Performance Data Management with TAU PerfExplorer
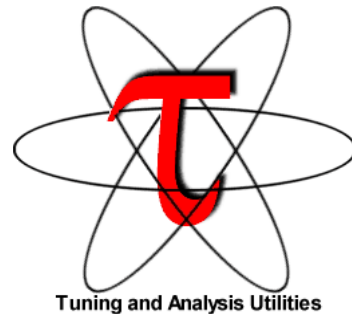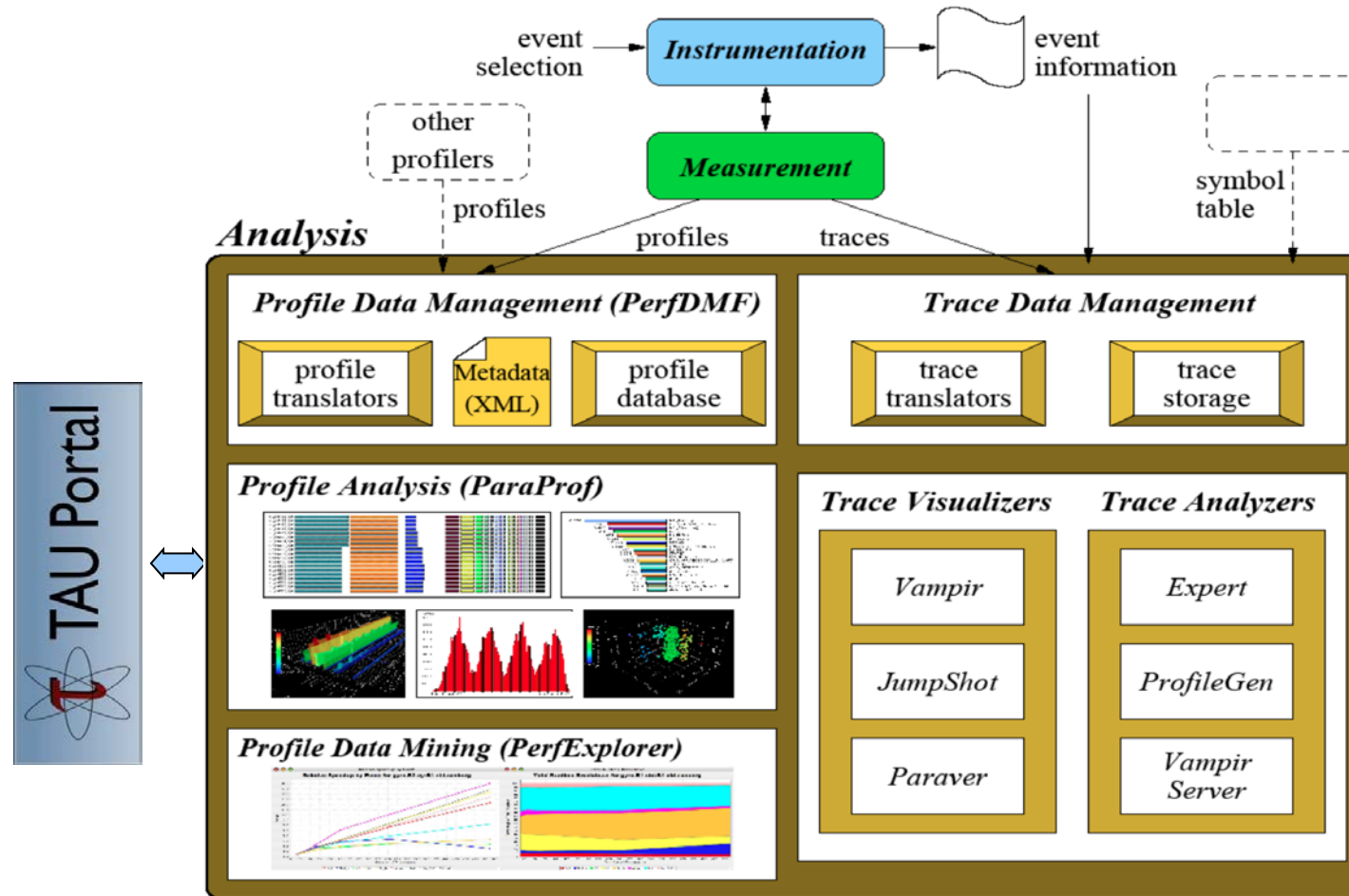
Sameer Shende
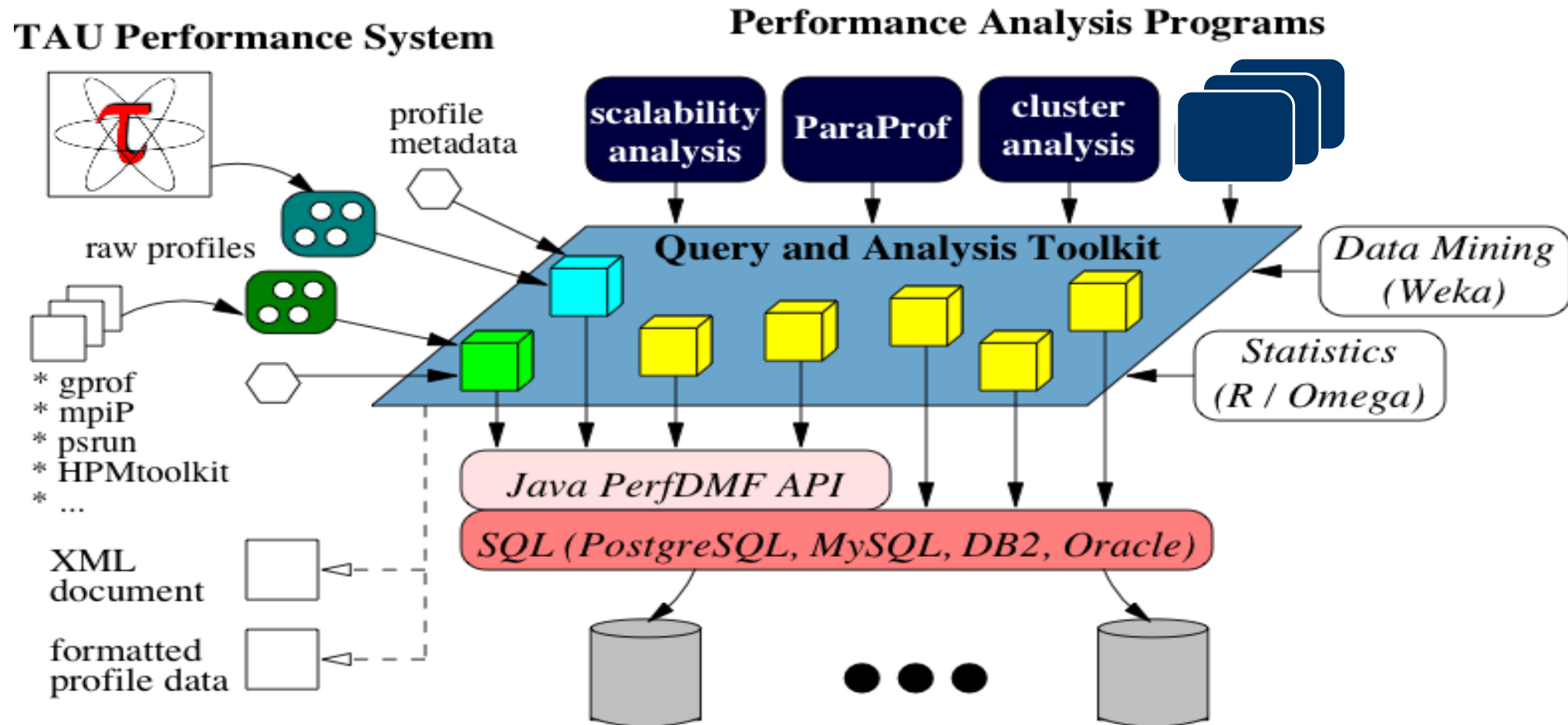
Performance Research Laboratory, University of Oregon

