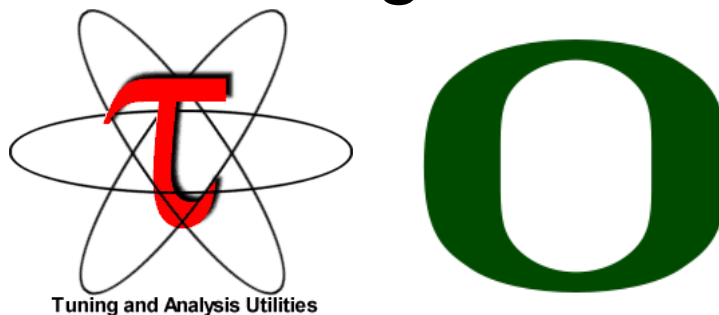




## Profile Data Mining with PerfExplorer

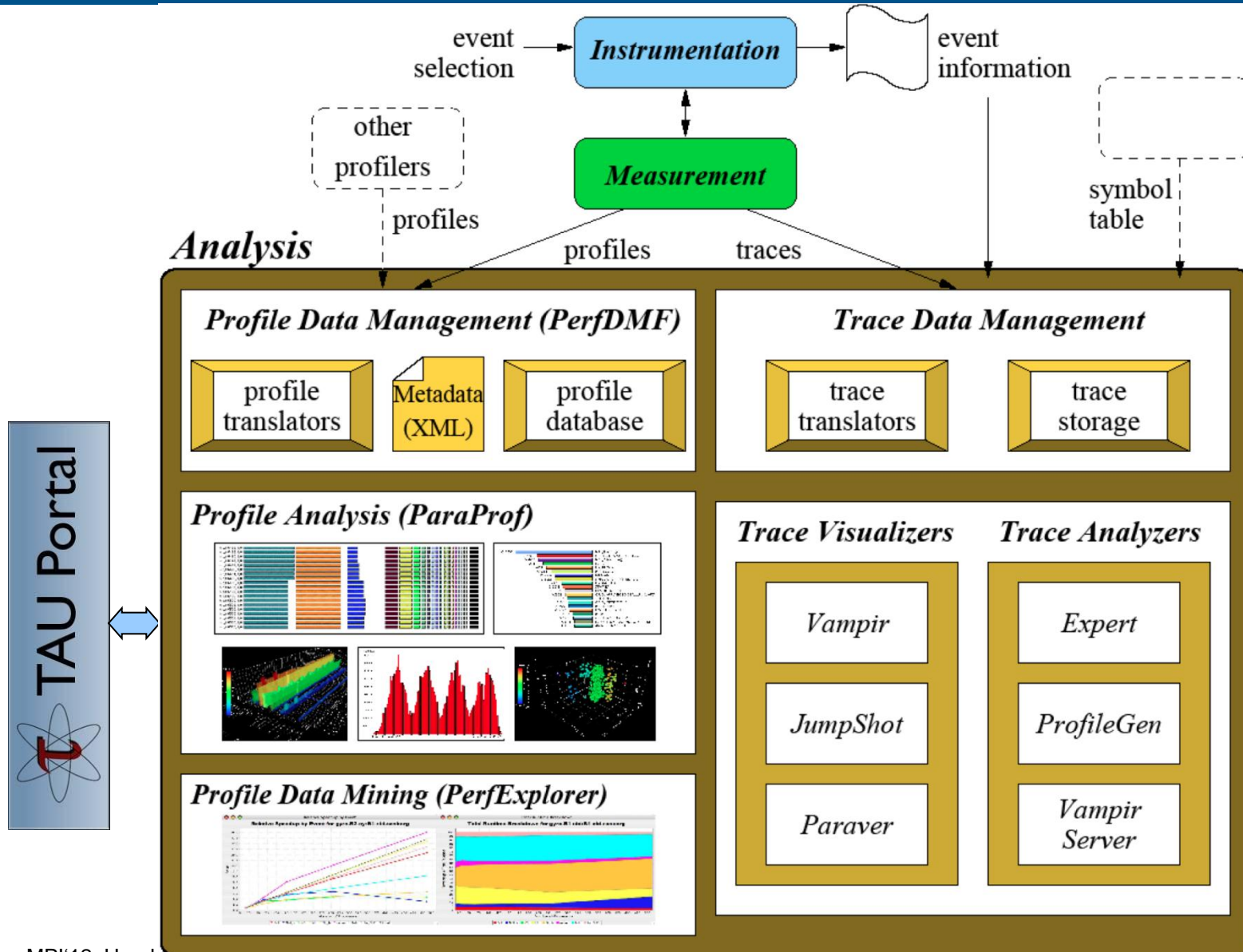


Tuning and Analysis Utilities

Sameer Shende

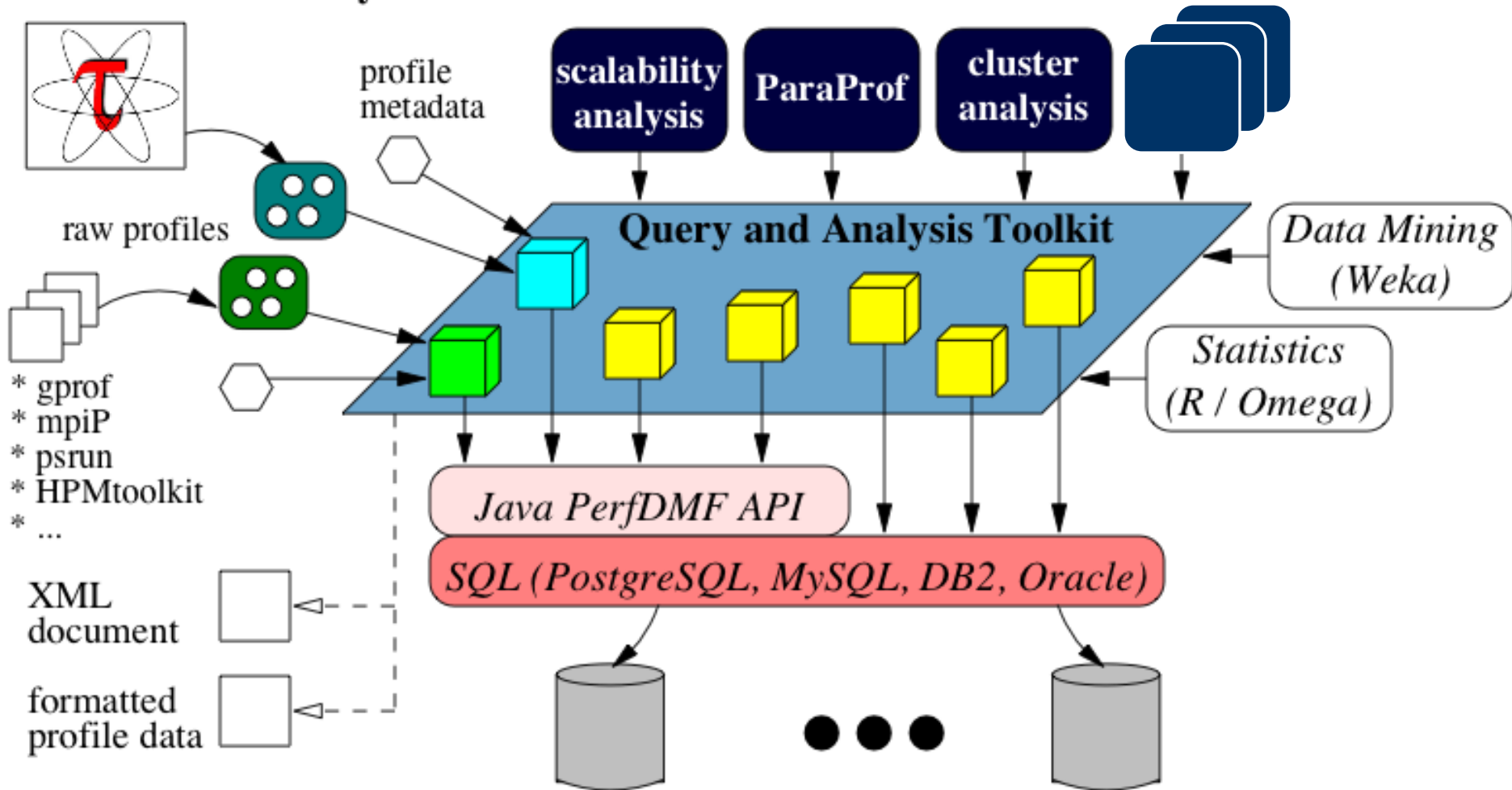
Performance Research Lab, University of Oregon

<http://TAU.uoregon.edu>



## TAU Performance System

## Performance Analysis Programs



- **Configure PerfDMF (Done by each user)**

- % perfdmf\_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/perfdmf.cfg file

- **Configure PerfExplorer (Done by each user)**

- % perfexplorer\_configure

- **Execute PerfExplorer**

- % perfexplorer

```
% wget http://tau.uoregon.edu/data.tgz (Contains CUBE profiles from Score-P)
% perfdmf_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
  perfdmf_loadtrial -a "app" -x "experiment" -n "name" file.ppk
