Performance Measurement and Visualization on the Cray XT4

Luiz DeRose

Programming Environment Director

Cray Inc.

ldr@cray.com

NERSC
September 18-20, 2007

The Cray Tools Strategy

- Must be easy to use
  - Automatic program instrumentation
    - no source code or makefile modification needed
- Integrated performance tools solution
  - Multiple platforms
  - Multiple functionality
    - Measurements of user functions, MPI, I/O, memory, & HW Counters
- Strategy based on the three main steps normally used for application optimization and tuning:
  - Debug application
  - Single processor optimization
  - Parallel processing and I/O optimization
- Close interaction with user for feedback targeting functionality enhancements

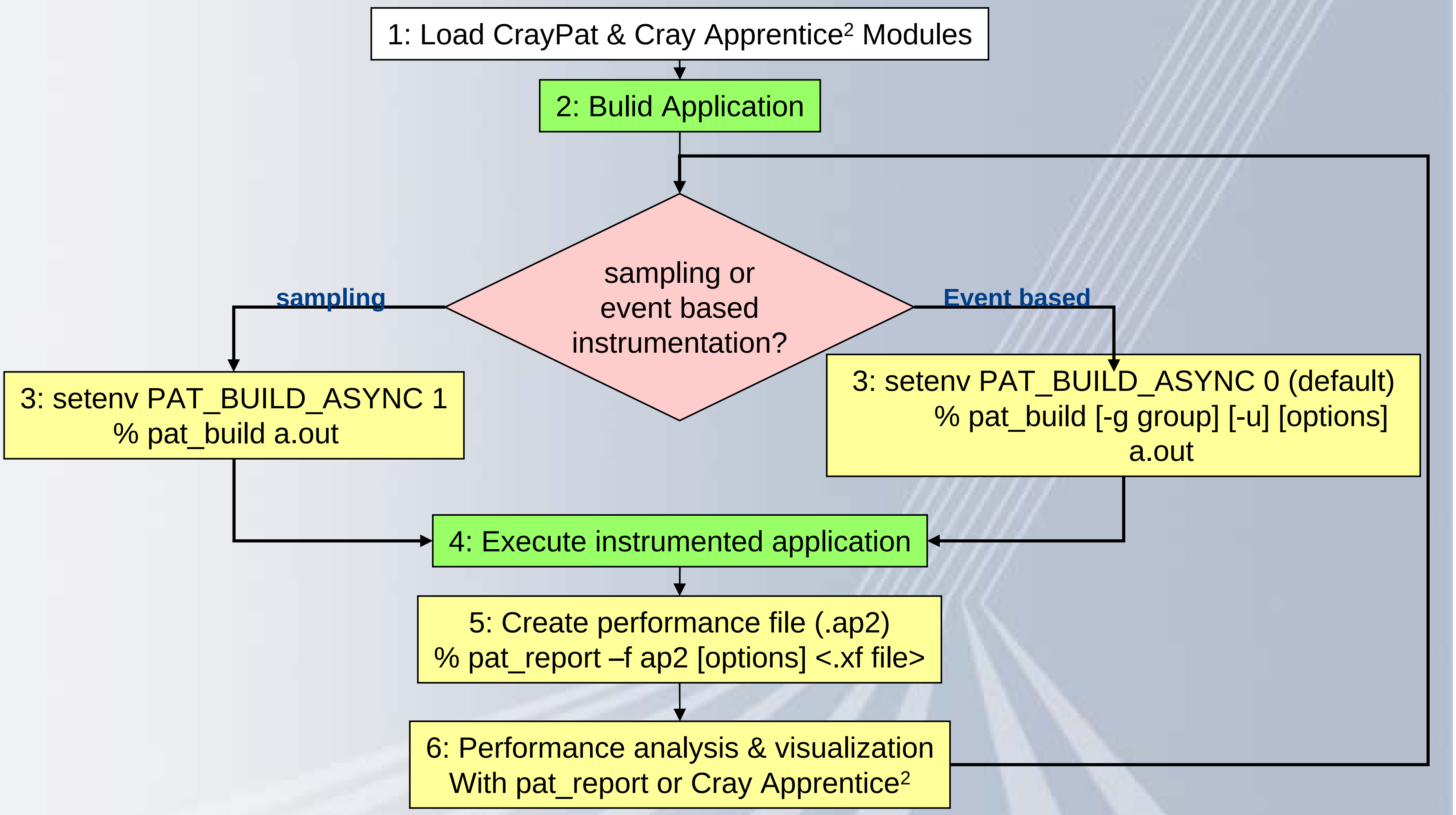
Cray Performance Analysis Infrastructure

- CrayPat
  - pat_hwpc:
    - Utility for hardware counters measurement of whole program
  - pat_build: Utility for application instrumentation
    - No source code modification required
  - run-time library for measurements
    - transparent to the user
  - pat_report:
    - Performance reports
    - Performance visualization file
  - pat_help
- Cray Apprentice^2
  - Graphical performance analysis and visualization tool
  - Can be used off-line on Linux system

Performance Data Collection

- Two dimensions
  - When performance collection is triggered
    - External agent (asynchronous)
      - Sampling
        - » Timer interrupt
        - » Hardware counters overflow
    - Internal agent (synchronous)
      - Code instrumentation
        - » Event based
        - » Automatic or manual instrumentation
  - How performance data is recorded
    - Profile ::= Summation of events over time
      - run time summarization (functions, call sites, loops, …)
    - Trace file ::= Sequence of events over time

Performance Analysis with CrayPat & Cray Apprentice^2



1: Load CrayPat & Cray Apprentice^2 Modules

2: Bulid Application

sampling or
event based
instrumentation?

sampling

3: setenv PAT_BUILD_ASYNC 1
% pat_build a.out

Event based

3: setenv PAT_BUILD_ASYNC 0 (default)
% pat_build [-g group] [-u] [options] a.out

4: Execute instrumented application

5: Create performance file (.ap2)
% pat_report –f ap2 [options] <.xf file>

6: Performance analysis & visualization
With pat_report or Cray Apprentice^2

When Should I Use Sampling?

- Sampling is useful to determine where (functions and lines) the program spends most of its time
- The environment variable PAT_RT_EXPERIMENT allows the specification of the type of experiment prior to execution
  - samp_pc_prof (default)
    - Samples the PC by time
    - This option has the lowest sampling overhead
    - Does not allow collection of hardware counters
  - samp_pc_time
    - Samples the PC at intervals specified in microseconds with PAT_RT_INTERVAL
    - The default interval is 10,000 microseconds.
    - Optionally record the values of hardware counters specified with PAT_RT_HWPC
  - samp_cs_time
    - collect call stack, which is useful when generating reports
    - Does not allow collection of hardware counters

Why Should I generate an ".ap2" file?

- The ".ap2" file is a self contained compressed performance file
  - Normally it is about 5 times smaller than the ".xf" file
  - Contains the information needed from the application binary
    - Can be reused, even if the application binary is no longer available or if it was rebuilt
  - It is the only input format accepted by Cray Apprentice^2

Application Instrumentation with pat_build

- No source code or makefile modification required
  - Automatic instrumentation at group (function) level
    - Groups: mpi, io, heap, user function (-u), main (-w) …
  - API provided for instrumentation at a finer granularity
- Performs binary rewrite
  - Relink application
  - Requires object files
  - Generates a stand alone instrumented program
  - Supports sampling and event based instrumentation
    - Selection with environment variable (PAT_BUILD_ASYNC)
      - Default is event based
- Runtime environment variable defines if profile or trace file will be generated
  - PAT_RT_SUMMARY
    - Default is 1 (for runtime summarization)

Function Selection

- By default “-u” instrument all user functions.
- “-T” and “-t” flags can be used to select a subset of functions
  - “-T funcList”: instruments program to trace the function entry point references in “funcList”
    - This option may requires “-w” (instrument main)
    - The exclamation point (“!”) indicates negation. Functions in “funcList” preceeded by “!” are not traced
      - This option can be used to exclude both user-defined functions and functions that are part of “-g” groups
    - Examples:
      - Instrument only user functions “f1” and “f2”
        - » pat_build –w –T f1,f2
      - Instrument all user functions, with the exception of “f1” and “f2”
        - » pat_build –u –T !f1,!f2
  - “-t <filename>”: Instrument program to trace all function entry point references listed in the <filename>

CrayPat API

- CrayPat performs automatic instrumentation at function level
- The CrayPat API can be used for fine grain instrumentation
  - Fortran
    - call PAT_region_begin(id, “label”, ierr)
    - DO Work
    - call PAT_region_end(id, ierr)
  - C
    - include <pat_api.h>
    - …
    - ierr = PAT_region_begin(id, “label”);
    - DO_Work();
    - ierr = PAT_region_end(id);

Runtime Environment Variables

- The following runtime environment variables affect how the data is collected:
  - PAT_RT_EXPFILE_PER_PROCESS
    - If it is not set or set to zero, a single experiment data file is created
  - PAT_RT_SUMMARY
    - Enables run-time summarization
      - Includes the aggregation of data during run-time
      - Runtime summarization is enabled by default
  - PAT_RT_HWPC <groups #>
    - Activate collection of hardware performance counters
      - There are 10 groups on the XT

pat_report Options

- Reformating the performance file (Cray Apprentice^2 input)
  - pat_report [-V] [-i dir|instrprog] [-o output_file]
    -f ap2 |txt |xml data_directory | data_file.xf
- Generating performance reports
  - pat_report [-V] [-i dir|instrprog] [-o output_file]
    [-O keyword] [-b b-opts] [-d d-opts] [-s key=value] [-P] [-T]
    data_directory | data_file.xf | data_file.ap2
- Main options:
  - -i is only if the instrumented program has a different name or is in a different directory path than when it was executed
  - **-O provides shortcuts for common reports**
  - -b, -d, -s can be used to further customize the report

Pat_report Output

```
CrayPat/X:  Version 3.2 Revision 926 (xf 860)  07/16/07 08:31:54

Experiment:  trace

Experiment data file:
  /lus/nid00008/ldr/Apps/sweep3d/sweep3d+pat+51td.xf  (RTS)

Current path to data file:
  /lus/nid00008/ldr/Apps/sweep3d/sweep3d+pat+51td.ap2  (RTS)

Original program:  /lus/nid00008/ldr/Apps/sweep3d/sweep3d

Instrumented with:  pat_build -f -u -g mpi,io,heap sweep3d

Instrumented program:  /lus/nid00008/ldr/Apps/sweep3d/./sweep3d+pat

Program invocation:  ./sweep3d+pat

Number of PEs:  48

Exit Status:  0  PEs:  0-47

Runtime environment variables:
  MPICHBASEDIR=/opt/xt-mpt/1.5.57/mpich2-64
  MPICH_DIR=/opt/xt-mpt/1.5.57/mpich2-64/P2
  MPICH_DIR_FTN_DEFAULT64=/opt/xt-mpt/1.5.57/mpich2-64/P2W

Report time environment variables:
  PAT_ROOT=/opt/xt-tools/craypat/3.2.3/cpatx

Report command line options:  <none>

System name, type, and speed:  guppy1 x86_64  2400 MHz

Operating system:  catamount 1.0 2.0

Estimated minimum overhead per call of a traced function,
  which was subtracted from the data shown in this report
  (for raw data, use the option:  -s overhead=include):
    Time    0.646  microseconds

Number of traced functions: 161
. . .
```

List of instrumented functions is available when the flag:
-s traced_functions=show
is set on pat_report.

