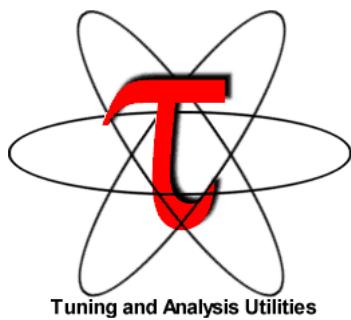


TAU Performance System®

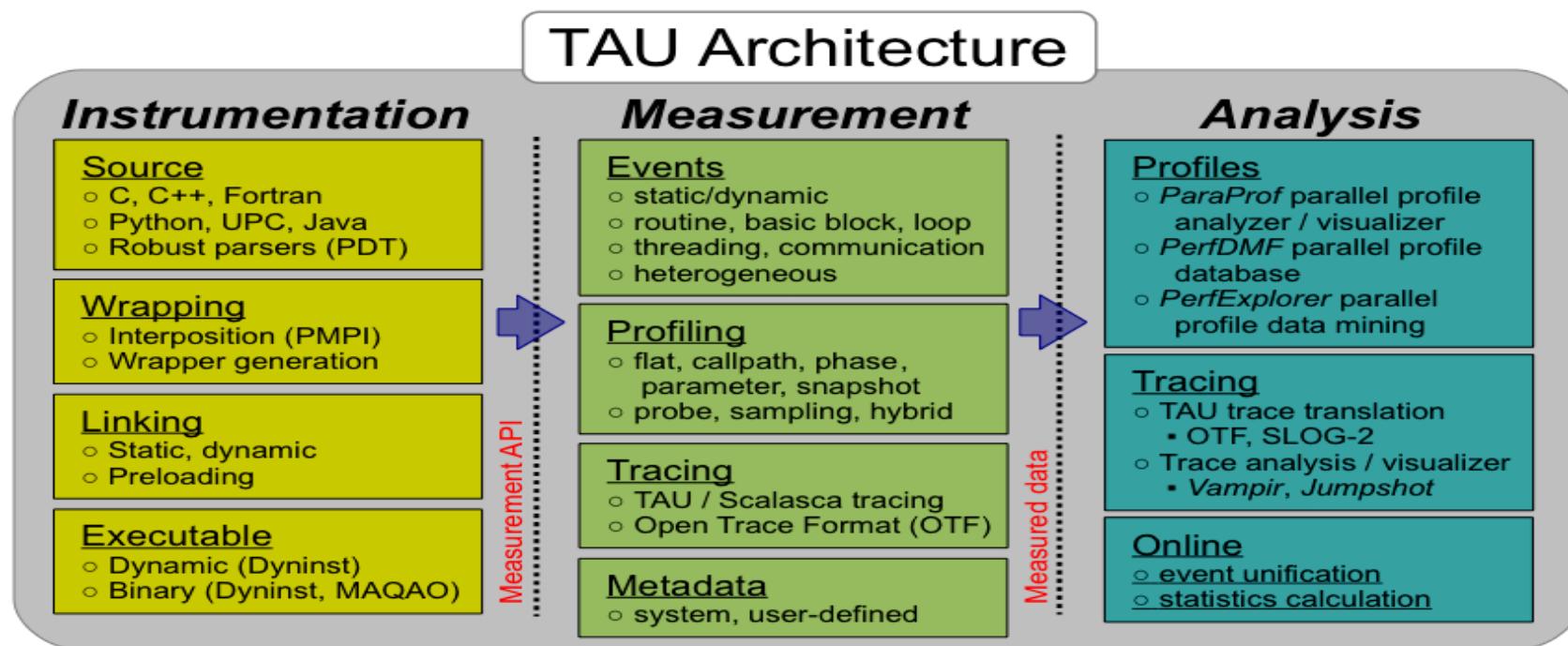
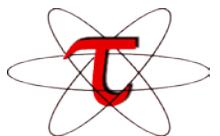


Sameer Shende
sameer@cs.uoregon.edu
Performance Research Laboratory
University of Oregon
<http://tau.uoregon.edu>



TAU Performance System®

- Parallel performance framework and toolkit
 - Supports all HPC platforms, compilers, runtime system
 - Provides portable instrumentation, measurement, analysis



TAU Performance System

- Instrumentation
 - Fortran, C++, C, UPC, Java, Python, Chapel
 - Automatic instrumentation
- Measurement and analysis support
 - MPI, OpenSHMEM, ARMCI, PGAS, DMAPP
 - pthreads, OpenMP, OMPT interface, hybrid, other thread models
 - GPU, CUDA, OpenCL, OpenACC
 - Parallel profiling and tracing
 - Use of Score-P for native OTF2 and CUBEX generation
 - Efficient callpath profiles and trace generation using Score-P
- Analysis
 - Parallel profile analysis (ParaProf), data mining (PerfExplorer)
 - Performance database technology (TAUdb)
 - 3D profile browser

TAU Performance System

- Instrumentation
 - Fortran, C++, C, UPC, Java, Python, Chapel
 - Automatic instrumentation
- Measurement and analysis support
 - MPI, OpenSHMEM, ARMCI, PGAS, DMAPP
 - pthreads, OpenMP, OMPT interface, hybrid, other thread models
 - GPU, CUDA, OpenCL, OpenACC
 - Parallel profiling and tracing
 - Use of Score-P for native OTF2 and CUBEX generation
 - Efficient callpath profiles and trace generation using Score-P
- Analysis
 - Parallel profile analysis (ParaProf), data mining (PerfExplorer)
 - Performance database technology (TAUdb)
 - 3D profile browser

TAU Performance System

- TAU supports both sampling and direct instrumentation
- Memory debugging as well as I/O performance evaluation
- Profiling as well as tracing
- Interfaces with Score-P for more efficient measurements
- TAU's instrumentation covers:
 - Runtime library interposition (`tau_exec`)
 - Compiler-based instrumentation
 - PDT based Source level instrumentation: routine & loop
 - Event based sampling (`TAU_SAMPLING=1` or `tau_exec -ebs`)
 - Callstack unwinding with sampling (`TAU_EBS_UNWIND=1`)
 - OpenMP Tools Interface (OMPT, `tau_exec -T ompt`)
 - CUDA CUPTI, OpenCL (`tau_exec -T cupti -cupti`)

Application Performance Engineering using TAU

- How much time is spent in each application routine and outer *loops*? Within loops, what is the contribution of each *statement*? What is the time spent in OpenMP loops?
- How many instructions are executed in these code regions?
Floating point, Level 1 and 2 *data cache misses*, hits, branches taken? What is the extent of vectorization for loops on Intel MIC?
- What is the memory usage of the code? When and where is memory allocated/de-allocated? Are there any memory leaks? What is the memory footprint of the application? What is the memory high water mark?
- How much energy does the application use in Joules? What is the peak power usage?
- What are the I/O characteristics of the code? What is the peak read and write *bandwidth* of individual calls, total volume?
- What is the contribution of each *phase* of the program? What is the time wasted/spent waiting for collectives, and I/O operations in Initialization, Computation, I/O phases?
- How does the application *scale*? What is the efficiency, runtime breakdown of performance across different core counts?

Using TAU

- TAU supports several compilers, measurement, and thread options

Intel compilers, profiling with hardware counters using PAPI, MPI library, CUDA...

Each measurement configuration of TAU corresponds to a unique stub makefile (configuration file) and library that is generated when you configure it

- To instrument source code automatically using PDT

Choose an appropriate TAU stub makefile in <arch>/lib:

```
% module load tau
% export TAU_MAKEFILE=$TAU/Makefile.tau-icpc-papi-mpi-pdt
% export TAU_OPTIONS=' -optVerbose ...' (see tau_compiler.sh )
```

Use tau_f90.sh, tau_cxx.sh, tau_upc.sh, or tau_cc.sh as F90, C++, UPC, or C compilers respectively:

```
% mpif90 foo.f90      changes to
% tau_f90.sh foo.f90
```

- Set runtime environment variables, execute application and analyze performance data:

```
% pprof  (for text based profile display)
% paraprof (for GUI)
```

Automatic Source Instrumentation using PDT

```
% module load tau; ls $TAU/Makefile.*  
Makefile.tau-icpc-papi-mpi-pdt  
Makefile.tau-icpc-papi-mpi-pthread-pdt  
Makefile.tau-icpc-papi-ompt-mpi-pdt-openmp  
Makefile.tau-gnu-papi-mpi-pdt  
Makefile.tau-papi-mpi-pdt-openmp-o pari-scorep  
Makefile.tau-papi-mpi-pdt-scorep  
Makefile.tau-papi-mpi-pthread-pdt  
Makefile.tau-papi-pthread-pdt  
Makefile.tau-papi-shmem-mpi-pdt
```

- For an MPI+F90 application with Intel MPI, you may choose

Makefile.tau-mpi-pdt

- Supports MPI instrumentation & PDT for automatic source instrumentation

```
% export TAU_MAKEFILE=$TAU/Makefile.tau-icpc-papi-ompt-mpi-pdt-openmp
```

```
% tau_f90.sh matmult.f90 -o matmult
```

```
% mpirun -np 256 ./matmult
```

```
% paraprof
```

Installing and Configuring TAU

■ Installing PDT:

- wget http://tau.uoregon.edu/pdt_lite.tgz
- ./configure –prefix=<dir>; make ; make install

■ Installing TAU:

- wget http://tau.uoregon.edu/tau.tgz
- ./configure –arch=x86_64 -bfd=download -pdt=<dir> -papi=<dir> ...
- For MIC:
- ./configure –arch=mic_linux –pdt=<dir> -pdt_c++=g++ -papi=dir ...
- make install

■ Using TAU:

- export TAU_MAKEFILE=<taudir>/x86_64/lib/Makefile.tau-<TAGS>
- make CC=tau_cc.sh CXX=tau_cxx.sh F90=tau_f90.sh

Compile-Time Options

Optional parameters for the TAU_OPTIONS environment variable:

% tau_compiler.sh	
-optVerbose	Turn on verbose debugging messages
-optComInst	Use compiler based instrumentation
-optNoComInst	Do not revert to compiler instrumentation if source instrumentation fails.
-optTrackIO	Wrap POSIX I/O call and calculates vol/bw of I/O operations (configure TAU with <i>–iowrapper</i>)
-optTrackGOMP	Enable tracking GNU OpenMP runtime layer (used without <i>–opari</i>)
-optMemDbg	Enable runtime bounds checking (see TAU_MEMDBG_* env vars)
-optKeepFiles	Does not remove intermediate .pdb and .inst.* files
-optPreProcess	Preprocess sources (OpenMP, Fortran) before instrumentation
-optTauSelectFile=" <i><file></i> "	Specify selective instrumentation file for <i>tau_instrumentor</i>
-optTauWrapFile=" <i><file></i> "	Specify path to <i>link_options.tau</i> generated by <i>tau_gen_wrapper</i>
-optHeaderInst	Enable Instrumentation of headers
-optTrackUPCR	Track UPC runtime layer routines (used with <i>tau_upc.sh</i>)
-optLinking=""	Options passed to the linker. Typically \$(TAU_MPI_FLIBS) \$(TAU_LIBS) \$(TAU_CXXLIBS)
-optCompile=""	Options passed to the compiler. Typically \$(TAU_MPI_INCLUDE) \$(TAU_INCLUDE) \$(TAU_DEFS)
-optPdtF95Opts=""	Add options for Fortran parser in PDT (f95parse/gfparse) ...

Compile-Time Options (contd.)

- Optional parameters for the TAU_OPTIONS environment variable:

% tau_compiler.sh

-optMICOffload	Links code for Intel MIC offloading, requires both host and MIC TAU libraries
-optShared	Use TAU's shared library (libTAU.so) instead of static library (default)
-optPdtCxxOpts=""	Options for C++ parser in PDT (cxxparse).
-optPdtF90Parser=""	Specify a different Fortran parser
-optPdtCleanscapeParser	Specify the Cleanscape Fortran parser instead of GNU gparser
-optTau=""	Specify options to the tau_instrumentor
-optTrackDMAPP	Enable instrumentation of low-level DMAPP API calls on Cray
-optTrackPthread	Enable instrumentation of pthread calls

See tau_compiler.sh for a full list of TAU_OPTIONS.

...

Compiling Fortran Codes with TAU

- If your Fortran code uses free format in .f files (fixed is default for .f), you may use:
% export TAU_OPTIONS= '-optPdtF95Opts="-R free" -optVerbose'
- To use the compiler based instrumentation instead of PDT (source-based):
% export TAU_OPTIONS= '-optComplInst -optVerbose'
- If your Fortran code uses C preprocessor directives (#include, #ifdef, #endif):
% export TAU_OPTIONS= '-optPreProcess -optVerbose -optDetectMemoryLeaks'
- To use an instrumentation specification file:
% export TAU_OPTIONS= '-optTauSelectFile=select.tau -optVerbose -optPreProcess'
% cat select.tau
BEGIN_INSTRUMENT_SECTION
loops routine="#"
this statement instruments all outer loops in all routines. # is wildcard as well as comment in first column.
END_INSTRUMENT_SECTION

Runtime Environment Variables

Environment Variable	Default	Description
TAU_TRACE	0	Setting to 1 turns on tracing
TAU_CALLPATH	0	Setting to 1 turns on callpath profiling
TAU_TRACK_MEMORY_FOOTPRINT	0	Setting to 1 turns on tracking memory usage by sampling periodically the resident set size and high water mark of memory usage
TAU_TRACK_POWER	0	Tracks power usage by sampling periodically.
TAU_CALLPATH_DEPTH	2	Specifies depth of callpath. Setting to 0 generates no callpath or routine information, setting to 1 generates flat profile and context events have just parent information (e.g., Heap Entry: foo)
TAU_SAMPLING	1	Setting to 1 enables event-based sampling.
TAU_TRACK_SIGNALS	0	Setting to 1 generate debugging callstack info when a program crashes
TAU_COMM_MATRIX	0	Setting to 1 generates communication matrix display using context events
TAU_THROTTLE	1	Setting to 0 turns off throttling. Throttles instrumentation in lightweight routines that are called frequently
TAU_THROTTLE_NUMCALLS	100000	Specifies the number of calls before testing for throttling
TAU_THROTTLE_PERCALL	10	Specifies value in microseconds. Throttle a routine if it is called over 100000 times and takes less than 10 usec of inclusive time per call
TAU_COMPENSATE	0	Setting to 1 enables runtime compensation of instrumentation overhead
TAU_PROFILE_FORMAT	Profile	Setting to “merged” generates a single file. “snapshot” generates xml format
TAU_METRICS	TIME	Setting to a comma separated list generates other metrics. (e.g., ENERGY,TIME,P_VIRTUAL_TIME,PAPI_FP_INS,PAPI_NATIVE_<event>:<subevent>)

Runtime Environment Variables (contd.)