Then,
% perfexplorer
(Select experiment, Menu: Charts -> Speedup)
```

- Development of the TAU portal
  - Common repository for collaborative data sharing
  - Profile uploading, downloading, user management
  - Paraprof, PerfExplorer can be launched from the portal using Java Web Start (no TAU installation required)
- Portal URL  
<http://tau.nic.uoregon.edu>

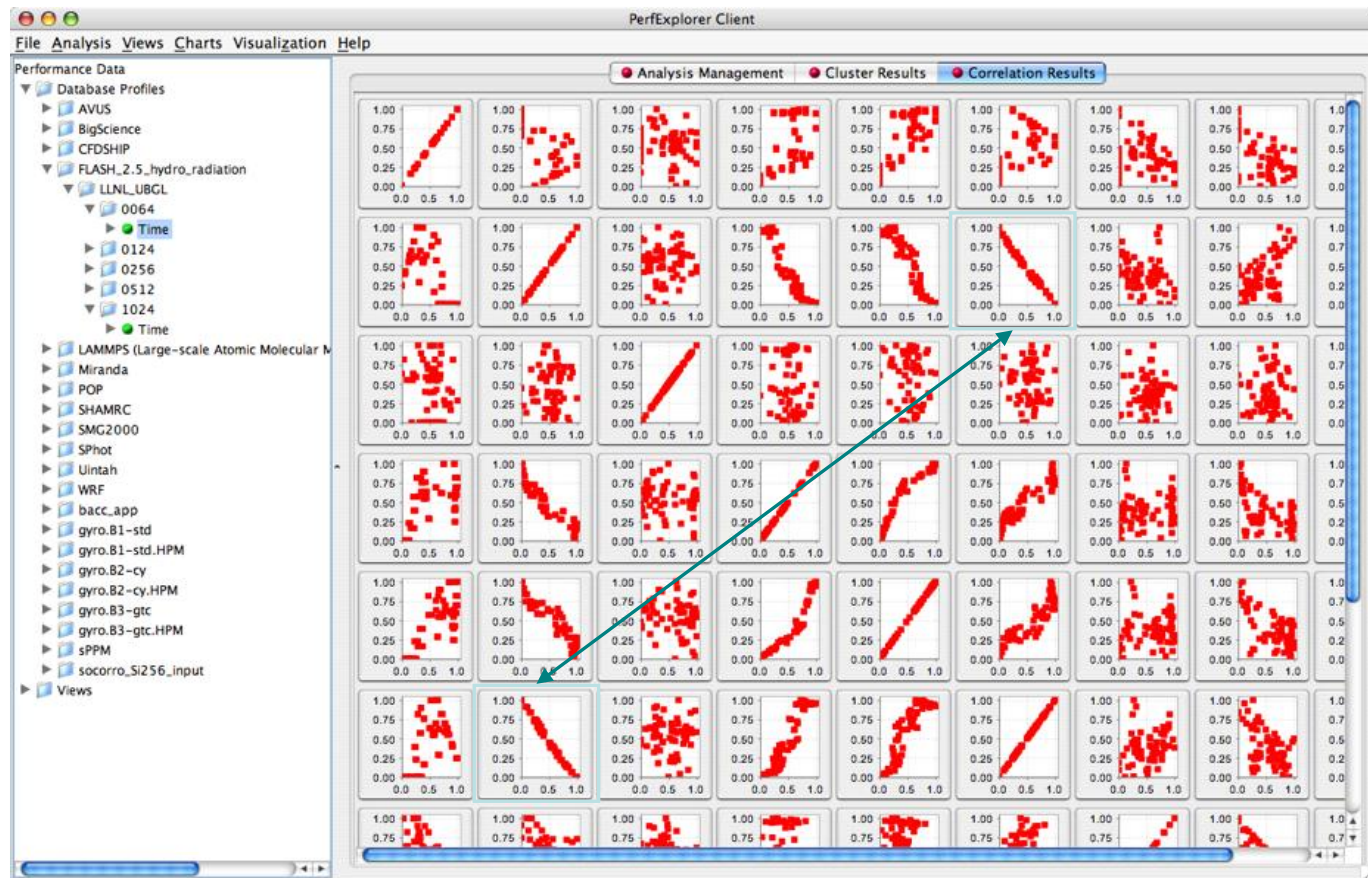
- Performance knowledge discovery framework
  - Data mining analysis applied to parallel performance data
    - comparative, clustering, correlation, dimension reduction, ...
  - Use the existing TAU infrastructure
    - TAU performance profiles, PerfDMF
  - Client-server based system architecture
-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

