



# HPCPI/Xtools Performance Analysis Toolset

David LaFrance-Linden  
High Performance Computing  
Division

Gelato ICE, April 2007



# Overview

- HPCPI
  - Statistical sampling profiler
  - From DCPI (Digital Continuous Profiling Infrastructure)
  - Compare (vaguely) to:
    - OProfile: conceptually based on DCPI
    - Caliper: has many other modes/features
    - Vtune: from Intel, with GUI
- Xtools
  - Performance visualization tools
  - `xclus`: cluster-wide visualization tool
  - `xperf`: node-specific visualization tool

# HPCPI – Standard sampling

- Set default database location

```
% setenv HPCPIDB ~/hpcpidb
```

- Start daemon:

```
% hpcpid
```

- Run programs:

```
% time ./mb_pi.O.exe -iters 100
3.1415926535897932384626433832795028
3.913u 0.000s 0:03.91 100.0%    0+0k 0+0io 0pf+0w
% time ./mb_pi.g.exe -iters 100
3.1415926535897932384626433832795028
37.752u 0.001s 0:37.76 99.9%    0+0k 0+0io 0pf+0w
```

- Flush database to disk

```
% hpcpictl flush
hpcpictl flush successful
```

- Analyze

```
% hpcpiprof
% hpcpiprof ./mb_pi.g.exe ./mb_pi.O.exe
% hpcpilist mb_fill_in_data ./mb_pi.O.exe
```

# hccpiprof (by image)

```
% hccpiprof
```

```
Event Name      Events      Period Samples
-----
CPU_CYCLES     202980300000 60000 3383005
```

```
CPU_CYCLES      %      cum%  image
-----
135815e06 66.9% 66.9% vmlinux-2.6.9-34.7hp.XCsmpt
60180e06 29.6% 96.6% mb_pi.g.exe
6238e06 3.1% 99.6% mb_pi.0.exe
569040e03 0.3% 99.9% ipmi_si.ko
48660e03 0.0% 99.9% libperl.so
38040e03 0.0% 100.0% emacs
28260e03 0.0% 100.0% libc-2.3.4.so
11640e03 0.0% 100.0% ld-2.3.4.so
9420e03 0.0% 100.0% mdmpd
```

```
...
```

# hpcpiprof (by procedure)

```
% hpcpiprof ./mb_pi.0.exe ./mb_pi.g.exe
```

| Event Name | Events      | Period | Samples |
|------------|-------------|--------|---------|
| CPU_CYCLES | 66419940000 | 60000  | 1106999 |

| CPU_CYCLES | %     | cum%   | procedure       | image       |
|------------|-------|--------|-----------------|-------------|
| 59195e06   | 89.1% | 89.1%  | mandel_val      | mb_pi.g.exe |
| 6238e06    | 9.4%  | 98.5%  | mandel_val      | mb_pi.0.exe |
| 985800e03  | 1.5%  | 100.0% | mb_fill_in_data | mb_pi.g.exe |

# hpcpulist (by source/assembly)

- Unfortunately, mandel\_val got inlined, so you have to know to look in mb\_fill\_in\_data

```
% hpcpulist mb_fill_in_data ./mb_pi.O.exe
```

```
Event Name   Events   Period
-----
CPU_CYCLES 623958000 60000
```

```
Could not find source file for routine mb_fill_in_data
try the -f option to specify the source file to use
```