http://TAU.uoregon.edu

TAU Analysis

TAUdb: Performance Data Management Framework

Using TAUdb

- Configure TAUdb (Done by each user)
  % taudb_configure --create-default
  - Choose derby, PostgreSQL, MySQL, Oracle or DB2
  - Hostname
  - Username
  - Password
  - Say yes to downloading required drivers (we are not allowed to distribute these)
  - Stores parameters in your ~/.ParaProf/taudb.cfg file
- Configure PerfExplorer (Done by each user)
  % perfexplorer_configure
- Execute PerfExplorer
  % perfexplorer

Using PerfExplorer

```
% wget http://tau.uoregon.edu/data.tgz  (Contains CUBE profiles from Score-P)
% taudb_configure --create-default
(Chooses derby, blank user/passwd, yes to save passwd, defaults)
% perfexplorer_configure
(Yes to load schema, defaults)
% paraprof
(load each trial: DB -> Add Trial -> Type (Paraprof Packed Profile) -> OK)
OR use taudb_loadtrial –a "app" –x "experiment" –n "name" file.ppk
Then,
% tar zxf $TAU/data.tgz; cd data/tau;
% taudb_loadtrial –a BT_MZ –x "Class_B" bt-mz_B.*.ppk
% perfexplorer
(Select experiment, Menu: Charts -> Speedup)
```

Performance Data Mining (PerfExplorer)

- ## Performance knowledge discovery framework
  - Data mining analysis applied to parallel performance data
    - comparative, clustering, correlation, dimension reduction, …
  - Use the existing TAU infrastructure
    - TAU performance profiles, taudb
  - Client-server based system architecture
- ## Technology integration
  - Java API and toolkit for portability
  - taudb
  - R-project/Omegahat, Octave/Matlab statistical analysis
  - WEKA data mining package
  - JFreeChart for visualization, vector output (EPS, SVG)

PerfExplorer: Using Cluster Analysis

- Performance data represented as vectors - each dimension is the cumulative time for an event
- k-means: k random centers are selected and instances are grouped with the "closest" (Euclidean) center
- New centers are calculated and the process repeated until stabilization or max iterations
- Dimension reduction necessary for meaningful results
- Virtual topology, summaries constructed
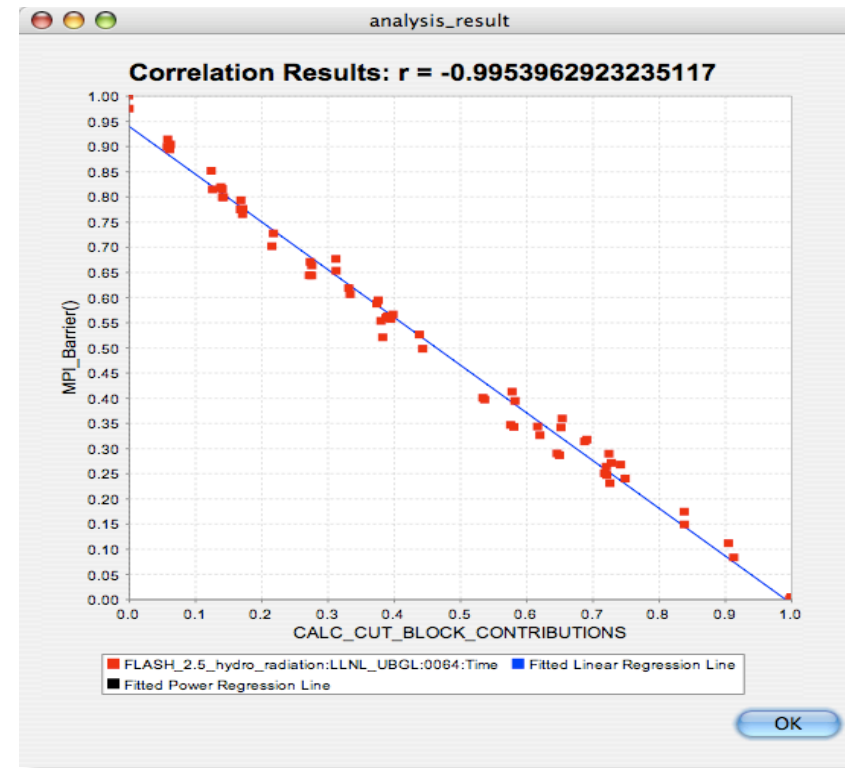
PerfExplorer - Cluster Analysis (sPPM)
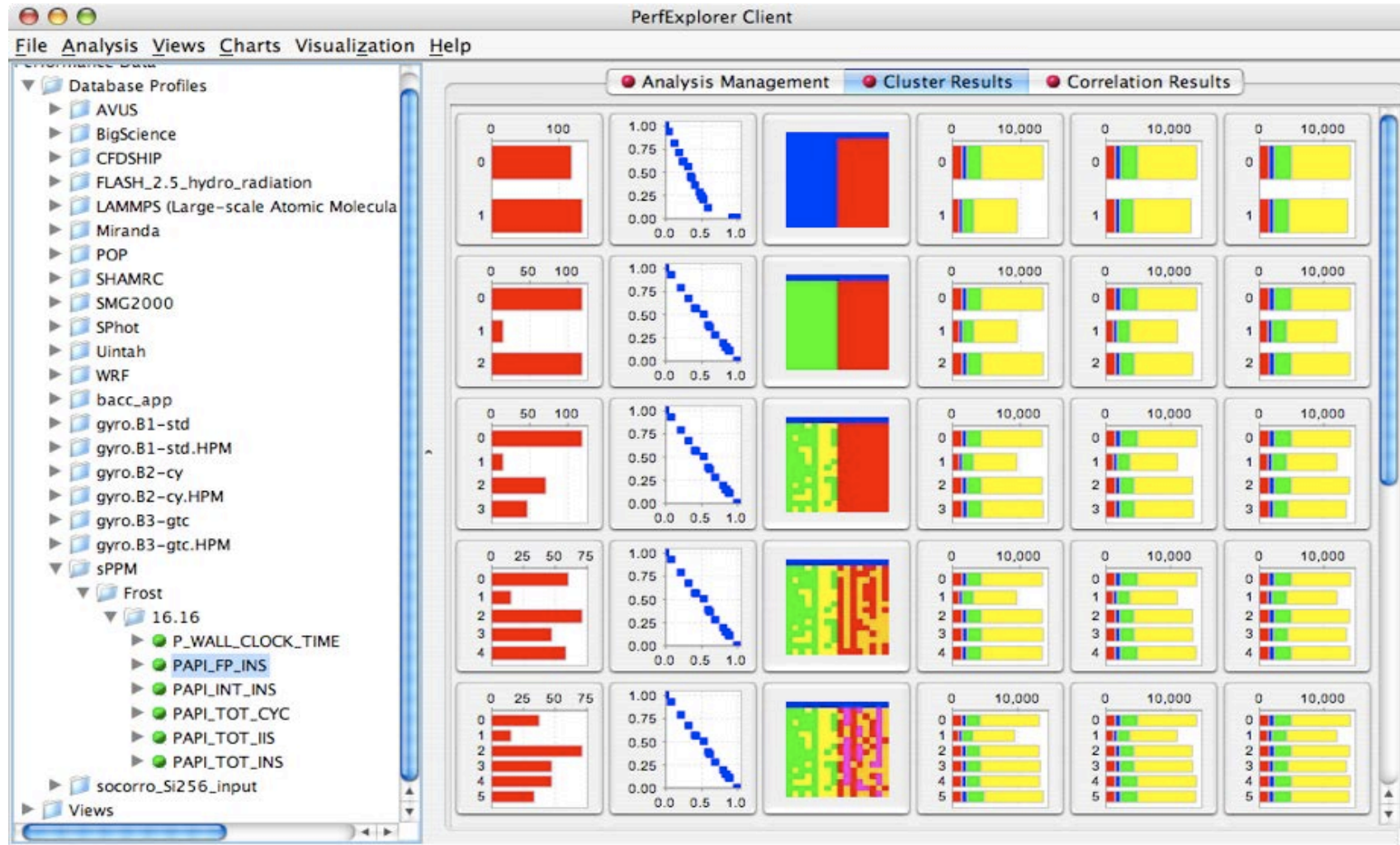
PerfExplorer - Correlation Analysis (Flash)