- 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

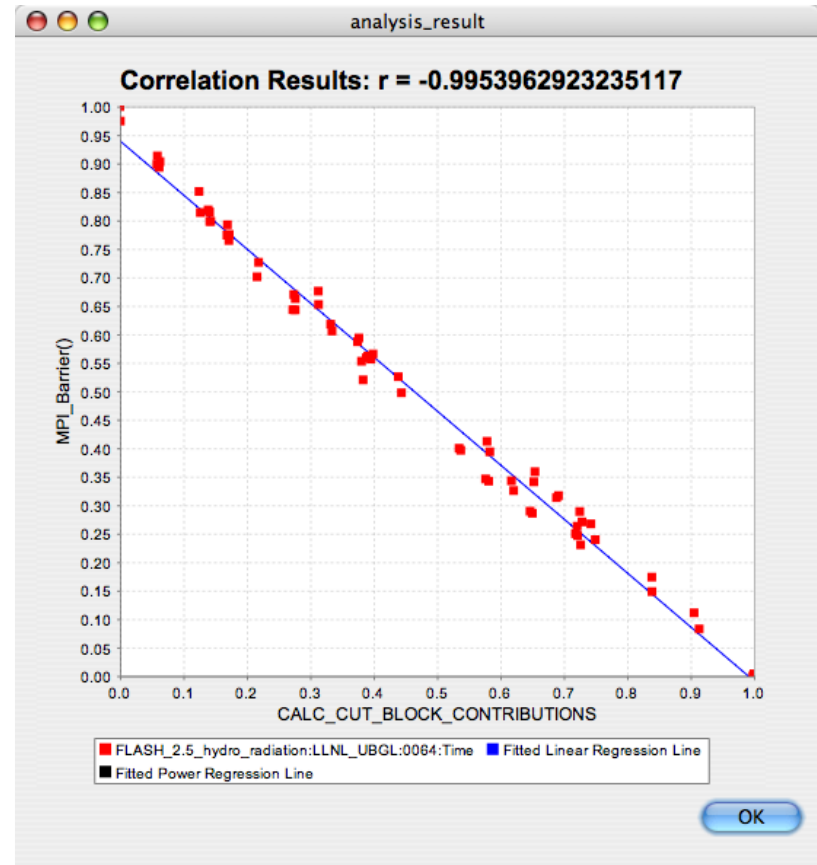




- Describes strength and direction of a linear relationship between two variables (events) in the data