Sampling Output (Default table)

```
Notes for table 1:

  Table option:
    -O samp_prof
  Options implied by table option:
    -d sa%@0.05,cum_sa%,sa,imb_sa,imb_sa% -b fu,pe=HIDE

  This table shows only lines with Samp% > 0.05.

  Percentages at each level are relative
    (for absolute percentages, specify:  -s percent=a).


Table 1:  Profile by Function

 Samp % |  Cum.  |  Samp |  Imb. |   Imb. |Function
        | Samp % |       |  Samp | Samp % | PE='HIDE'

 100.0% | 100.0% | 22387 |    -- |     -- |Total
|----------------------------------------------------
|  66.1% |  66.1% | 14790 |  9.88 |   3.2% |sweep_
|  15.4% |  81.5% |  3454 | 12.04 |  14.6% |PtlEQPeek
|   4.7% |  86.2% |  1045 | 11.23 |  34.8% |PtlEQGet
|   3.1% |  89.2% |   686 | 10.71 |  43.7% |PtlEQGet_internal
|   2.3% |  91.5% |   519 |  7.19 |  40.8% |__c_mcopy8
|   2.1% |  93.6% |   462 | 11.38 |  55.3% |ptl_hndl2nal
|   1.5% |  95.1% |   339 |  5.94 |  46.6% |check_eqs_for_event
|   1.1% |  96.2% |   238 |  2.04 |  29.8% |source_
|   1.0% |  97.2% |   219 |  6.44 |  59.8% |poll
|   0.7% |  97.8% |   151 |  4.85 |  62.0% |lock
|   0.7% |  98.5% |   150 |  3.88 |  56.5% |unlock
|   0.3% |  98.8% |    71 |  2.52 |  64.4% |__c_mzero8
|   0.2% |  99.0% |    47 |  3.02 |  77.1% |_P3_getmsg
|   0.2% |  99.2% |    42 |  3.12 |  79.8% |memcpy
|   0.1% |  99.4% |    33 |  2.31 |  78.7% |flux_err_
|   0.1% |  99.5% |    32 |  3.33 |  85.1% |control_message_in
|   0.1% |  99.6% |    19 |  1.60 |  81.9% |memset
|====================================================
```

Sampling Output (with samp_cs_time & -O ct+src)

```
Notes for table 1:

  Table option:
    -O calltree+src
  Options implied by table option:
    -d ti%@0.05,cum_ti%,ti,tr -b ct,pe=HIDE -s show_ca='fu,so,li' \
    -s source_limit='1'

  This table shows only lines with Samp% > 0.05.

  Percentages at each level are relative
    (for absolute percentages, specify:  -s percent=a).


Table 1:  Calltree View with Callsite Line Numbers

 Samp % |   Cum. |  Samp |Calltree
        | Samp % |       | PE='HIDE'

 100.0% | 100.0% | 22359 |Total
|-----------------------------------
|  96.5% |  96.5% | 21575 |main:...:line.0
||----------------------------------
||  99.8% |  99.8% | 21536 |MAIN_:driver.f:line.174
3|        |        |       | inner_auto_:inner_auto.f:line.69
||||--------------------------------
4|||  98.6% |  98.6% | 21226 |inner_:inner.f:line.102
|||||-------------------------------
5||||  18.8% |  18.8% |  3995 |sweep_:sweep.f:line.237
6||||        |        |       | rcv_real_:mpi_stuff.f:line.167
7||||        |        |       |  mpi_recv_:...:line.0
||||||||----------------------------
8|||||||  98.9% |  98.9% |  3953 |PtlEQPoll:api-eq.c:line.154
|||||||||---------------------------
9||||||||  96.0% |  96.0% |  3794 |poll:qkapi.c:line.86
||||||||||--------------------------
10||||||||  94.5% |  94.5% |  3587 |check_eqs_for_event:api-eq.c:line.169
|||||||||||-------------------------
11|||||||||  76.7% |  76.7% |  2750 |PtlEQGet:api-eq.c:line.120
||||||||||||------------------------
12||||||||||  87.6% |  87.6% |  2409 |PtlEQGet_internal:api-eq.c:line.87
|||||||||||||-----------------------
13|||||||||||  72.4% |  72.4% |  1745 |PtlEQPeek:internal.h:line.55
13|||||||||||   7.8% |  80.2% |   187 |PtlEQPeek:pool.h:line.62
13|||||||||||   3.2% |  83.4% |    76 |PtlEQPeek:api-eq.c:line.57
13|||||||||||   2.2% |  85.5% |    52 |PtlEQPeek:api-eq.c:line.50
||||||||||||||----------------------
. . .
```

Table 1: Flat Profile (Default)

```
Notes for table 1:

  High level option:  -O profile
  Low level options:  -d ti%@0.05,ti,imb_ti,imb_ti%,tr \
    -b exp,gr,fu,pe=HIDE

  This table shows only lines with Time% > 0.05.

  Percentages at each level are relative
    (for absolute percentages, specify:  -s percent=a).


Table 1:  Profile by Function Group and Function

 Time % |     Time |Imb. Time |   Imb.  | Calls |Group
        |          |          | Time %  |       | Function
        |          |          |         |       |  PE='HIDE'

 100.0% | 4.609555 |       -- |      -- | 11947 |Total
|-----------------------------------------------------------
|  72.5% | 3.340820 |       -- |      -- |  5112 |USER
||----------------------------------------------------------
||  97.8% | 3.267208 | 0.045394 |   1.4% |    12 |sweep_
||   1.5% | 0.049875 | 0.000763 |   1.5% |    12 |source_
||   0.3% | 0.009005 | 0.000226 |   2.5% |    12 |flux_err_
||   0.2% | 0.007557 | 0.000917 |  11.0% |  2460 |snd_real_
||   0.1% | 0.003155 | 0.000552 |  15.2% |  2460 |rcv_real_
||==========================================================
```

By default, the report will only show functions with at least 0.05% of the time

Table 1: Flat Profile (Continuation)

```
|  22.9% | 1.053745 |       -- |     -- |  4963 |MPI
||-------------------------------------------------------
||  94.2% | 0.992757 | 0.287228 |  22.9% |  2460 |mpi_recv_
||   5.6% | 0.058613 | 0.010351 |  15.3% |  2460 |mpi_send_
||   0.2% | 0.002107 | 0.000663 |  24.5% |    32 |mpi_allreduce_
||=======================================================
|   4.5% | 0.205414 |       -- |     -- |    39 |MPI_SYNC
||-------------------------------------------------------
||  81.5% | 0.167507 | 0.183974 |  53.5% |    32 |mpi_allreduce_(sync)
||  13.7% | 0.028122 | 0.000784 |   2.8% |     3 |mpi_barrier_(sync)
||   4.8% | 0.009785 | 0.000263 |   2.7% |     4 |mpi_bcast_(sync)
||=======================================================
|   0.1% | 0.004985 |       -- |     -- |  1825 |HEAP
||-------------------------------------------------------
||  58.2% | 0.002901 | 0.001326 |  32.0% |   914 |malloc
||  38.4% | 0.001914 | 0.000685 |  26.9% |   910 |free
||   3.4% | 0.000169 | 0.000001 |   0.7% |     1 |calloc
||=======================================================
|   0.1% | 0.004591 |       -- |     -- |     8 |IO
||-------------------------------------------------------
||  89.3% | 0.004098 | 0.036538 |  91.8% |     6 |fwrite
||   7.2% | 0.000331 | 0.015561 | 100.0% |     0 |fputc
||   3.2% | 0.000149 | 0.007006 | 100.0% |     1 |getc
||   0.2% | 0.000009 | 0.000400 | 100.0% |     0 |fopen
||   0.1% | 0.000003 | 0.000156 | 100.0% |     0 |fclose
|========================================================
```

