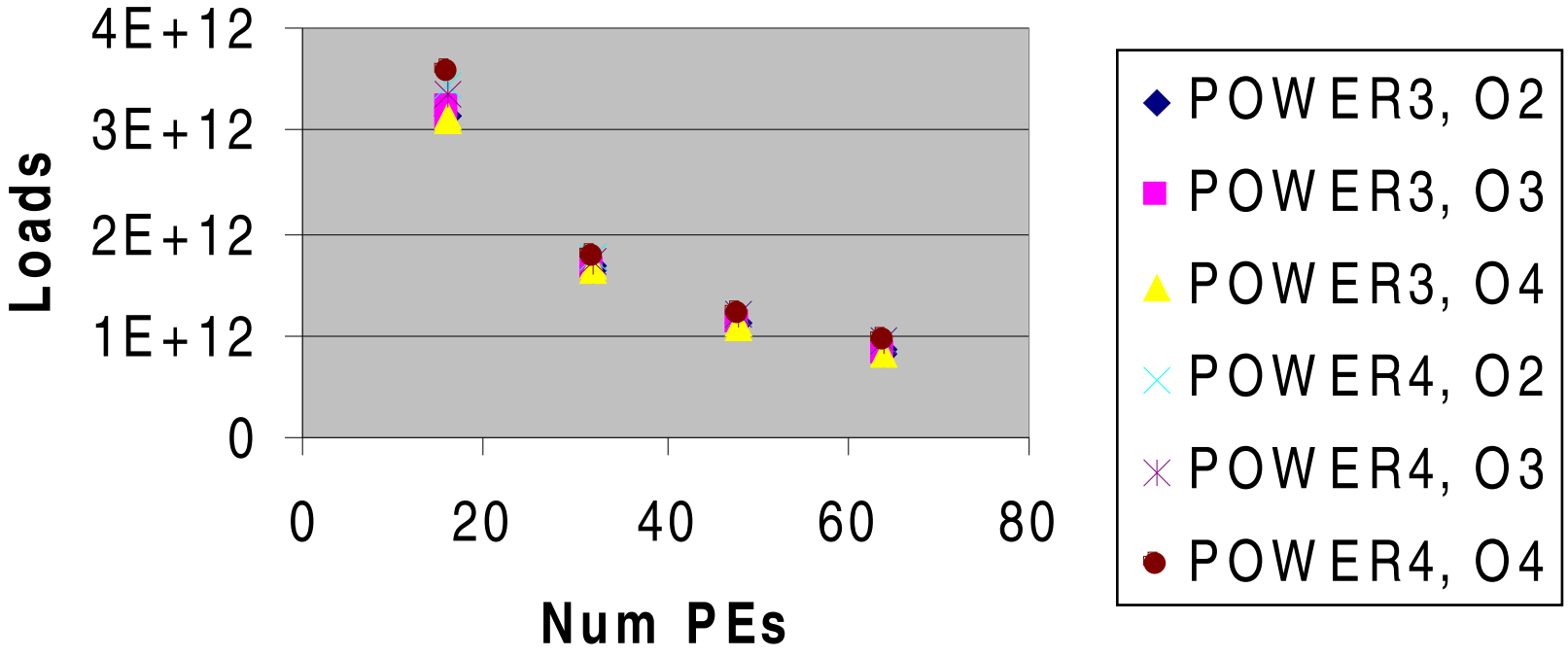

GAMESS Runtimes
100000
10000
1000
time (sec)
10
100
Num PEs
POWER3, O2
POWER3, O3
POWER3, O4
O3900, O2
O3900, O3
O3900, O4