Environment Variable	Default	Description
TAU_TRACK_MEMORY_LEAKS	0	Tracks allocates that were not de-allocated (needs –optMemDbg or tau_exec –memory)
TAU_EBS_SOURCE	TIME	Allows using PAPI hardware counters for periodic interrupts for EBS (e.g., TAU_EBS_SOURCE=PAPI_TOT_INS when TAU_SAMPLING=1)
TAU_EBS_PERIOD	100000	Specifies the overflow count for interrupts
TAU_MEMDBG_ALLOC_MIN/MAX	0	Byte size minimum and maximum subject to bounds checking (used with TAU_MEMDBG_PROTECT_*)
TAU_MEMDBG_OVERHEAD	0	Specifies the number of bytes for TAU's memory overhead for memory debugging.
TAU_MEMDBG_PROTECT_BELOW/ABOVE	0	Setting to 1 enables tracking runtime bounds checking below or above the array bounds (requires –optMemDbg while building or tau_exec –memory)
TAU_MEMDBG_ZERO_MALLOC	0	Setting to 1 enables tracking zero byte allocations as invalid memory allocations.
TAU_MEMDBG_PROTECT_FREE	0	Setting to 1 detects invalid accesses to deallocated memory that should not be referenced until it is reallocated (requires –optMemDbg or tau_exec –memory)
TAU_MEMDBG_ATTEMPT_CONTINUE	0	Setting to 1 allows TAU to record and continue execution when a memory error occurs at runtime.
TAU_MEMDBG_FILL_GAP	Undefined	Initial value for gap bytes
TAU_MEMDBG_ALIGNMENT	Sizeof(int)	Byte alignment for memory allocations
TAU_EVENT_THRESHOLD	0.5	Define a threshold value (e.g., .25 is 25%) to trigger marker events for min/max

Simplifying TAU's usage (tau_exec)

- Uninstrumented execution
 - % mpirun -np 4 ./a.out

- Track MPI performance
 - % mpirun -np 4 tau_exec ./a.out

- Track POSIX I/O and MPI performance (MPI enabled by default)

- % mpirun -np 4 tau_exec -T mpi,pdt -io ./a.out

- Track memory operations
 - % export TAU_TRACK_MEMORY_LEAKS=1
 - % mpirun -np 8 tau_exec -memory_debug ./a.out (bounds check)

- Use event based sampling (compile with -g)
 - % mpirun -np 8 tau_exec -ebs ./a.out
 - Also -ebs_source=<PAPI_COUNTER> -ebs_period=<overflow_count>

- Load wrapper interposition library
 - % mpirun -np 8 tau_exec -loadlib=<path/libwrapper.so> ./a.out

- Track GPGPU operations
 - % mpirun -np 8 tau_exec -cupti ./a.out
 - % mpirun -np 8 tau_exec -opencl ./a.out
 - % mpirun -np 8 tau_exec -openacc ./a.out

Binary Rewriting Instrumentation

- Support for both static and dynamic executables
- Specify a list of routines to instrument
- Specify the TAU measurement library to be injected
- MAQAO [UVSQ, Intel Exascale Labs]:

```
% tau_rewrite -T [tags] a.out -o a.inst
```

- DyninstAPI [U. Maryland and U. Wisconsin, Madison]:

```
% tau_run -T [tags] a.out -o a.inst
```

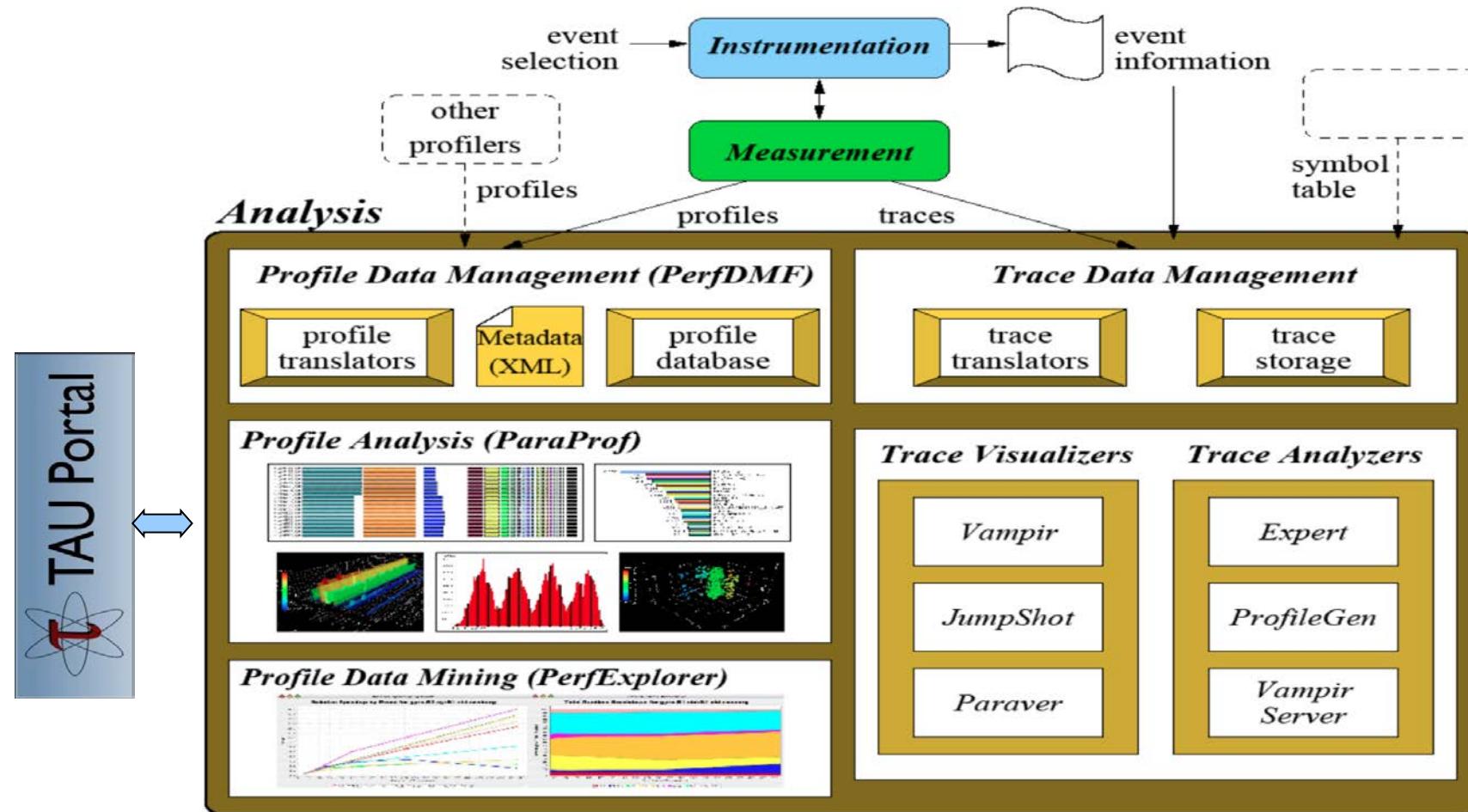
- Pebil [UC San Diego]:

```
% tau_pebil_rewrite -T [tags] a.out -o a.inst
```

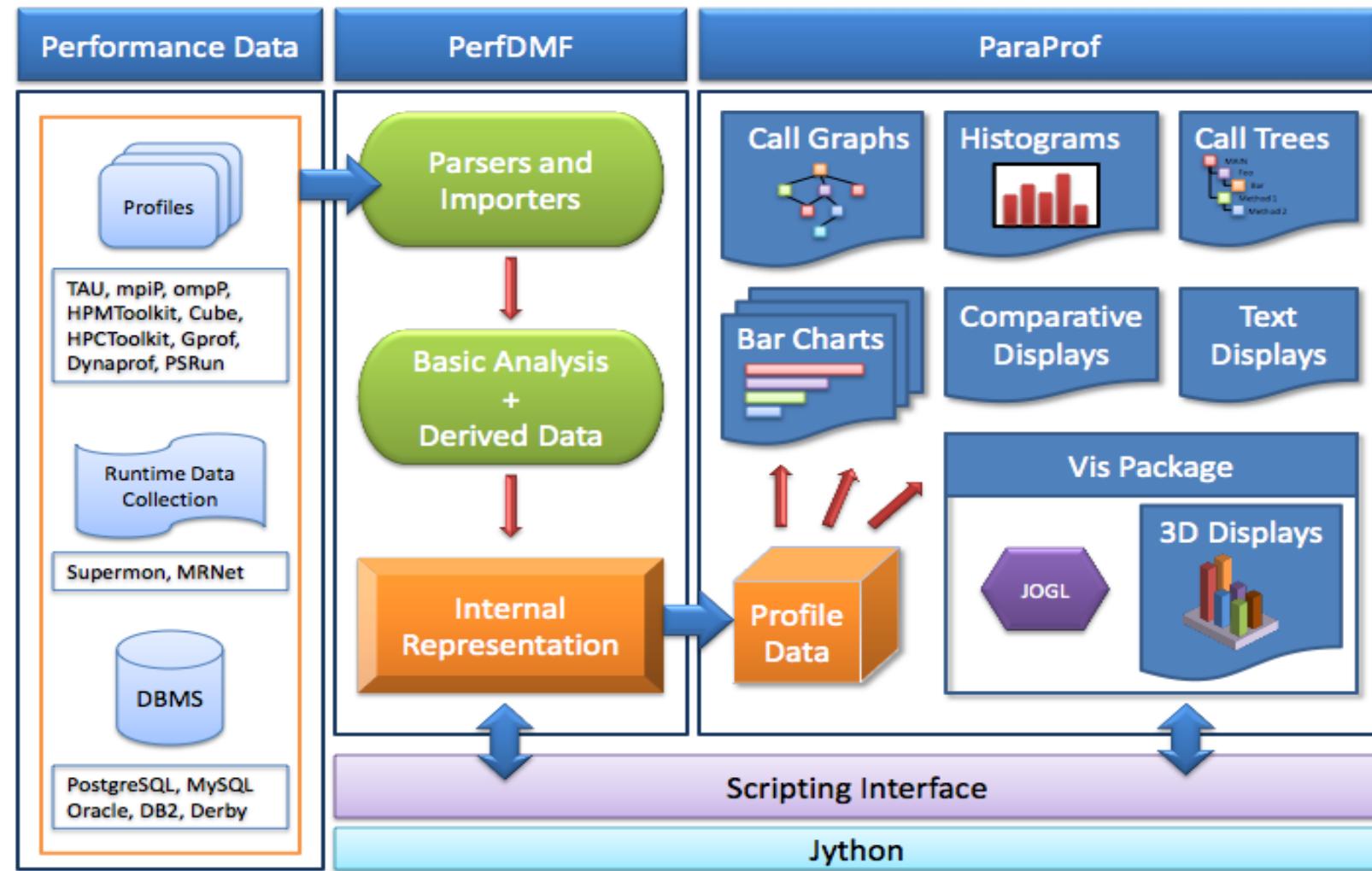
- Execute the application to get measurement data:

```
% mpirun -np 256 ./a.inst
```

