Performance Tools for Petascale Systems

The BMBF project SILC

Zellescher Weg 12

Willers-Bau A114

Tel. +49 351 - 463 - 38323

Andreas Knüpfer (andreas.knuefer@tu-dresden.de)

ZIH
Zentrum für Informationsdienste und Hochleistungsrechnen

Overview

- Performance tools for HPC
  - Periscope
  - Scalasca
  - Vampir Tools
- Project goals
  - The tool's perspective
  - The user's perspective
- Status and outlook

Overview

- **Performance tools for HPC**
  - **Periscope**
  - **Scalasca**
  - **Vampir Tools**
- Project goals
  - The tool's perspective
  - The user's perspective
- Status and outlook

Why HPC Performance Tools matter

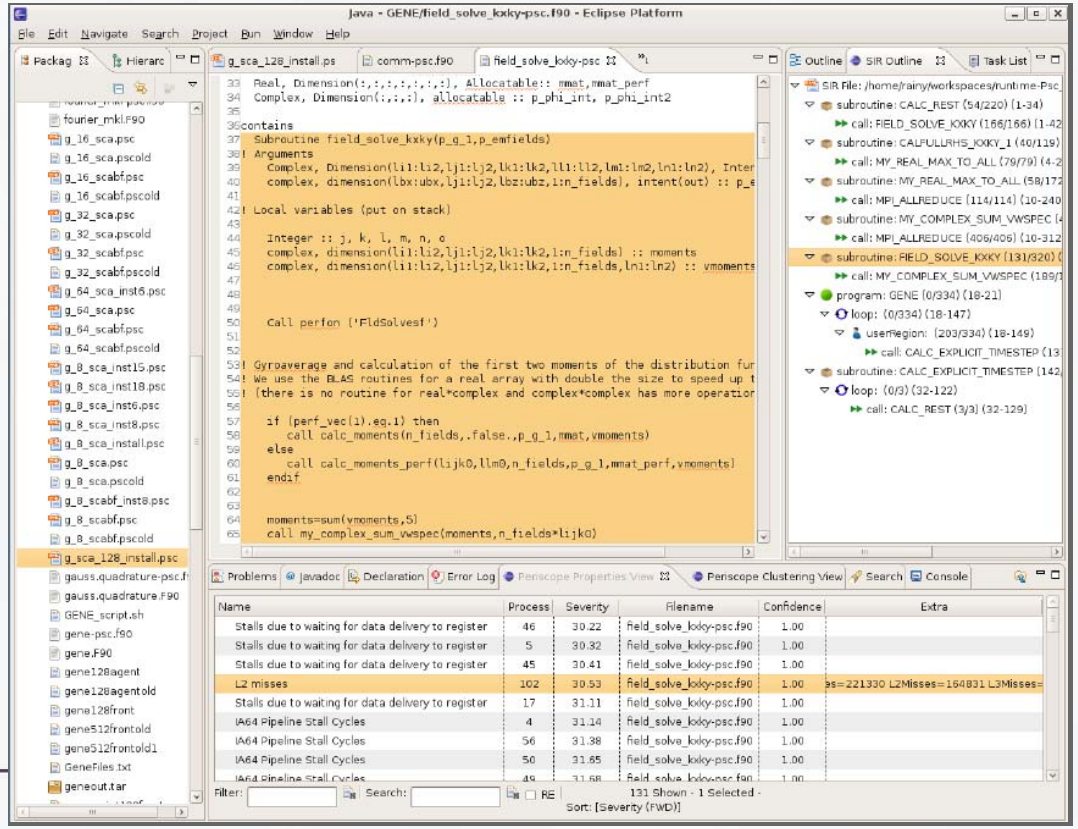
- HPC performance analysis and optimization is hard
- So is the scalability challenge
- Need to “look inside parallel application runs“

- Use appropriate tools!
- No do-it-yourself solutions with **`MPI_WTime`**, **`printf`**, and **`write`**

- There are specific support tools:
  - For MPI, OpenMP, Pthreads, and hybrid applications
  - Experimental support for Cell, OpenCL, CUDA
  - Including timer management, performance counters, I/O monitoring, etc.
  - For all relevant HPC platforms