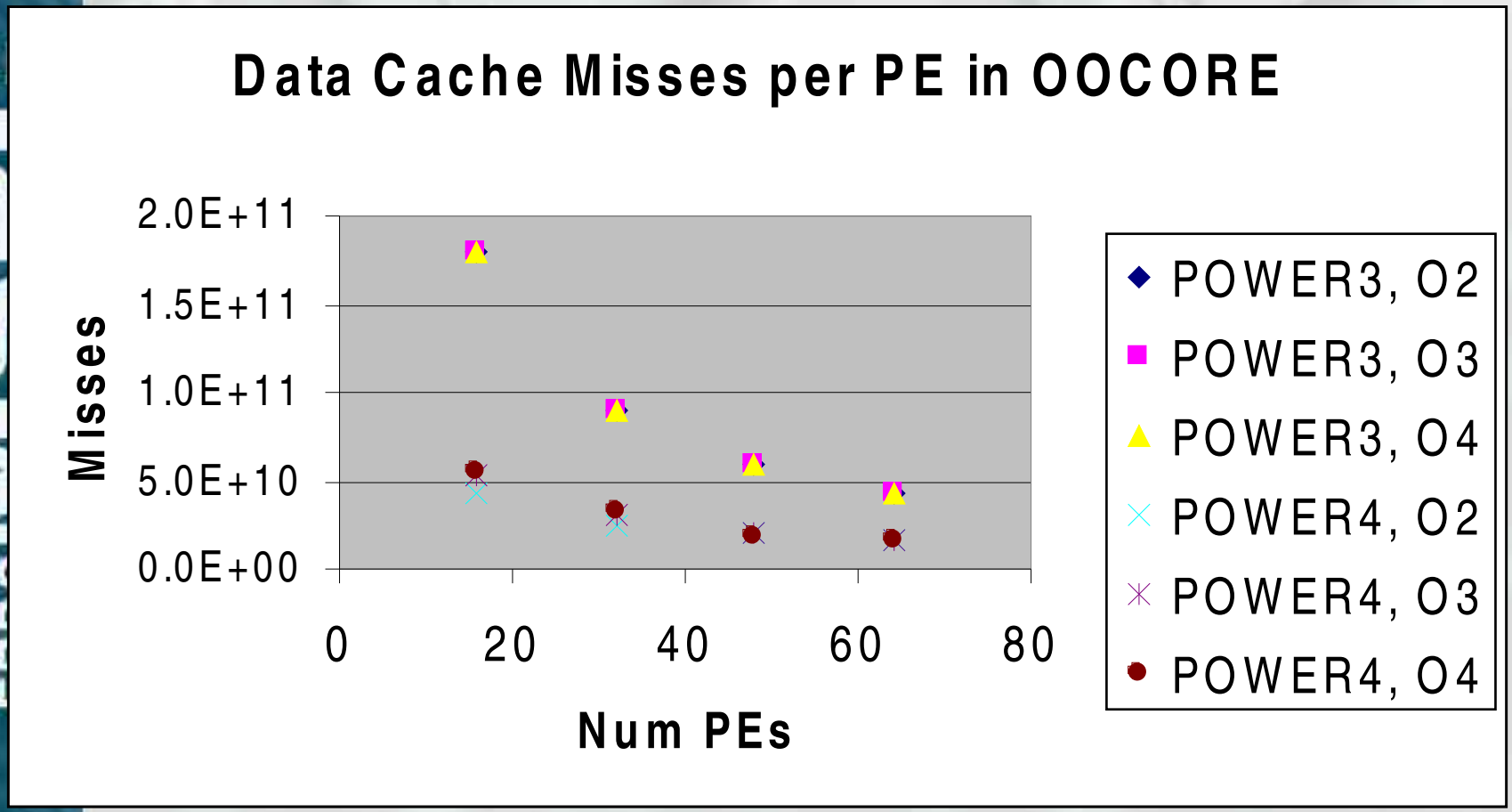

Branches per PE in OOCORE
Branches
6E+11
4E+11
2E+11
0
0
20
40
60
80
Num PEs
POWER3, O2
POWER3, O3
POWER3, O4