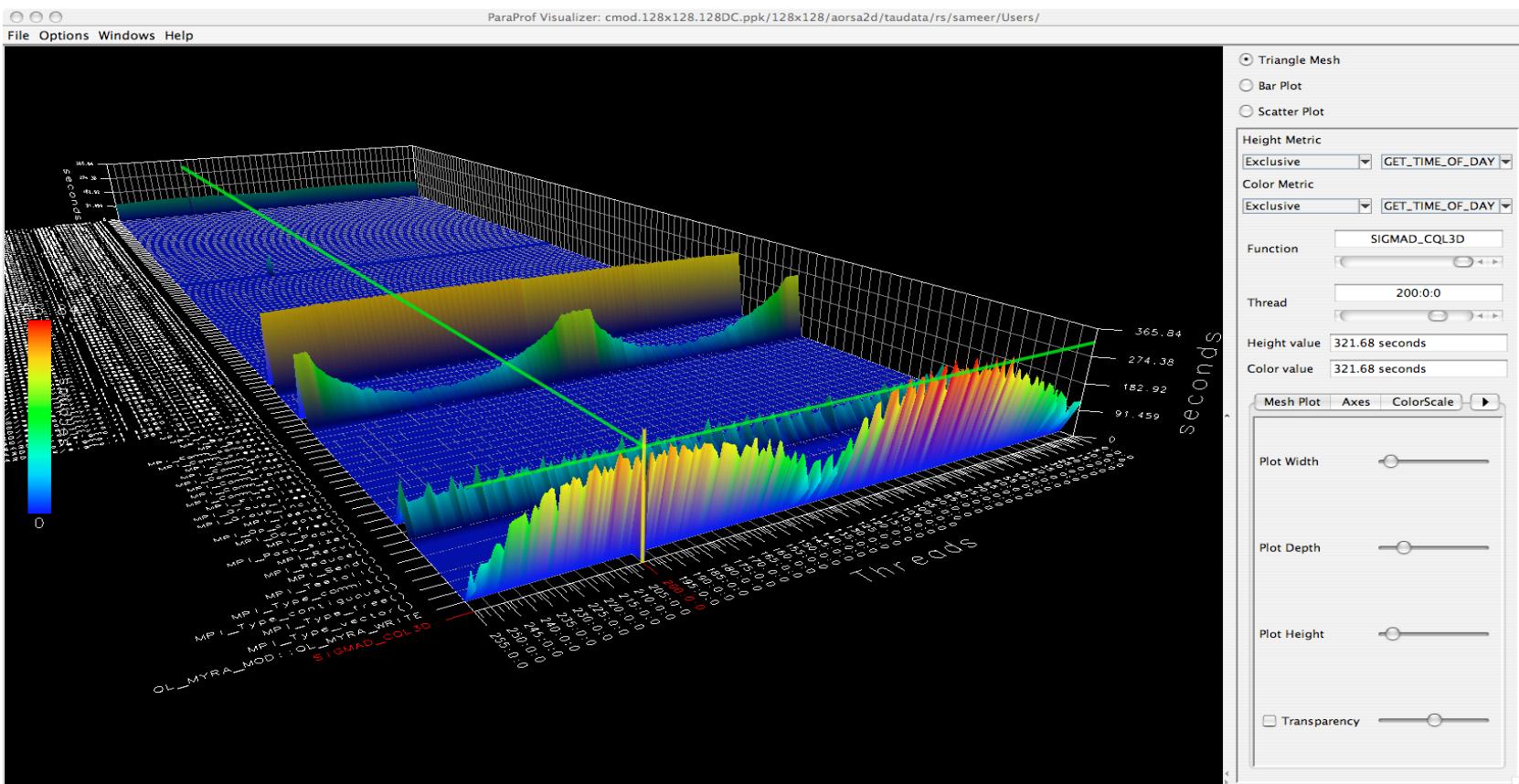
TAU Analysis



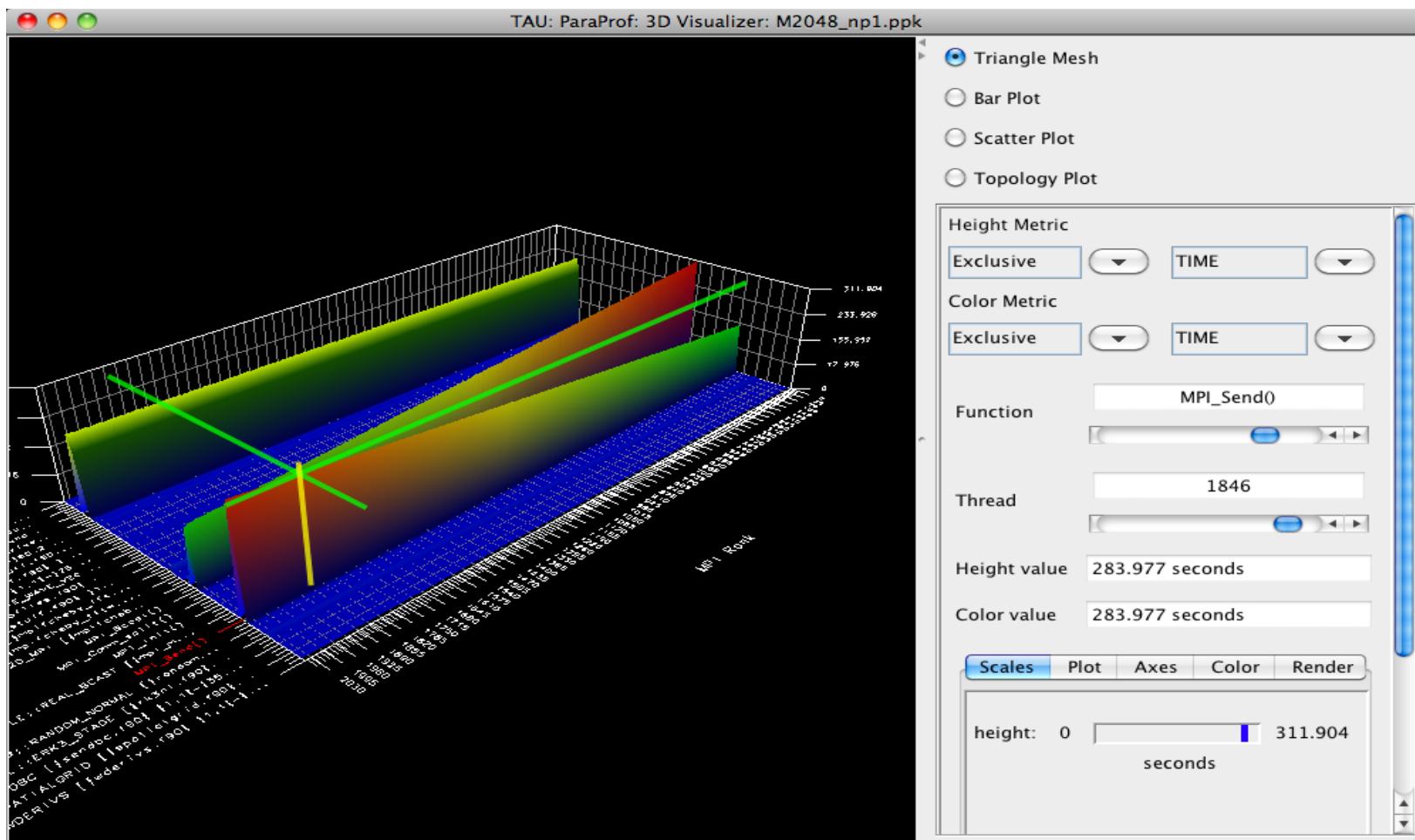
ParaProf Profile Analysis Framework



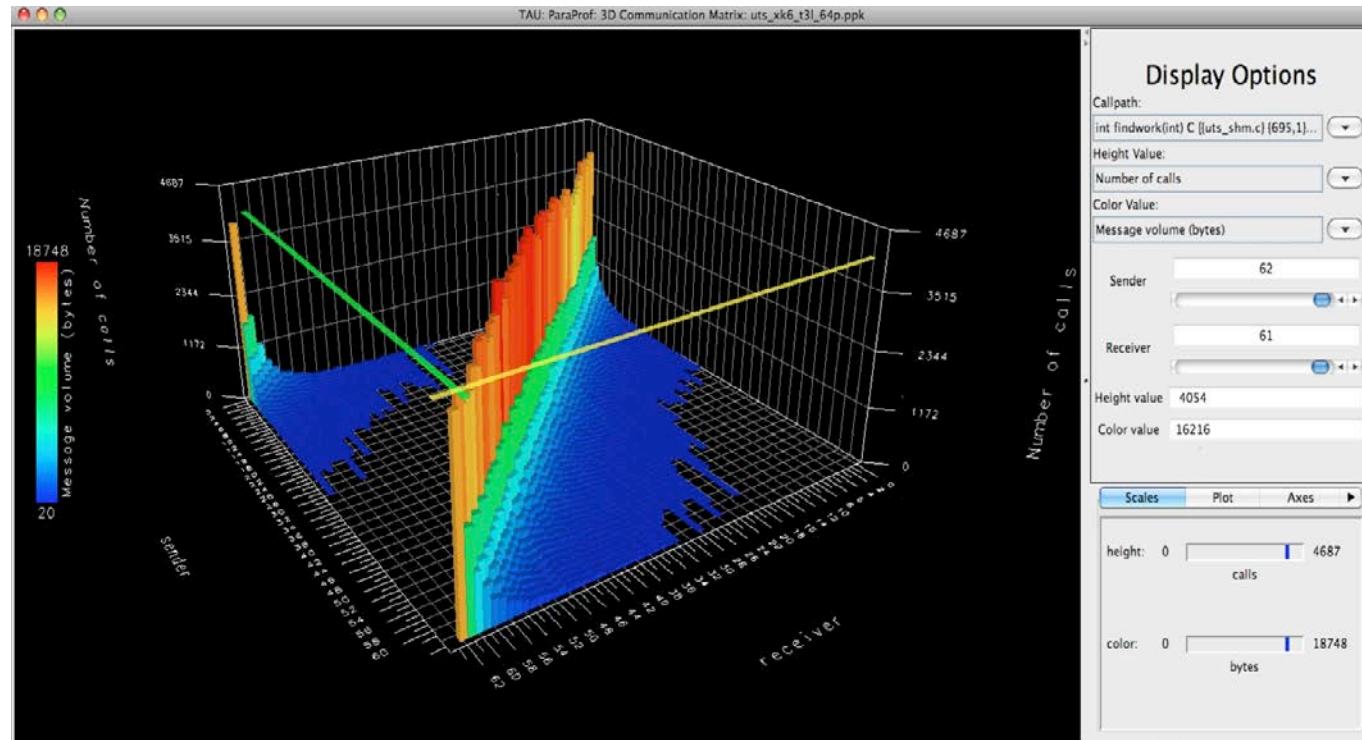
Parallel Profile Visualization: ParaProf



Parallel Profile Visualization: ParaProf

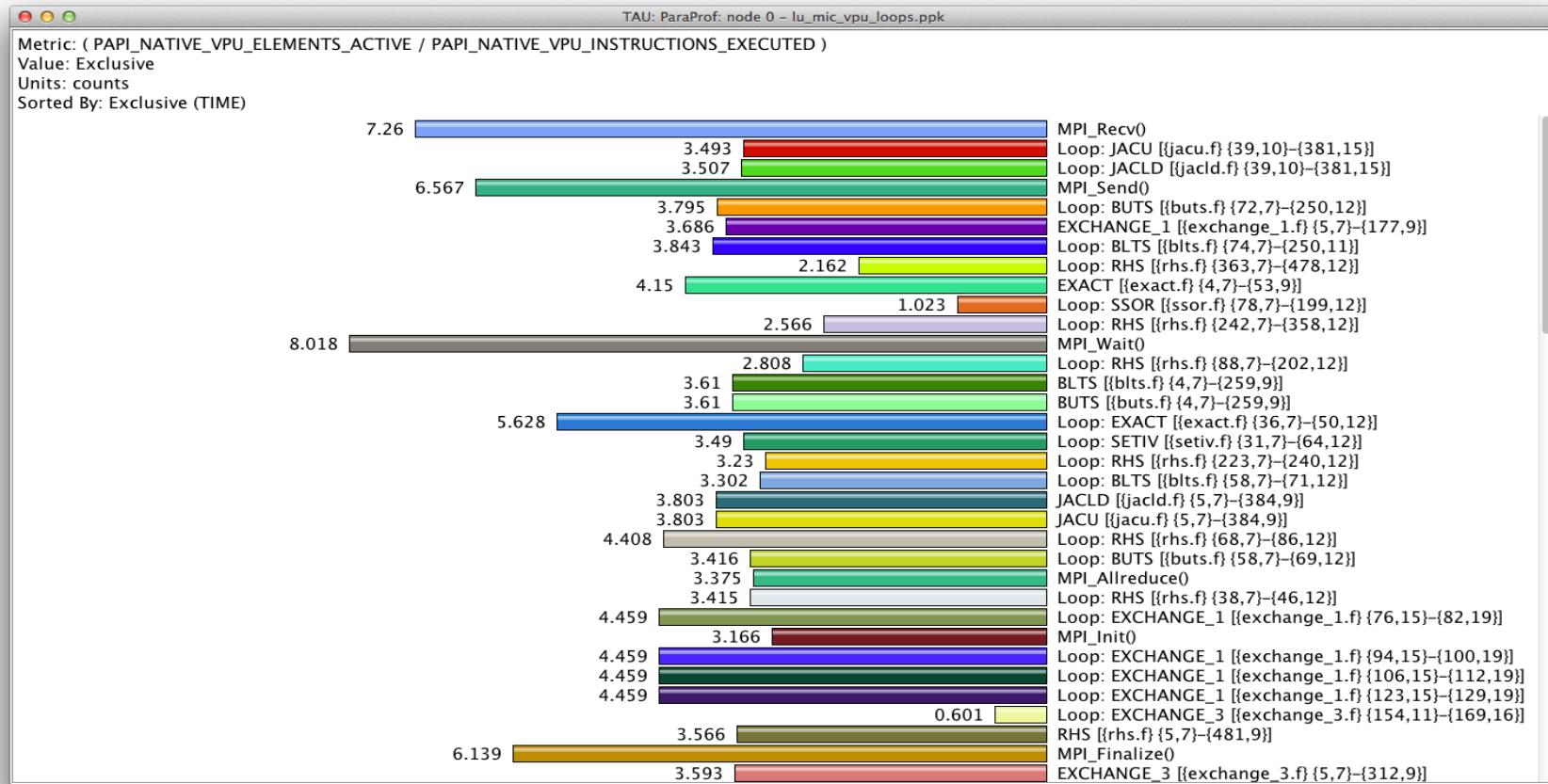


ParaProf 3D Communication Matrix



```
% export TAU_COMM_MATRIX=1
```

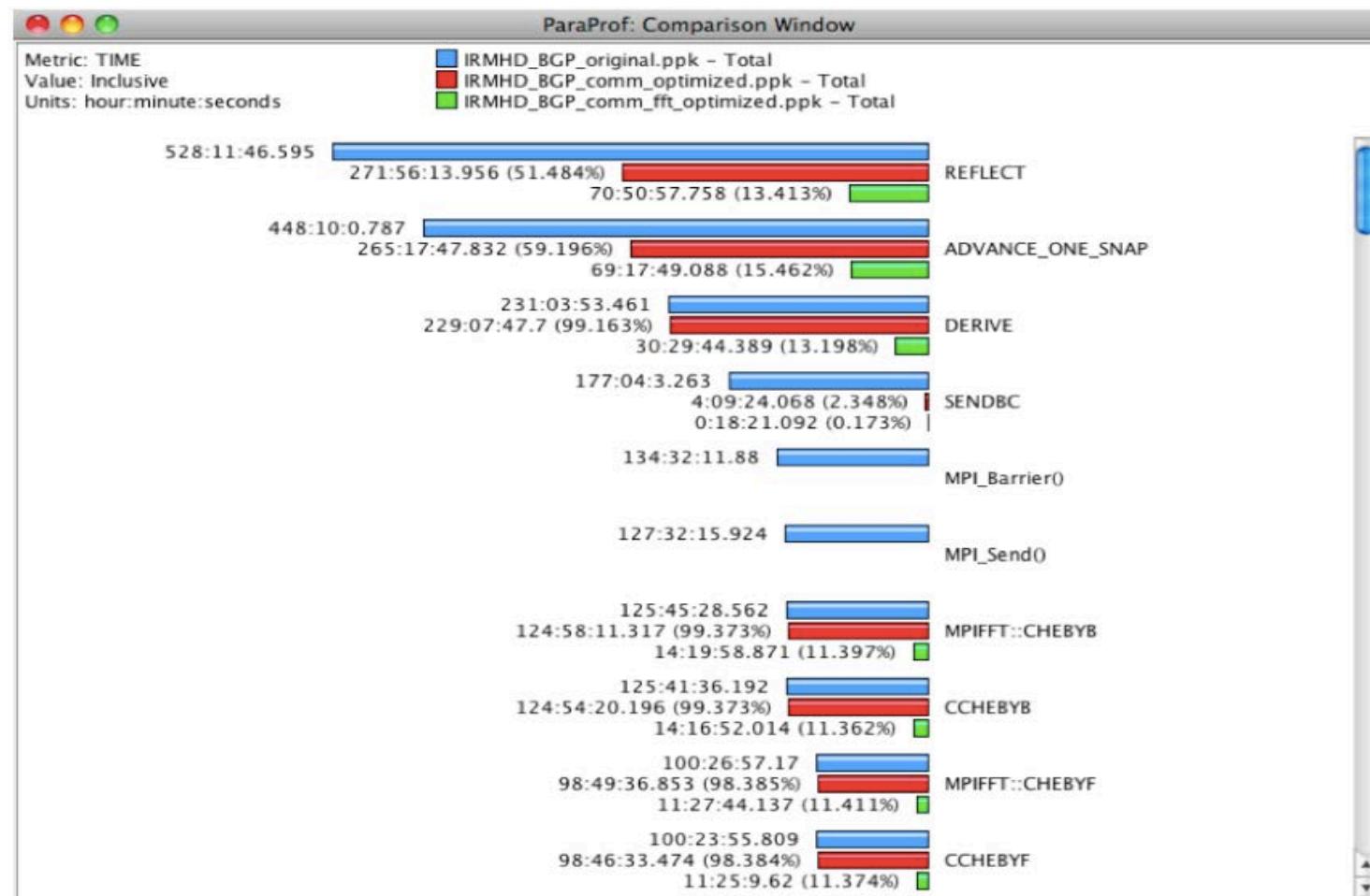
ParaProf Derived Metric Window: Intel MIC