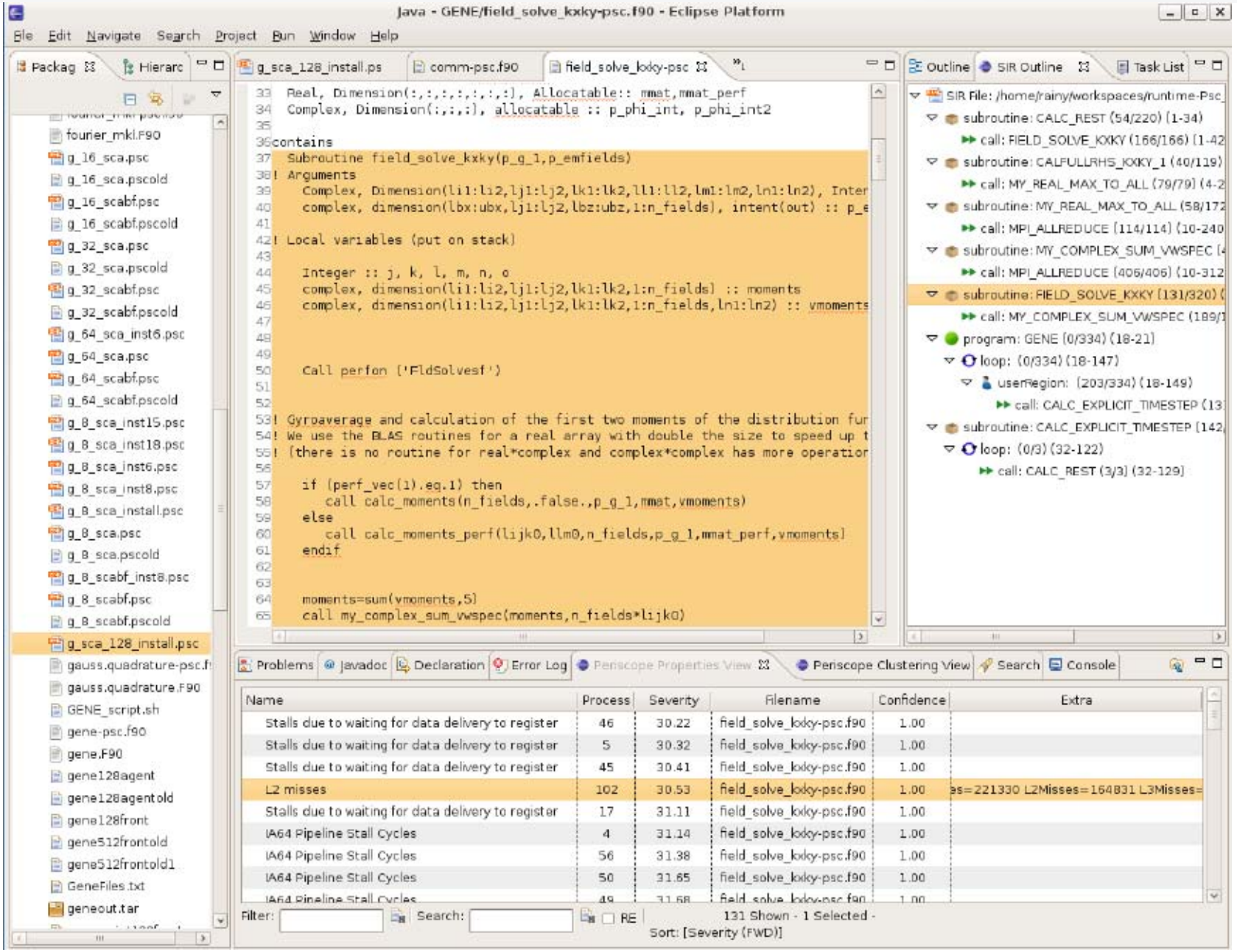
Periscope

- Automatic online performance analysis
  - Start with performance flaw hypotheses
  - Investigate hypotheses with measurement during one/few iterations
  - Revise hypotheses after each step
- For parallel applications
- Local and global evaluation
- Run for successive iterations, restart application if needed
- Eclipse integration