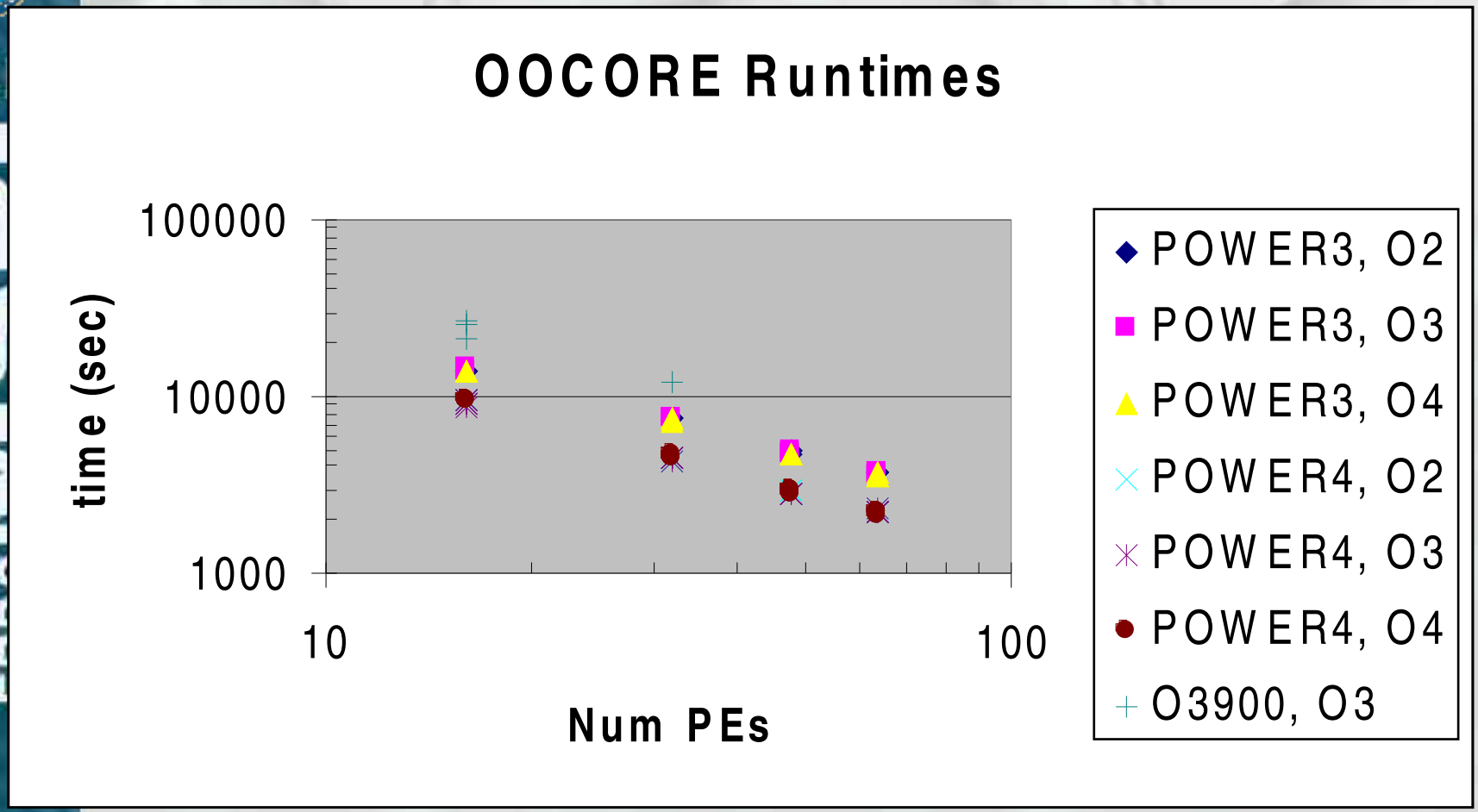
Floating Pt. Inst. per PE in OOCORE

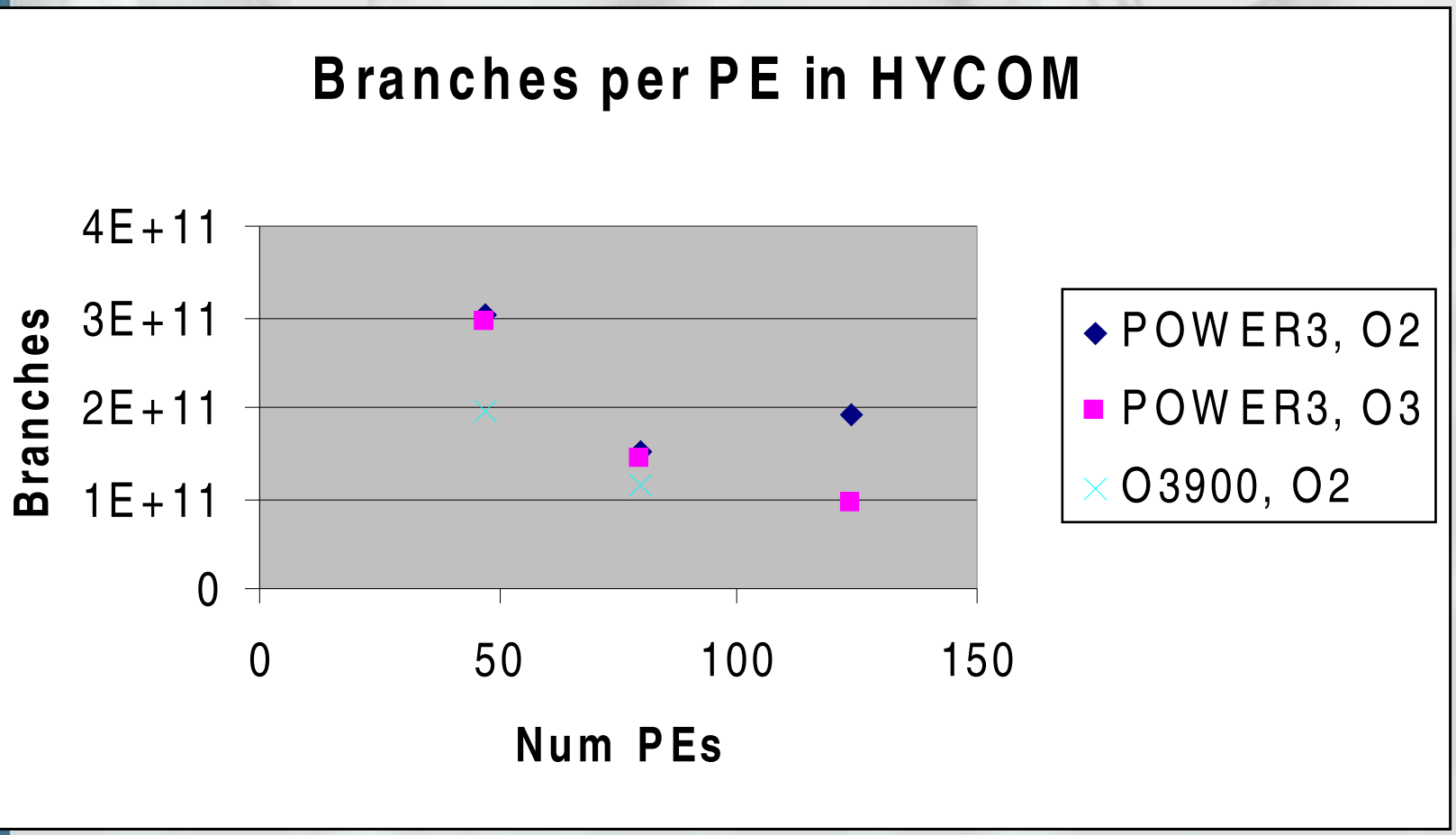


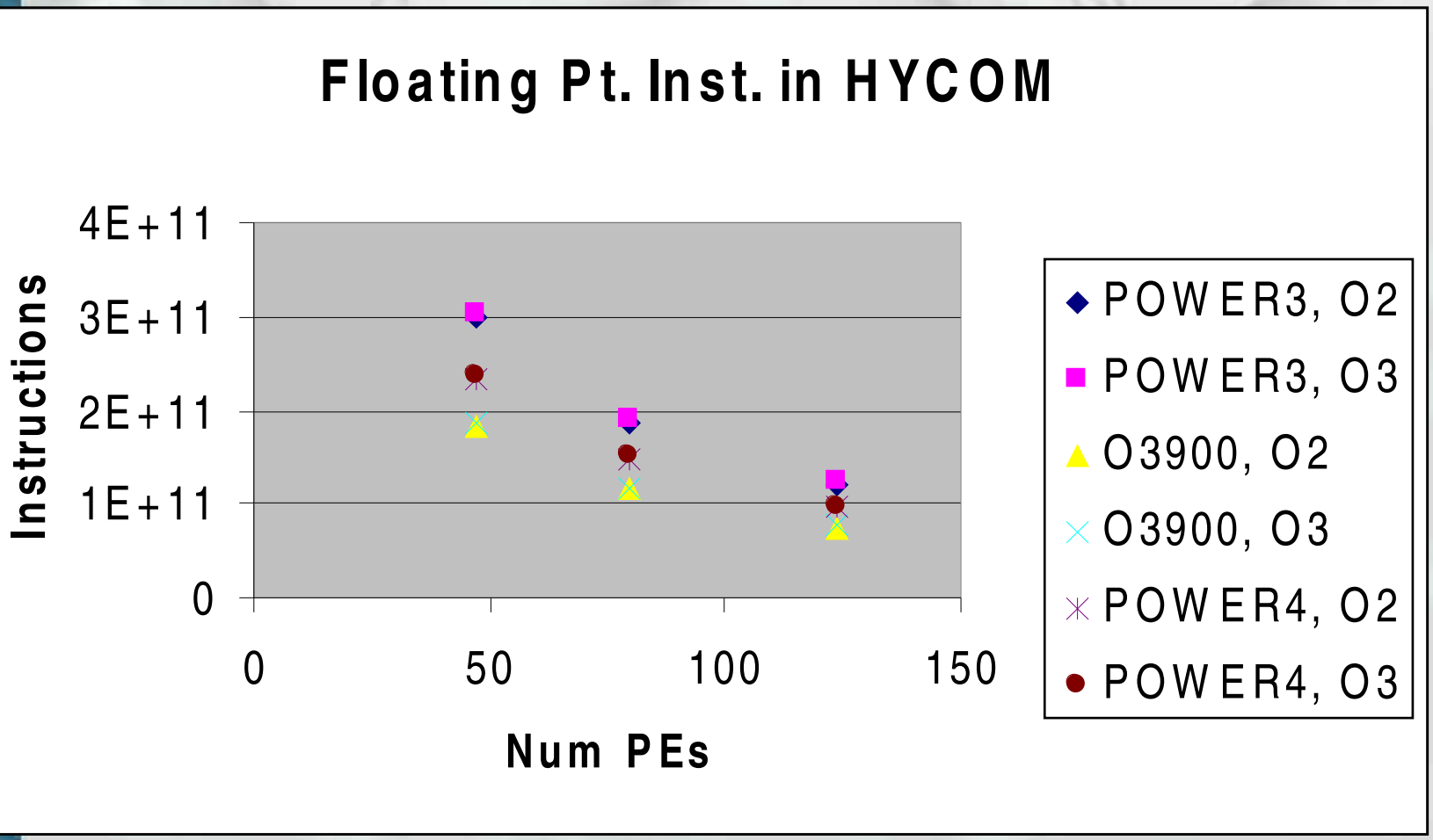

Instructions per PE in OOCORE
Instructions
1.2E+13
1E+13
8E+12
6E+12
4E+12
2E+12
0
0
20
40
60
80
Num PEs
POWER3, O2
POWER3, O3
POWER3, O4
POWER4, O2
POWER4, O3
POWER4, O4