- Describes strength and direction of a linear relationship between two variables (events) in the data
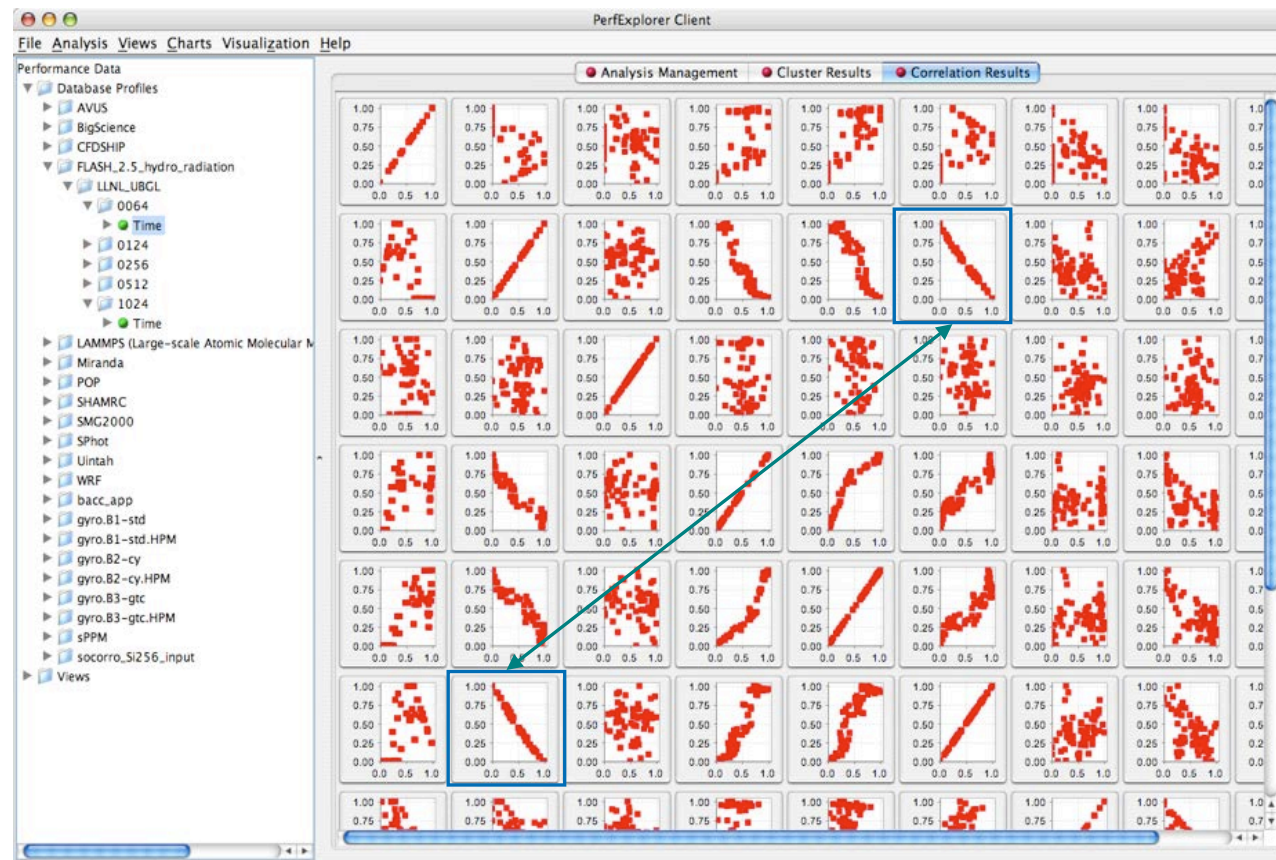
PerfExplorer - Correlation Analysis (Flash)