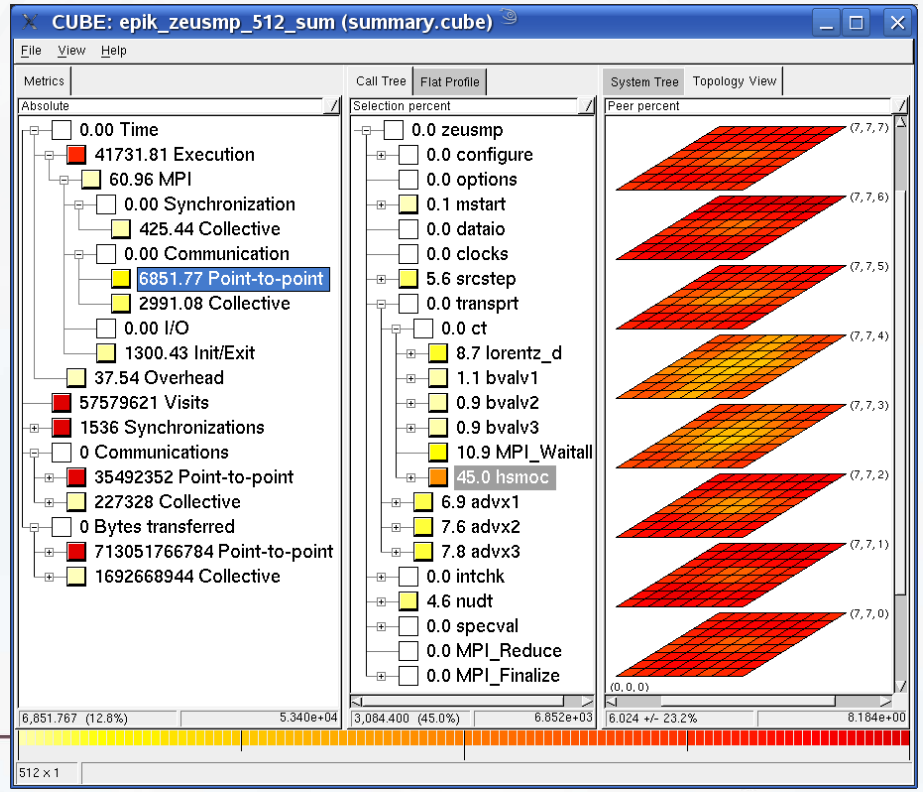
TECHNISCHE UNIVERSITÄT DRESDEN

Periscope

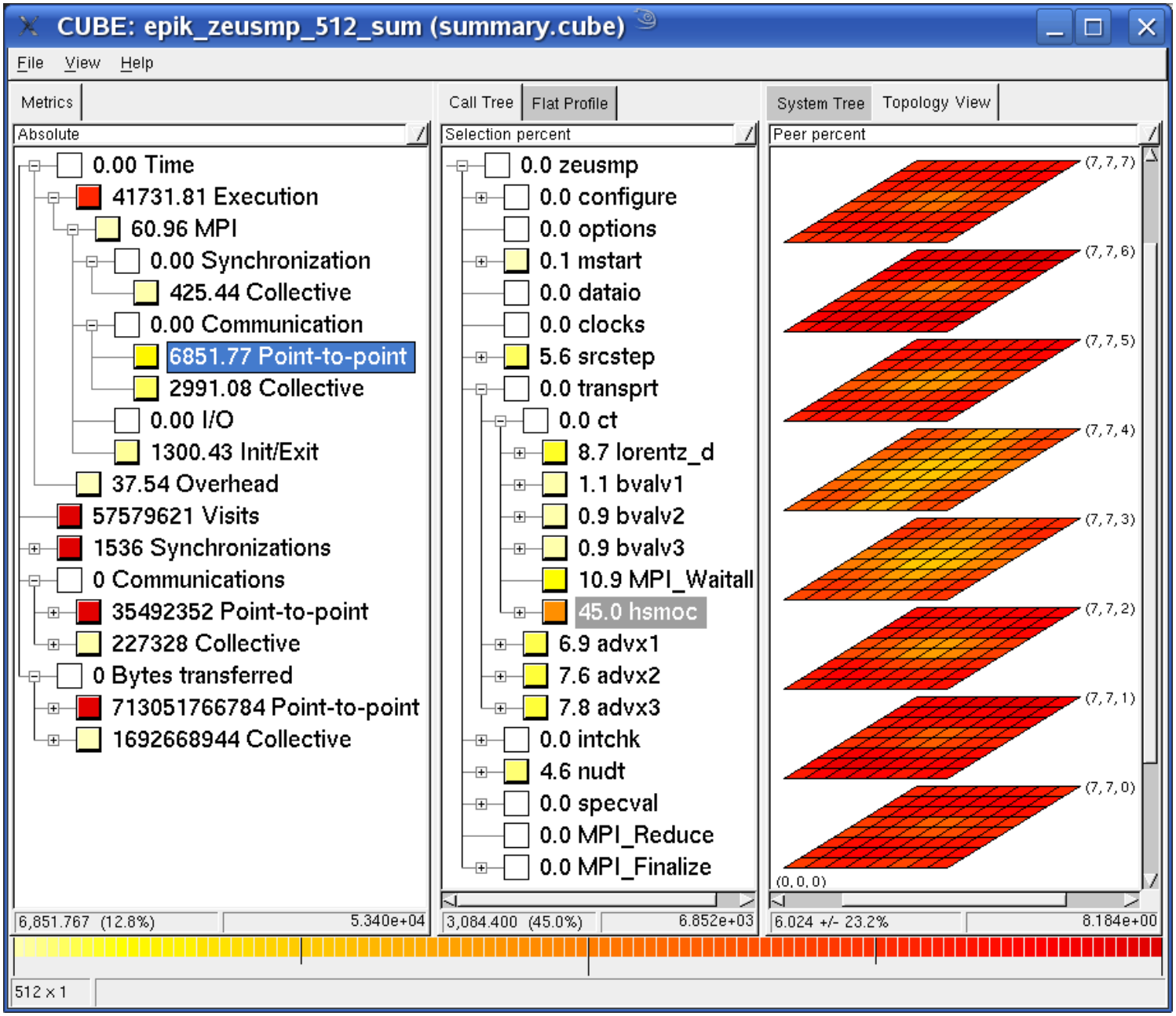

Java - GENE/field_solve_kxky-psc.f90 - Eclipse Platform
File Edit Navigate Search Project Run Window Help
Packag Hierarc
fourier_mkl.F90
g_16_sca.psc
g_16_sca.pscold
g_16_scabf.psc
g_16_scabf.pscold
g_32_sca.psc
g_32_sca.pscold
g_32_scabf.psc
g_32_scabf.pscold
g_64_sca_inst6.psc
g_64_sca.psc
g_64_scabf.psc
g_64_scabf.pscold
g_8_sca_inst15.psc
g_8_sca_inst18.psc
g_8_sca_inst6.psc
g_8_sca_inst8.psc
g_8_sca_install.psc
g_8_sca.psc
g_8_sca.pscold
g_8_scabf_inst8.psc
g_8_scabf.psc
g_8_scabf.pscold
g_sca_128_install.psc
gauss.quadrature-psc.f
gauss.quadrature.F90
GENE_script.sh
gene-psc.f90
gene.F90
gene128agent
gene128agentold
gene128front
gene512frontold
gene512frontold1
GeneFiles.txt
geneout.tar
g_sca_128_install.ps
comm-psc.f90
field_solve_kxky-psc
33 Real, Dimension(:,:,:,:,:,:,:), Allocatable:: mmat,mmat_perf
34 Complex, Dimension(:,:,:), allocatable :: p_phi_int, p_phi_int2
35
36contains
37 Subroutine field_solve_kxky(p_g_1,p_emfields)
38! Arguments
39 Complex, Dimension(li1:li2,lj1:lj2,lk1:lk2,ll1:ll2,lm1:lm2,ln1:ln2), Inter
40 complex, dimension(lbx:ubx,lj1:lj2,lbz:ubz,1:n_fields), intent(out) :: p_e
41
42! Local variables (put on stack)
43
44 Integer :: j, k, l, m, n, o
45 complex, dimension(li1:li2,lj1:lj2,lk1:lk2,1:n_fields) :: moments
46 complex, dimension(li1:li2,lj1:lj2,lk1:lk2,1:n_fields,ln1:ln2) :: vmoments
47
48
49
50 Call perfon ('FldSolvesf')
51
52
53! Gyroaverage and calculation of the first two moments of the distribution fur
54! We use the BLAS routines for a real array with double the size to speed up t
55! (there is no routine for real*complex and complex*complex has more operatior
56
57 if (perf_vec(1).eq.1) then
58 call calc_moments(n_fields,.false.,p_g_1,mmat,vmoments)
59 else
60 call calc_moments_perf(lijk0,llm0,n_fields,p_g_1,mmat_perf,vmoments)
61 endif
62
63
64 moments=sum(vmoments,5)