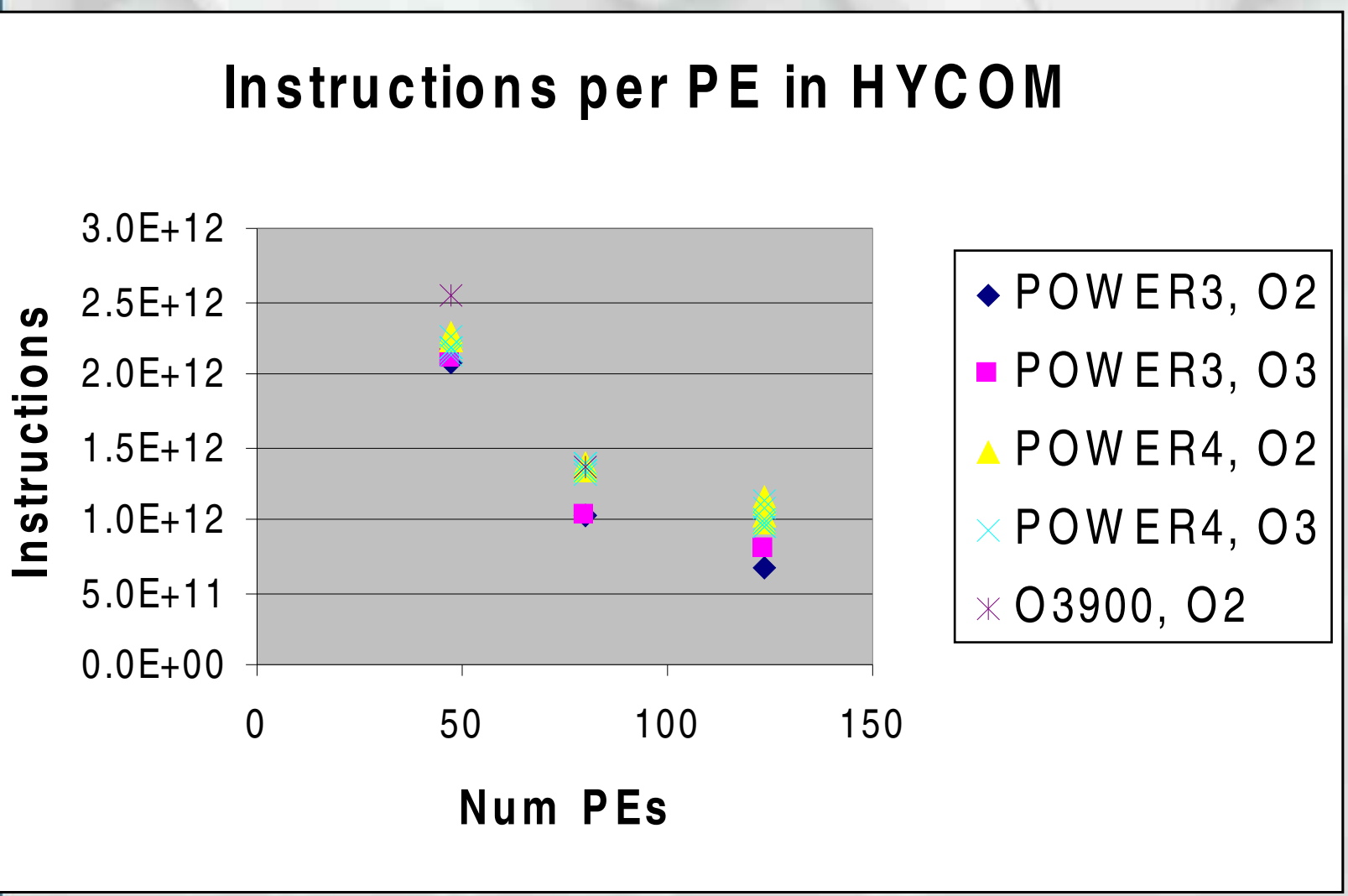

Loads per PE in OOCORE
Loads
4E+12
3E+12
2E+12
1E+12
0
0
20
40
60
80
Num PEs
POWER3, O2
POWER3, O3
POWER3, O4
POWER4, O2
POWER4, O3
POWER4, O4