| CPU_CYCLES | PC                     | B    | ASM              | Source                 |             |
|------------|------------------------|------|------------------|------------------------|-------------|
| ...        |                        |      |                  |                        |             |
| 1714e06    | mb_fill_in_data+0x0270 | :    | nop.m            | 0                      | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x0271 |      | fma.d.s0         | f36=farg0,farg0,f0     | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x0272 |      | nop.b            | 0                      | mb_pi.c:164 |
| 966e03     | mb_fill_in_data+0x0280 |      | nop.m            | 0                      | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x0281 |      | fma.d.s0         | f37=farg2,farg2,f0     | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x0282 |      | adds             | ret0=1,ret0            | mb_pi.c:164 |
| 169800e03  | mb_fill_in_data+0x0290 |      | nop.m            | 0                      | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x0291 |      | nop.m            | 0                      | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x0292 |      | fma.d.s0         | f35=farg0,farg2,f0;;   | mb_pi.c:164 |
| 223320e03  | mb_fill_in_data+0x02a0 |      | cmp4.lt          | p4,p5=ret0,r38         | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x02a1 |      | nop.m            | 0                      | unknown_src |
| 0          | mb_fill_in_data+0x02a2 |      | nop.f            | 0;;                    | unknown_src |
| 1451e06    | mb_fill_in_data+0x02b0 | (p4) | addl             | r14=1,r0               | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x02b1 |      | fma.d.s0         | f34=f36,f1,f37         | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x02b2 |      | nop.i            | 0                      | unknown_src |
| 0          | mb_fill_in_data+0x02c0 |      | nop.m            | 0                      | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x02c1 |      | fms.d.s0         | f33=f36,f1,f37         | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x02c2 | (p5) | adds             | r14=0,r0               | mb_pi.c:164 |
| 155460e03  | mb_fill_in_data+0x02d0 |      | nop.m            | 0                      | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x02d1 |      | nop.m            | 0                      | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x02d2 |      | fma.d.s0         | f32=f35,f1,f35;;       | mb_pi.c:164 |
| 1623e06    | mb_fill_in_data+0x02e0 |      | cmp4.eq          | p8,p9=0,r14            | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x02e1 |      | nop.m            | 0                      | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x02e2 |      | fcmp.le.s0       | p2,p3=f34,f3           | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x02f0 |      | nop.m            | 0                      | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x02f1 |      | fma.d.s0         | farg0=f33,f1,farg3     | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x02f2 |      | nop.b            | 0                      | mb_pi.c:164 |
| 253200e03  | mb_fill_in_data+0x0300 |      | nop.m            | 0                      | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x0301 |      | nop.m            | 0                      | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x0302 |      | fma.d.s0         | farg2=f32,f1,f2;;      | mb_pi.c:164 |
| 158940e03  | mb_fill_in_data+0x0310 |      | nop.b            | 0                      | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x0311 | (p3) | br.cond.dpnt.few | mb_fill_in_data+0x0320 | mb_pi.c:164 |
| 0          | mb_fill_in_data+0x0312 | (p9) | br.cond.dptk.few | mb_fill_in_data+0x0270 | mb_pi.c:164 |
| ...        |                        |      |                  |                        |             |

# HPCPI's differentiators

- Sample rate
- Overhead
- Features
  - Can sample more than 1 event
  - Can sample arbitrary number of events
  - The 'label' feature
- Attention to accuracy

# Sample rate and overhead

- Default sample rate higher (interval lower), minimum sample rate high in comparison:

| Tool         | Default interval | Max interval |
|--------------|------------------|--------------|
| OProfile     | 100K             | n/a          |
| HP Caliper   | 500K             |              |
| <b>HPCPI</b> | <b>60K</b>       | <b>64K</b>   |

- Low overhead:

| Tool            |                  | cycles per sample interrupt | Incremental overhead sampling CPU_CYCLES at various intervals |              |              |              |
|-----------------|------------------|-----------------------------|---|--------------|--------------|--------------|
|                 |                  |                             | 60K   | 20K          | 10K          | 5K           |
| <b>Standard</b> | OProfile         | 1070                        | 2.13%   | 6.34%        | 11.35%       | 19.56%       |
|                 | HP Caliper       | 1660                        | 2.80%   | 8.67%        | 16.35%       | 27.82%       |
|                 | HPCPI/C          | 657                         | 1.09%   | 3.07%        | 6.02%        | 11.19%       |
| <b>asm</b>      | <b>HPCPI/asm</b> | <b>208</b>                  | <b>0.35%</b>  | <b>0.99%</b> | <b>2.22%</b> | <b>3.70%</b> |

# Feature: Can sample more than one event



- Useful for deriving metrics at image, routine or loop level
- So can OProfile and Vtune, but not yet Caliper
- IPC
  - CPU\_CYCLES
  - NOPS\_RETIRED
  - PREDICATE\_SQUASHED\_RETIRED
  - IA64\_INST\_RETIRED
- HPC
  - CPU\_CYCLES
  - BACK\_END\_BUBBLE.ALL
  - DATA\_EAR\_EVENTS.CACHE\_MISS.GE64
  - BUS\_MEMORY.ALL.SELF
- Server
  - CPU\_CYCLES
  - BACK\_END\_BUBBLE.ALL
  - DATA\_EAR\_EVENTS.CACHE\_MISS.GE64
  - IA64\_INST\_RETIRED