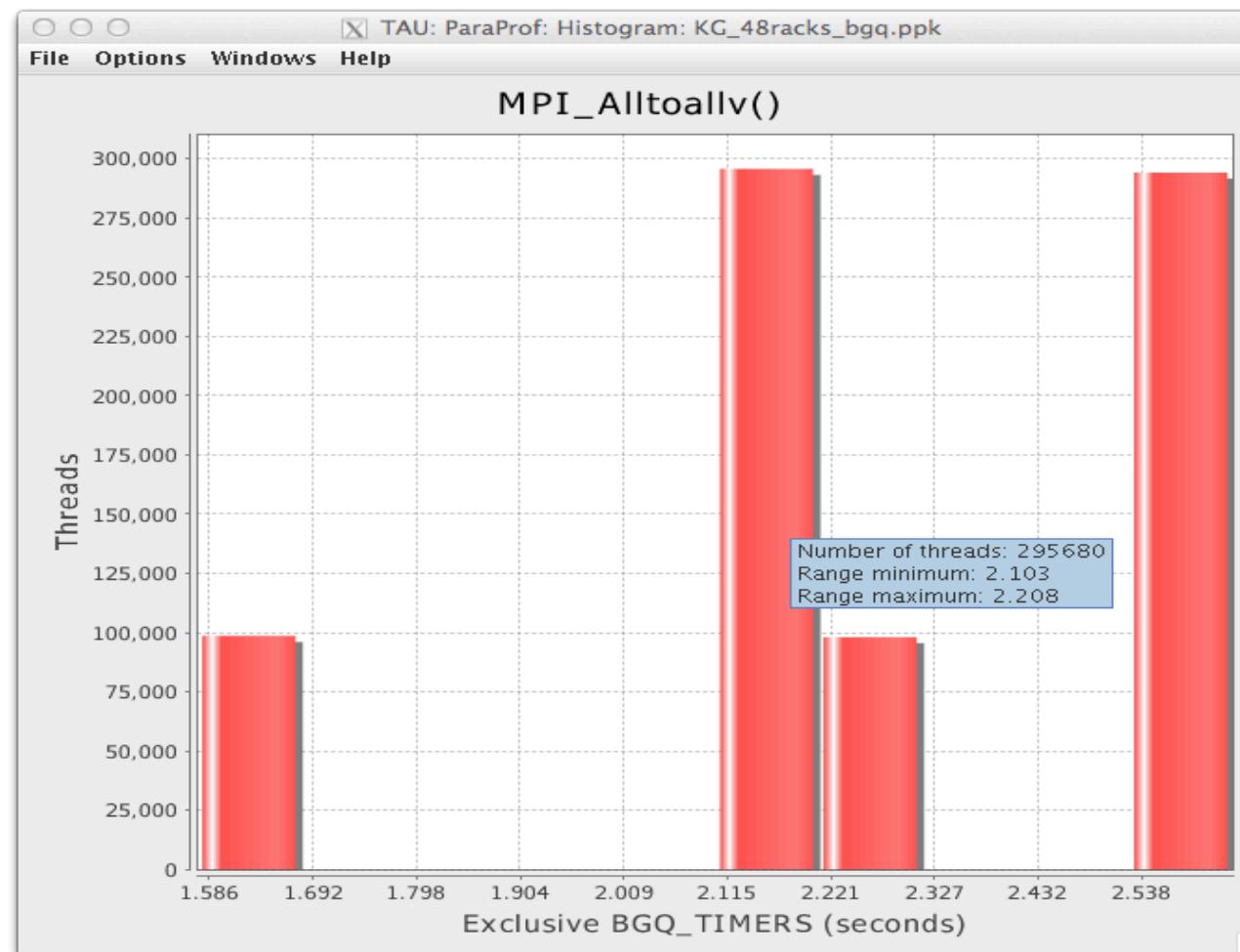
```
% export TAU_MAKEFILE=$TAUROOT/mic_linux/lib/Makefile.tau-icpc-papi-mpi-pdt
```

```
% export TAU_METRICS=TIME,PAPI_NATIVE_VPU_ELEMENTS_ACTIVE,PAPI_NATIVE_VPU_INSTRUCTIONS_EXECUTED
```

ParaProf Comparison Window



ParaProf Histogram Window

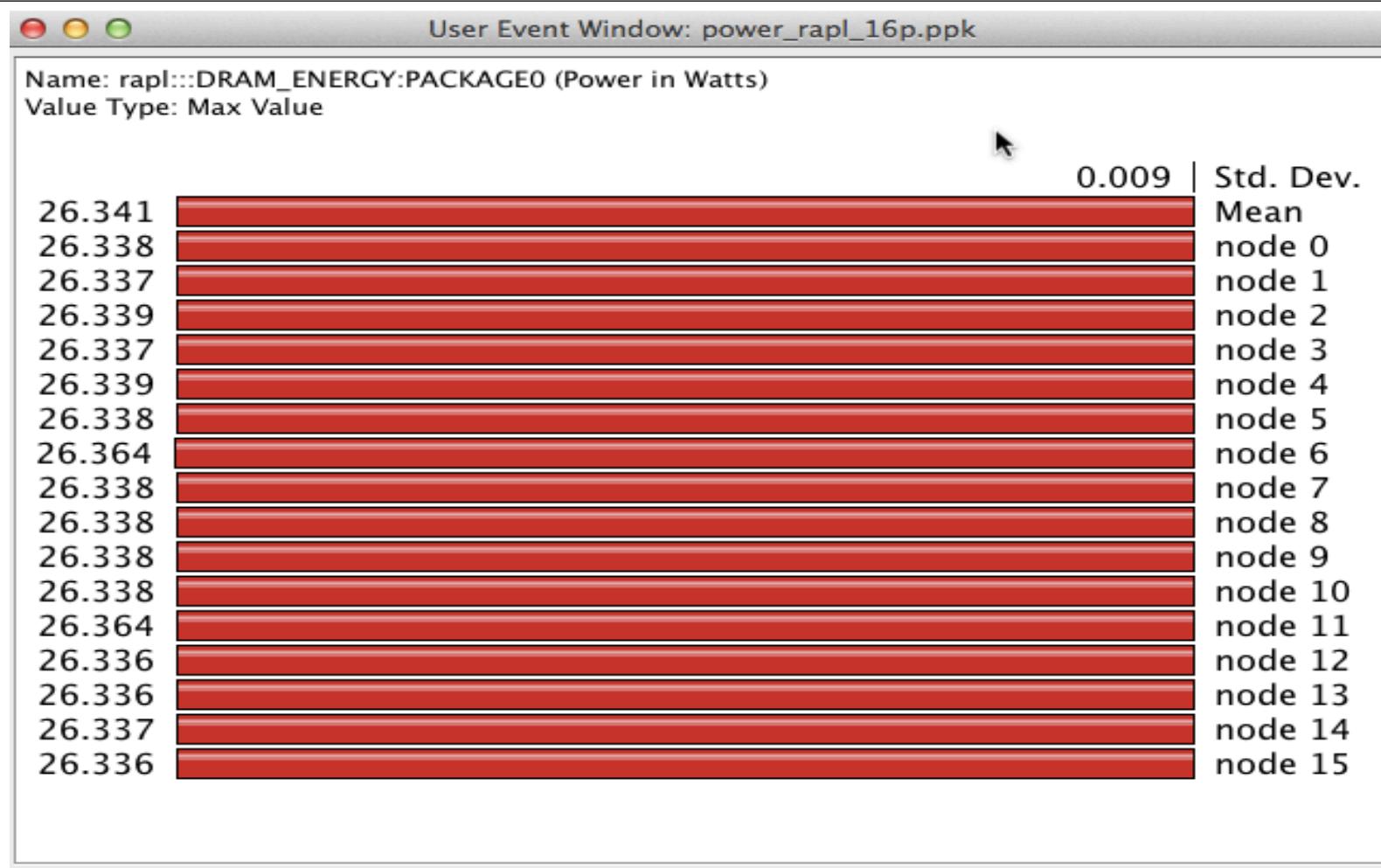


Marker Events in TAU Show Sudden Spikes

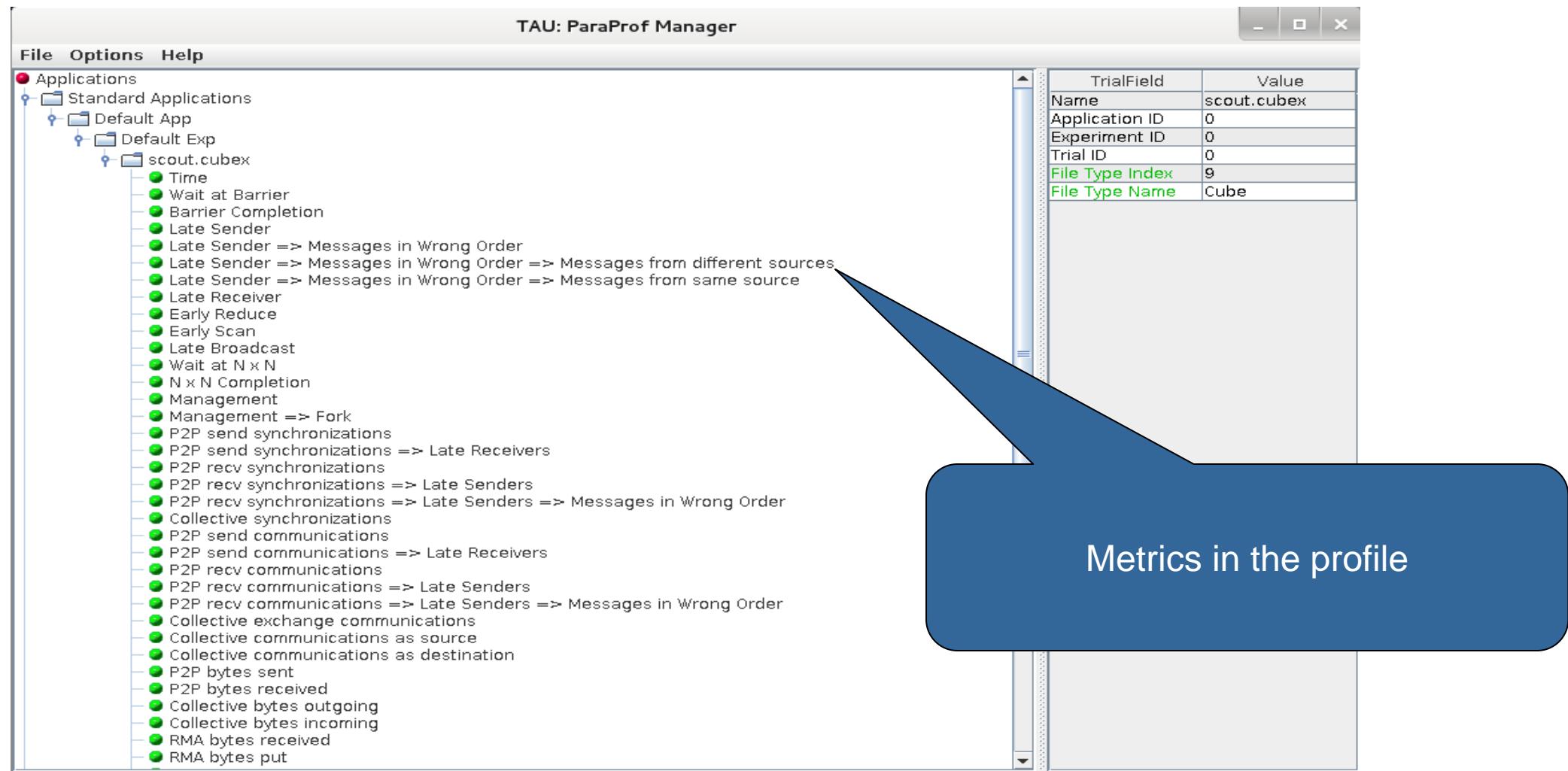
Name	MaxValue	MinValue	NumSamples	MeanValue	Std. Dev.
TAU: ParaProf: Context Events for: node 2 – rapl_marker_16p.ppk					
int main(int, char **) C [{matmult.c}{165,1}–{237,1}]					
double do_work(void) C [{matmult.c}{126,1}–{157,1}]					
void compute(double **, double **, double **, int, int, int) C [{matmult.c}{84,1}–{103,1}]					
[GROUP=MAX_MARKER] rapl:::DRAM_ENERGY:PACKAGE0 (Power in Watts)	17.585	17.469	5	17.521	0.037
[GROUP=MAX_MARKER] rapl:::DRAM_ENERGY:PACKAGE1 (Power in Watts)	15.261	15.218	4	15.237	0.016
[GROUP=MAX_MARKER] rapl:::PACKAGE_ENERGY:PACKAGE0 (Power in Watts)	118.903	114.923	22	116.98	1.201
[GROUP=MAX_MARKER] rapl:::PACKAGE_ENERGY:PACKAGE1 (Power in Watts)	113.466	110.207	22	111.778	0.996
[GROUP=MAX_MARKER] rapl:::PPO_ENERGY:PACKAGE0 (Power in Watts)	100.138	96.266	24	98.206	1.13
[GROUP=MAX_MARKER] rapl:::PPO_ENERGY:PACKAGE1 (Power in Watts)	95.846	92.758	24	94.319	0.937
[GROUP=MIN_MARKER] rapl:::DRAM_ENERGY:PACKAGE0 (Power in Watts)	17.397	17.303	4	17.358	0.035
[GROUP=MIN_MARKER] rapl:::DRAM_ENERGY:PACKAGE1 (Power in Watts)	15.048	15.042	2	15.045	0.003
int mysleep(int) C [{matmult.c}{46,1}–{49,1}]					
[GROUP=MIN_MARKER] rapl:::DRAM_ENERGY:PACKAGE0 (Power in Watts)	15.84	15.84	1	15.84	0
[GROUP=MIN_MARKER] rapl:::DRAM_ENERGY:PACKAGE1 (Power in Watts)	14.275	14.275	1	14.275	0
[GROUP=MIN_MARKER] rapl:::PACKAGE_ENERGY:PACKAGE1 (Power in Watts)	96.853	96.853	1	96.853	0
[GROUP=MIN_MARKER] rapl:::PACKAGE_ENERGY:PACKAGE0 (Power in Watts)	93.125	93.125	1	93.125	0
[GROUP=MIN_MARKER] rapl:::PPO_ENERGY:PACKAGE0 (Power in Watts)	75.096	75.096	1	75.096	0
[GROUP=MIN_MARKER] rapl:::PPO_ENERGY:PACKAGE1 (Power in Watts)	79.646	79.646	1	79.646	0
void compute_interchange(double **, double **, double **, int, int, int) C [{matmult.c}{105,1}–{124,1}]					
[GROUP=MAX_MARKER] rapl:::DRAM_ENERGY:PACKAGE0 (Power in Watts)	26.064	25.711	2	25.887	0.176
[GROUP=MAX_MARKER] rapl:::DRAM_ENERGY:PACKAGE1 (Power in Watts)	24.373	23.965	4	24.232	0.159
[GROUP=MAX_MARKER] rapl:::PACKAGE_ENERGY:PACKAGE0 (Power in Watts)	126.872	125.182	6	125.732	0.557
[GROUP=MAX_MARKER] rapl:::PACKAGE_ENERGY:PACKAGE1 (Power in Watts)	124.377	116.689	5	122.428	2.885
[GROUP=MAX_MARKER] rapl:::PPO_ENERGY:PACKAGE0 (Power in Watts)	103.981	102.21	6	102.769	0.584
[GROUP=MAX_MARKER] rapl:::PPO_ENERGY:PACKAGE1 (Power in Watts)	102.615	101.693	4	102.115	0.33
rapl:::DRAM_ENERGY:PACKAGE0 (Power in Watts)	26.064	15.84	36	19.053	3.39
rapl:::DRAM_ENERGY:PACKAGE1 (Power in Watts)	24.373	14.275	36	16.435	3.155
rapl:::PACKAGE_ENERGY:PACKAGE0 (Power in Watts)	126.872	93.125	36	117.729	5.403
rapl:::PACKAGE_ENERGY:PACKAGE1 (Power in Watts)	124.377	96.853	36	112.961	4.776
rapl:::PPO_ENERGY:PACKAGE0 (Power in Watts)	103.981	75.096	36	98.208	4.466
rapl:::PPO_ENERGY:PACKAGE1 (Power in Watts)	102.615	79.646	36	94.872	3.662

```
% export TAU_EVENT_THRESHOLD 0.5
```