65 call my_complex_sum_vwspec(moments,n_fields*lijk0)
Outline
SIR Outline
Task List
SIR File: /home/rainy/workspaces/runtime-Psc_
subroutine: CALC_REST (54/220) (1-34)
call: FIELD_SOLVE_KXKY (166/166) (1-42
subroutine: CALFULLRHS_KXKY_1 (40/119)
call: MY_REAL_MAX_TO_ALL (79/79) (4-2
subroutine: MY_REAL_MAX_TO_ALL (58/172
call: MPI_ALLREDUCE (114/114) (10-240
subroutine: MY_COMPLEX_SUM_VWSPEC (4
call: MPI_ALLREDUCE (406/406) (10-312
subroutine: FIELD_SOLVE_KXKY (131/320) (
call: MY_COMPLEX_SUM_VWSPEC (189/1
program: GENE (0/334) (18-21)
loop: (0/334) (18-147)
userRegion: (203/334) (18-149)
call: CALC_EXPLICIT_TIMESTEP (13
subroutine: CALC_EXPLICIT_TIMESTEP (142,
loop: (0/3) (32-122)
call: CALC_REST (3/3) (32-129)
Problems
Javadoc
Declaration
Error Log
Periscope Properties View
Periscope Clustering View
Search
Console
Name
Process
Severity
Filename
Confidence
Extra
Stalls due to waiting for data delivery to register 46 30.22 field_solve_kxky-psc.f90 1.00
Stalls due to waiting for data delivery to register 5 30.32 field_solve_kxky-psc.f90 1.00
Stalls due to waiting for data delivery to register 45 30.41 field_solve_kxky-psc.f90 1.00
L2 misses 102 30.53 field_solve_kxky-psc.f90 1.00 es=221330 L2Misses=164831 L3Misses=
Stalls due to waiting for data delivery to register 17 31.11 field_solve_kxky-psc.f90 1.00
IA64 Pipeline Stall Cycles 4 31.14 field_solve_kxky-psc.f90 1.00
IA64 Pipeline Stall Cycles 56 31.38 field_solve_kxky-psc.f90 1.00
IA64 Pipeline Stall Cycles 50 31.65 field_solve_kxky-psc.f90 1.00
Filter:
Search:
RE
131 Shown - 1 Selected -
Sort: [Severity (FWD)]

Scalasca

- Automatic post-mortem performance analysis
  - Record per-process event traces first
  - Automatically analyse for known performance problems then
- Analysis on original processes at original scale
- Communication replay to exchange performance/timing data
- Finally summarize global results, map to source code
- Graphical interface CUBE
- Scales to > 290 000 cores
- Developed for more than 10 years