# Example: IPC for 'mb\_pi'

- Collect and report:

```
% hpcpid -events IPC
% ./mb_pi.O.exe -iters 100
% ./mb_pi.g.exe -iters 100
% hpcpictl flush
% hpcpiprof ./mb_pi.O.exe ./mb_pi.g.exe
```

| Event Name                | Events      | Period | Samples |
|---------------------------|-------------|--------|---------|
| CPU_CYCLES                | 66372780000 | 60000  | 1106213 |
| NOPS_RETIRE               | 25517160000 | 60000  | 425286  |
| PREDICATE_SQUASHED_RETIRE | 890844000   | 6000   | 148474  |
| IA64_INST_RETIRE          | 61668360000 | 60000  | 1027806 |

| CPU_CYCLES | %     | cum%   | NOPS_RETIRE | PREDICATE_SQUASHED_RETIRE | IA64_INST_RETIRE | procedure       | image       |
|------------|-------|--------|-------------|---------------------------|------------------|-----------------|-------------|
| 59160e06   | 89.1% | 89.1%  | 17316e06    | 438846e03                 | 45712e06         | mandel_val      | mb_pi.g.exe |
| 6236e06    | 9.4%  | 98.5%  | 7740e06     | 451974e03                 | 14642e06         | mandel_val      | mb_pi.O.exe |
| 976140e03  | 1.5%  | 100.0% | 460620e03   | 24000                     | 1312e06          | mb_fill_in_data | mb_pi.g.exe |
| 60000      | 0.0%  | 100.0% | 0           | 0                         | 0                | __divdi3        | mb_pi.O.exe |
| 60000      | 0.0%  | 100.0% | 0           | 0                         | 0                | main            | mb_pi.g.exe |
| 0          | 0.0%  | 100.0% | 0           | 0                         | 60000            | __divdi3        | mb_pi.g.exe |

- Compute IPC of mandel\_val

```
- In .g.: 45712e06 / 59160e06 = 0.773
- In .O.: 14642e06 / 6236e06 = 2.348
```

# Multiplex arbitrary events

- Typical stall chase:
  - CPU\_CYCLES
  - BACK\_END\_BUBBLES.ALL
    - BE\_FLUSH\_BUBBLE
    - BE\_L1D\_FPU\_BUBBLE.ALL
      - BE\_L1D\_FPU\_BUBBLE.FPU
      - BE\_L1D\_FPU\_BUBBLE.L1D
        - ... variety of L1D causes ...
    - BE\_EXE\_BUBBLE.ALL
      - BE\_EXE\_BUBBLE.GRALL
        - ...variety of cache events...
      - BE\_EXE\_BUBBLE.GRGR
      - BE\_EXE\_BUBBLE.FRALL
    - BE\_RSE\_BUBBLE
    - BACK\_END\_BUBBLE.FE
- Why not just do them all?
  - And more!
- Unique to HPCPI

# HelpMe event set on bench12

- Setup:

```
% hpcpid -events HelpMe
```

- Then one run per table size, followed by post-processing hpcpiprof output:

All values are in things/update.  
 item ltabsize:

|                                   | 17     | 18     | 19     | 20      | 21      | 22      | ... | 26      | 27      | 28      | 29      |
|-----------------------------------|--------|--------|--------|---------|---------|---------|-----|---------|---------|---------|---------|
| usecs:                            | 0.018  | 0.023  | 0.072  | 0.138   | 0.173   | 0.190   | ... | 0.209   | 0.216   | 0.254   | 0.322   |
| CPU_CYCLES                        | 23.832 | 29.923 | 92.389 | 177.427 | 221.993 | 245.242 | ... | 270.756 | 280.362 | 328.287 | 415.196 |
| IA64_INST_RETIRED                 | 17.498 | 17.543 | 17.766 | 18.074  | 18.353  | 18.272  | ... | 19.394  | 22.735  | 39.640  | 58.850  |
| NOPS_RETIRED                      | 4.009  | 4.007  | 4.039  | 4.115   | 4.161   | 4.103   | ... | 4.301   | 5.464   | 11.261  | 17.844  |
| PREDICATE_SQUASHED_RETIRED        | 0.500  | 0.502  | 0.501  | 0.508   | 0.520   | 0.506   | ... | 0.538   | 1.008   | 3.266   | 5.827   |
| LOADS_RETIRED                     | 2.006  | 2.044  | 2.557  | 2.801   | 2.964   | 2.976   | ... | 3.106   | 3.349   | 4.614   | 6.052   |
| STORES_RETIRED                    | 1.005  | 1.001  | 1.000  | 1.023   | 1.093   | 1.116   | ... | 1.435   | 1.412   | 1.406   | 1.418   |
| BACK_END_BUBBLE.ALL               | 16.813 | 22.878 | 84.729 | 169.341 | 214.881 | 237.803 | ... | 262.682 | 270.515 | 308.295 | 383.796 |
| BE_FLUSH_BUBBLE.ALL               | 0.002  | 0.004  | 0.022  | 0.041   | 0.057   | 0.048   | ... | 0.106   | 0.996   | 5.264   | 10.109  |
| BE_FLUSH_BUBBLE.BRU               | 0.001  | 0.002  | 0.008  | 0.013   | 0.022   | 0.014   | ... | 0.025   | 0.073   | 0.326   | 0.616   |
| BE_FLUSH_BUBBLE.XPN               | 0.001  | 0.002  | 0.013  | 0.021   | 0.034   | 0.033   | ... | 0.079   | 0.926   | 4.955   | 9.518   |
| BACK_END_BUBBLE.L1D_FPU_RSE       | 5.531  | 6.586  | 14.576 | 18.667  | 21.422  | 23.595  | ... | 29.689  | 33.685  | 47.813  | 61.156  |
| BE_L1D_FPU_BUBBLE.ALL             | 5.523  | 6.580  | 14.611 | 18.751  | 21.479  | 23.495  | ... | 29.602  | 33.344  | 46.805  | 59.367  |
| BE_L1D_FPU_BUBBLE.L1D             | 5.523  | 6.580  | 14.611 | 18.748  | 21.475  | 23.495  | ... | 29.602  | 33.344  | 46.806  | 59.367  |
| BE_L1D_FPU_BUBBLE.L1D_HPW         | 0.000  | 0.972  | 10.814 | 16.211  | 19.590  | 21.876  | ... | 27.523  | 31.026  | 43.208  | 54.326  |
| BE_L1D_FPU_BUBBLE.L1D_TLB         | 0.893  | 0.907  | 0.454  | 0.235   | 0.127   | 0.071   | ... | 0.021   | 0.072   | 0.324   | 0.613   |
| BE_L1D_FPU_BUBBLE.L1D_DCURECIR    | 3.654  | 4.665  | 12.911 | 17.330  | 20.228  | 22.353  | ... | 28.119  | 31.986  | 45.634  | 58.432  |
| BE_L1D_FPU_BUBBLE.FPU             | 0.000  | 0.000  | 0.000  | 0.001   | 0.000   | 0.000   | ... | 0.000   | 0.000   | 0.000   | 0.000   |
| BE_EXE_BUBBLE.ALL                 | 11.276 | 16.281 | 69.944 | 150.406 | 193.157 | 213.963 | ... | 232.511 | 235.165 | 252.980 | 308.444 |