- -0.995 indicates strong, negative relationship
- As CALC_CUT_ BLOCK_CONTRIBUTIONS() increases in execution time, MPI_Barrier() decreases

PerfExplorer - Comparative Analysis

- Relative speedup, efficiency
  - total runtime, by event, one event, by phase
- Breakdown of total runtime
- Group fraction of total runtime
- Correlating events to total runtime
- Timesteps per second

PerfExplorer - Interface
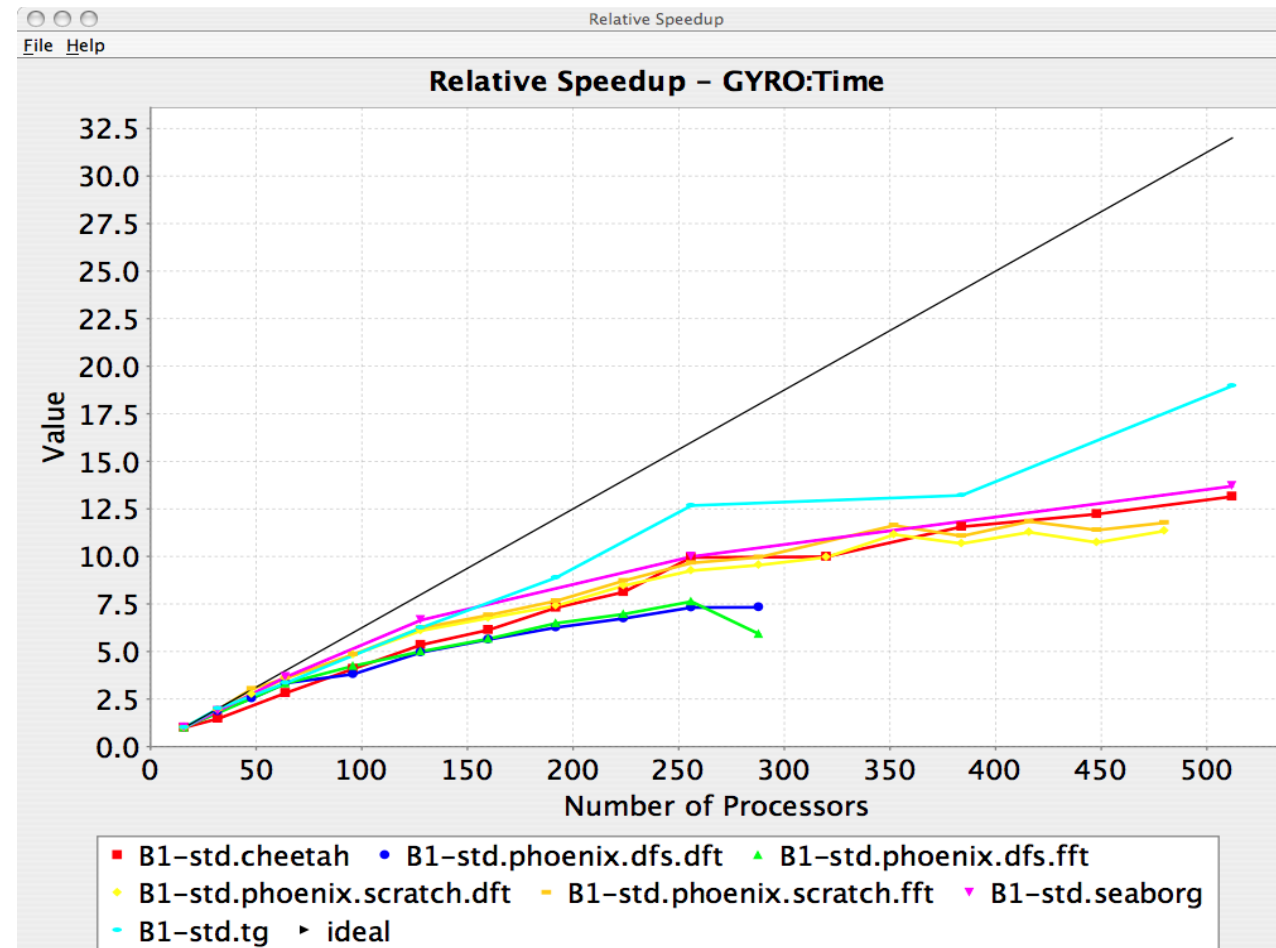
PerfExplorer - Interface
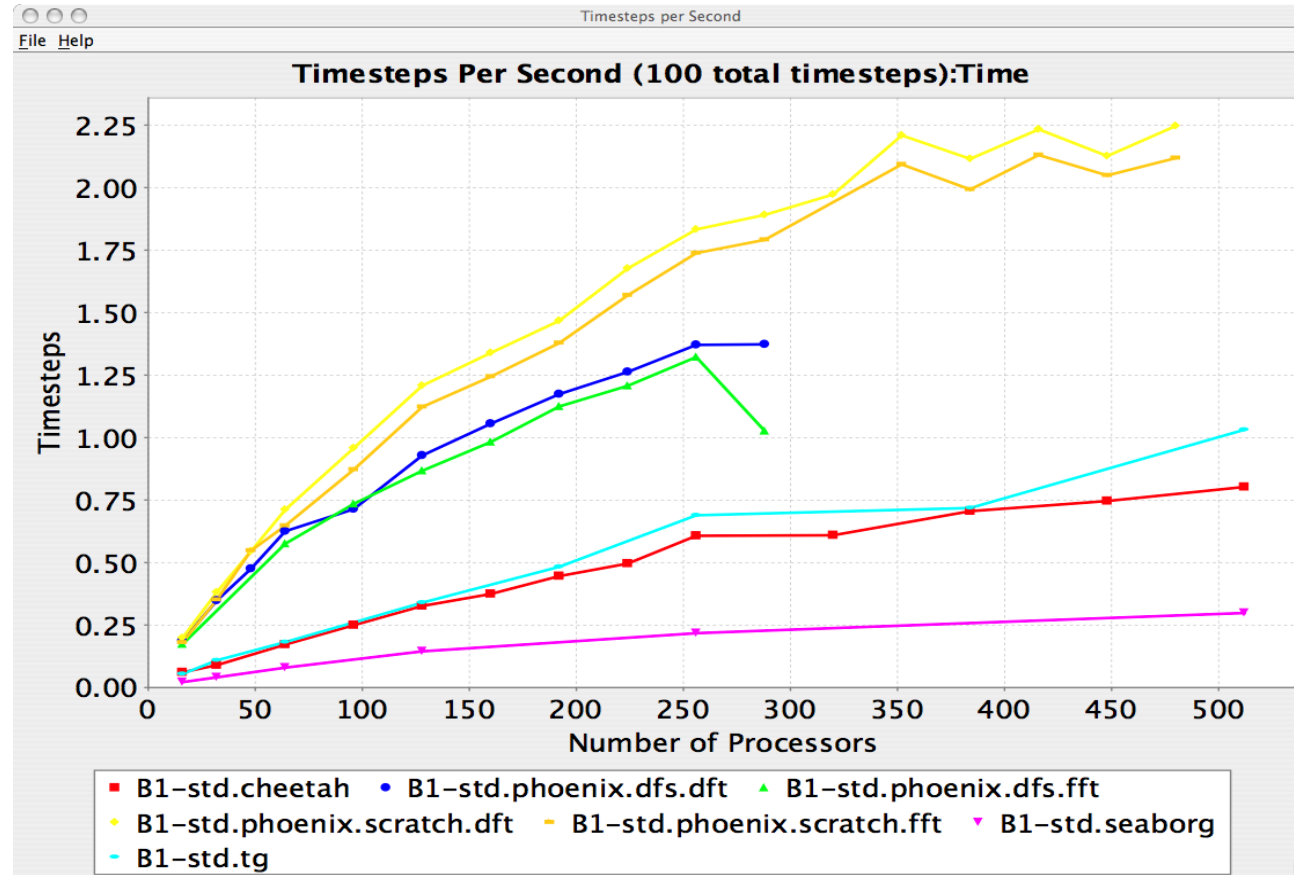
PerfExplorer - Relative Efficiency Plots