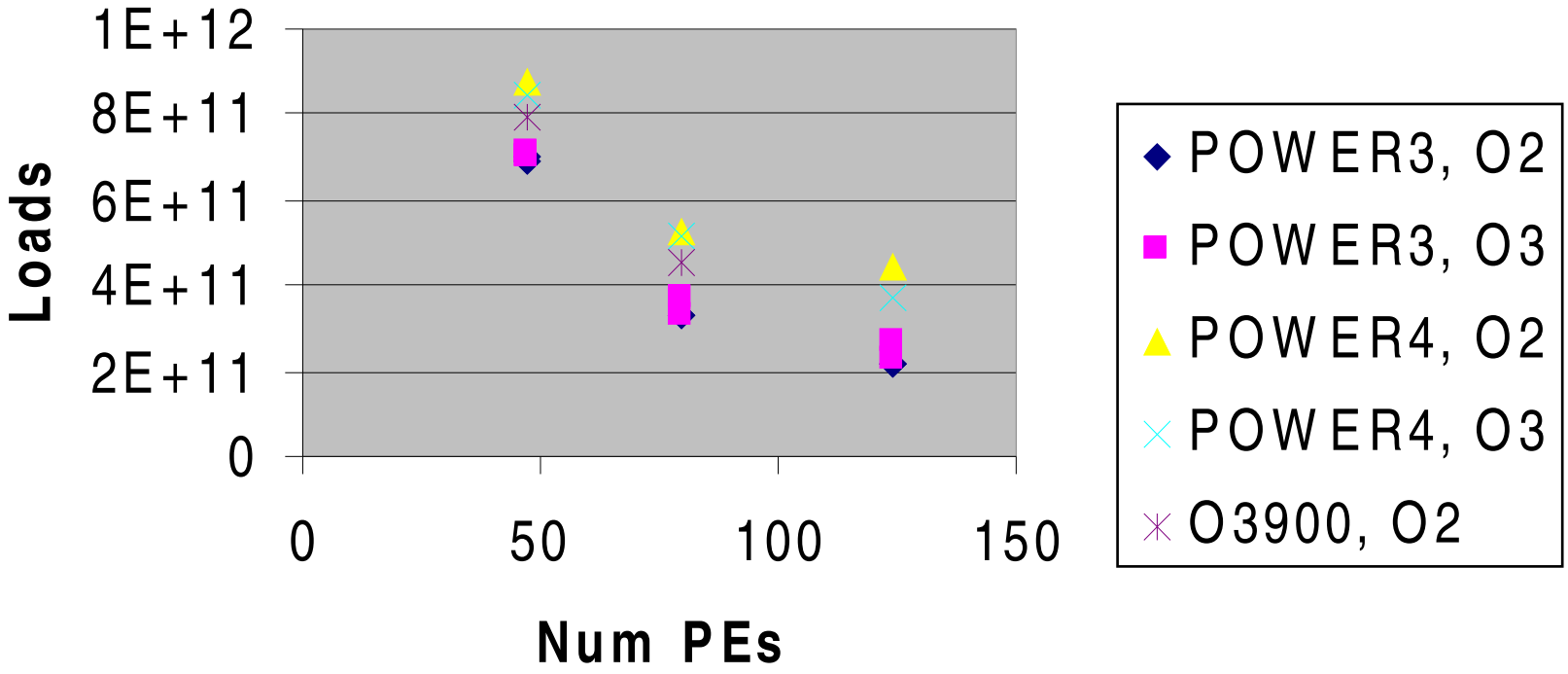

Data Cache Misses per PE in OOCORE
Misses
2.0E+11
1.5E+11
1.0E+11
5.0E+10
0.0E+00
0
20
40
60
80
Num PEs
POWER3, O2
POWER3, O3
POWER3, O4
POWER4, O2
POWER4, O3
POWER4, O4