| BE_EXE_BUBBLE.FRALL               | 0.000  | 0.000  | 0.006  | 0.000   | 0.012   | 0.012   | ... | 0.012   | 0.015   | 0.017   | 0.019   |
| BE_EXE_BUBBLE.GRALL               | 11.294 | 16.296 | 69.990 | 150.697 | 193.215 | 214.044 | ... | 232.575 | 235.128 | 252.504 | 307.724 |
| BE_EXE_BUBBLE.GRGR                | 0.000  | 0.000  | 0.000  | 0.000   | 0.000   | 0.000   | ... | 0.000   | 0.001   | 0.002   | 0.003   |
| BE_EXE_BUBBLE.ARCR_PR_CANCEL_BANK | 0.003  | 0.005  | 0.017  | 0.029   | 0.037   | 0.036   | ... | 0.054   | 0.148   | 0.545   | 0.775   |
| BE_RSE_BUBBLE.ALL                 | 0.001  | 0.001  | 0.003  | 0.006   | 0.010   | 0.007   | ... | 0.011   | 0.011   | 0.012   | 0.015   |
| BACK_END_BUBBLE.FE                | 0.004  | 0.007  | 0.084  | 0.113   | 0.129   | 0.114   | ... | 0.261   | 0.552   | 2.075   | 3.836   |
| DATA_REFERENCES_SET0              | 3.003  | 3.005  | 3.069  | 3.138   | 3.150   | 3.150   | ... | 3.510   | 3.859   | 5.636   | 7.655   |
| L1DTLB_TRANSFER                   | 0.892  | 0.908  | 0.456  | 0.238   | 0.126   | 0.072   | ... | 0.020   | 0.071   | 0.323   | 0.610   |
| L2DTLB_MISSES                     | 0.000  | 0.042  | 0.528  | 0.761   | 0.875   | 0.935   | ... | 0.995   | 0.997   | 0.997   | 0.997   |
| DTLB_INSERTS_HPW                  | 0.000  | 0.042  | 0.524  | 0.762   | 0.874   | 0.935   | ... | 0.991   | 0.941   | 0.690   | 0.405   |
| HPW_DATA_REFERENCES               | 0.000  | 0.042  | 0.525  | 0.764   | 0.875   | 0.935   | ... | 0.994   | 0.996   | 0.997   | 0.997   |
| L1D_READS_SET0                    | 2.002  | 2.001  | 2.046  | 2.057   | 2.092   | 2.046   | ... | 2.104   | 2.451   | 4.216   | 6.231   |
| L1D_READ_MISSES.ALL               | 1.080  | 1.083  | 1.088  | 1.081   | 1.090   | 1.062   | ... | 1.089   | 1.395   | 2.917   | 4.484   |
| L2_REFERENCES                     | 2.082  | 2.130  | 2.646  | 2.935   | 3.057   | 3.119   | ... | 3.546   | 3.716   | 4.741   | 5.745   |
| L2_MISSES                         | 0.756  | 0.900  | 1.010  | 1.112   | 1.212   | 1.352   | ... | 1.988   | 1.991   | 1.984   | 2.004   |
| L3_REFERENCES                     | 1.506  | 1.773  | 1.946  | 2.070   | 2.184   | 2.556   | ... | 2.991   | 2.996   | 2.989   | 3.009   |
| L3_MISSES                         | 0.000  | 0.018  | 0.270  | 0.635   | 0.816   | 0.905   | ... | 1.029   | 1.051   | 1.170   | 1.451   |
| BUS_MEMORY.ALL_SELF               | 0.000  | 0.036  | 0.527  | 1.245   | 1.597   | 1.800   | ... | 2.010   | 2.043   | 2.164   | 2.446   |
| DATA_EAR_EVENTS.CACHE_MISS.GE4    | 0.135  | 0.135  | 0.136  | 0.136   | 0.136   | 0.132   | ... | 0.136   | 0.167   | 0.325   | 0.483   |
| DATA_EAR_EVENTS.CACHE_MISS.GE64   | 0.000  | 0.002  | 0.033  | 0.078   | 0.101   | 0.112   | ... | 0.124   | 0.125   | 0.130   | 0.156   |
| CPU_CPL_CHANGES                   | 0.000  | 0.000  | 0.000  | 0.000   | 0.000   | 0.001   | ... | 0.005   | 0.112   | 0.614   | 1.184   |