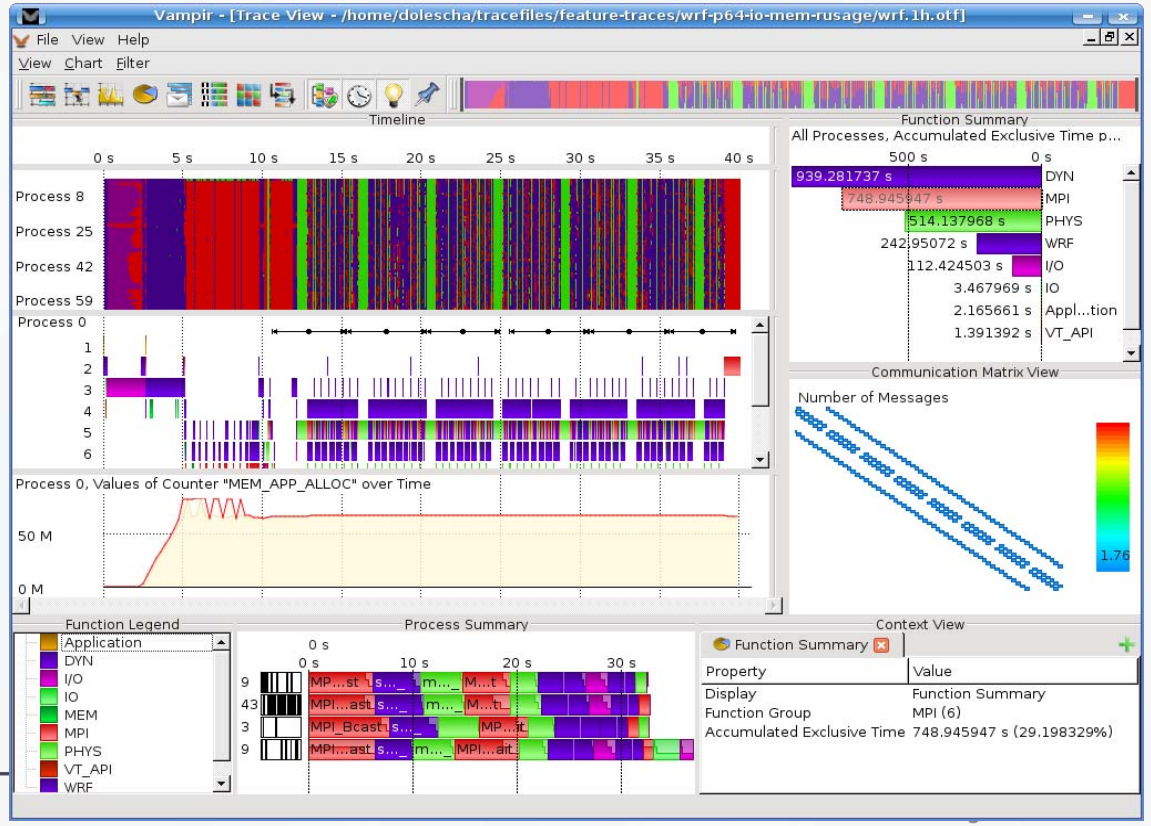


Scalasca


CUBE: epik_zeusmp_512_sum (summary.cube)
File View Help
Metrics
Call Tree
Flat Profile
System Tree
Topology View
Absolute
Selection percent
Peer percent
0.00 Time
41731.81 Execution
60.96 MPI
0.00 Synchronization
425.44 Collective
0.00 Communication
6851.77 Point-to-point
2991.08 Collective
0.00 I/O
1300.43 Init/Exit
37.54 Overhead
57579621 Visits
1536 Synchronizations
0 Communications
35492352 Point-to-point
227328 Collective
0 Bytes transferred
713051766784 Point-to-point
1692668944 Collective
0.0 zeusmp
0.0 configure
0.0 options
0.1 mstart
0.0 dataio
0.0 clocks
5.6 srcstep
0.0 transprt
0.0 ct
8.7 lorentz_d
1.1 bvalv1
0.9 bvalv2
0.9 bvalv3
10.9 MPI_Waitall
45.0 hsmoc
6.9 advx1
7.6 advx2
7.8 advx3
0.0 intchk
4.6 nudt
0.0 specval
0.0 MPI_Reduce
0.0 MPI_Finalize
(7, 7, 7)
(7, 7, 6)
(7, 7, 5)
(7, 7, 4)
(7, 7, 3)
(7, 7, 2)
(7, 7, 1)
(7, 7, 0)
(0, 0, 0)
6,851.767 (12.8%)
5.340e+04
3,084.400 (45.0%)
6.852e+03
6.024 +/- 23.2%
8.184e+00
512 x 1