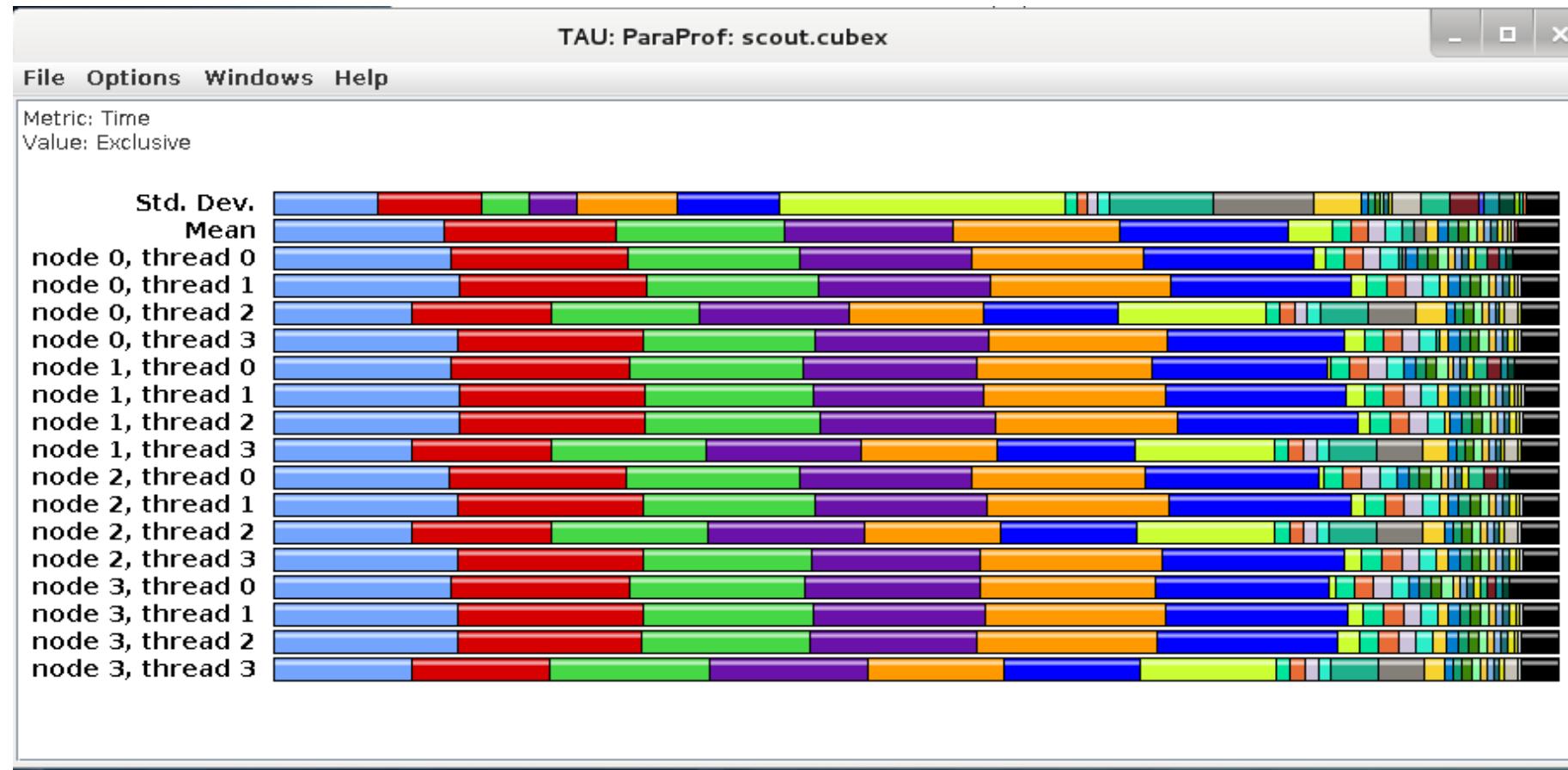
Energy Profiling in TAU



ParaProf Manager Widow: scout.cubex



ParaProf: Main Window



ParaProf: Thread Statistics Table

TAU: ParaProf: Statistics for: node 0, thread 0 – scout.cubex

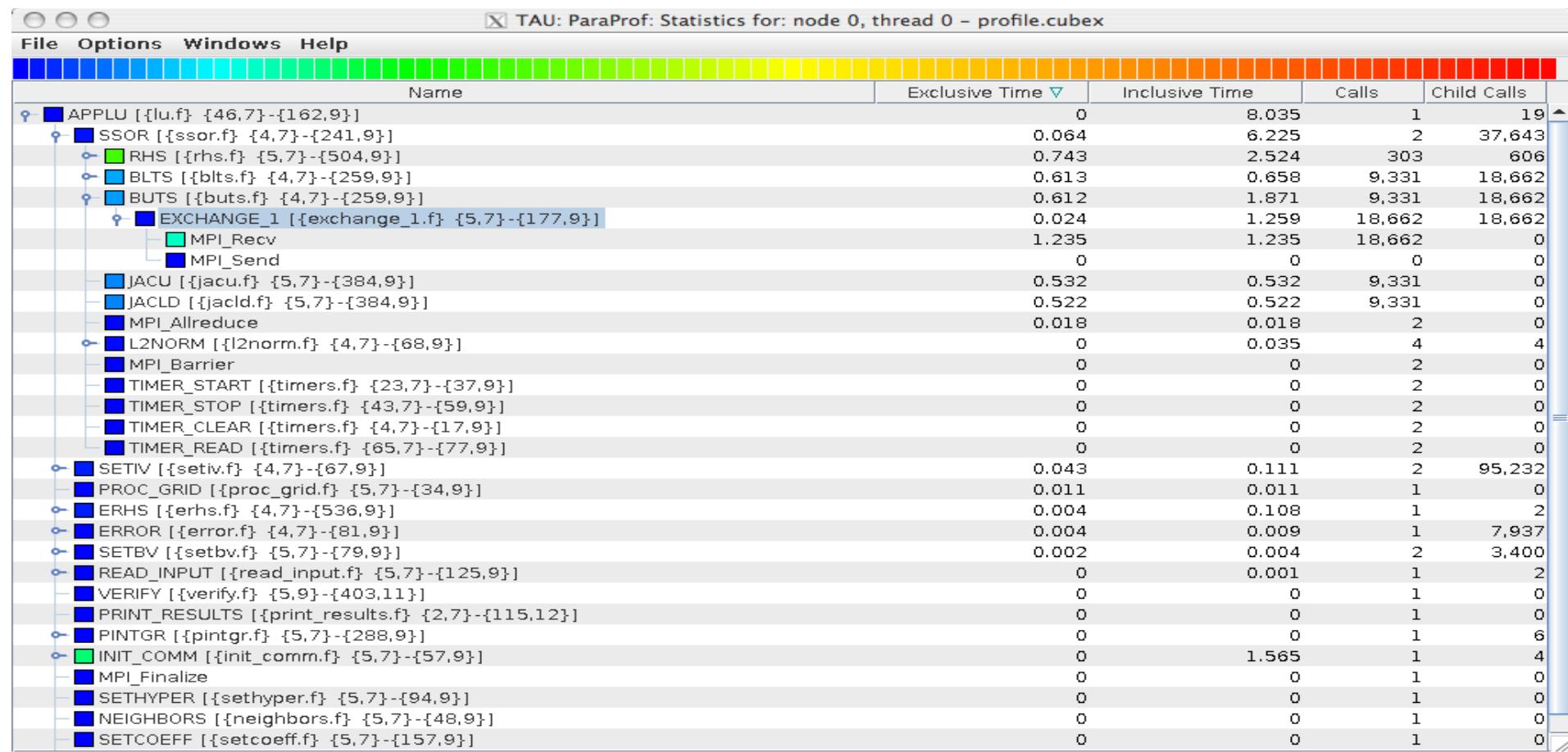
File Options Windows Help

Time

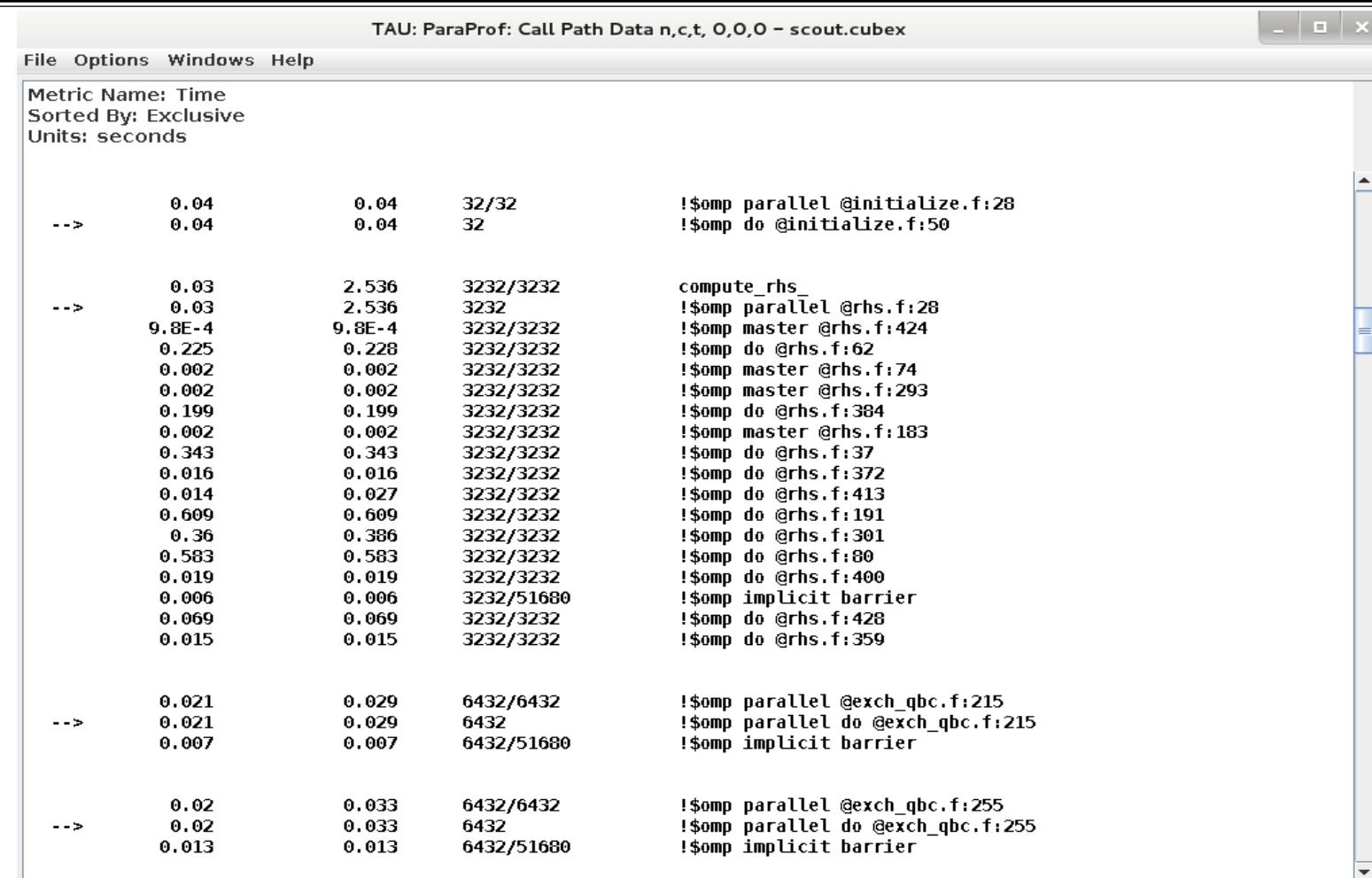
Name	Exclusive Time	Inclusive Time	Calls	Child Calls
!\$omp do @y_solve.f:52	5.817	5.817	3,216	0
!\$omp do @z_solve.f:52	5.657	5.657	3,216	0
!\$omp do @x_solve.f:54	5.609	5.609	3,216	0
!\$omp do @rhs.f:191	0.609	0.609	3,232	0
!\$omp do @rhs.f:80	0.583	0.583	3,232	0
MPI_Waitall	0.402	0.402	603	0
!\$omp implicit barrier	0.402	0.402	0	0
!\$omp do @rhs.f:301	0.36	0.36	0	0
!\$omp implicit barrier	0.026	0.026	0	0
!\$omp implicit barrier	0	0	0	0
!\$omp do @rhs.f:37	0.343	0.343	0	0
!\$omp do @rhs.f:62	0.225	0.225	16	0
!\$omp implicit barrier	0.004	0.004	3,216	0
!\$omp implicit barrier	0	0	0	0
MPI_Init_thread	0.218	0.218	1	0
!\$omp do @rhs.f:384	0.199	0.199	3,232	0
!\$omp parallel do @add.f:22	0.099	0.111	3,216	3,216
!\$omp do @rhs.f:428	0.069	0.069	3,232	0
MPI_Isend	0.043	0.043	603	0
!\$omp do @initialize.f:50	0.04	0.04	32	0
!\$omp parallel @rhs.f:28	0.03	2.536	3,232	51,712
!\$omp parallel do @exch_qbc.f:215	0.021	0.029	6,432	6,432
!\$omp parallel do @exch_qbc.f:255	0.02	0.033	6,432	6,432
!\$omp parallel @exch_qbc.f:255	0.02	0.053	6,432	6,432
!\$omp parallel @exch_qbc.f:244	0	0	0	0