Table 2: Load Balance

```
Notes for table 2:

  High level option:  -O load_balance_sm
  Low level options:  -d ti%@0.05,ti,sc,sm,sz -b gr,pe=[mmm]

Table 2:  Load Balance with MPI Sent Message Stats

  Time % |     Time |  Sent | Sent Msg | Avg Sent |Group
         |          |   Msg |    Total | Msg Size |PE[mmm]
         |          | Count |    Bytes |          |

 100.0% | 4.617618 |  2460 | 25920000 | 10536.59 |Total
|---------------------------------------------------------
|  72.4% | 3.344121 |    -- |       -- |       -- |USER
||--------------------------------------------------------
||   2.1% | 3.391748 |    -- |       -- |       -- |pe.31
||   2.1% | 3.376433 |    -- |       -- |       -- |pe.34
||   2.0% | 3.233403 |    -- |       -- |       -- |pe.43
||========================================================
|  22.9% | 1.056950 |  2460 | 25920000 | 10536.59 |MPI
||--------------------------------------------------------
||   2.6% | 1.321445 |  1440 | 14860800 | 10320.00 |pe.47
||   2.0% | 1.034243 |  2880 | 30412800 | 10560.00 |pe.21
||   1.6% | 0.835843 |  1440 | 15206400 | 10560.00 |pe.0
||========================================================
|   4.4% | 0.205439 |    -- |       -- |       -- |MPI_SYNC
||--------------------------------------------------------
||   3.6% | 0.358704 |    -- |       -- |       -- |pe.6
||   2.1% | 0.205843 |    -- |       -- |       -- |pe.26
||   0.4% | 0.041313 |    -- |       -- |       -- |pe.47
||========================================================
|   0.1% | 0.006163 |    -- |       -- |       -- |HEAP
||--------------------------------------------------------
||   3.0% | 0.008870 |    -- |       -- |       -- |pe.36
||   2.0% | 0.005901 |    -- |       -- |       -- |pe.27
||   1.5% | 0.004345 |    -- |       -- |       -- |pe.10
||========================================================
|   0.1% | 0.004597 |    -- |       -- |       -- |IO
||--------------------------------------------------------
||  29.2% | 0.064487 |    -- |       -- |       -- |pe.0
||   1.5% | 0.003319 |    -- |       -- |       -- |pe.9
||   0.8% | 0.001690 |    -- |       -- |       -- |pe.21
|=========================================================
```

Table 3: MPI Send Stats by Bucket

```
Notes for table 3:

  Table option:
    -O mpi_callers
  Options implied by table option:
    -d sm,sc@,mb1..7 -b fu,ca,pe=[mmm]

  Options for related tables not shown by default:
    -O mpi_dest_bytes
    -O mpi_dest_counts

  This table shows only lines with Sent Msg Count > 0.


Table 3:  MPI Sent Message Stats by Caller

 Sent Msg |   Sent |  4KB<= |Function
    Total |    Msg |  MsgSz | Caller
    Bytes |  Count |  <64KB |  PE[mmm]
          |        |  Count |

 25920000 |   2460 |   2460 |Total
|-----------------------------------
|  25920000 |   2460 |   2460 |mpi_send_
|           |        |        | snd_real_
3           |        |        |  sweep_
4           |        |        |   inner_
5           |        |        |    inner_auto_
6           |        |        |     MAIN_
7           |        |        |      main
||||||||----------------------------
8||||||| 30412800 |   2880 |   2880 |pe.33
8||||||| 23846400 |   2160 |   2160 |pe.17
8||||||| 14860800 |   1440 |   1440 |pe.47
|===================================
```

Table 4: Heap Usage

```
Notes for table 4:

  Table option:
    -O heap_program
  Options implied by table option:
    -d IU,IF,NF,FM -b pe=[mmm]


Table 4:  Heap Usage at Start and End of Main Program

MB Heap |   MB Heap |   Heap | Max Free |PE[mmm]
Used at |   Free at |    Not |Object at |
  Start |     Start |  Freed |       End |
        |           |     MB |          |

 99.699 | 3872.301 | 1.042 | 3871.233 |Total
|---------------------------------------------
| 99.728 | 3872.272 | 1.076 | 3870.062 |pe.0
| 99.698 | 3872.302 | 1.041 | 3871.257 |pe.39
| 99.697 | 3872.303 | 1.041 | 3871.259 |pe.19
|=============================================
```

Table 5: Heap Statistics

```
Notes for table 5:

  Table option:
    -O heap_hiwater
  Options implied by table option:
    -d am@,ub,ta,ua,tf,nf,ac,ab -b pe=[mmm]

  This table shows only lines with Tracked Heap HiWater MBytes > 0.


Table 5:  Heap Stats during Main Program

 Tracked |  Total | Total | Tracked | Tracked |PE[mmm]
    Heap | Allocs | Frees | Objects |  MBytes |
 HiWater |        |       |     Not |     Not |
  MBytes |        |       |   Freed |   Freed |

   9.794 |    915 |   910 |       4 |   1.011 |Total
|-----------------------------------------------------
||   9.943 |   1170 |  1103 |      68 |   1.046 |pe.0
||   9.909 |    715 |   712 |       3 |   1.010 |pe.22
||   9.446 |   1278 |  1275 |       3 |   1.010 |pe.43
|=====================================================
```

Table 6: Heap Leaks

```
Notes for table 6:

  Table option:
    -O heap_leaks
  Options implied by table option:
    -d lb%@1,lb@0.0005,lc -b ca,pe=[mmm]

  This table shows only lines with:
    Tracked MBytes Not Freed% > 1
    Tracked MBytes Not Freed > 0.0005
  Percentages at each level are relative
    (for absolute percentages, specify:  -s percent=a).


Table 6:  Heap Leaks during Main Program

 Tracked | Tracked | Tracked |Caller
  MBytes |  MBytes | Objects | PE[mmm]
     Not |     Not |     Not |
 Freed % |   Freed |   Freed |

  100.0% |   1.011 |       3 |Total
|-------------------------------------
|   99.0% |   1.000 |       1 |allhdr
|         |         |         | MAIN_
3         |         |         |  main
||||----------------------------------
4|||    2.1% |   1.000 |       1 |pe.33
4|||    2.1% |   1.000 |       1 |pe.22
4|||    2.1% |   1.000 |       1 |pe.5
|=====================================
```

Table 7: I/O (Read) Statistics

```
Notes for table 7:

  Table option:
    -O read_stats
  Options implied by table option:
    -d rt,rb,rR,rd@,rC -b fi,pe=[mmm],fd

  This table shows only lines with Reads > 0.


Table 7:  File Input Stats by Filename

  Read |   Read MB |   Read Rate | Reads |    Read |File Name
  Time |           |      MB/sec |       |  B/Call | PE[mmm]
       |           |             |       |         |  File Desc

 0.000 | 0.000065 | 124.611408 |     1 |   68.00 |Total
|-----------------------------------------------------------
|  0.000 | 0.000065 | 124.611408 |     1 |   68.00 |input
||----------------------------------------------------------
||  0.000 | 0.000065 |   2.596288 |    68 |    1.00 |pe.0
3|        |          |            |       |         | fd.6
||  0.000 |       -- |         -- |    -- |      -- |pe.22
||  0.000 |       -- |         -- |    -- |      -- |pe.5
|===========================================================
```

Table 8: I/O (Write) Statistics

```
Notes for table 8:

  Table option:
    -O write_stats
  Options implied by table option:
    -d wt,wb,wR,wr@,wC -b fi,pe=[mmm],fd

  This table shows only lines with Writes > 0.


Table 8:  File Output Stats by Filename

 Write | Write MB |  Write Rate | Writes |  Write |File Name
  Time |          |      MB/sec |        | B/Call | PE[mmm]
       |          |             |        |        |  File Desc

 0.000 | 0.002596 |  653.125440 |      7 | 388.86 |Total
|-----------------------------------------------------------
|  0.000 | 0.002001 |  573.708592 |      6 | 349.67 |stdout
||----------------------------------------------------------
||  0.000 | 0.002001 |   11.952411 |    269 |    7.80 |pe.0
3|        |          |             |        |         | fd.1
||  0.000 |       -- |          -- |     -- |      -- |pe.22
||  0.000 |       -- |          -- |     -- |      -- |pe.5
||==========================================================
|  0.000 | 0.000595 | 1221.747354 |      1 | 624.00 |stderr
||----------------------------------------------------------
||  0.000 | 0.000012 |   21.750467 |      1 |   13.00 |pe.45
3|        |          |             |        |         | fd.2
||  0.000 | 0.000012 |   25.067092 |      1 |   13.00 |pe.1
3|        |          |             |        |         | fd.2
||  0.000 | 0.000012 |   49.262647 |      1 |   13.00 |pe.0
3|        |          |             |        |         | fd.2
|===========================================================
```

Table 9: Wall Clock Time

```
Notes for table 9:

  Table option:
    -O program_time
  Options implied by table option:
    -d pt -b pe=[mmm]


Table 9:  Program Wall Clock Time

  Process |PE[mmm]
     Time |

 5.814715 |Total
|-----------------
|  5.979116 |pe.0
|  5.811115 |pe.9
|  5.650106 |pe.21
|=================
```

Call Tree Profile (Top Down)

```
Notes for table 1:

  High level option:  -O calltree
  Low level options:  -d ti%@0.05,cum_ti%,ti,tr -b exp,ct,pe=HIDE

  This table shows only lines with Time% > 0.05.

  Percentages at each level are relative
    (for absolute percentages, specify:  -s percent=a).


Table 1:  Function Calltree View

 Time % |   Cum. |      Time |     Calls |Experiment=1
        | Time % |           |           |Calltree
        |        |           |           | PE='HIDE'

 100.0% | 100.0% | 90.217759 | 637231917 |Total
|-----------------------------------------------------
| 100.0% | 100.0% | 90.175202 | 637205576 |MAIN_
||----------------------------------------------------
||  99.7% |  99.7% | 89.922750 | 637194666 |runhyd_
|||---------------------------------------------------
|||  15.4% |  15.4% | 13.864217 | 106169040 |zysweep_
||||--------------------------------------------------
||||  87.3% |  87.3% | 12.097038 | 106168320 |sppm2_
|||||-------------------------------------------------
|||||  49.4% |  49.4% |  5.980766 |  11796480 |sppm2_(exclusive)
|||||  24.1% |  73.6% |  2.920440 |  11796480 |difuze_
|||||  19.0% |  92.6% |  2.296747 |  58982400 |interf_
|||||   7.4% | 100.0% |  0.899084 |  23592960 |dintrf_
|||||=================================================
||||  12.7% | 100.0% |  1.767180 |       720 |zysweep_(exclusive)
||||==================================================
|||  15.4% |  30.8% | 13.854807 | 106169040 |xysweep_
||||--------------------------------------------------
||||  87.0% |  87.0% | 12.049373 | 106168320 |sppm2_
|||||-------------------------------------------------
|||||  49.5% |  49.5% |  5.970403 |  11796480 |sppm2_(exclusive)
|||||  24.0% |  73.6% |  2.894189 |  11796480 |difuze_
```