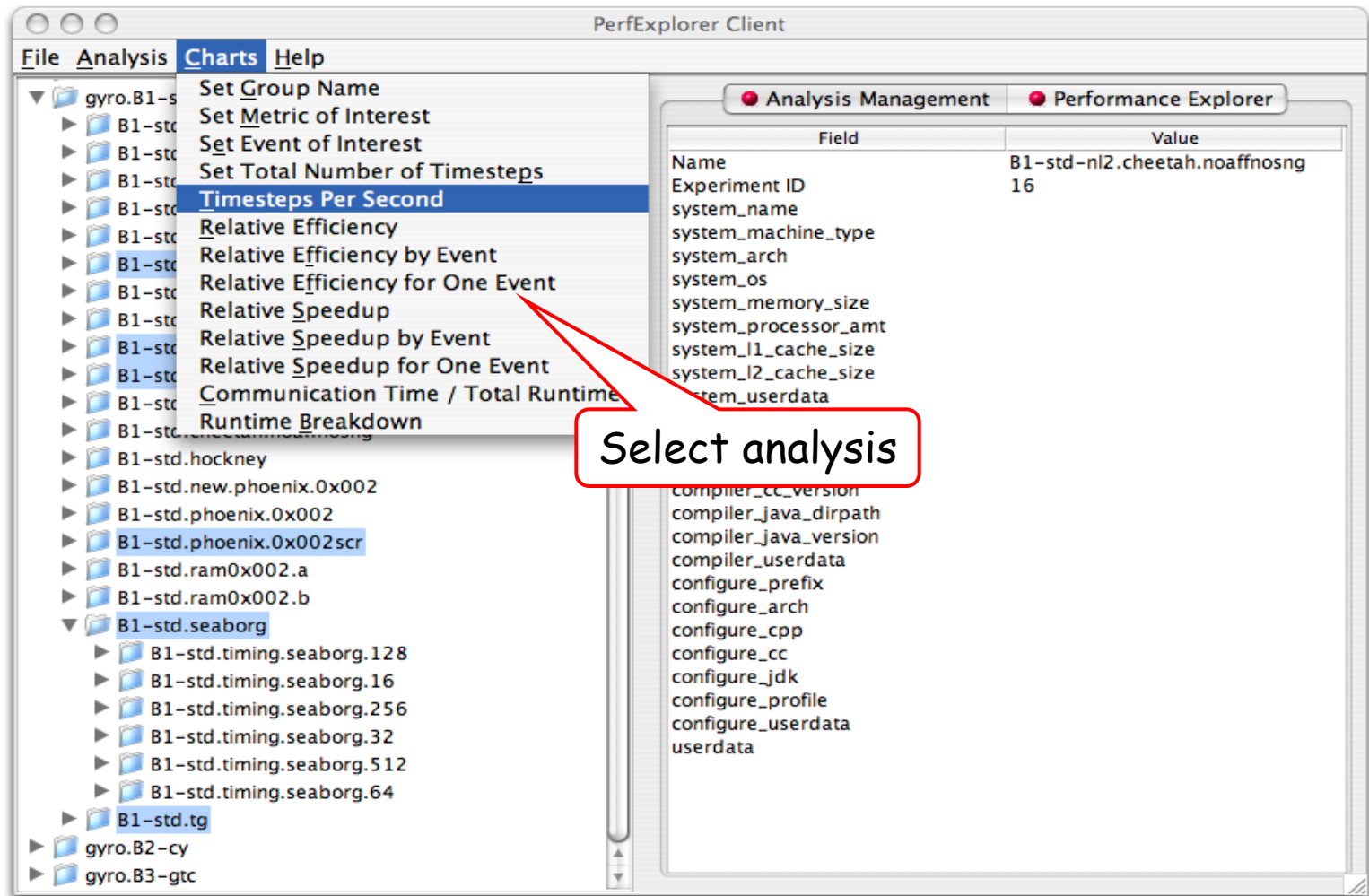


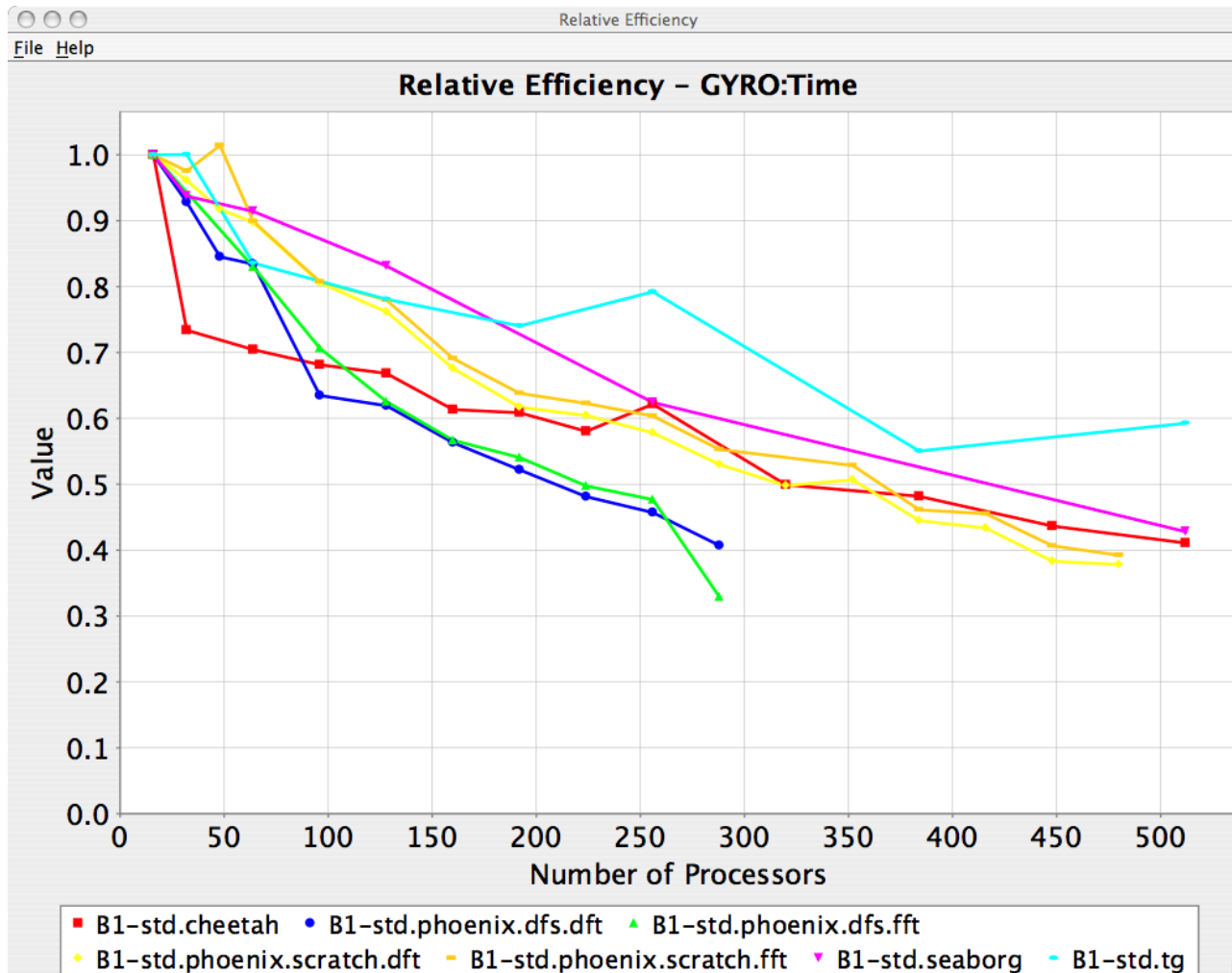
- -0.995 indicates strong, negative relationship
- As CALC\_CUT\_BLOCK\_CONTRIBUTIONS increases in execution time, MPI\_Barrier() decreases