PerfExplorer - Relative Efficiency by Routine
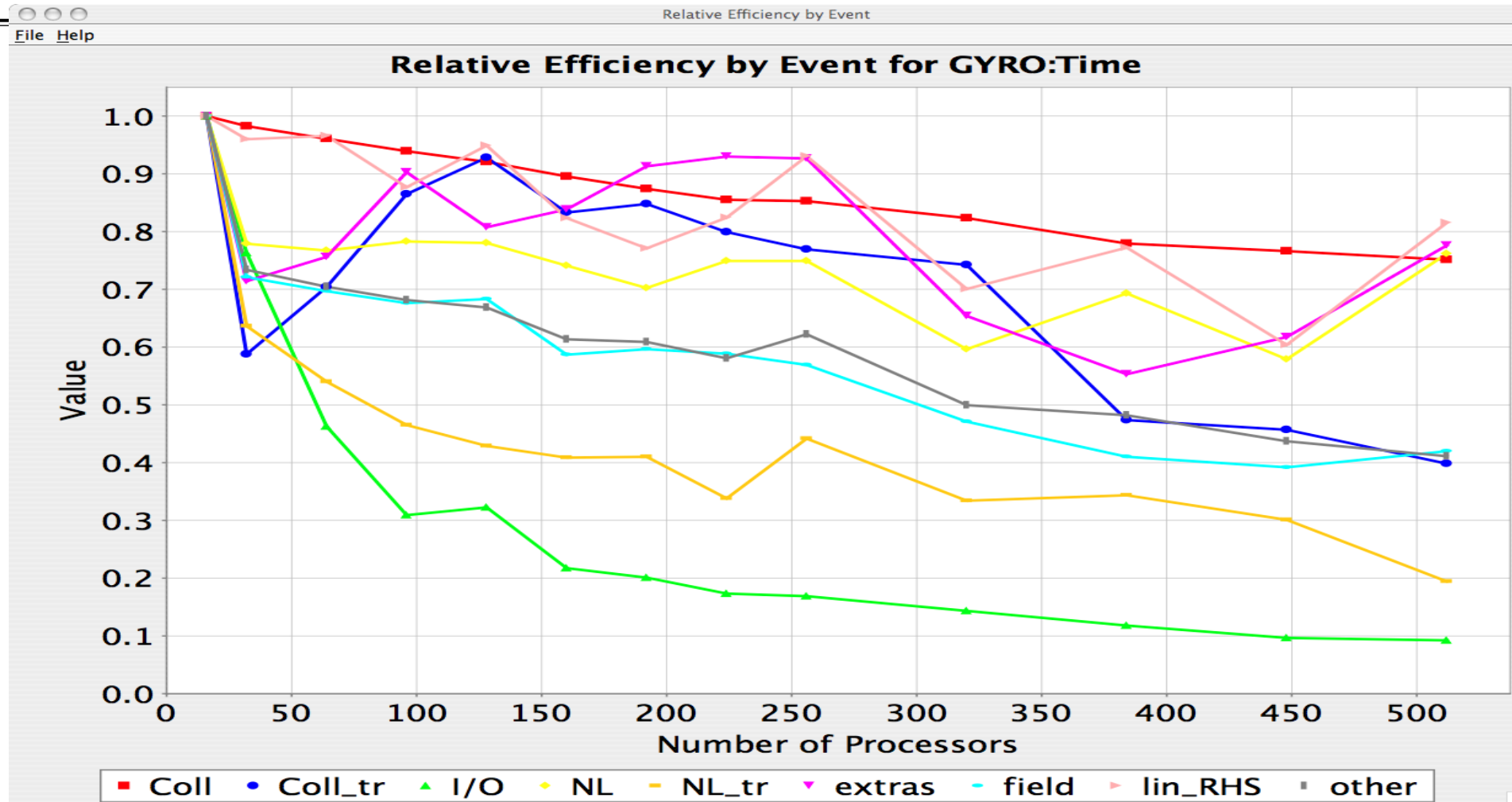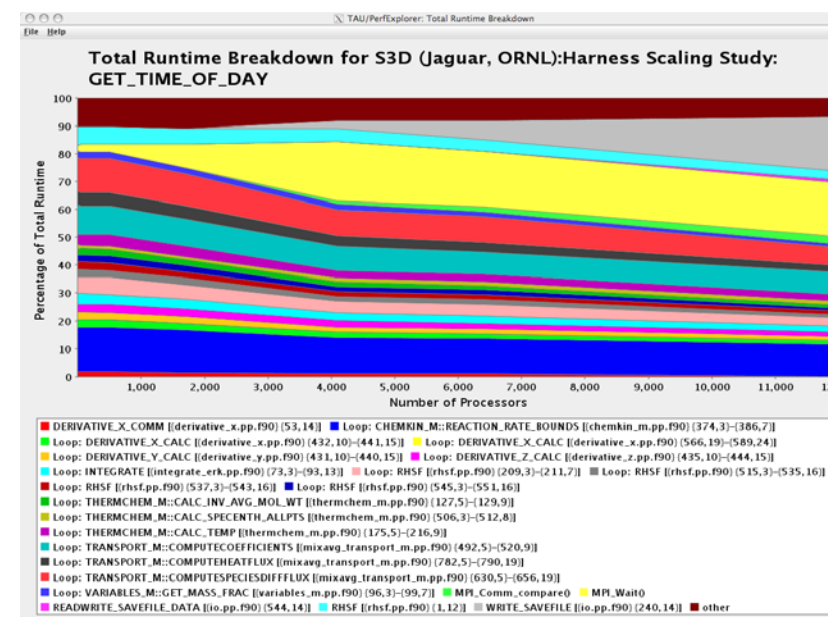
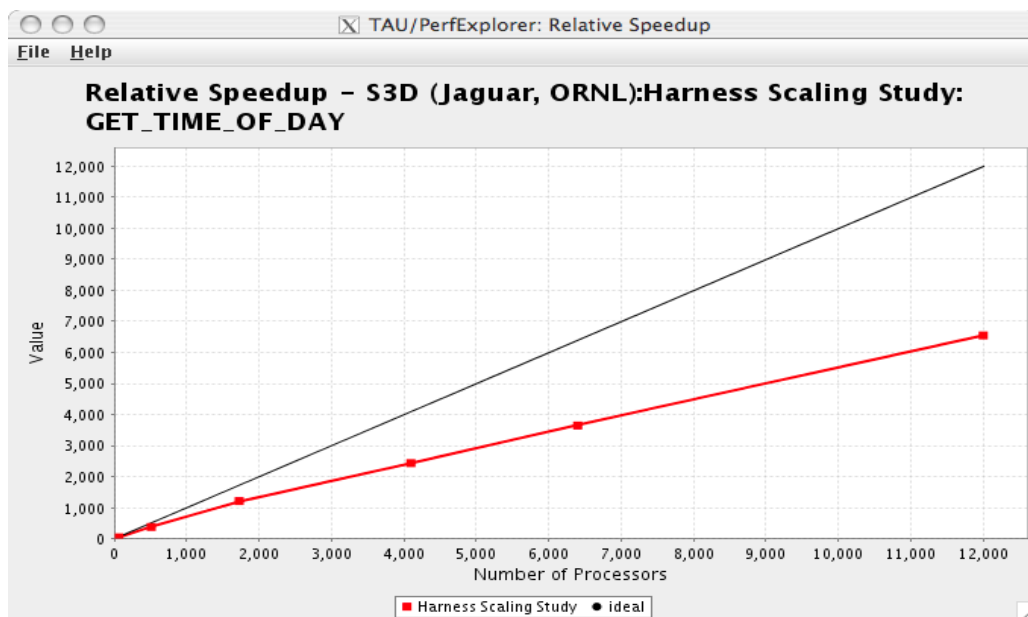# PerfExplorer - Relative Speedup