Callers Profile (Bottom Up)

```
 Notes for table 1:

  High level option:  -O callers
  Low level options:  -d ti%@0.05,cum_ti%,ti,tr -b exp,gr,fu,ca,pe=HIDE

  This table shows only lines with Time% > 0.05.

Table 1:  Profile by Function and Callers

 Time % |   Cum.  |      Time |     Calls |Experiment=1
        |  Time % |           |           |Group
        |         |           |           | Function
        |         |           |           |  Caller
        |         |           |           |   PE='HIDE'

 100.0% | 100.0% | 90.217759 | 637231917 |Total
|------------------------------------------------------
|  92.3% |  92.3% | 83.265853 | 637033288 |USER
||-----------------------------------------------------
||  43.1% |  43.1% | 35.864107 |  70778880 |sppm2_
|||----------------------------------------------------
|||  16.7% |  16.7% |  5.986173 |  11796480 |yxsweep_
|||        |        |           |           | runhyd_
|||        |        |           |           |  MAIN_
|||  16.7% |  33.4% |  5.980851 |  11796480 |yzsweep_
|||        |        |           |           | runhyd_
|||        |        |           |           |  MAIN_
|||  16.7% |  50.0% |  5.980766 |  11796480 |zysweep_
|||        |        |           |           | runhyd_
|||        |        |           |           |  MAIN_
|||  16.7% |  66.7% |  5.973496 |  11796480 |zzsweep_
|||        |        |           |           | runhyd_
|||        |        |           |           |  MAIN_
|||  16.7% |  83.4% |  5.972417 |  11796480 |xxsweep_
|||        |        |           |           | runhyd_
|||        |        |           |           |  MAIN_
|||  16.6% | 100.0% |  5.970403 |  11796480 |xysweep_
|||        |        |           |           | runhyd_
|||        |        |           |           |  MAIN_
|||====================================================
||  21.0% |  64.0% | 17.447719 |  70778880 |difuze_
||        |        |           |           | sppm2_
```

Callers Profile – MPI (Cont.)

```
||===================================================
||   7.7% |  99.9% |  6.906194 |    106344 |MPI
||---------------------------------------------------
|||  70.2% |  70.2% |  4.851312 |     51840 |mpi_wait_
|||--------------------------------------------------
|||   41.2% |  41.2% |  1.997854 |     17280 |zbdrys_
||||-------------------------------------------------
||||        |        |          |           |runhyd_
|||||------------------------------------------------
|||||        |        |          |           |MAIN_
||||=================================================
|||   34.3% |  75.5% |  1.664276 |     17280 |ybdrys_
|||         |        |           |           | runhyd_
|||         |        |           |           |  MAIN_
|||   24.5% | 100.0% |  1.189183 |     17280 |xbdrys_
|||         |        |           |           | runhyd_
|||         |        |           |           |  MAIN_
|||==================================================
||   29.7% |  99.9% |  2.048254 |      2232 |mpi_allreduce_
|||--------------------------------------------------
|||   96.6% |  96.6% |  1.978537 |       720 |glblmax_
|||         |        |           |           | runhyd_
|||         |        |           |           |  MAIN_
|||    3.4% | 100.0% |  0.069717 |      1512 |glbldsum_
||||-------------------------------------------------
||||   98.5% |  98.5% |  0.068700 |       792 |trace_
||||         |        |           |           | MAIN_
||||    1.5% | 100.0% |  0.001017 |       720 |runhyd_
||||         |        |           |           | MAIN_
|||==================================================
||    0.1% | 100.0% |  0.004263 |     25920 |mpi_isend_
|||--------------------------------------------------
|||   33.7% |  33.7% |  0.001436 |      8640 |xbdrys_
|||         |        |           |           | runhyd_
|||         |        |           |           |  MAIN_
|||   33.2% |  66.9% |  0.001416 |      8640 |zbdrys_
|||         |        |           |           | runhyd_
|||         |        |           |           |  MAIN_
|||   33.1% | 100.0% |  0.001410 |      8640 |ybdrys_
|||         |        |           |           | runhyd_
|||         |        |           |           |  MAIN_
|====================================================
```

Callers Profile with Line Numbers

```
Notes for table 1:

  High level option:  -O callers+src
  Low level options:  -d ti%@0.05,cum_ti%,ti,tr \
    -b exp,gr,fu,ca,pe=HIDE -s show_ca='fu,so,li' \
    -s source_limit='1'

  This table shows only lines with Time% > 0.05.

  Percentages at each level are relative
    (for absolute percentages, specify:  -s percent=a).


Table 1:  Profile by Function and Callers, with Line Numbers

 Time % |   Cum. |       Time |      Calls |Experiment=1
        | Time % |            |            |Group
        |        |            |            | Function
        |        |            |            |  Caller
        |        |            |            |   PE='HIDE'

 100.0% | 100.0% | 90.217759 | 637231917 |Total
|-----------------------------------------------------
|  92.3% |  92.3% | 83.265853 | 637033288 |USER
||----------------------------------------------------
||  43.1% |  43.1% | 35.864107 |  70778880 |sppm2_
|||---------------------------------------------------
|||  16.7% |  16.7% |  5.986173 |  11796480 |yxsweep_:sweeps.F:line.1400
|||        |        |           |           | runhyd_:main.F:line.1080
|||        |        |           |           |  MAIN_:main.F:line.226
|||  16.7% |  33.4% |  5.980851 |  11796480 |yzsweep_:sweeps.F:line.518
|||        |        |           |           | runhyd_:main.F:line.1056
|||        |        |           |           |  MAIN_:main.F:line.226
|||  16.7% |  50.0% |  5.980766 |  11796480 |zysweep_:sweeps.F:line.1106
|||        |        |           |           | runhyd_:main.F:line.1072
|||        |        |           |           |  MAIN_:main.F:line.226
|||  16.7% |  66.7% |  5.973496 |  11796480 |zzsweep_:sweeps.F:line.812
|||        |        |           |           | runhyd_:main.F:line.1064
|||        |        |           |           |  MAIN_:main.F:line.226
|||  16.7% |  83.4% |  5.972417 |  11796480 |xxsweep_:sweeps.F:line.1694
|||        |        |           |           | runhyd_:main.F:line.1088
|||        |        |           |           |  MAIN_:main.F:line.226
|||  16.6% | 100.0% |  5.970403 |  11796480 |xysweep_:sweeps.F:line.219
|||        |        |           |           | runhyd_:main.F:line.1048
|||        |        |           |           |  MAIN_:main.F:line.226
||===================================================
||  21.0% |  64.0% | 17.447719 |  70778880 |difuze_
||        |        |           |           | sppm2_:sppm.F:line.630
. . .
```

Load Balancing Function per PE

```
Notes for table 1:

  High level option:  -O load_balance_program
  Low level options:  -d ti%@0.05,cum_ti%,ti,tr -b exp,pe

  This table shows only lines with Time% > 0.05.

  Percentages at each level are relative
    (for absolute percentages, specify:  -s percent=a).


Table 1:  Load Balance across PE's

 Time % |   Cum. |     Time |  Calls |Experiment=1
        | Time % |          |        |PE

 100.0% | 100.0% | 3.798177 | 579653 |Total
|-------------------------------------------------
|   2.1% |   2.1% | 3.823080 |   7160 |pe.0
|   2.1% |   4.2% | 3.799148 |  13753 |pe.8
| ...
|   2.1% |  97.9% | 3.796151 |   7683 |pe.5
|   2.1% | 100.0% | 3.796144 |  10431 |pe.29
|=================================================
```

Table 2: LB Across PE’s by Group

```
Notes for table 2:

  High level option:  -O load_balance_group
  Low level options:  -d ti%@0.05,cum_ti%,ti,tr -b exp,gr,pe

  . . .

Table 2:  Load Balance across PE's by FunctionGroup

 Time % |   Cum. |     Time |  Calls |Experiment=1
        | Time % |          |        |Group
        |        |          |        | PE

 100.0% | 100.0% | 3.798177 | 579653 |Total
|-------------------------------------------------
|  70.9% |  70.9% | 2.692783 | 245380 |USER
||------------------------------------------------
||   2.2% |   2.2% | 2.833001 |   3076 |pe.0
|| ...
||   2.0% | 100.0% | 2.597093 |   4512 |pe.43
|=================================================
|  28.8% |  99.7% | 1.092307 | 238224 |MPI
||------------------------------------------------
||   2.3% |   2.3% | 1.188383 |   4363 |pe.43
|| ...
||   1.6% | 100.0% | 0.859333 |   2923 |pe.0
||================================================
|   0.2% |  99.8% | 0.007329 |  95597 |HEAP
||------------------------------------------------
||   2.7% |   2.7% | 0.009363 |   2482 |pe.12
|| ...
||   0.6% | 100.0% | 0.002062 |    803 |pe.0
||================================================
|   0.2% | 100.0% | 0.005758 |    452 |IO
||------------------------------------------------
||  46.6% |  46.6% | 0.128685 |    358 |pe.0
|| ...
||   0.6% | 100.0% | 0.001644 |      2 |pe.29
|=================================================
```

Table 3: LB Across PE’s by Function

```
Notes for table 3:

  High level option:  -O load_balance_function
  Low level options:  -d ti%@0.05,cum_ti%,ti,tr -b exp,gr,fu,pe

  This table shows only lines with Time% > 0.05.

  Percentages at each level are relative
    (for absolute percentages, specify:  -s percent=a).


Table 3:  Load Balance across PE's by Function

 Time % |   Cum. |     Time |  Calls |Experiment=1
        | Time % |          |        |Group
        |        |          |        | Function
        |        |          |        |  PE

 100.0% | 100.0% | 3.798177 | 579653 |Total
|-------------------------------------------------
|  70.9% |  70.9% | 2.692783 | 245380 |USER
||------------------------------------------------
||  97.1% |  97.1% | 2.615916 |    576 |sweep_
|||-----------------------------------------------
|||   2.2% |   2.2% | 2.753279 |     12 |pe.0
|||   2.1% |   4.3% | 2.654725 |     12 |pe.5
||| . . .
|||   2.0% |  98.0% | 2.525587 |     12 |pe.43
|||   2.0% | 100.0% | 2.523325 |     12 |pe.37
|||===============================================
 . . .
|||===============================================
||   0.4% |  99.2% | 0.010300 | 118080 |snd_real_
|||-----------------------------------------------
|||   2.4% |   2.4% | 0.011699 |   2880 |pe.26
|||   2.3% |   4.7% | 0.011475 |   2880 |pe.27
||| . . .
|||   1.5% |  98.6% | 0.007266 |   1440 |pe.0
|||   1.4% | 100.0% | 0.006907 |   1440 |pe.5
|||===============================================
```