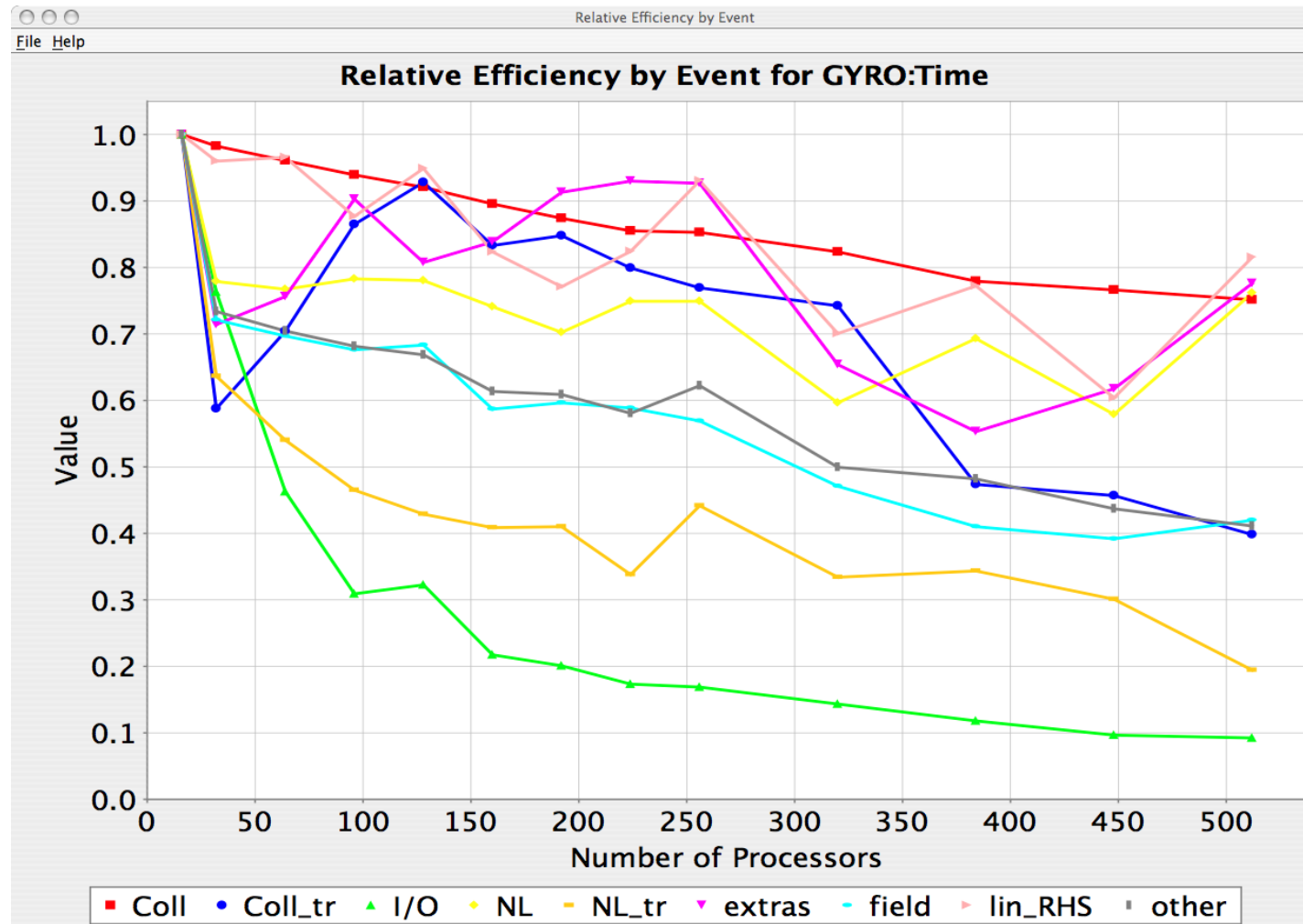


- Relative speedup, efficiency
  - total runtime, by event, one event, by phase
- Breakdown of total runtime
- Group fraction of total runtime
- Correlating events to total runtime
- Timesteps per second