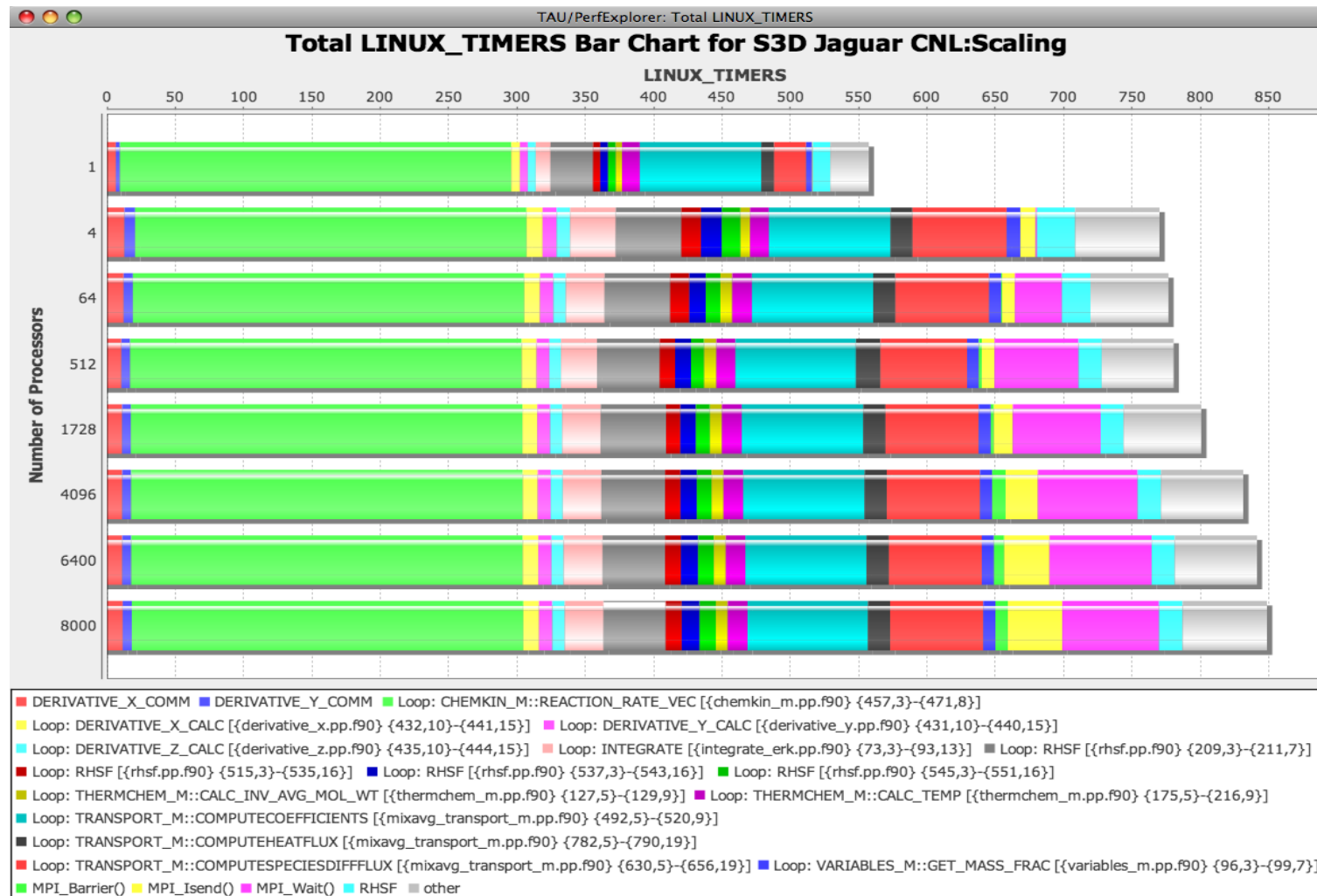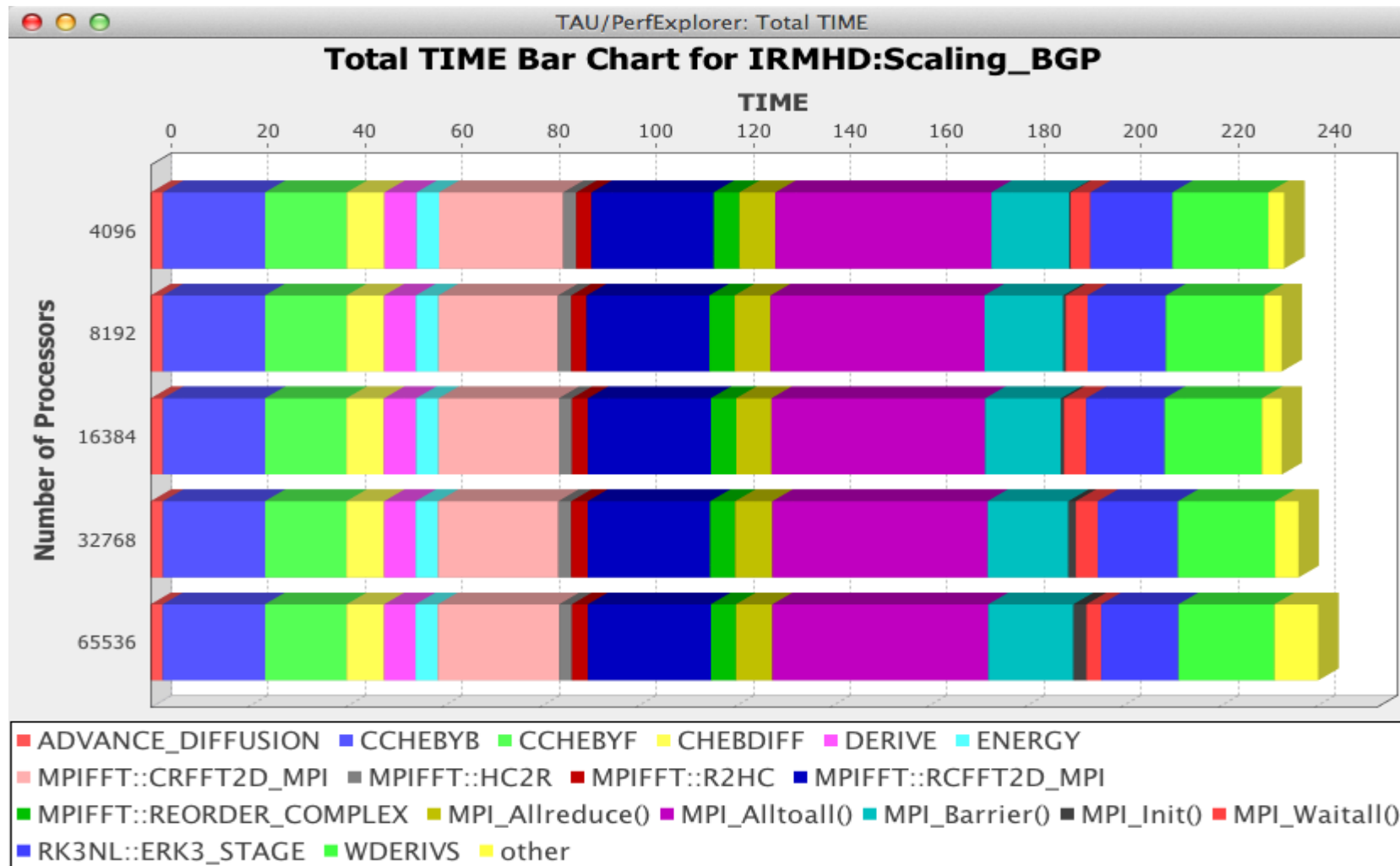# PerfExplorer - Timesteps Per Second