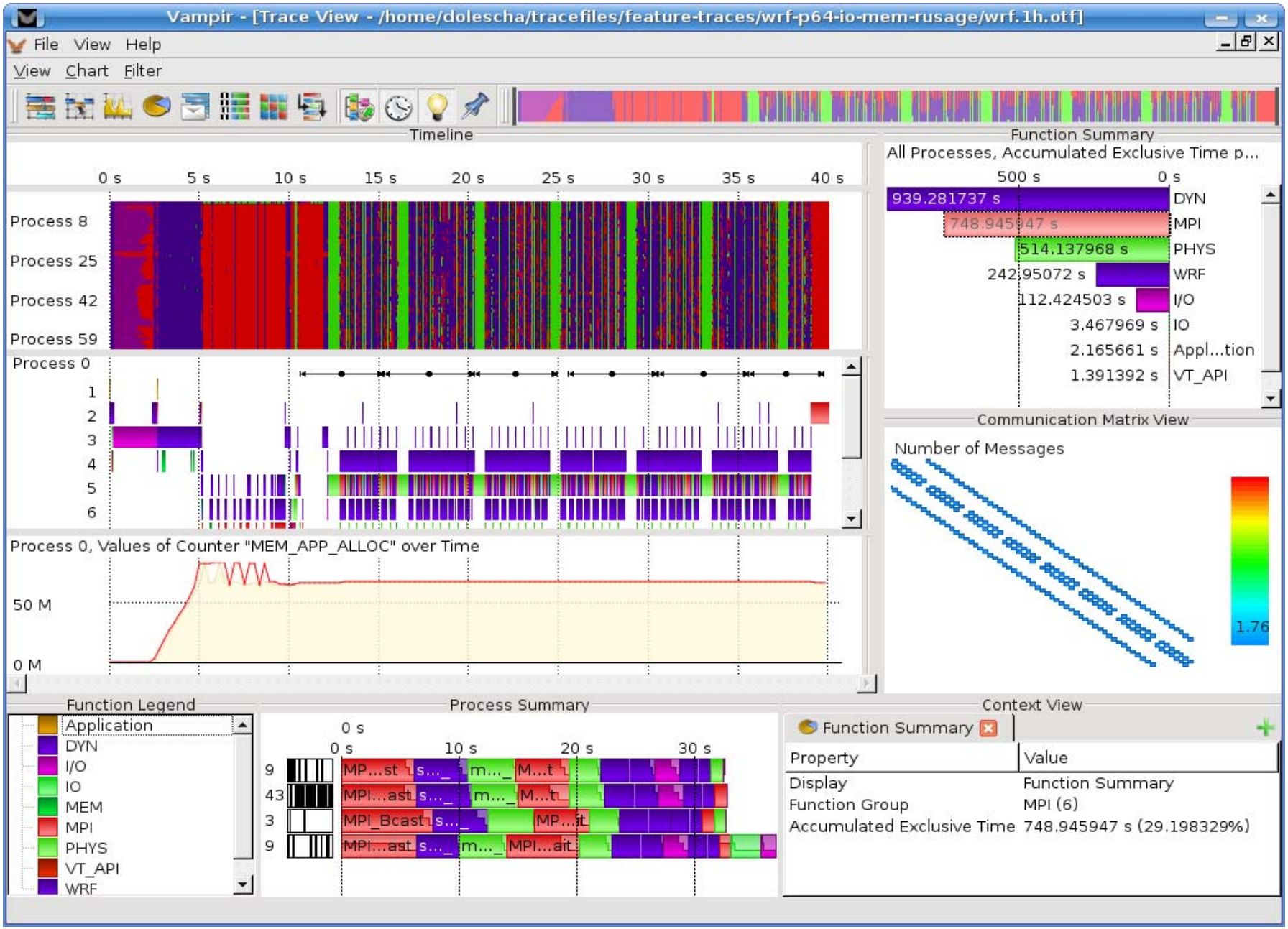
VampirTrace and Vampir GUI

- Code instrumentation and event tracing with VampirTrace
  - Also official part of Open MPI
- Event trace visualization with Vampir
- Interactive analysis
  - Parallel evaluation
  - Distributed trace data
- Scalable to
  - Ten thousands of processes/threads
  - Hundreds of gigabytes of trace data
- Since more than 15 years



TECHNISCHE UNIVERSITÄT DRESDEN

Vampir


Vampir - [Trace View - /home/dolescha/tracefiles/feature-traces/wrf-p64-io-mem-rusage/wrf.1h.otf]
File View Help
View Chart Filter
Timeline
0 s 5 s 10 s 15 s 20 s 25 s 30 s 35 s 40 s
Process 8
Process 25
Process 42
Process 59
Process 0
1
2
3
4
5
6
Process 0, Values of Counter "MEM_APP_ALLOC" over Time
50 M
0 M
Function Summary
All Processes, Accumulated Exclusive Time p...
500 s 0 s
939.281737 s DYN
748.945947 s MPI
514.137968 s PHYS
242.95072 s WRF
112.424503 s I/O
3.467969 s IO
2.165661 s Appl...tion
1.391392 s VT_API
Communication Matrix View
Number of Messages
1.76
Function Legend
Application
DYN
I/O
IO