Table 3 (Cont.)

```
||================================================
|  28.8% |  99.7% | 1.092307 | 238224 |MPI
||------------------------------------------------
||  76.1% |  76.1% | 0.831311 | 118080 |mpi_recv_
|||-----------------------------------------------
|||    2.7% |    2.7% | 1.066077 |    1440 |pe.47
|||    2.6% |    5.3% | 1.034307 |    2160 |pe.41
|||  . . .
|||    1.8% |   98.6% | 0.700970 |    2160 |pe.1
|||    1.4% |  100.0% | 0.573420 |    1440 |pe.0
|||===============================================
. . .
||================================================
|   0.2% |  99.8% | 0.007329 |  95597 |HEAP
||------------------------------------------------
||  61.1% |  61.1% | 0.004481 |  47861 |malloc
|||-----------------------------------------------
|||    2.7% |    2.7% | 0.005884 |    1242 |pe.12
|||    2.6% |    5.4% | 0.005658 |    1226 |pe.19
|||  . . .
|||    1.3% |   99.5% | 0.002827 |     618 |pe.34
|||    0.5% |  100.0% | 0.001164 |     417 |pe.0
|||===============================================
||  38.9% | 100.0% | 0.002848 |  47735 |free
|||-----------------------------------------------
|||    2.7% |    2.7% | 0.003748 |    1422 |pe.37
|||    2.7% |    5.5% | 0.003706 |    1469 |pe.43
|||  . . .
|||    1.4% |   99.3% | 0.001867 |     616 |pe.34
|||    0.7% |  100.0% | 0.000896 |     385 |pe.0
||================================================
|   0.2% | 100.0% | 0.005758 |    452 |IO
||------------------------------------------------
||  81.3% |  81.3% | 0.004679 |    309 |fwrite
|||-----------------------------------------------
|||   34.3% |   34.3% | 0.077141 |     262 |pe.0
|||    2.1% |   36.4% | 0.004615 |       1 |pe.8
|||  . . .
```

Load Balance: Max, Median, Min

```
Notes for table 1:

  High level option:  -O load_balance_program
  Low level options:  -d ti%@0.05,cum_ti%,ti,tr -b exp,pe=[mmm]

  This table shows only lines with Time% > 0.05.

  Percentages at each level are relative
    (for absolute percentages, specify:  -s percent=a).


Table 1:  Load Balance across PE's

 Time % |   Cum. |      Time |  Calls |Experiment=1
        | Time % |           |        |PE[mmm]

 100.0% | 100.0% | 3.798177 | 579653 |Total
|-------------------------------------------------
|   2.1% |   2.1% | 3.823080 |   7160 |pe.0
|   2.1% |  52.1% | 3.797671 |  10695 |pe.3
|   2.1% | 100.0% | 3.796144 |  10431 |pe.29
|=================================================
```

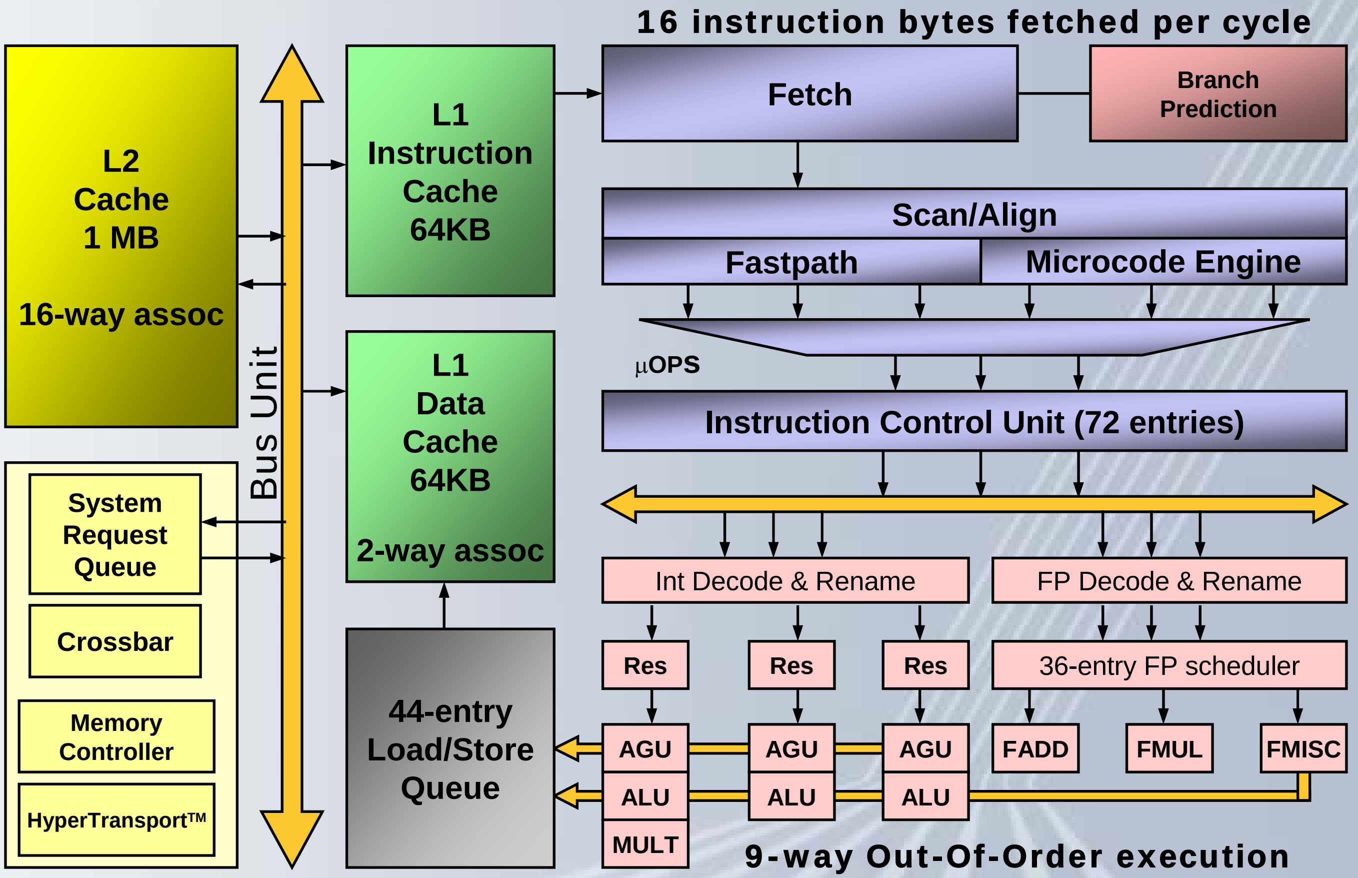
LB [MMM] Table 2

```
Notes for table 2:

  High level option:  -O load_balance_group
  Low level options:  -d ti%@0.05,cum_ti%,ti,tr \
    -b exp,gr,pe=[mmm]

  This table shows only lines with Time% > 0.05.

  Percentages at each level are relative
    (for absolute percentages, specify:  -s percent=a).


Table 2:  Load Balance across PE's by FunctionGroup

 Time % |   Cum. |     Time |   Calls |Experiment=1
        | Time % |          |         |Group
        |        |          |         | PE[mmm]

 100.0% | 100.0% | 3.798177 | 579653 |Total
|------------------------------------------------
|  70.9% |  70.9% | 2.692783 | 245380 |USER
||-----------------------------------------------
||    2.2% |    2.2% | 2.833001 |    3076 |pe.0
||    2.1% |   52.8% | 2.717019 |    4512 |pe.12
||    2.0% |  100.0% | 2.597093 |    4512 |pe.43
||===============================================
|  28.8% |  99.7% | 1.092307 | 238224 |MPI
||-----------------------------------------------
||    2.3% |    2.3% | 1.188383 |    4363 |pe.43
||    2.0% |   53.7% | 1.069314 |    5803 |pe.7
||    1.6% |  100.0% | 0.859333 |    2923 |pe.0
||===============================================
|   0.2% |  99.8% | 0.007329 |  95597 |HEAP
||-----------------------------------------------
||    2.7% |    2.7% | 0.009363 |    2482 |pe.12
||    2.2% |   59.6% | 0.007614 |    2192 |pe.40
||    0.6% |  100.0% | 0.002062 |     803 |pe.0
||===============================================
|   0.2% | 100.0% | 0.005758 |    452 |IO
||-----------------------------------------------
||   46.6% |   46.6% | 0.128685 |     358 |pe.0
||    1.1% |   80.3% | 0.003144 |       2 |pe.47
||    0.6% |  100.0% | 0.001644 |       2 |pe.29
|================================================
```