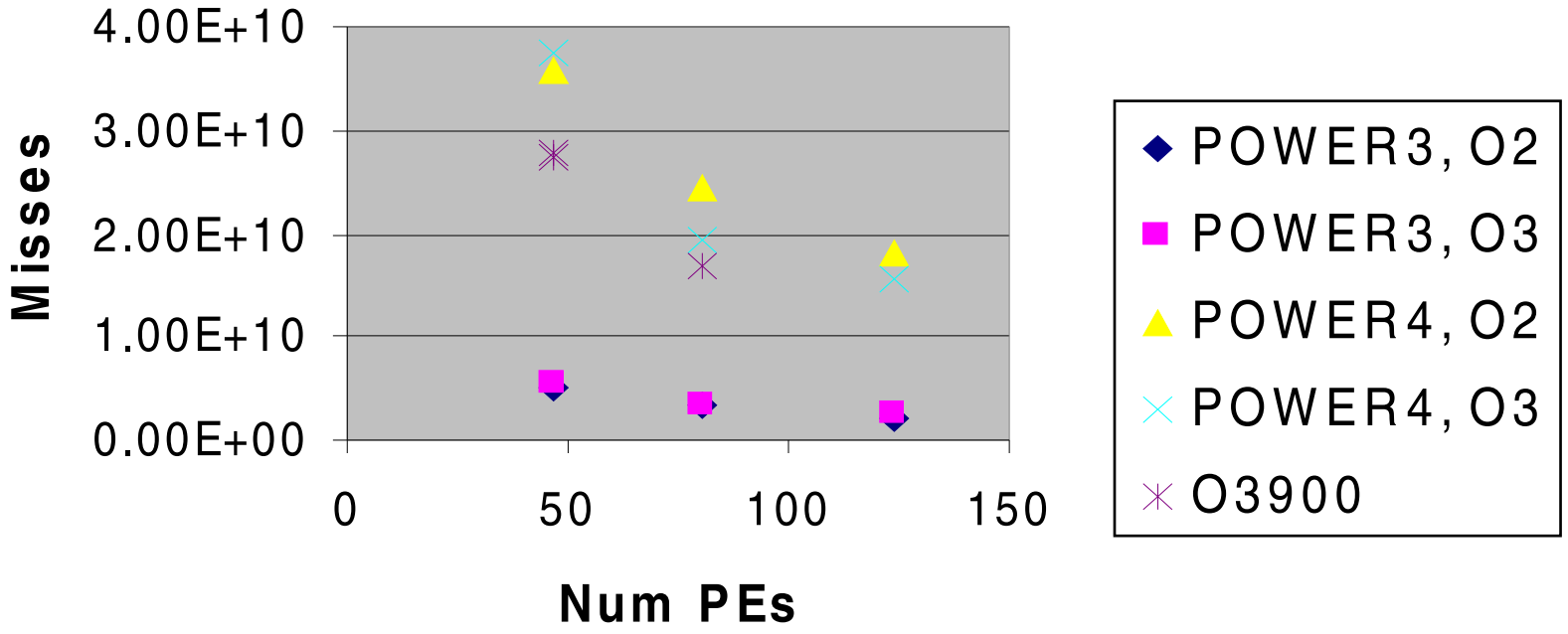

OOCORE Runtimes
100000
10000
1000
time (sec)
10
100
Num PEs
POWER3, O2
POWER3, O3
POWER3, O4
POWER4, O2
POWER4, O3
POWER4, O4
O3900, O3