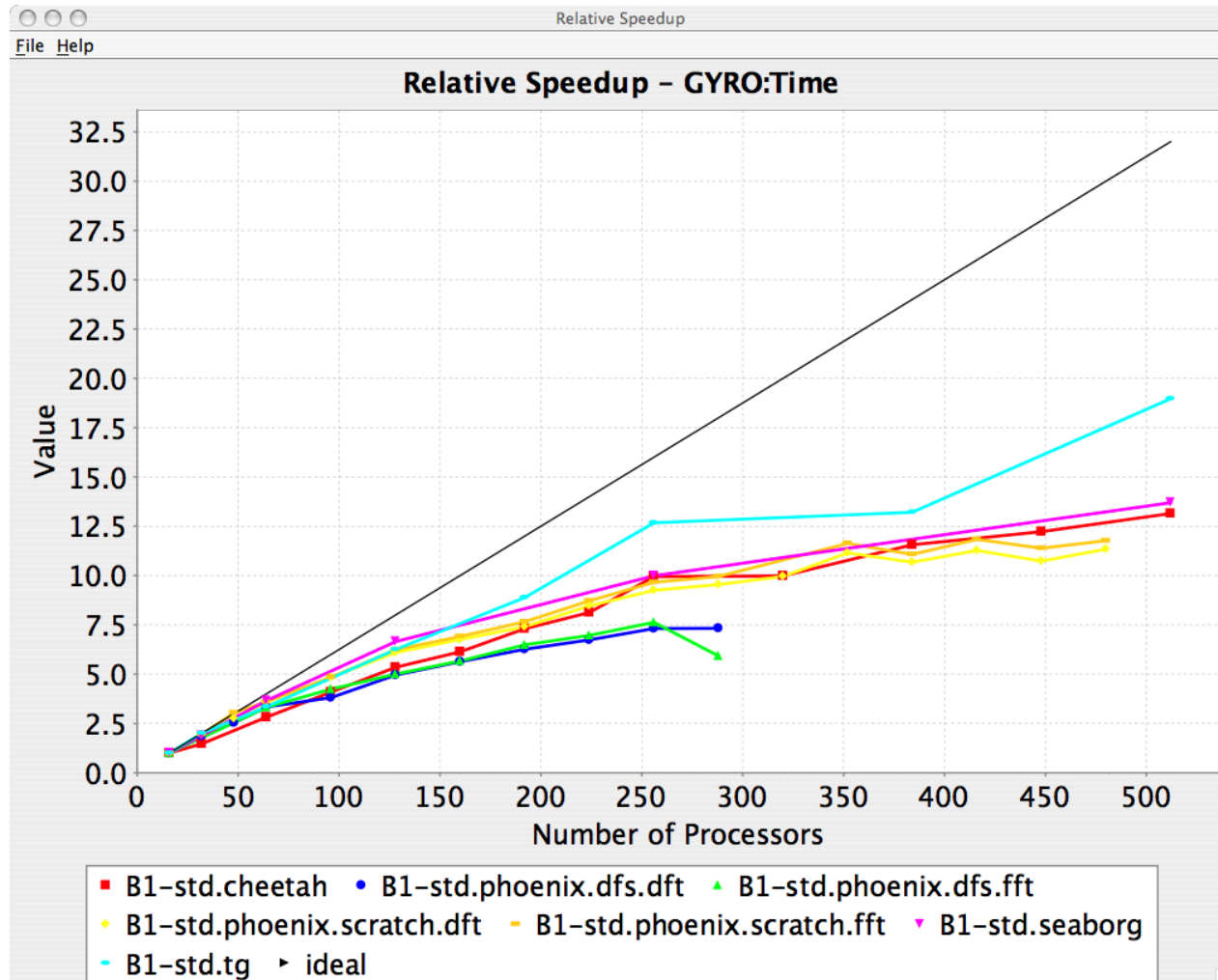


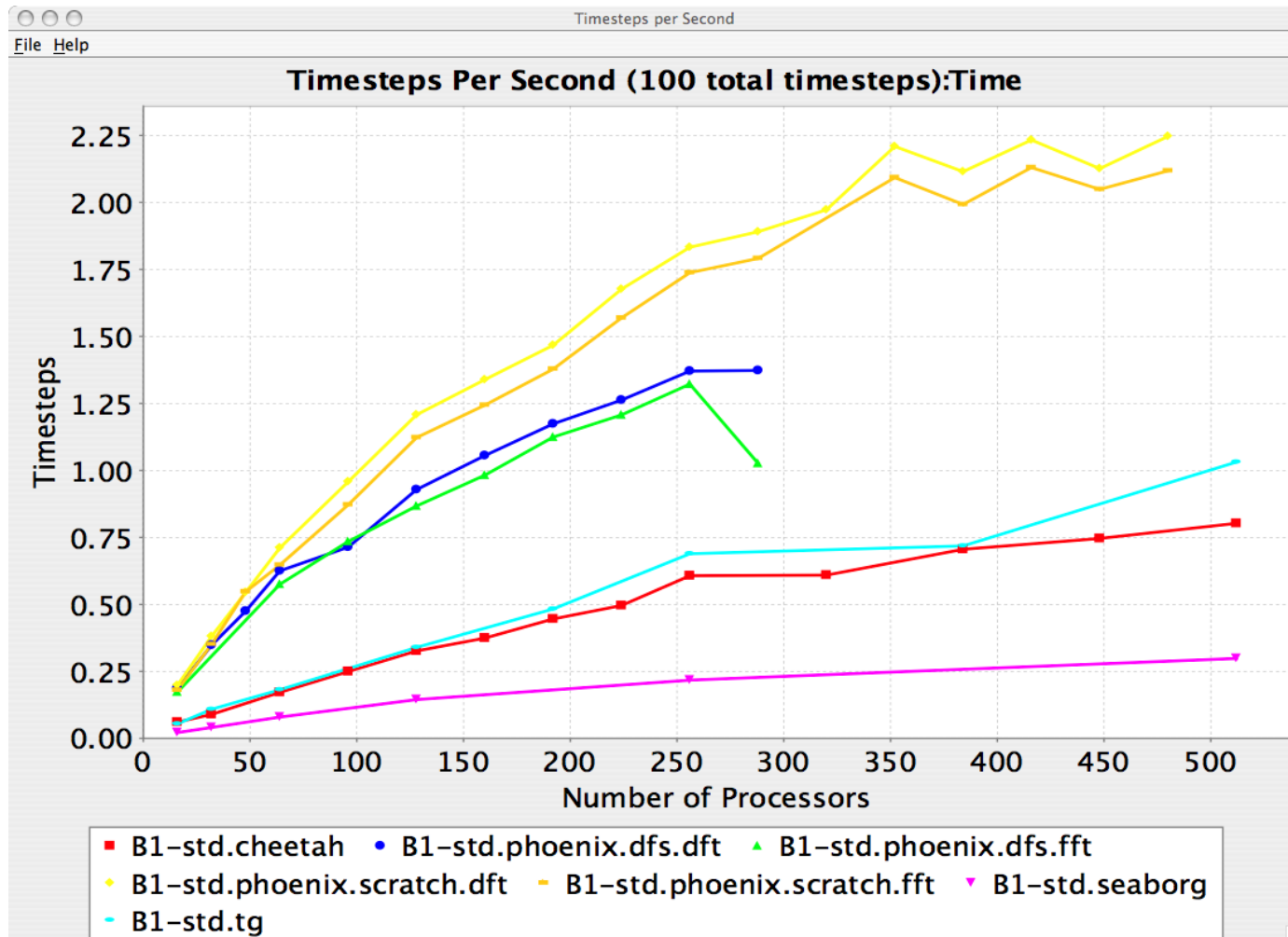






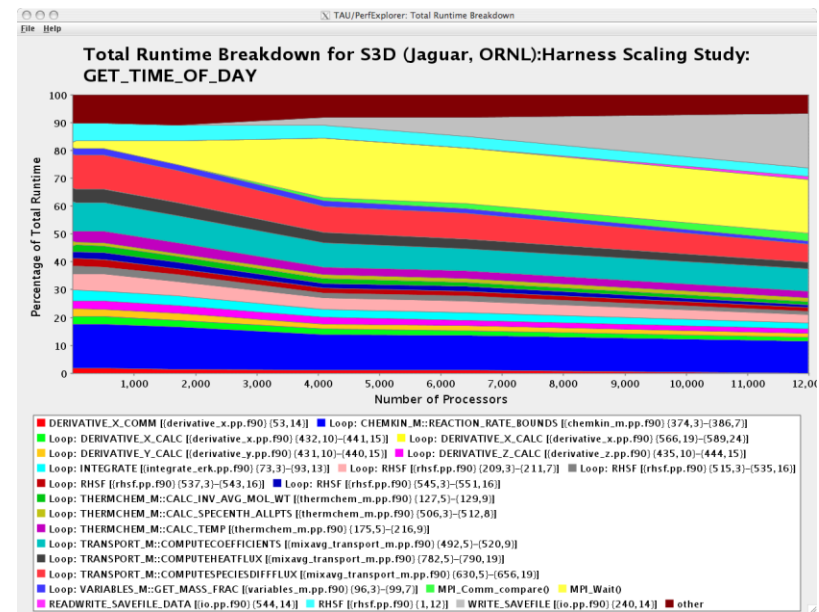
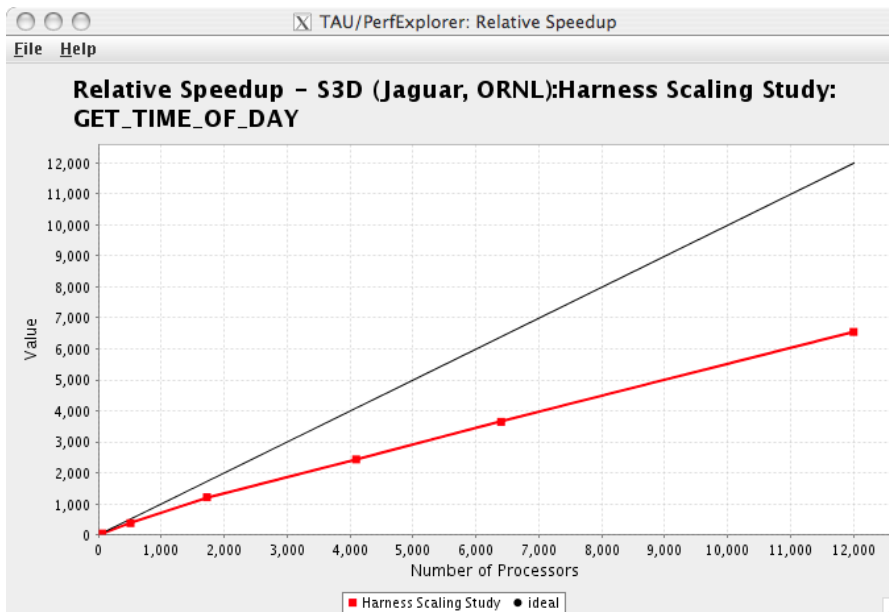