LB [MMM] Table 3

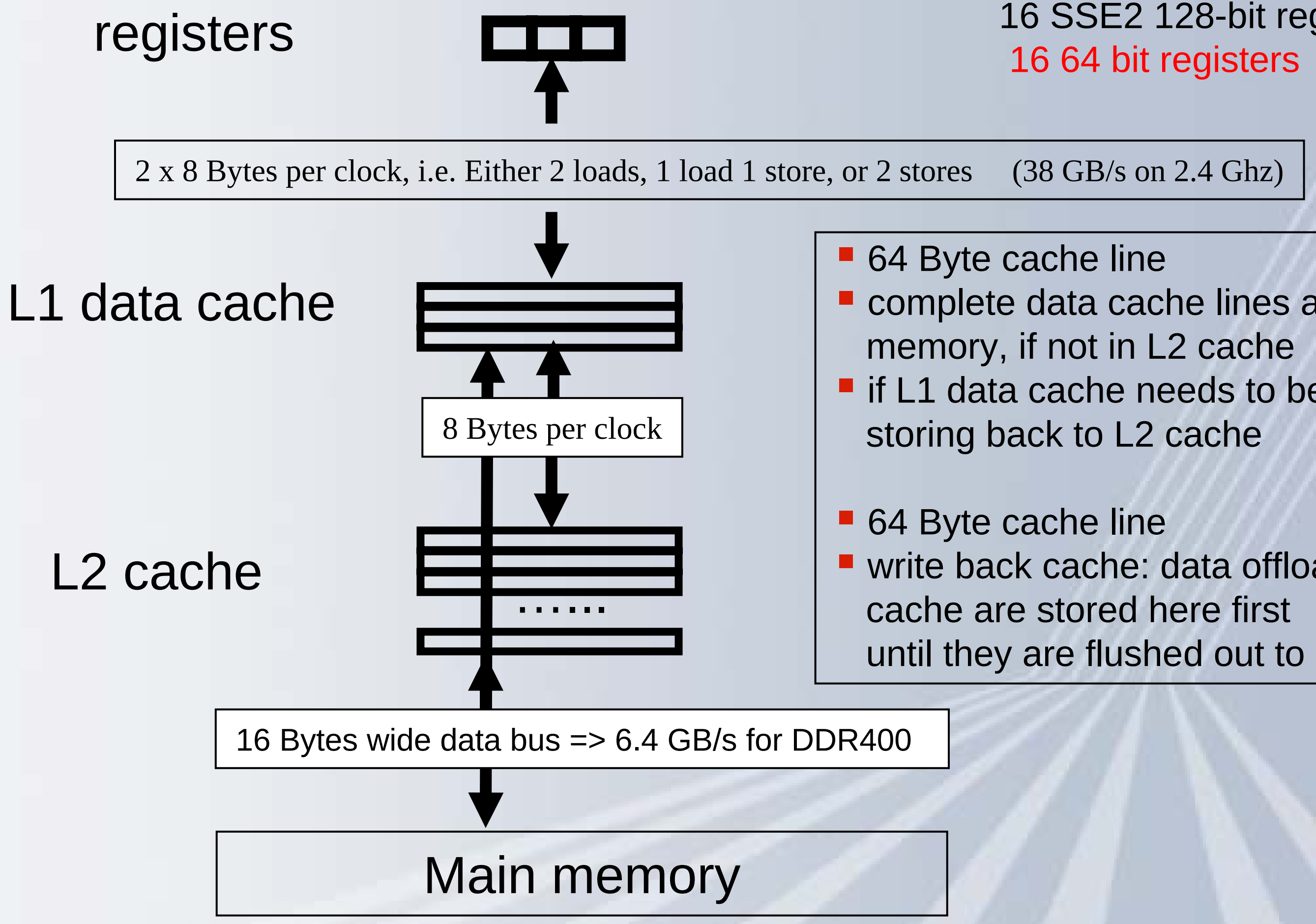
```
Notes for table 3:

  High level option:  -O load_balance_function
  Low level options:  -d ti%@0.05,cum_ti%,ti,tr \
    -b exp,gr,fu,pe=[mmm]

  This table shows only lines with Time% > 0.05.

  Percentages at each level are relative
    (for absolute percentages, specify:  -s percent=a).


Table 3:  Load Balance across PE's by Function

 Time % |   Cum. |     Time |  Calls |Experiment=1
        | Time % |          |        |Group
        |        |          |        | Function
        |        |          |        |  PE[mmm]

 100.0% | 100.0% | 3.798177 | 579653 |Total
|-----------------------------------------------
|  70.9% |  70.9% | 2.692783 | 245380 |USER
||  97.1% |  97.1% | 2.615916 |    576 |sweep_
||----------------------------------------------
|||   2.2% |    2.2% | 2.753279 |     12 |pe.0
|||   2.1% |   52.8% | 2.638898 |     12 |pe.16
|||   2.0% |  100.0% | 2.523325 |     12 |pe.37
|||=============================================
. . .
||==============================================
|  28.8% |  99.7% | 1.092307 | 238224 |MPI
||----------------------------------------------
||  76.1% |  76.1% | 0.831311 | 118080 |mpi_recv_
||----------------------------------------------
|||   2.7% |    2.7% | 1.066077 |   1440 |pe.47
|||   2.0% |   56.9% | 0.801256 |   2880 |pe.21
|||   1.4% |  100.0% | 0.573420 |   1440 |pe.0
|||=============================================
 . .
||  0.2% |  99.8% | 0.007329 |  95597 |HEAP
||----------------------------------------------
. . .
||==============================================
|  0.2% | 100.0% | 0.005758 |    452 |IO
. . .
```

AMD Opteron Processor



- 36 entry FPU instruction scheduler
- 64-bit/80-bit FP Realized throughput (1 Mul + 1 Add)/cycle: 1.9 FLOPs/cycle
- 32-bit FP Realized throughput (2 Mul + 2 Add)/cycle: 3.4+ FLOPs/cycle

Simplified memory hierachy on the AMD Opteron



registers

2 x 8 Bytes per clock, i.e. Either 2 loads, 1 load 1 store, or 2 stores (38 GB/s on 2.4 Ghz)

L1 data cache

8 Bytes per clock

L2 cache

……

16 Bytes wide data bus => 6.4 GB/s for DDR400

Main memory

16 SSE2 128-bit registers
16 64 bit registers

- 64 Byte cache line
- complete data cache lines are loaded from main memory, if not in L2 cache
- if L1 data cache needs to be refilled, then storing back to L2 cache

- 64 Byte cache line
- write back cache: data offloaded from L1 data cache are stored here first until they are flushed out to main memory

Hardware Performance Counters

- AMD Opteron Hardware Performance Counters
  - Four 48-bit performance counters.
    - Each counter can monitor a single event
      - Count specific processor events
        - » the processor increments the counter when it detects an occurrence of the event
        - » (e.g., cache misses)
      - Duration of events
        - » the processor counts the number of processor clocks it takes to complete an event
        - » (e.g., the number of clocks it takes to return data from memory after a cache miss)
  - Time Stamp Counters (TSC)
    - Cycles (user time)

PAPI Predefined Events

- Common set of events deemed relevant and useful for application performance tuning
  - Accesses to the memory hierarchy, cycle and instruction counts, functional units, pipeline status, etc.
  - The "papi_avail" utility shows which predefined events are available on the system

- PAPI also provides access to native events
  - The "papi_native_avail" utility list all AMD native events available on the system

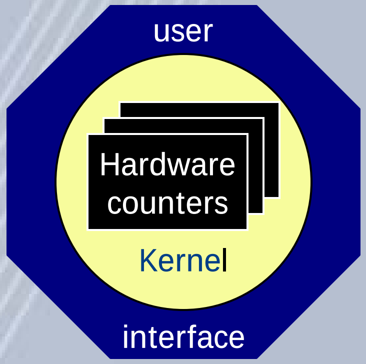
- Information on PAPI and AMD native events
  - pat_help counters
  - man papi_counters
  - For more information on AMD counters:
    - http://www.amd.com/us-en/assets/content_type/white_papers_and_tech_docs/26049.PDF

Hardware Counters Selection

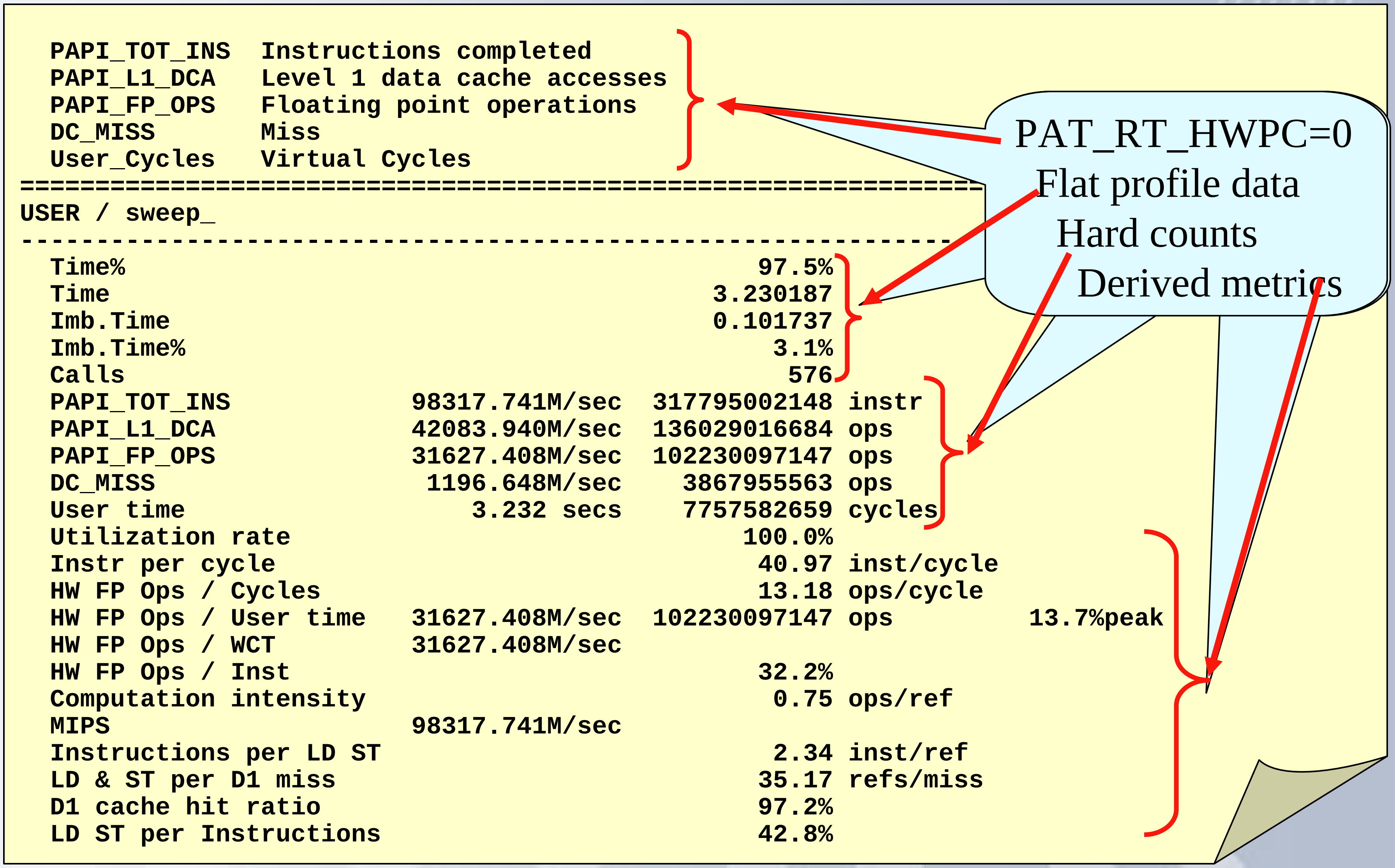
- PAT_RT_HWPC <set number> | <event list>
  - Specifies hardware counter events to be monitored
    - A set number can be used to select a group of predefined hardware counters events (recommended)
      - CrayPat provides 10 group on the Cray XT systems
        - with CrayPat 3.2.1
    - Alternatively a list of hardware performance counter event names can be used
      - Maximum of 4 events
    - Both formats can be specified at the same time, with later definitions overriding previous definitions
    - Hardware counter events are not collected by default
    - Hardware counters collection is not supported with sampling on systems running Catamount on the compute nodes