# The 'label' feature

- Partitions samples, usually based on process(es)
- See the man page for `hpcpilabel`
- DCPI classic label:
  - `% hpcpictl label run1 a.out one 1 uno`
  - `% hpcpictl label run2 a.out two 2 dos`
- Restrict to a script and its children:
  - `% hpcpictl label specs -pgid this runSpec`
- Snapshot a system-wide interval:
  - `% hpcpictl label oneMinute -pid -1 -not sleep 60`
- “Attach” to a process
  - `% hpcpictl label attached -pid desiredPID sleep 99999`
- Monitor the idle process on CPU 0 for 5 minutes:
  - `% hpcpictl label pid0cpu0 -pid 0 -cpu 0 -and sleep 300`
- Can be initiated and managed by programs
  - Use `popen()` of `hpcpictl` with `'-pgid this'` or `'-pid parent'`
- Don't forget to `hpcpictl flush`
- Use `'-label labelName'` with the analysis tools

# The 'label' feature -- example

- HPCPI labelling its own 'flush' activity:

```
% hpcpictl label daemonFlush -pid $pidOfDaemon hpcpictl flush
hpcpictl flush successful
```

```
% hpcpictl flush
hpcpictl flush successful
```

```
% hpcpiprof -label daemonFlush
```

| Event Name | Events    | Period | Samples |
|------------|-----------|--------|---------|
| CPU_CYCLES | 222130000 | 5000   | 44426   |

| CPU_CYCLES | %     | cum%   | image                          |
|------------|-------|--------|--------------------------------|
| 169635e03  | 76.4% | 76.4%  | hpcpid.exe                     |
| 38255e03   | 17.2% | 93.6%  | vmlinux-2.4.21-15.14hp.XCsmp   |
| 11530e03   | 5.2%  | 98.8%  | libc-2.3.2.so                  |
| 1880e03    | 0.8%  | 99.6%  | libstdc++.so.5.0.3             |
| 400000     | 0.2%  | 99.8%  | ecount.2.4.21-15.14hp.XCsmp.ko |
| 200000     | 0.1%  | 99.9%  | libpthread-0.60.so             |
| 120000     | 0.1%  | 100.0% | ipmi_kcs_drv.o                 |
| 70000      | 0.0%  | 100.0% | ld-2.3.2.so                    |
| 30000      | 0.0%  | 100.0% | scsi_mod.o                     |
| 5000       | 0.0%  | 100.0% | ipmi_msghandler.o              |
| 5000       | 0.0%  | 100.0% | libgcc_s-3.2.3-20040414.so.1   |

# The 'label' feature – example (cont)



- Inside the kernel:

```
% hpcpiprof -label daemonFlush /boot/vmlinux-2.4.21-15.14hp.XCsmpt | head -20
```

```
Event Name      Events      Period      Samples
-----
CPU_CYCLES      38255000    5000        7651
```

```
CPU_CYCLES      %      cum%      procedure      image
-----
1865e03         4.9%    4.9%      ext3_find_entry      vmlinux-2.4.21-15.14hp.XCsmpt
1735e03         4.5%    9.4%      ext3_check_dir_entry  vmlinux-2.4.21-15.14hp.XCsmpt
1530e03         4.0%   13.4%      clear_page           vmlinux-2.4.21-15.14hp.XCsmpt
1405e03         3.7%   17.1%      link_path_walk_it    vmlinux-2.4.21-15.14hp.XCsmpt
1225e03         3.2%   20.3%      unlock_buffer        vmlinux-2.4.21-15.14hp.XCsmpt
1210e03         3.2%   23.4%      memset               vmlinux-2.4.21-15.14hp.XCsmpt
1170e03         3.1%   26.5%      do_get_write_access  vmlinux-2.4.21-15.14hp.XCsmpt
1095e03         2.9%   29.4%      ext3_add_entry       vmlinux-2.4.21-15.14hp.XCsmpt
 975000         2.5%   31.9%      journal_cancel_revoke vmlinux-2.4.21-15.14hp.XCsmpt
 910000         2.4%   34.3%      get_hash_table       vmlinux-2.4.21-15.14hp.XCsmpt
 900000         2.4%   36.6%      d_lookup             vmlinux-2.4.21-15.14hp.XCsmpt
 860000         2.2%   38.9%      journal_dirty_metadata vmlinux-2.4.21-15.14hp.XCsmpt
 860000         2.2%   41.1%      journal_add_journal_head vmlinux-2.4.21-15.14hp.XCsmpt
 825000         2.2%   43.3%      ext3_do_update_inode vmlinux-2.4.21-15.14hp.XCsmpt
```

# The 'label' feature – example (cont)



- In libc:

```
% hpcpiprof -label daemonFlush /lib/tls/libc-2.3.2.so | head -20
```

```
Event Name      Events      Period      Samples
-----
CPU_CYCLES     11530000    5000        2306
```

```
CPU_CYCLES      %      cum%      procedure      image
-----
2165e03         18.8%   18.8%     __GI_memset     libc-2.3.2.so
1470e03         12.7%   31.5%     _IO_vfprintf_internal  libc-2.3.2.so
 815000         7.1%   38.6%     _IO_fwrite_internal  libc-2.3.2.so
 585000         5.1%   43.7%     _IO_new_file_xsputn  libc-2.3.2.so
 380000         3.3%   47.0%     _int_malloc     libc-2.3.2.so
 365000         3.2%   50.1%     __GI_getenv     libc-2.3.2.so
 290000         2.5%   52.6%     __GI_strftime   libc-2.3.2.so
 285000         2.5%   55.1%     __find_specmb   libc-2.3.2.so
 285000         2.5%   57.6%     __GC__libc_write libc-2.3.2.so
 265000         2.3%   59.9%     _wordcopy_fwd_aligned  libc-2.3.2.so
 260000         2.3%   62.1%     __GI_strlen     libc-2.3.2.so
 255000         2.2%   64.4%     _IO_default_xsputn_internal  libc-2.3.2.so
 240000         2.1%   66.4%     __tzfile_compute libc-2.3.2.so
 235000         2.0%   68.5%     _IO_str_overflow_internal  libc-2.3.2.so
```