Click to sort by a given metric, drag and move to rearrange columns

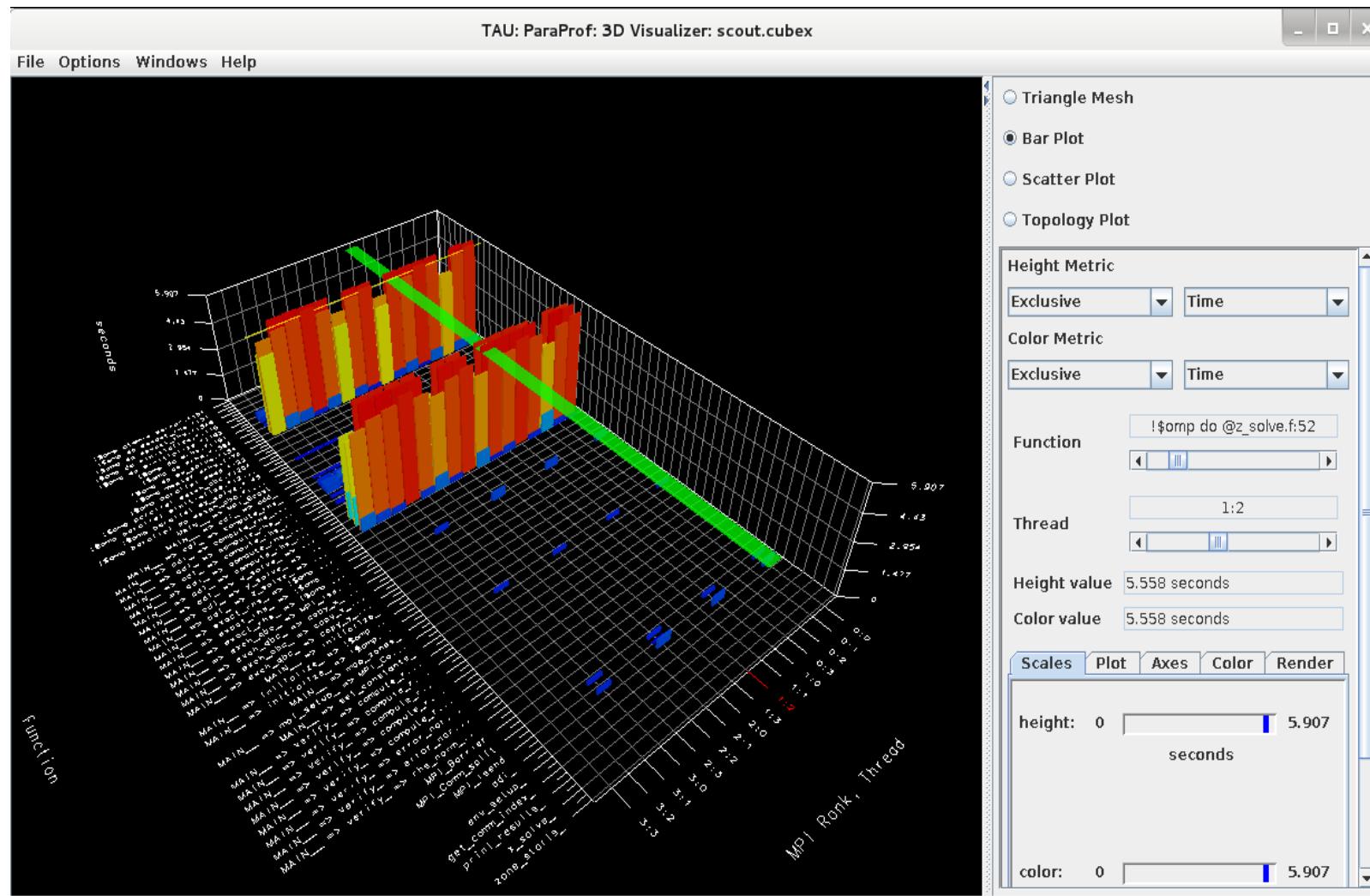
Example: Score-P with TAU (NPB LU)



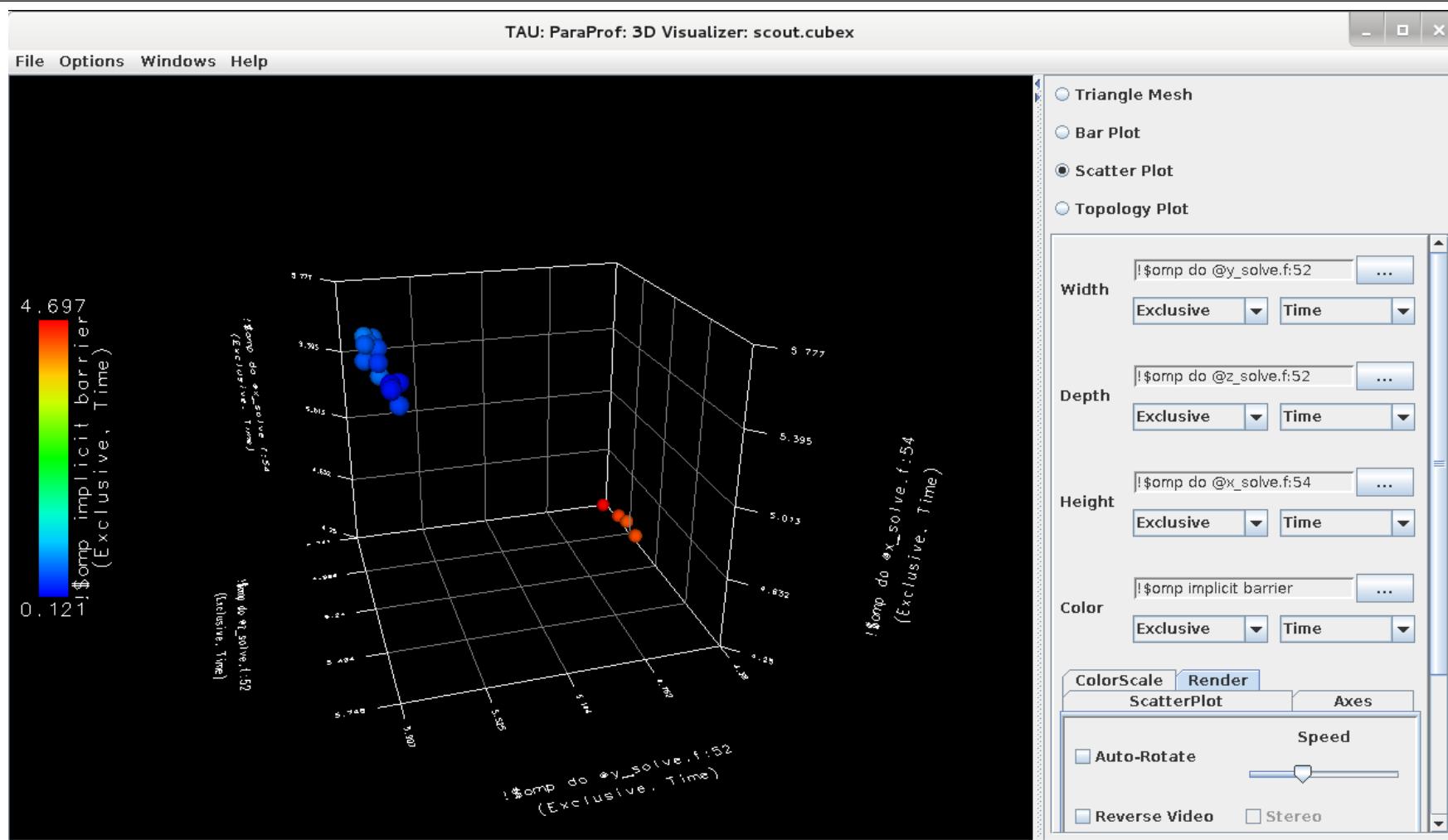
ParaProf: Callpath Thread Relations Window