Accuracy Issues

- Granularity of the measured code
  - If not sufficiently large enough, overhead of the counter interfaces may dominate
- Pay attention to what is not measured:
  - Out-of-order processors
  - Speculation
  - Lack of standard on what is counted
    - Microbenchmarks can help determine accuracy of the hardware counters
- For more information on AMD counters:
  - architecture manuals:
    - http://www.amd.com/us-en/assets/content_type/white_papers_and_tech_docs/26049.PDF



Hardware Performance Counters



```
  PAPI_TOT_INS  Instructions completed
  PAPI_L1_DCA   Level 1 data cache accesses
  PAPI_FP_OPS   Floating point operations
  DC_MISS       Miss
  User_Cycles   Virtual Cycles
================================================================
USER / sweep_
--------------------------------------------------------------
  Time%                                          97.5%
  Time                                        3.230187
  Imb.Time                                    0.101737
  Imb.Time%                                       3.1%
  Calls                                            576
  PAPI_TOT_INS            98317.741M/sec  317795002148 instr
  PAPI_L1_DCA             42083.940M/sec  136029016684 ops
  PAPI_FP_OPS             31627.408M/sec  102230097147 ops
  DC_MISS                  1196.648M/sec    3867955563 ops
  User time                   3.232 secs    7757582659 cycles
  Utilization rate                              100.0%
  Instr per cycle                                40.97 inst/cycle
  HW FP Ops / Cycles                             13.18 ops/cycle
  HW FP Ops / User time   31627.408M/sec  102230097147 ops         13.7%peak
  HW FP Ops / WCT         31627.408M/sec
  HW FP Ops / Inst                               32.2%
  Computation intensity                           0.75 ops/ref
  MIPS                    98317.741M/sec
  Instructions per LD ST                          2.34 inst/ref
  LD & ST per D1 miss                            35.17 refs/miss
  D1 cache hit ratio                             97.2%
  LD ST per Instructions                         42.8%
```

PAT_RT_HWPC=0

Flat profile data

Hard counts

Derived metrics

Hardware Performance Counters

```
  PAPI_TLB_DM  Data translation lookaside buffer misses
  PAPI_L1_DCA  Level 1 data cache accesses
  PAPI_FP_OPS  Floating point operations
  DC_MISS      Data Cache Miss
  User_Cycles  Virtual Cycles
========================================================================
USER / sweep_
------------------------------------------------------------------------
  Time%                                         97.5%
  Time                                       3.230138
  Imb.Time                                   0.102759
  Imb.Time%                                      3.1%
  Calls                                           576
  PAPI_TLB_DM               23.637M/sec      76402282 misses
  PAPI_L1_DCA            42084.031M/sec  136028363961 ops
  PAPI_FP_OPS            31627.272M/sec  102228944600 ops
  DC_MISS                 1196.211M/sec    3866518161 ops
  User time                  3.232 secs    7757528615 cycles
  Utilization rate                             100.0%
  HW FP Ops / Cycles                            13.18 ops/cycle
  HW FP Ops / User time  31627.272M/sec  102228944600 ops       13.7%peak
  HW FP Ops / WCT        31627.272M/sec
  Computation intensity                          0.75 ops/ref
  LD & ST per TLB miss                        1780.42 refs/miss
  LD & ST per D1 miss                           35.18 refs/miss
  D1 cache hit ratio                            97.2%
  % TLB misses / cycle                           0.0%
```

PAT_RT_HWPC=2 (Cache Info)

```
  PAPI_L1_DCA          Level 1 data cache accesses
  DC_L2_REFILL_MOESI   Refill from L2. Cache bits: Modified Owner Exclusive Shared Invalid
  DC_SYS_REFILL_MOESI  Refill from system. Cache bits: Modified Owner Exclusive Shared Invalid
  BU_L2_REQ_DC         Internal L2 request - DC fill
  User_Cycles          Virtual Cycles
========================================================================
USER / sweep_
------------------------------------------------------------------------
  Time%                                             97.6%
  Time                                           3.229766
  Imb.Time                                       0.102401
  Imb.Time%                                          3.1%
  Calls                                               576
  PAPI_L1_DCA             42089.216M/sec   136029595956 ops
  DC_L2_REFILL_MOESI       1194.502M/sec     3860550881 ops
  DC_SYS_REFILL_MOESI       275.938M/sec      891814098 ops
  BU_L2_REQ_DC             1228.504M/sec     3970443642 req
  User time                   3.232 secs     7756643255 cycles
  Utilization rate                                 100.0%
  L1 Data cache misses     1470.440M/sec     4752364979 misses
  LD & ST per D1 miss                               28.62 refs/miss
  D1 cache hit ratio                                96.5%
  LD & ST per D2 miss                              152.53 refs/miss
  D2 cache hit ratio                                77.5%
  L2 cache hit ratio                                81.2%
  Total cache hit ratio                             99.3%
  Effective Reuse                                    2.38 refs/byte
  Memory to D1 refill       275.938M/sec      891814098 lines
  Memory to D1 bandwidth  16841.928MB/sec   57076102272 bytes
  L2 to Dcache bandwidth  72906.587MB/sec  247075256384 bytes
========================================================================
```

PAT_RT_HWPC=3 (Bandwidth)

```
  DC_L2_REFILL_MOESI   Refill from L2. Cache bits: Modified Owner Exclusive Shared Invalid
  DC_SYS_REFILL_MOESI  Refill from system. Cache bits: Modified Owner Exclusive Shared Invalid
  DC_COPYBACK_MOESI    Copyback. Cache bits: Modified Owner Exclusive Shared Invalid
  SI_QUAD_WRITE        Quadwords Written to System: Quadword write transfer
  User_Cycles          Virtual Cycles
========================================================================
USER / sweep_
------------------------------------------------------------------------
  Time%                                           97.5%
  Time                                         3.229972
  Imb.Time                                     0.102654
  Imb.Time%                                        3.1%
  Calls                                             576
  DC_L2_REFILL_MOESI       1196.069M/sec     3865961067 ops
  DC_SYS_REFILL_MOESI       275.951M/sec      891933730 ops
  DC_COPYBACK_MOESI        1472.059M/sec     4758021002 ops
  SI_QUAD_WRITE            2130.392M/sec     6885899519 ops
  User time                   3.232 secs     7757332017 cycles
  Utilization rate                               100.0%
  L1 Data cache misses     1472.020M/sec     4757894797 misses
  L2 cache hit ratio                              81.3%
  Memory to D1 refill       275.951M/sec      891933730 lines
  Memory to D1 bandwidth  16842.692MB/sec   57083758720 bytes
  L2 to Dcache bandwidth  73002.276MB/sec  247421508288 bytes
  Dcache to L2 bandwidth  89847.351MB/sec  304513344128 bytes
  L2 to Memory bandwidth  16253.602MB/sec   55087196152 bytes
========================================================================
```

PAT_RT_HWPC=4 (HT Data Transfers)

```
  SI_QUAD_WRITE     Quadwords Written to System: Quadword write transfer
  HT_LL_MEM_XFR     HyperTransport data transfer from local memory to local memory
  HT_LL_IO_XFR      HyperTransport data transfer from local memory to local IO
  HT_LL_IO_MEM_XFR  HyperTransport data transfer from local IO to local memory
  User_Cycles       Virtual Cycles
========================================================================
USER / sweep_
------------------------------------------------------------------------
  Time%                                                97.6%
  Time                                              3.229909
  Imb.Time                                          0.102472
  Imb.Time%                                             3.1%
  Calls                                                  576
  SI_QUAD_WRITE               2130.809M/sec    6886944175 ops
  HT_LL_MEM_XFR               2229.315M/sec    7205322525 ops
  HT_LL_IO_XFR                   0.316M/sec       1022602 ops
  HT_LL_IO_MEM_XFR               4.333M/sec      14003499 ops
  User time                      3.232 secs    7756989744 cycles
  Utilization rate                                    100.0%
  L2 to Memory bandwidth  16256.785MB/sec  55095553400 bytes
```

PAT_RT_HWPC=5 (FP & Vectorization)

```
  PAPI_FML_INS            Floating point multiply instructions
  PAPI_FAD_INS            Floating point add instructions
  FR_FPU_SSE_SSE2_PACKED  Retired FPU instructions - Combined packed SSE and SSE2 instructions
  FR_FPU_SSE_SSE2_SCALAR  Retired FPU instructions - Combined scalar SSE and SSE2 instructions
  User_Cycles             Virtual Cycles
========================================================================
USER / sweep_
------------------------------------------------------------------------
  Time%                                          97.5%
  Time                                        3.230243
  Imb.Time                                    0.102413
  Imb.Time%                                       3.1%
  Calls                                            576
  PAPI_FML_INS            14751.121M/sec   47682418308 instr
  PAPI_FAD_INS            16874.595M/sec   54546466150 instr
  FR_FPU_SSE_SSE2_PACKED                             0 instr
  FR_FPU_SSE_SSE2_SCALAR  43000.358M/sec  138996966424 instr
  User time                   3.232 secs    7757905716 cycles
  Utilization rate                              100.0%
  HW FP Ops / Cycles                             13.18 ops/cycle
  HW FP Ops / User time   31625.716M/sec  102228884458 ops        13.7%peak
  HW FP Ops / WCT         31625.716M/sec
  FP Multiply / FP Ops                           46.6%
  FP Add / FP Ops                                53.4%
```