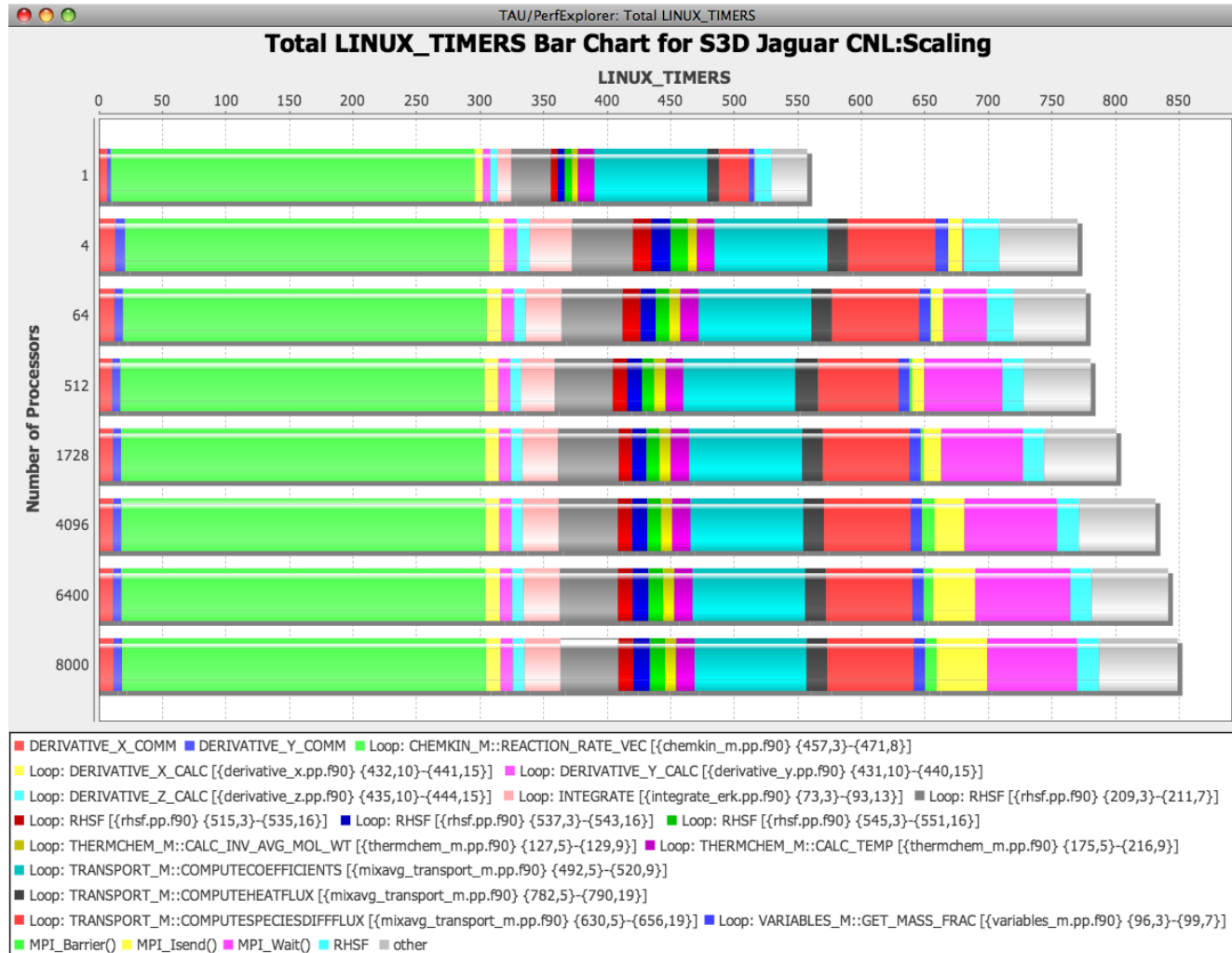


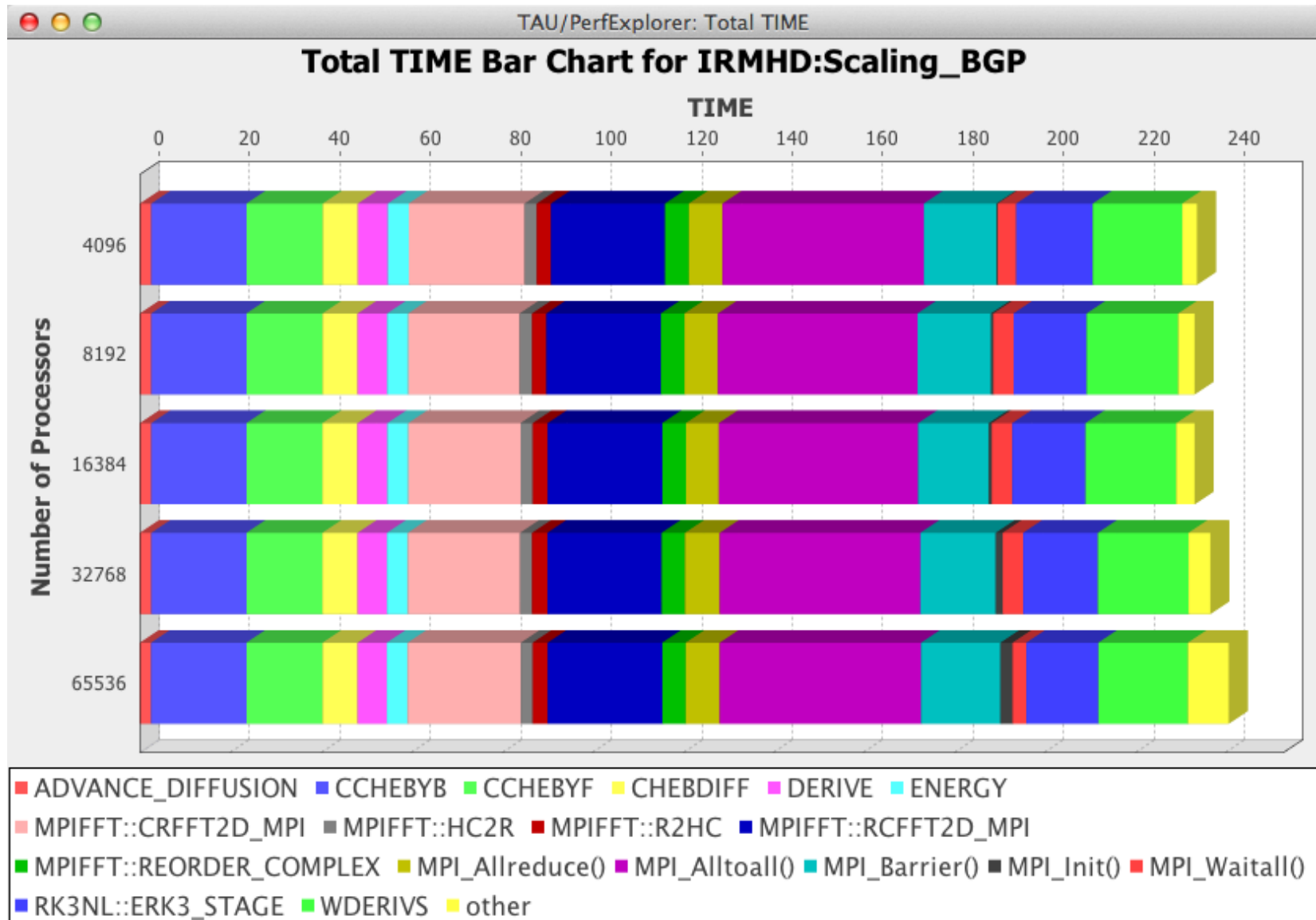
# Usage Scenarios: Evaluate Scalability HPS