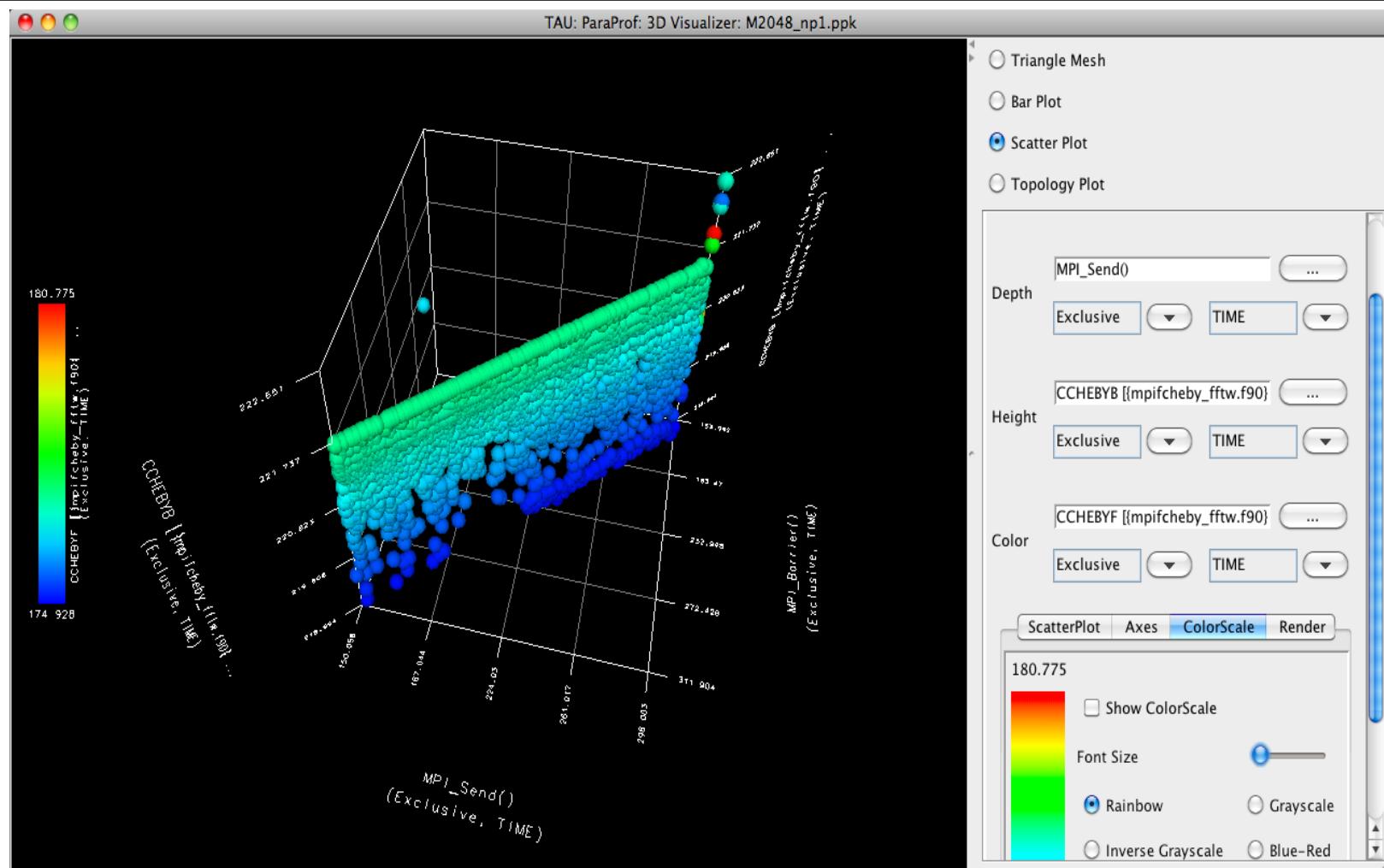
ParaProf: 3D Visualization Window



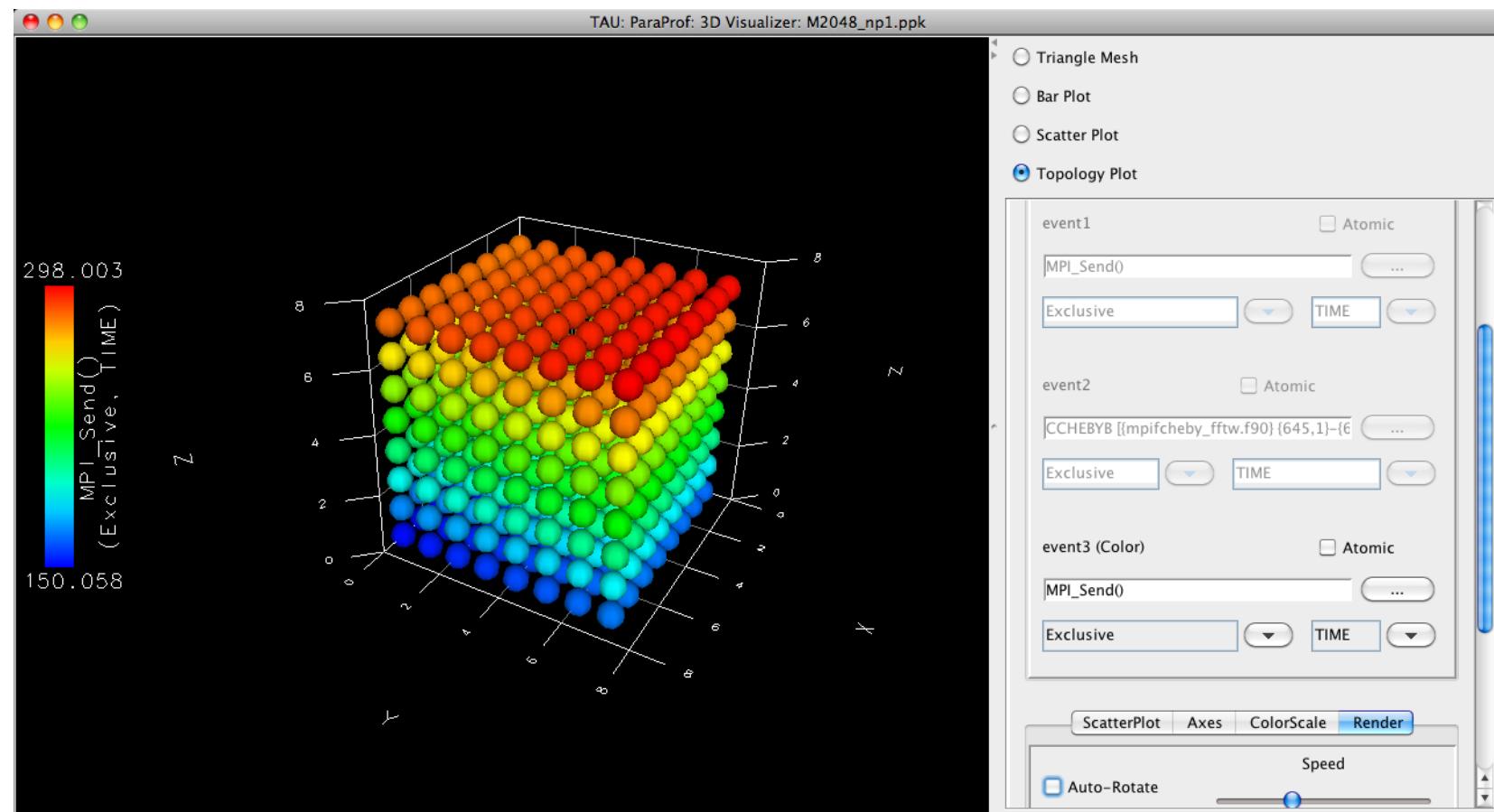
ParaProf: 3D Scatter Plot



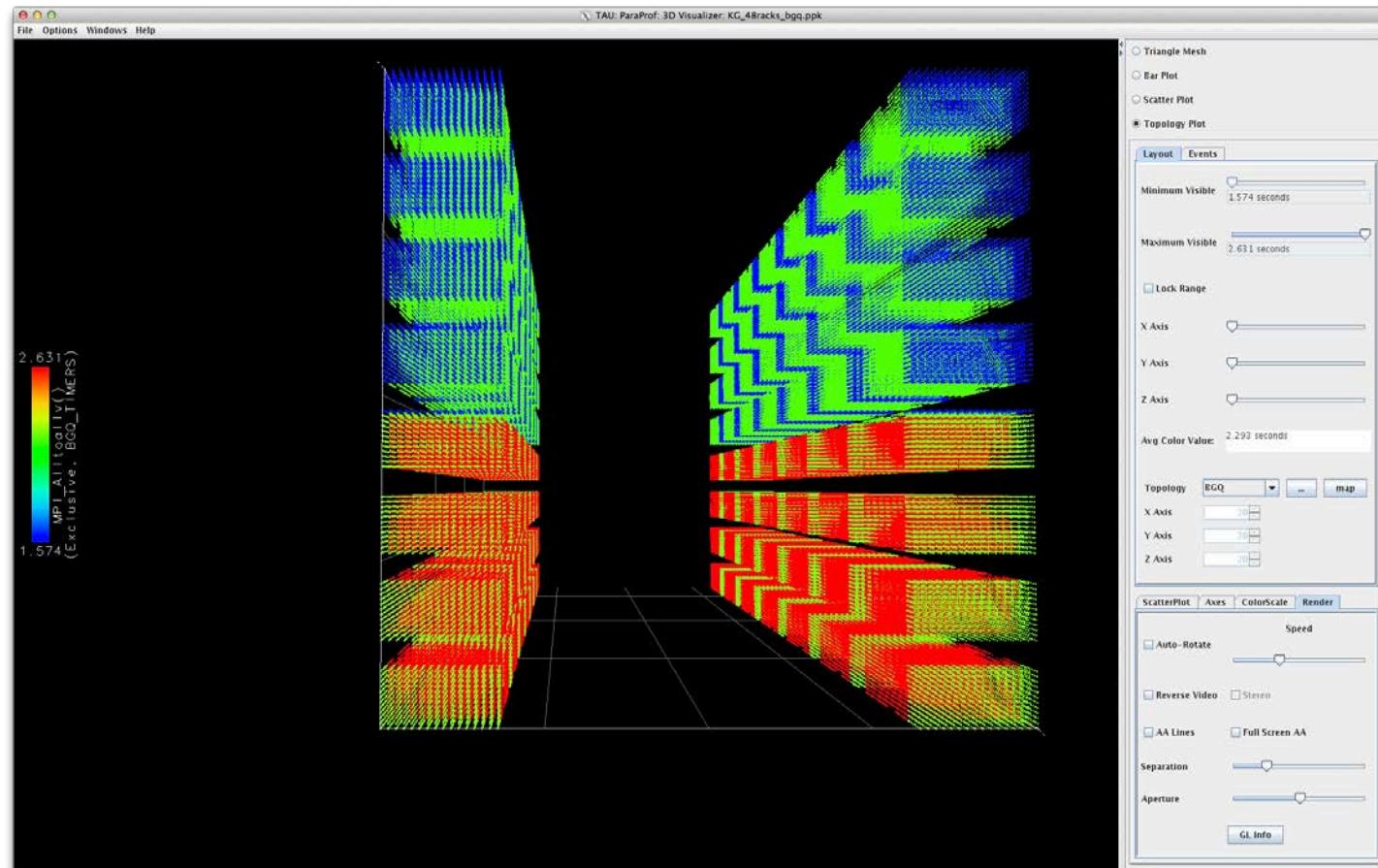
ParaProf: Scatter Plot



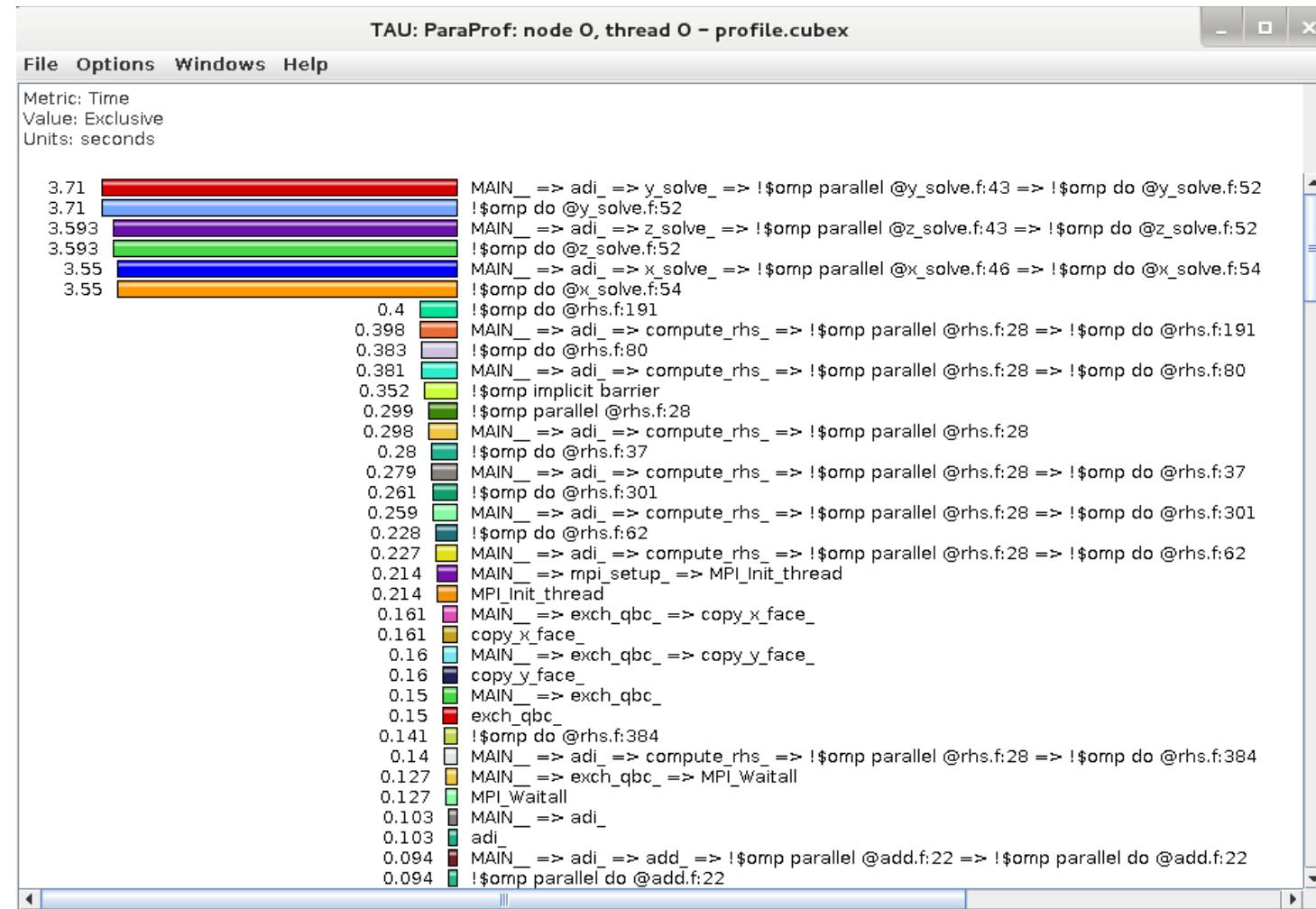
ParaProf: Topology 3D View (IBM BG/P)



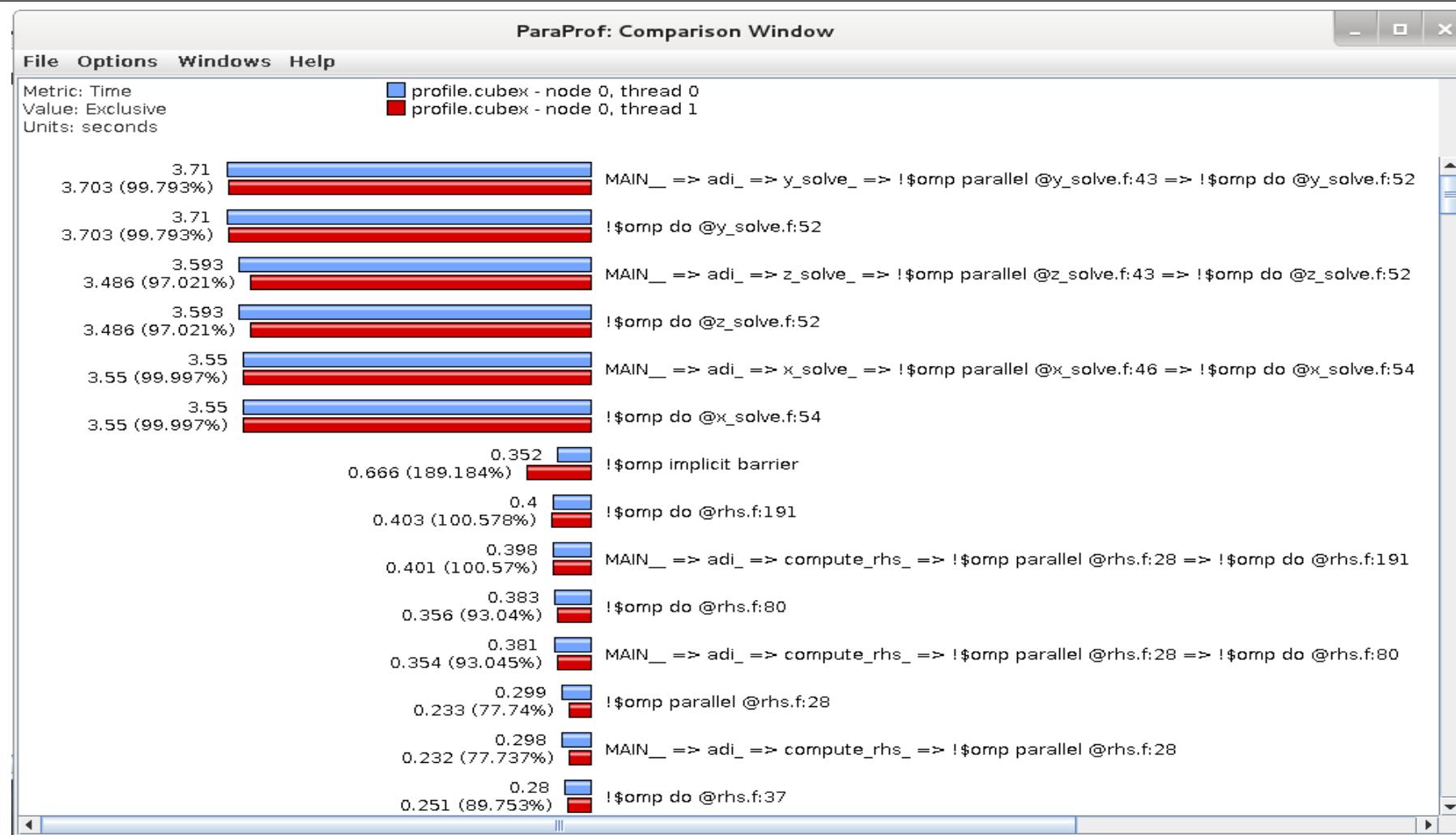
ParaProf: Topology View Torus (IBM BG/Q)