# Attention to accuracy

- Wrote micro-benchmarks with known behavior
- Eliminated post-unfreeze-pre-RFI event leaks
  - Micro-benchmark has no NOPS nor any predicate-squashed instructions

|             | IPC<br>(actual: 3) | NOPS<br>per sample | pred-<br>squashed<br>per sample |
|-------------|--------------------|--------------------|---------------------------------|
| problematic | 2.464              | 67.959             | 1.00                            |
| corrected   | 2.917              | 0.1998             | 0.009                           |

- Determined event-based multiplexing better than time-based

| Interval | Actual IPC | Non-muxed     | Time-muxed     | Modeled error | Event-muxed   |
|----------|------------|---------------|----------------|---------------|---------------|
| 40K      | 5.918      | 5.893(-0.37%) | 5.586(-5.61%)  | -6.81%        | 5.896(-0.37%) |
| 20K      |            | 5.883(-0.60%) | 5.295(-10.53%) | -12.50%       | 5.883(-0.59%) |
| 10K      |            | 5.858(-1.02%) | 4.874(-17.65%) | -21.43%       | 5.588(-1.03%) |
| 5K       |            | 5.807(-1.87%) | 4.247(-28.24%) | -33.33%       | 5.509(-1.85%) |

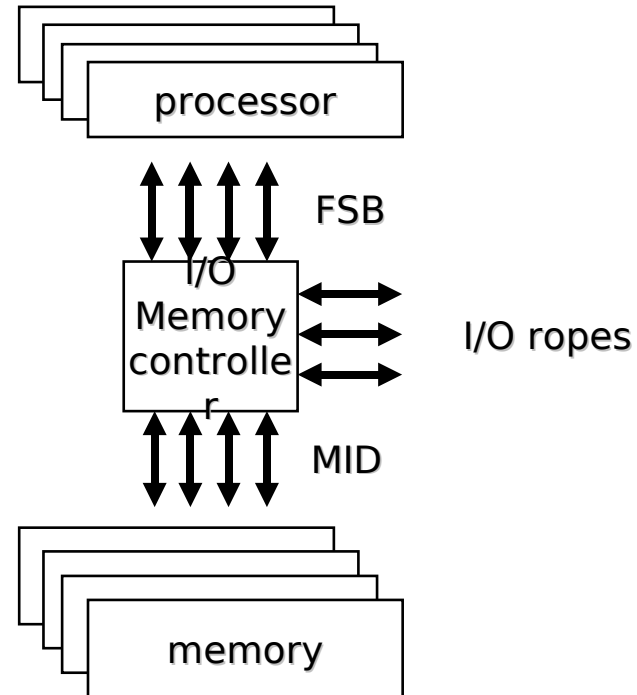
# Xtools

- Pair of visualization tools
- Separable and cooperative with HPCPI
- Xclus
  - Cluster-wide monitoring
  - Utilizations: CPU, FSB and MID bus I/O
- Xperf
  - Single-node monitoring
  - Graphs of derived events based on hardware counters
    - CPU utilization, IPC, cycle accounting, cache penalties, I/O activity, etc

# Basic structure of a system

For icon-design of xclus:

- Processors
- Front-Side Bus (FSB)
- (I/O Memory controller)
- I/O ropes
- MID bus
- Memory





# xclus node grouping (movie clip)



# Recap

- HPCPI
  - Sampling profiler
  - High frequency
  - Low overhead
  - Arbitrary events (auto-placement, auto-multiplexing)
  - Attention to accuracy
  - ‘label’ feature
- Xtools
  - Xclus: cluster-wide utilization visualizer
  - Xperf: node-specific time-graphs of counter-based metrics
    - Integrated with HPCPI

# Availability

- Available under an evaluation license
- Contact [Jim.Bovay@hp.com](mailto:Jim.Bovay@hp.com)