MEM
MPI
PHYS
VT_API
WRF
Process Summary
0 s
0 s 10 s 20 s 30 s
9
43
3
9
MP...st s..._ m..._ M...t
MPI...ast s..._ m..._ M...t
MPI_Bcast s..._ MP...it
MPI...ast s..._ m..._ MPI...ait
Context View
Function Summary
Property Value
Display Function Summary
Function Group MPI (6)
Accumulated Exclusive Time 748.945947 s (29.198329%)

Overview

- Performance tools for HPC
  - Periscope
  - Scalasca
  - Vampir Tools
- **Project goals**
  - **The tool's perspective**
  - **The user's perspective**
- Status and outlook

Project Goals

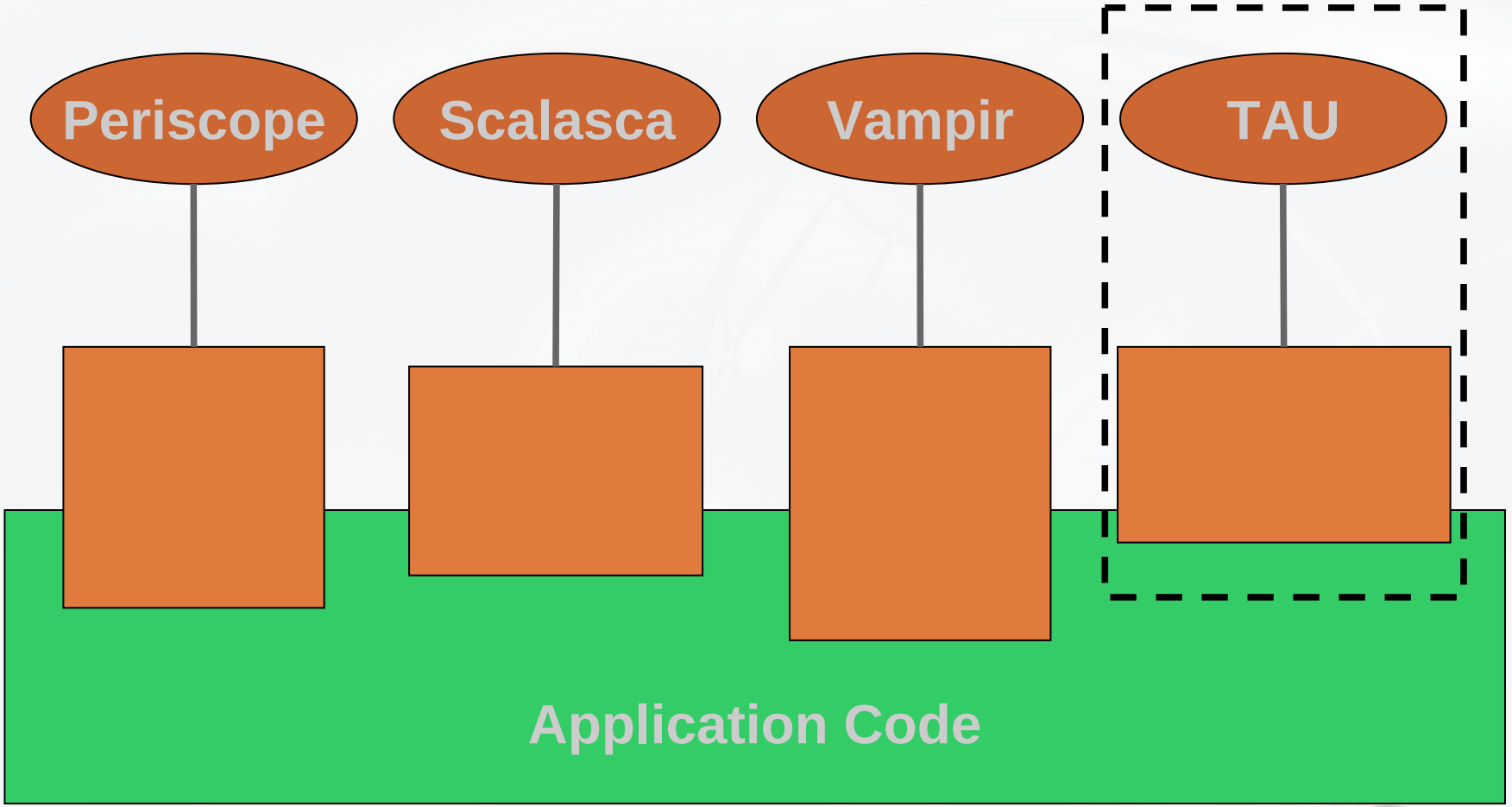
- Measurement Infrastructure
  - unification, support for long run times, enhance scalability, parallelization of tool's subtasks, common data formats, online interface, minimize overhead, support for OpenMP 3.0, wait time analysis
- Analysis
  - adapt analysis tools to new measurement infrastructure, enhance analysis functionality
- Software Quality
  - tests, documentation, code maintenance
- Evaluation and use cases
  - qualitative and quantitative evaluation
- Dissemination