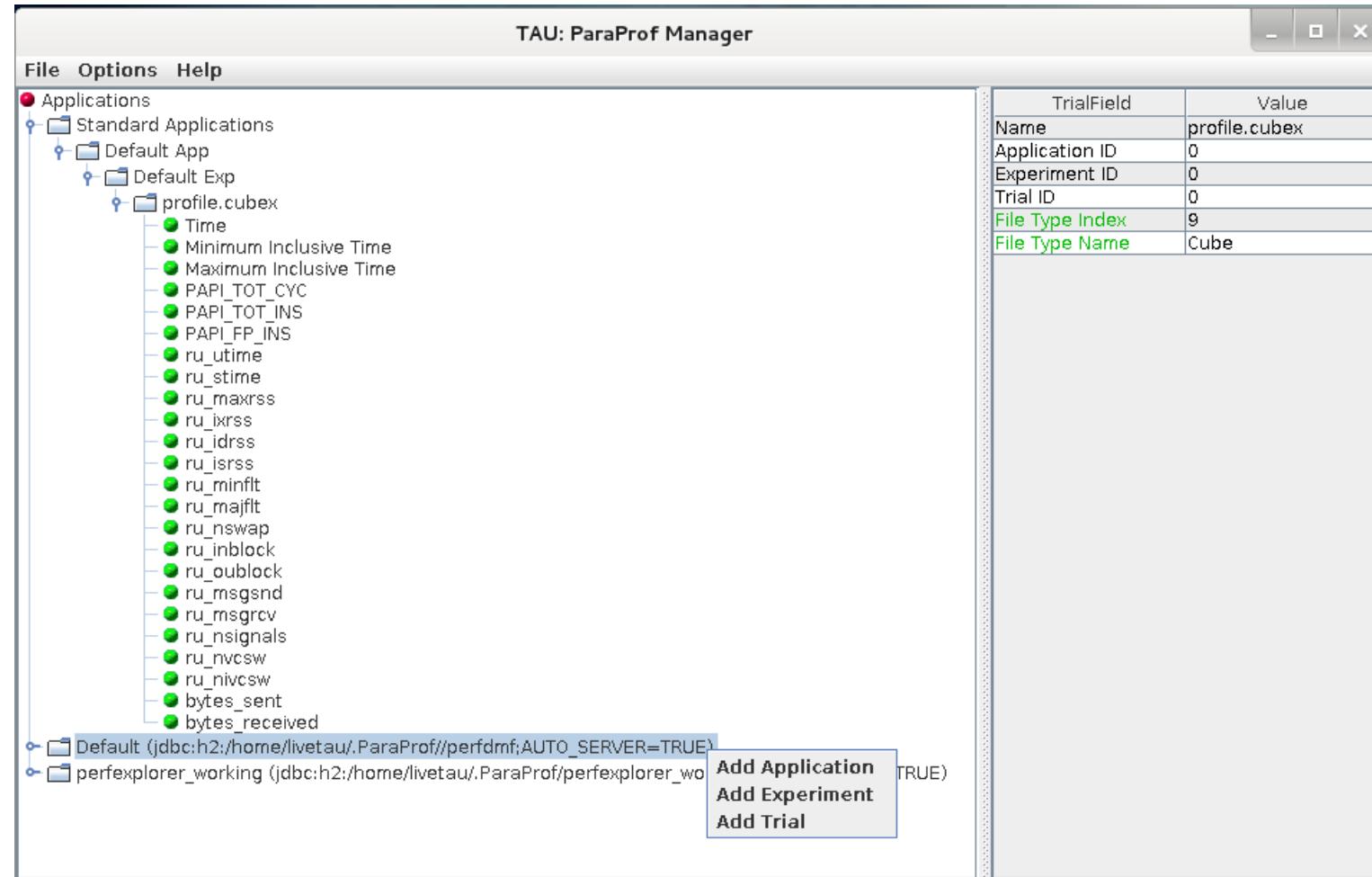
ParaProf: Node View



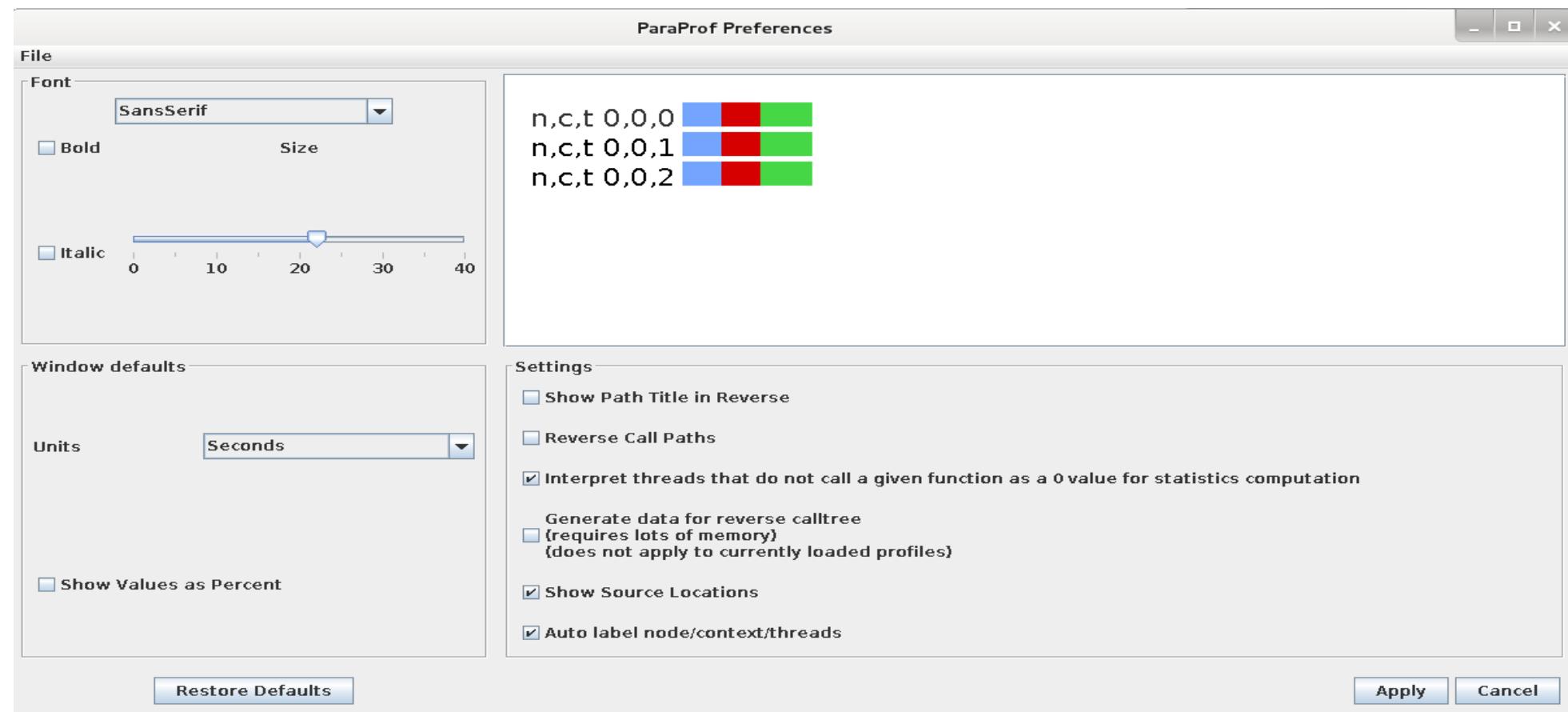
ParaProf: Add Thread to Comparison Window



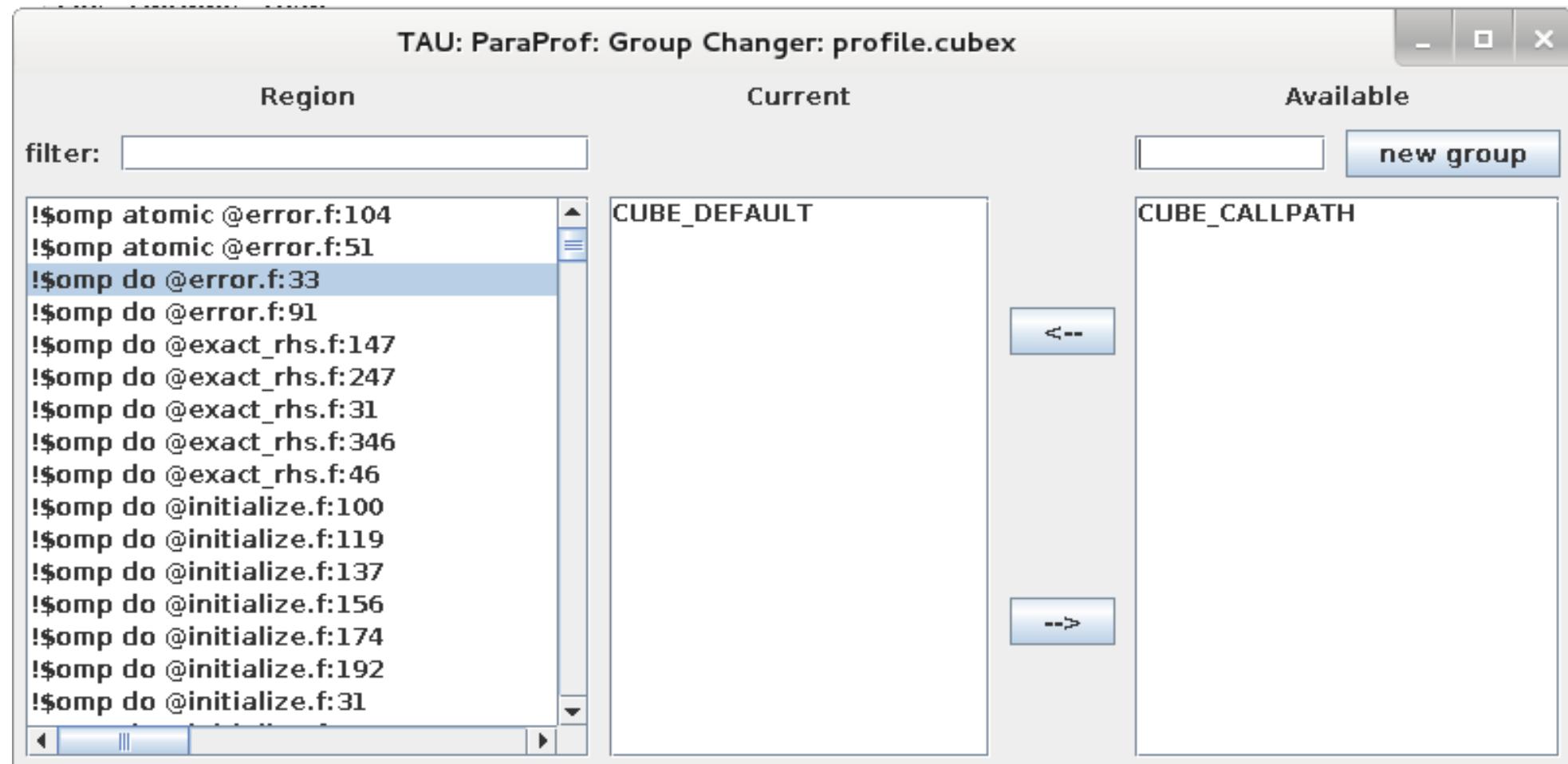
ParaProf: Score-P Profile Files, Database



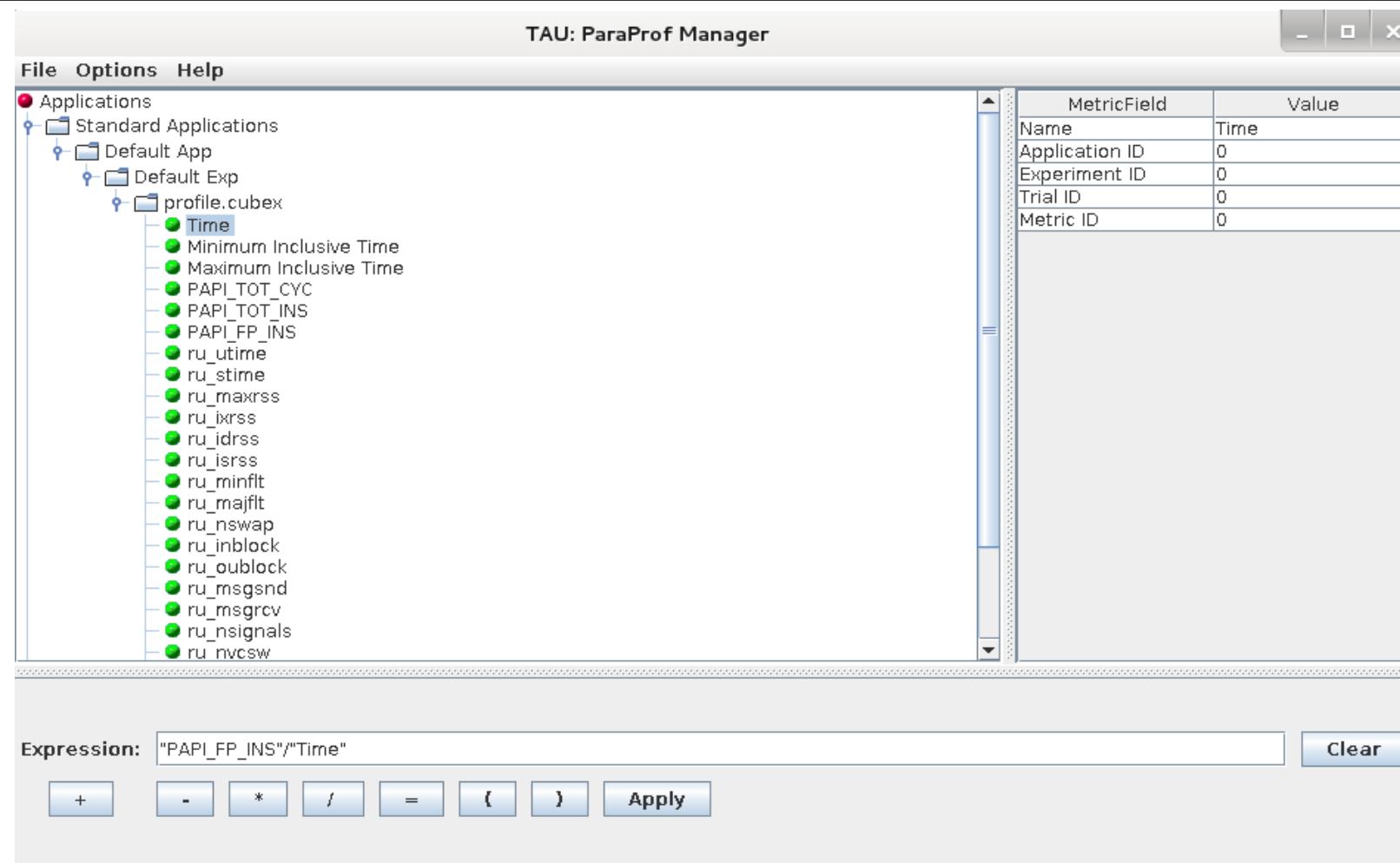
ParaProf: File Preferences Window



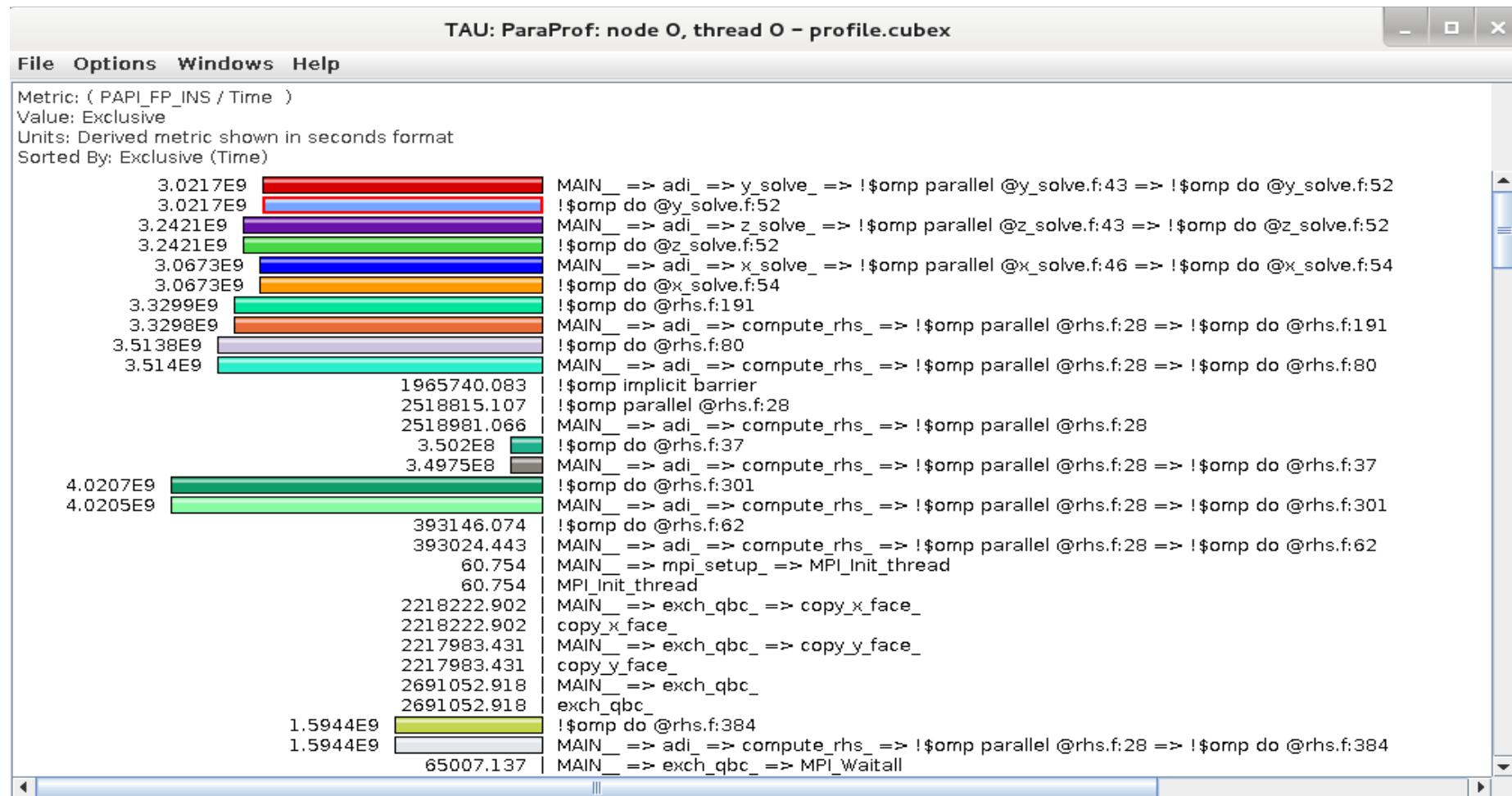
ParaProf: Group Changer Window



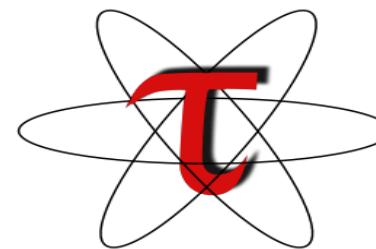
ParaProf: Derived Metric Panel in Manager Window



Sorting Derived FLOPS metric by Exclusive Time



Download TAU from U. Oregon



<http://tau.uoregon.edu>

<http://www.hpclinux.com> [LiveDVD, OVA]

Free download, open source, BSD license