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



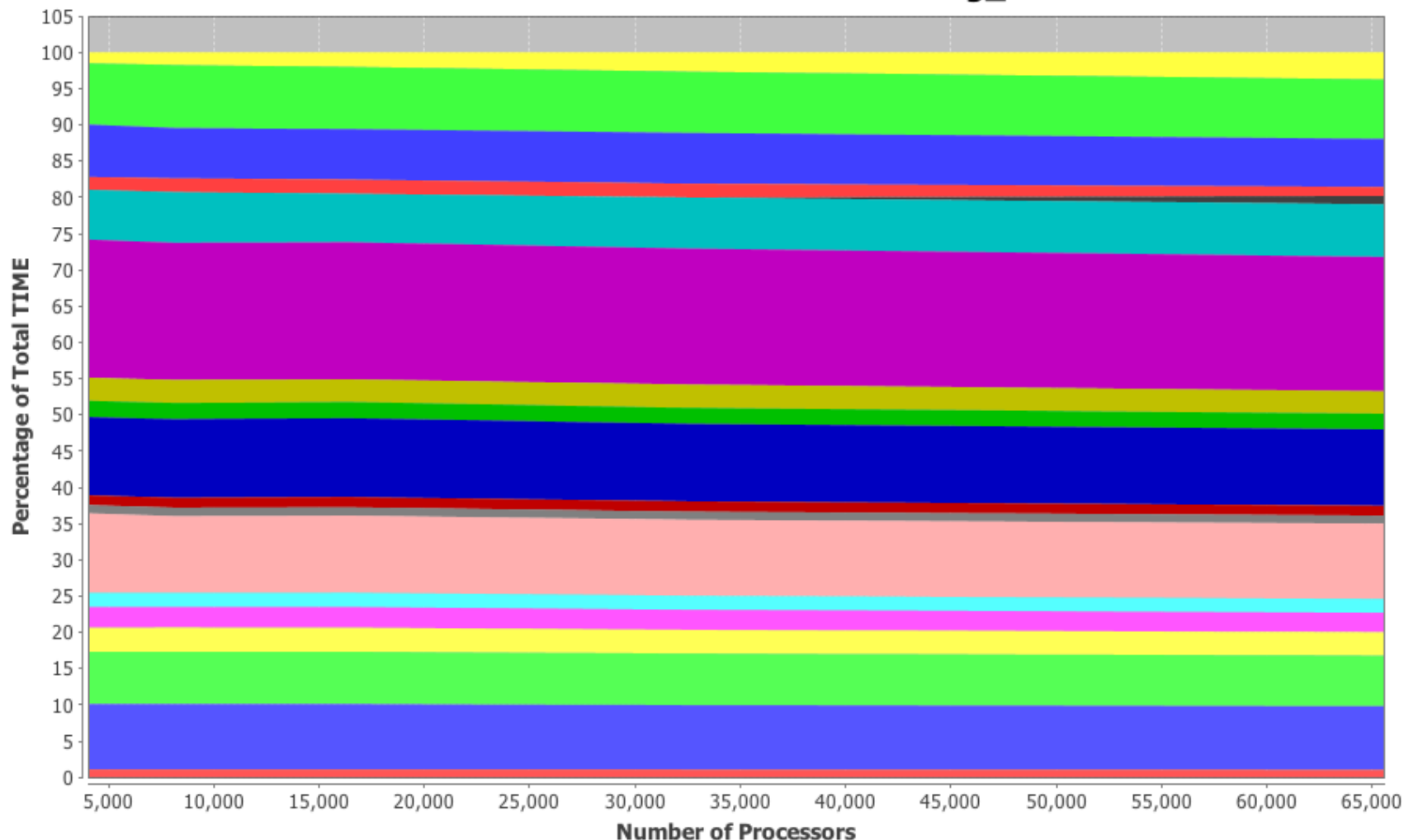
# Usage Scenarios: Evaluate Scalability **TAU/HPS**





TAU/PerfExplorer: Total TIME Breakdown

## Total TIME Breakdown for IRMHD:Scaling\_BGP



- ADVANCE\_DIFFUSION
- CCHEBYB
- CCHEBYF
- CHEBDIFF
- DERIVE
- ENERGY
- MPIFFT::CRFFT2D\_MPI
- MPIFFT::HC2R
- MPIFFT::R2HC
- MPIFFT::RCFFT2D\_MPI
- MPIFFT::REORDER\_COMPLEX
- MPI\_Allreduce()
- MPI\_Alltoall()
- MPI\_Barrier()
- MPI\_Init()
- MPI\_Waitall()
- RK3NL::ERK3\_STAGE
- WDERIVS
- other

# Performance Regression Testing VI-HPS