When compiled with fastsse:

```
========================================================================
USER / sweep_
------------------------------------------------------------------------
  Time%                                          97.0%
  Time                                        2.577571
  Imb.Time                                    0.101843
  Imb.Time%                                       3.9%
  Calls                                            576
  PAPI_FML_INS            16061.952M/sec   41438628312 instr
  PAPI_FAD_INS            18681.139M/sec   48195934483 instr
  FR_FPU_SSE_SSE2_PACKED   9315.154M/sec   24032397312 instr
  FR_FPU_SSE_SSE2_SCALAR  39220.314M/sec  101185461233 instr
  User time                   2.580 secs    6191819596 cycles
  Utilization rate                              100.0%
  HW FP Ops / Cycles                             14.48 ops/cycle
  HW FP Ops / User time   34743.091M/sec   89634562795 ops        15.1%peak
  HW FP Ops / WCT         34743.091M/sec
  FP Multiply / FP Ops                           46.2%
  FP Add / FP Ops                                53.8%
```

PAT_RT_HWPC=6 (Stalls / Resources Idle)

```
  PAPI_FPU_IDL    Cycles floating point units are idle
  PAPI_STL_ICY    Cycles with no instruction issue
  PAPI_RES_STL    Cycles stalled on any resource
  IC_FETCH_STALL  Instruction fetch stall
  User_Cycles     Virtual Cycles
===============================================================
USER / sweep_
---------------------------------------------------------------
  Time%                                             96.9%
  Time                                           2.619334
  Imb.Time                                       0.141884
  Imb.Time%                                          5.2%
  Calls                                               576
  PAPI_FPU_IDL          0.209 secs      500570926.75 cycles
  PAPI_STL_ICY          0.029 secs  70217803.2916667 cycles
  PAPI_RES_STL          1.725 secs     4140098264.75 cycles
  IC_FETCH_STALL        1.930 secs  4631824703.22917 cycles
  User time             2.622 secs  6292743345.83333 cycles
  Utilization rate                             100.0%
  Total time stalled    1.725 secs     4140098264.75 cycles  65.8%
  Time I Fetch Stalled  1.930 secs  4631824703.22917 cycles  73.6%
  Avg Time FPUs idle    0.104 secs     250285463.375 cycles   4.0%
  Time Decoder empty    0.029 secs  70217803.2916667 cycles   1.1%
===============================================================
```

PAT_RT_HWPC=7 (Stalls/ Resources Full)

```
  FR_DECODER_EMPTY             Nothing to dispatch - decoder empty
  FR_DISPATCH_STALLS           Dispatch stalls - D2h or DAh combined
  FR_DISPATCH_STALLS_FULL_FPU  Dispatch stall when FPU is full
  FR_DISPATCH_STALLS_FULL_LS   Dispatch stall when LS is full
  User_Cycles                  Virtual Cycles
========================================================================
USER / sweep_
------------------------------------------------------------------------
  Time%                                                 97.0%
  Time                                               2.618878
  Imb.Time                                           0.142062
  Imb.Time%                                              5.3%
  Calls                                                   576
  FR_DECODER_EMPTY             1281.985M/sec        3360773456 ops
  FR_DISPATCH_STALLS              1.725 secs    4139586865.875 cycles
  FR_DISPATCH_STALLS_FULL_FPU     1.118 secs  2683961106.04167 cycles
  FR_DISPATCH_STALLS_FULL_LS      0.438 secs  1050422214.02083 cycles
  User time                       2.622 secs   6291691461.4375 cycles
  Utilization rate                                     100.0%
  Total time stalled              1.725 secs    4139586865.875 cycles  65.8%
  Avg Time FPUs stalled           0.559 secs  1341980553.02083 cycles  21.3%
  Avg Time LSs stalled            0.219 secs  525211107.010417 cycles   8.3%
  Time Decoder empty              1.400 secs        3360773456 cycles  53.4%
========================================================================
```

PAT_RT_HWPC Other Sets

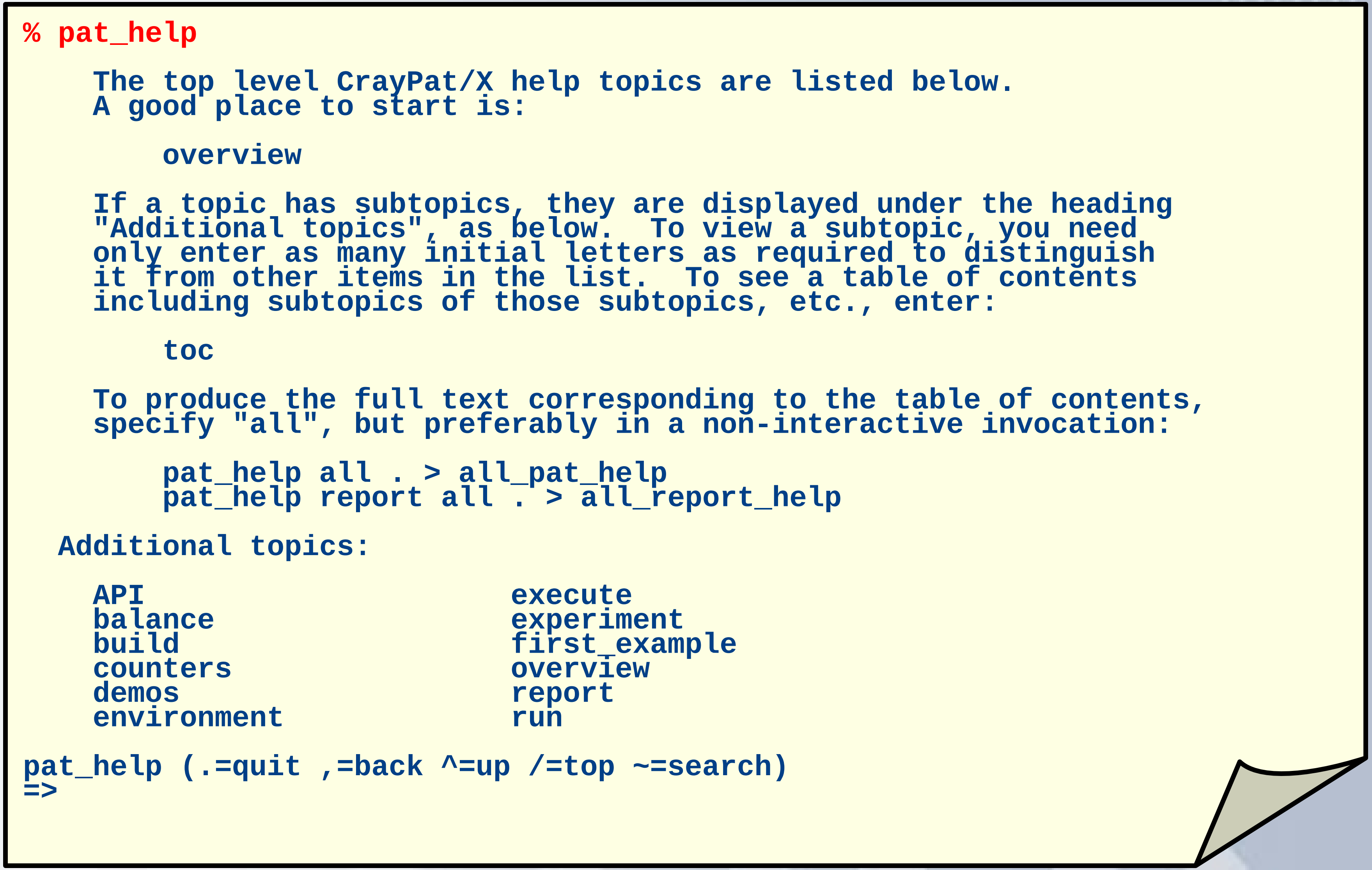
```
-------------------------------------------------------------------
Set 8: Branches
  PAPI_BR_TKN   Conditional branch instructions taken
  PAPI_BR_MSP   Conditional branch instructions mispredicted
  PAPI_TOT_INS  Instructions completed
  IC_MISS       IC Miss
  User_Cycles   Virtual Cycles


-------------------------------------------------------------------

Set 9: Instructions
  PAPI_L2_ICM   Level 2 instruction cache misses
  PAPI_L1_ICA   Level 1 instruction cache accesses
  IC_MISS       IC Miss
  IC_L2_REFILL  Refill from L2
  User_Cycles   Virtual Cycles
```

pat_help & Documentation

- The pat_help utility is an interactive viewer used to access information about and examples of using CrayPat
  - pat_help [topic [subtopic...]]

- See also man pages:
  - craypat
  - pat
  - pat_build
  - pat_report
  - pat_help
  - hwpc
  - papi_counters

pat_help Example



```
% pat_help

    The top level CrayPat/X help topics are listed below.
    A good place to start is:

        overview

    If a topic has subtopics, they are displayed under the heading
    "Additional topics", as below.  To view a subtopic, you need
    only enter as many initial letters as required to distinguish
    it from other items in the list.  To see a table of contents
    including subtopics of those subtopics, etc., enter:

        toc

    To produce the full text corresponding to the table of contents,
    specify "all", but preferably in a non-interactive invocation:

        pat_help all . > all_pat_help
        pat_help report all . > all_report_help

  Additional topics:

    API                         execute
    balance                     experiment
    build                       first_example
    counters                    overview
    demos                       report
    environment                 run

pat_help (.=quit ,=back ^=up /=top ~=search)
=>
```

Performance Measurement and Visualization on the Cray XT4

Questions / Comments

Thank You!

NERSC
September 18-20, 2007