Evaluate Scalability

- Goal: How does my application scale? What bottlenecks occur at what core counts?
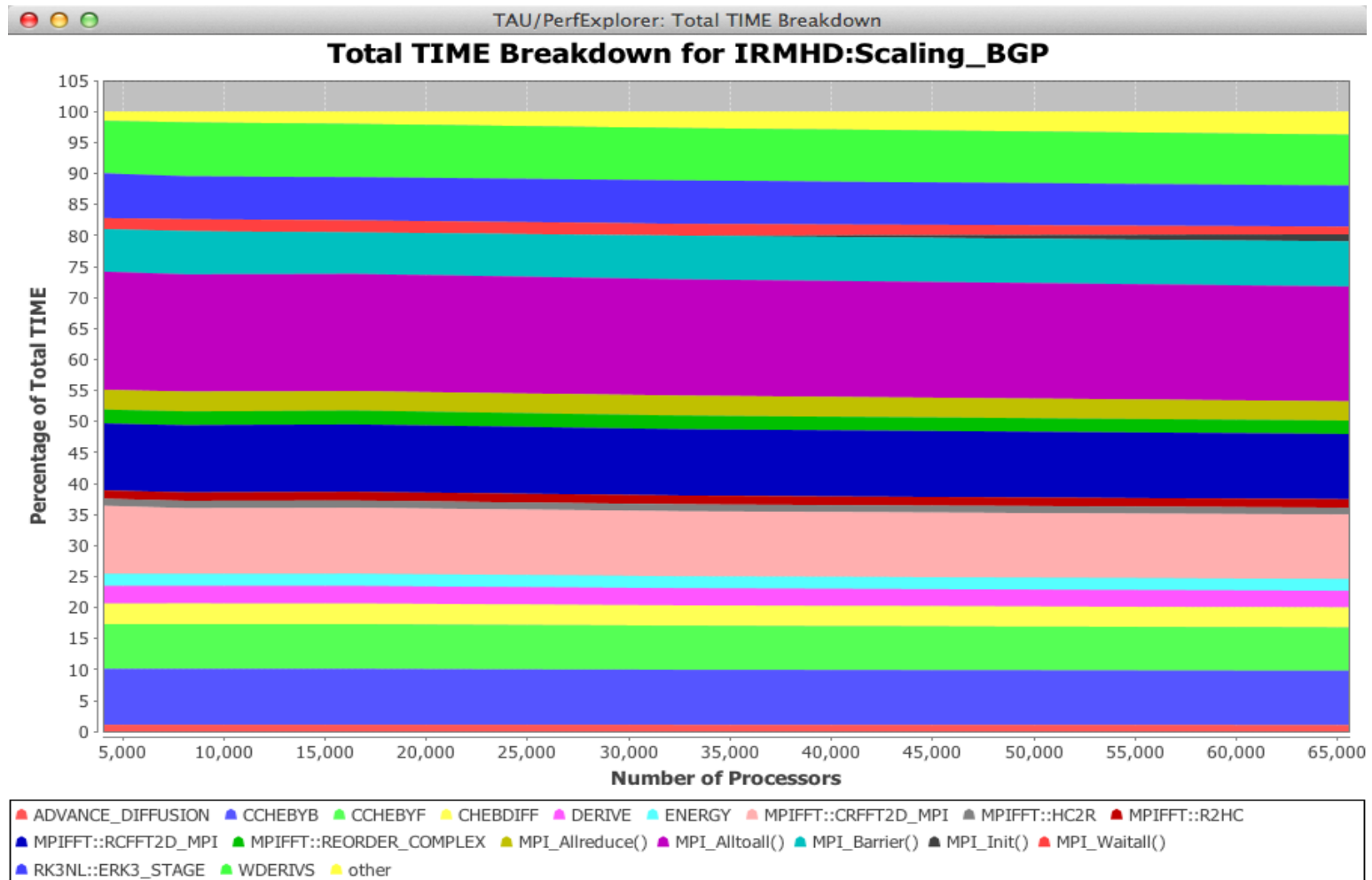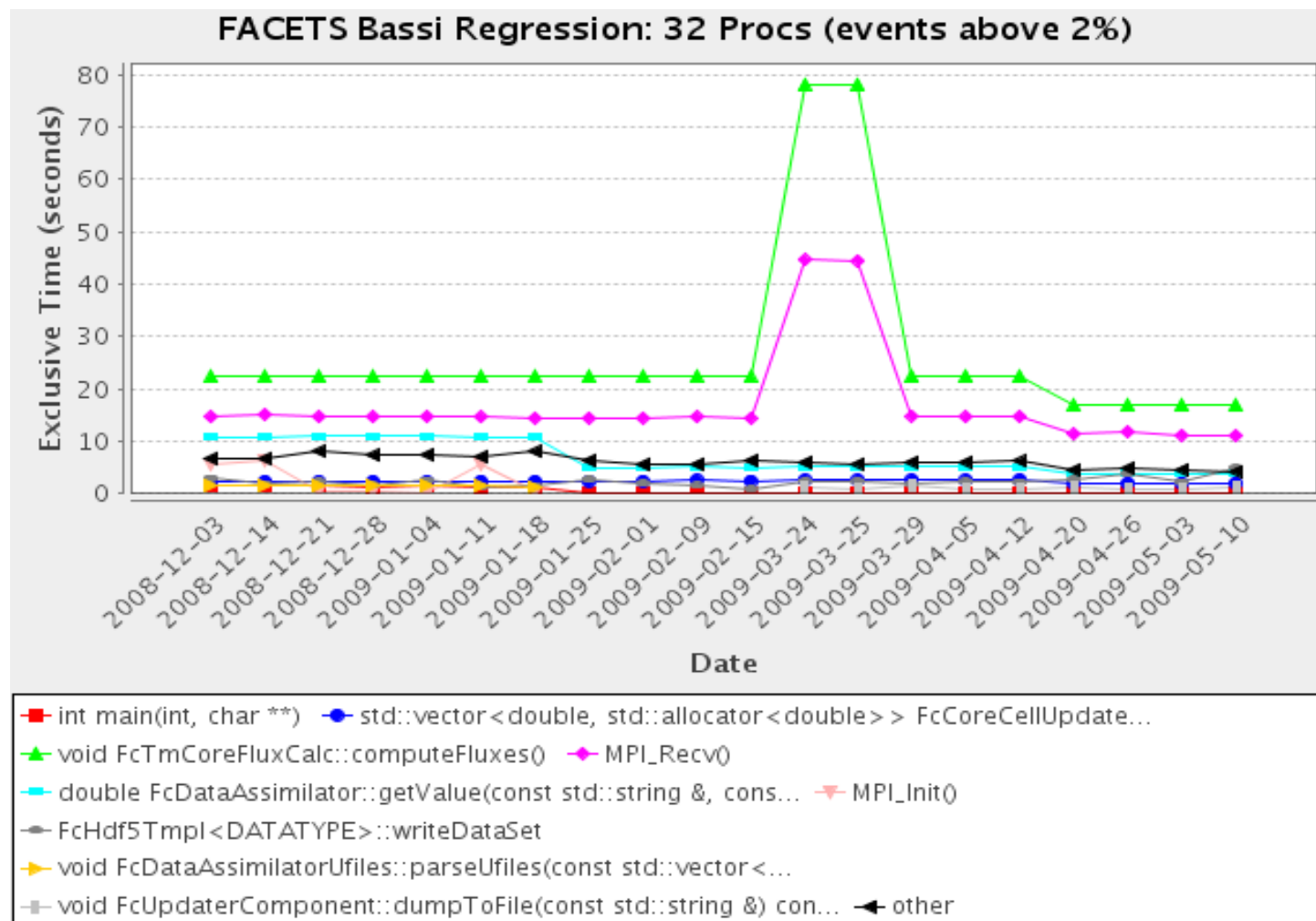- Load profiles in taudb database and examine with PerfExplorer

Evaluate Scalability

PerfExplorer

# PerfExplorer

Performance Regression Testing

Support Acknowledgments

- U.S. Department of Energy (DOE)
  - Office of Science
  - PNNL, LBL, ORNL
  - ASC/NNSA, Tri-labs (LLNL,LANL, SNL)
- U.S. Department of Defense (DoD)
  - HPC Modernization Office (HPCMO)
- NSF Software Development for Cyberinfrastructure (SDCI)
- Juelich Supercomputing Center, NIC
- Argonne National Laboratory
- T.U. Dresden
- ParaTools, Inc.

Download TAU from U. Oregon

Tuning and Analysis Utilities

http://tau.uoregon.edu
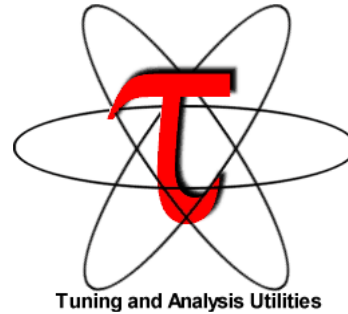
http://www.hpclinux.com [LiveDVD, OVA]

Free download, open source, BSD license