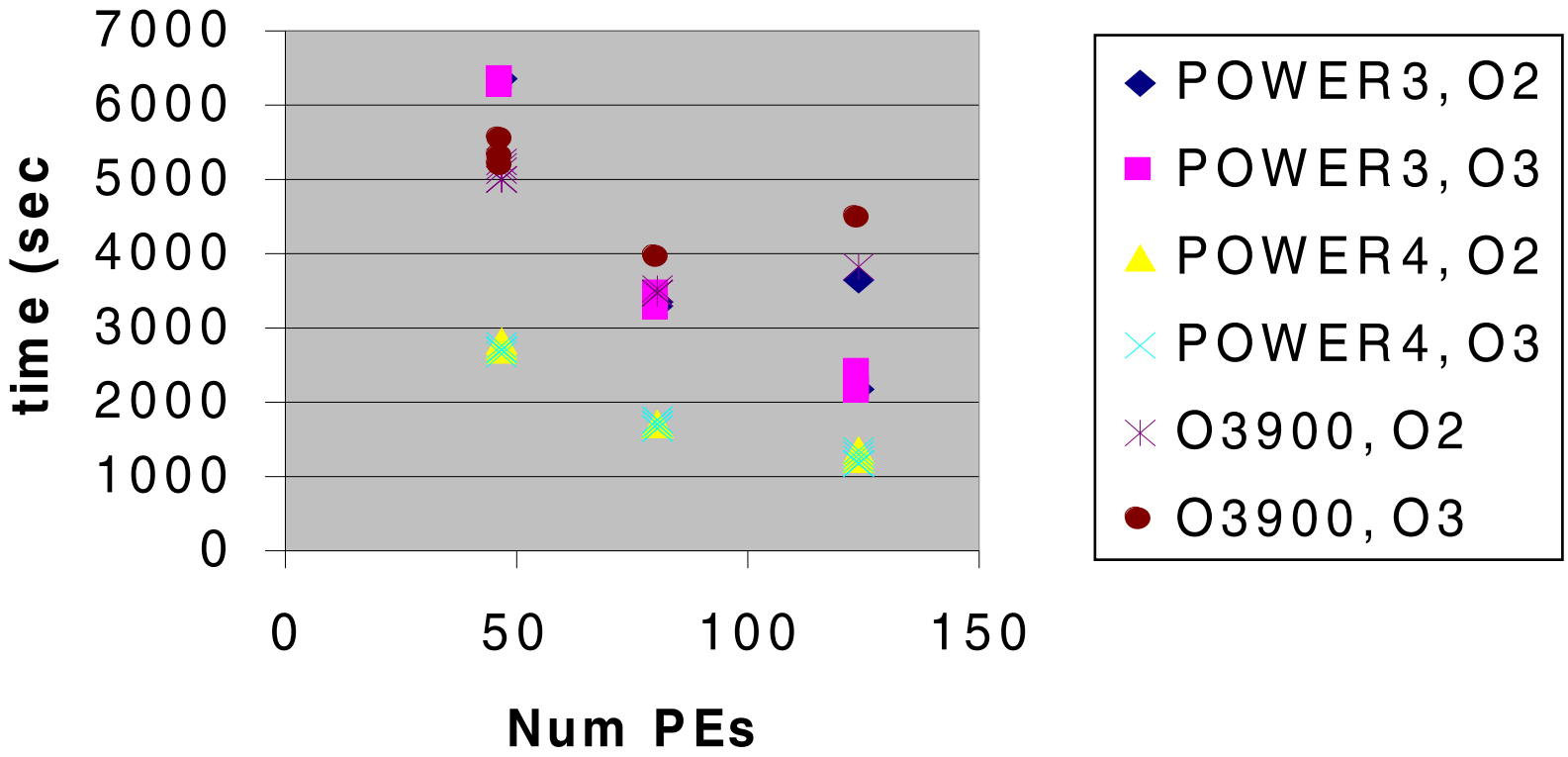

Branches per PE in HYCOM
Branches
4E+11
3E+11
2E+11
1E+11
0
0
50
100
150
Num PEs
POWER3, O2
POWER3, O3
O3900, O2


Floating Pt. Inst. in HYCOM
Instructions
4E+11
3E+11
2E+11
1E+11
0
0
50
100
150
Num PEs
POWER3, O2
POWER3, O3
O3900, O2
O3900, O3
POWER4, O2
POWER4, O3


Instructions per PE in HYCOM
Instructions
3.0E+12
2.5E+12
2.0E+12
1.5E+12
1.0E+12
5.0E+11
0.0E+00
0
50
100
150
Num PEs
POWER3, O2
POWER3, O3
POWER4, O2
POWER4, O3
O3900, O2


Loads per PE in HYCOM
Loads
1E+12
8E+11
6E+11
4E+11
2E+11
0
0
50
100
150
Num PEs
POWER3, O2
POWER3, O3
POWER4, O2
POWER4, O3
O3900, O2


Data cache misses per PE in HYCOM
4.00E+10
3.00E+10
2.00E+10
1.00E+10
0.00E+00
Misses
0
50
100
150
Num PEs
POWER3, O2
POWER3, O3
POWER4, O2
POWER4, O3
O3900


HYCOM Run Times
time (sec
7000
6000
5000
4000
3000
2000
1000
0
0
50
100
150
Num PEs
POWER3, O2
POWER3, O3
POWER4, O2
POWER4, O3
O3900, O2
O3900, O3