Tool's Perspective: Common Functionality

- Code Instrumentation
  - Compiler wrappers, instrumentation API, helper tools, ...
- Run-time recording
  - Timer management,
  - Performance counter support, ...
- Handling of data and meta data
- Data formats
  - Event tracing and profiling
  - Data interoperability
- Platform portability and production quality
- Extensive feature set and overhead reduction

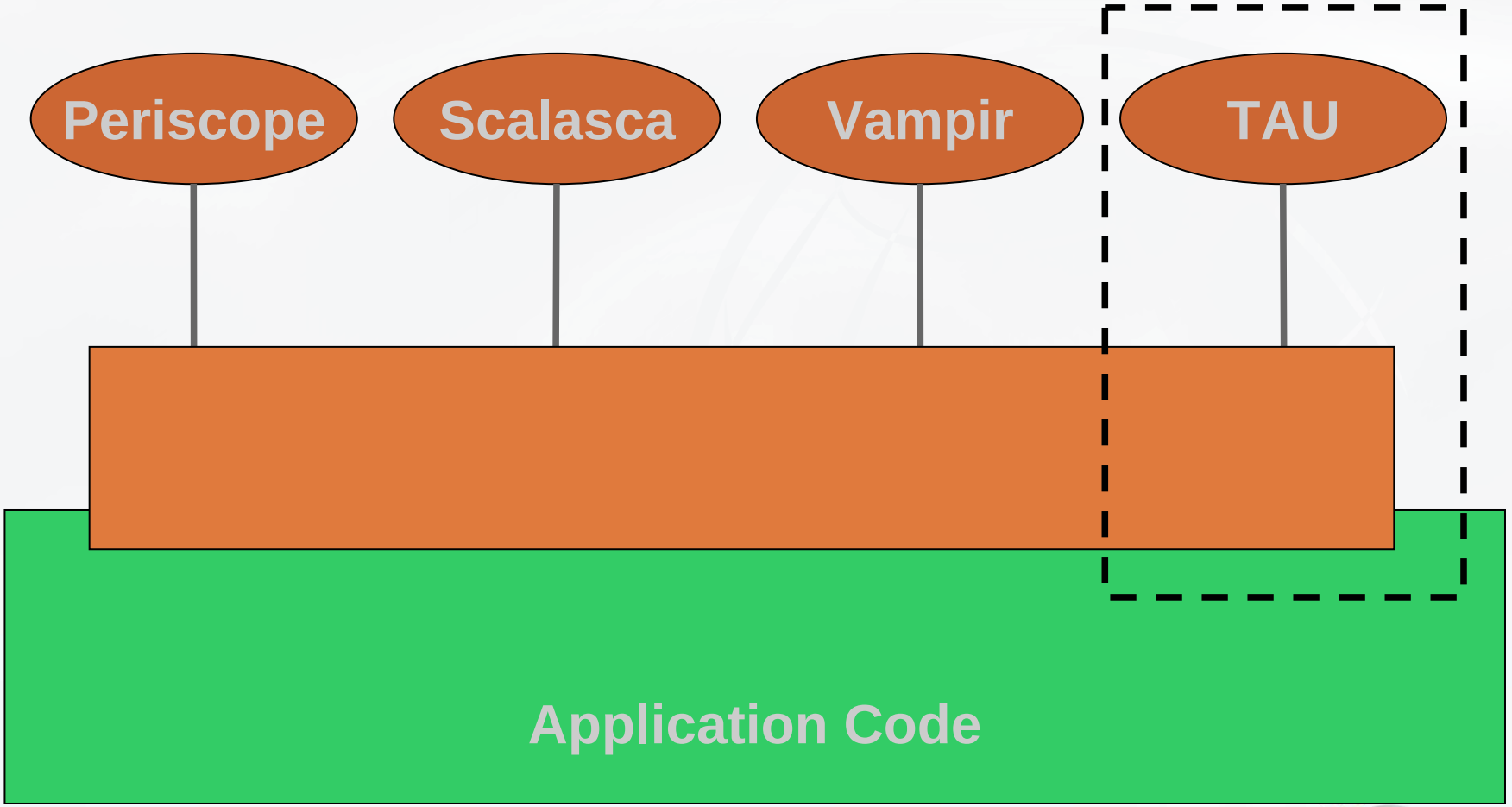
Common Functionality

- Currently, similar functionality with different infrastructure



Common Functionality

- Goal: common unified infrastructure



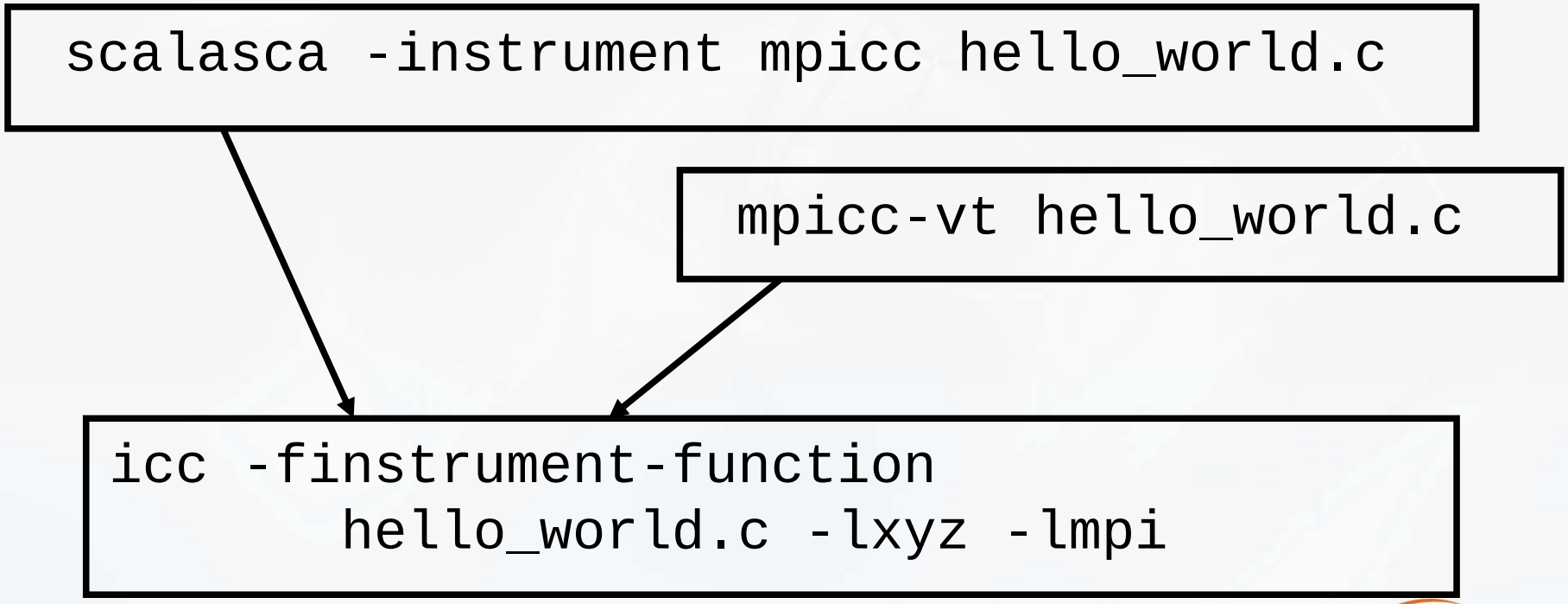
User's Perspective

- Different complimentary tools
- Learning curve for tools
- User friendly but not trivial

- Need to re-compile or re-link once or multiple times
- Investigate same experiment runs with several tools

User's Wishes

- Common user perspective
- Same command line tools
- Same configuration options



Overview

- Performance tools for HPC
  - Periscope
  - Scalasca
  - Vampir Tools
- Project goals
  - The tool's perspective
  - The user's perspective
- **Status and outlook**

Project Status

Status at middle of project time

- Thorough specification of requirements and design
- Combination/unification of existing functionalities
- Three main infrastructure components
  - Instrumentation and data collection layer (Adapters)
  - Run-time measurement and buffering layer (Core)
  - Open Trace Format 2 (OTF2)
- First working version

Dissemination

- Series of VI-HPS tuning workshops
- Announcement at SC BOF in New Orleans in November

SILC Partners

- Forschungszentrum Jülich, Jülich Supercomputing Centre (JSC)
  - Scalasca tools for event tracing and automatic analysis
- Gesellschaft für numerische Simulation mbH (GNS)
  - KMU, vendor for industrial simulation applications
  - Case studies with real world codes and scenarios
- Rheinisch-Westfälische Technische Hochschule Aachen (RWTH)
  - Member of OpenMP committee
- Technische Universität München,
  Lehrstuhl für Rechnertechnik und Rechnerorganisation
  - Periscope tools for automatic online performance analysis

SILC Partners

SILC Partners contd.

- Technische Universität Dresden, Zentrum für Informationsdienste und Hochleistungsrechnen (ZIH), SILC coordinator
  - Vampir tools for event tracing and interactive visualization

Associated partner

- Gesellschaft für Wissens- und Technologietransfer der TU Dresden mbH (GWT)
  - KMU, Vendor for industry simulation software

New associated partner

- University of Oregon
  - TAU tools for parallel profiling and event tracing

Conclusions

The SILC project:

- Performance tools integration
- Feature enhancements
- Scalability improvements
- Part of long term development plus user support
- Part of traditional strength of HPC research in Germany

See also:

- Virtual Institute High Productivity Supercomputing at http://www.vi-hps.org/
- SILC project at http://www.vi-hps.org/projects/silc/

- Thanks to BMBF and DLR!