Mitglied der Helmholtz-Gemeinschaft

JÜLICH
FORSCHUNGSZENTRUM

Recent Developments in Score-P and Scalasca V2

Aug 2015 | Bernd Mohr

9th Scalable Tools Workshop
Lake Tahoe

YOU KNOW YOU MADE IT …
… IF LARGE COMPANIES "STEAL" YOUR STUFF

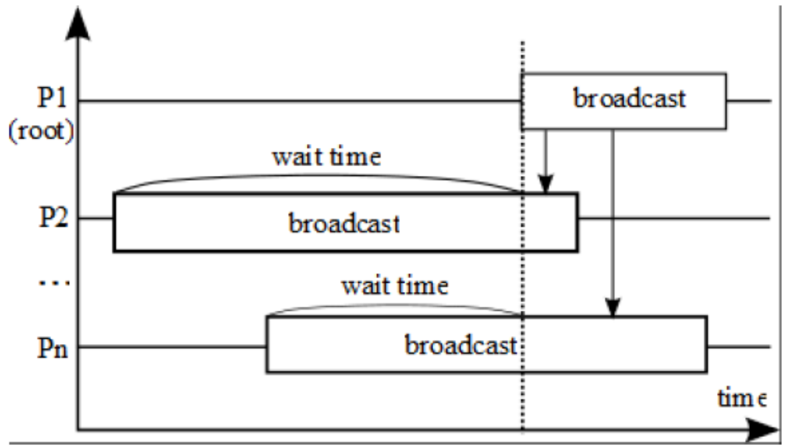
Introducing the Intel® Trace Analyzer and Collector Performance Assistant

Motivation: Improve method of performance analysis via the GUI

Solution:

- Define common/known performance problems
- Automate detection via the Intel® Trace Analyzer

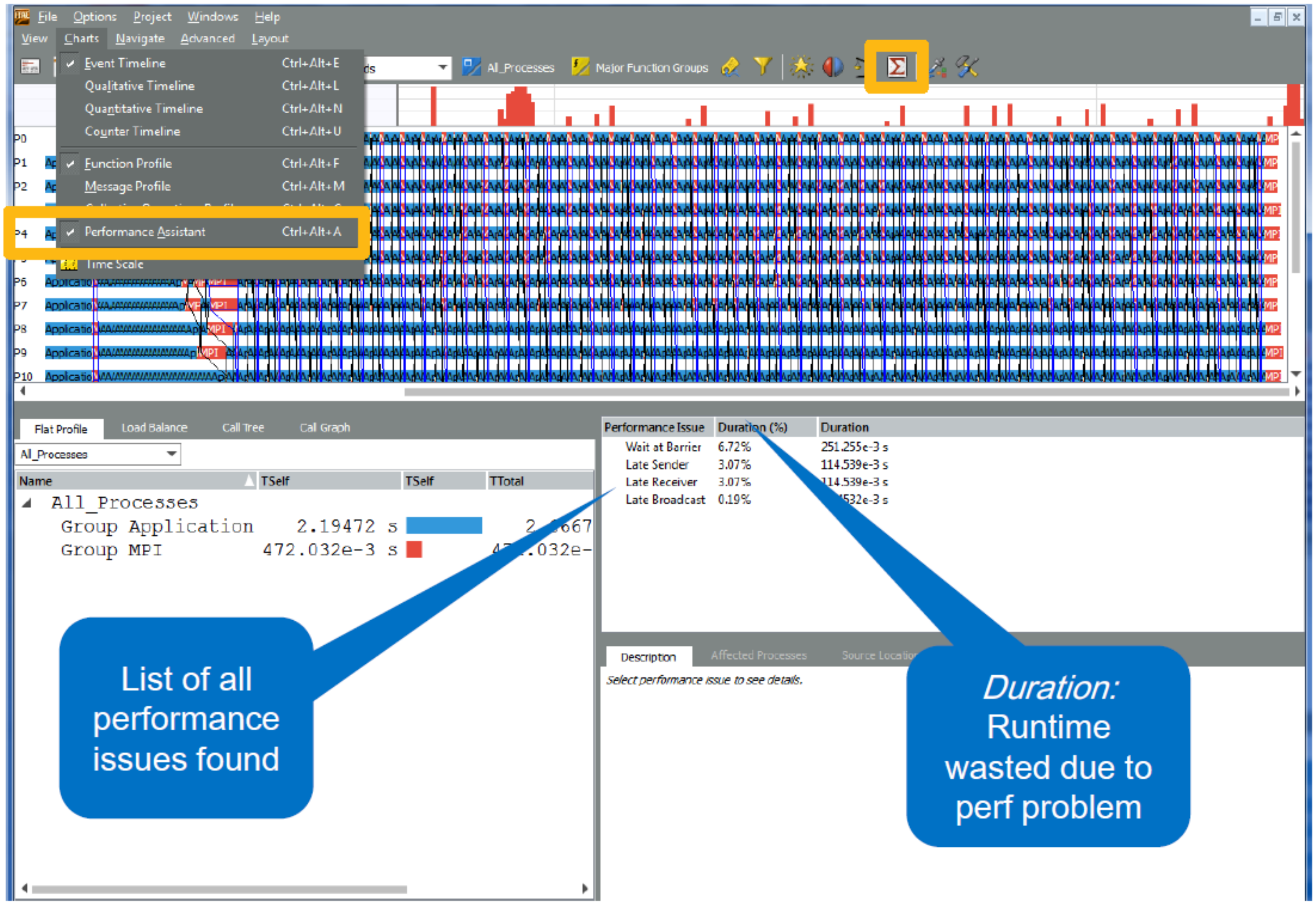
Example: A “Late Broadcast” is not easy to identify with existing views



Source:
https://software.intel.com/en-us/videos/quickly-discover-performance-issues-with-the-intel-trace-analyzer-and-collector-90-beta

Optimization Notice

New “Performance Assistant” Chart Added

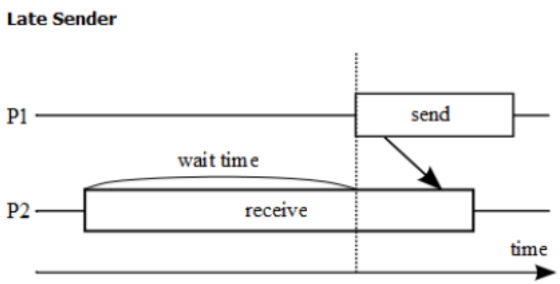


Source:
https://software.intel.com/en-us/videos/quickly-discover-performance-issues-with-the-intel-trace-analyzer-and-collector-90-beta

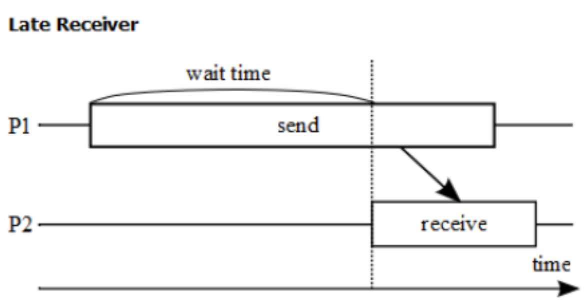
Which Performance Issues are automatically identified?

Point-to-point exchange problems:

- Late Sender

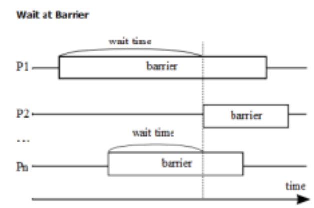


- Late Receiver

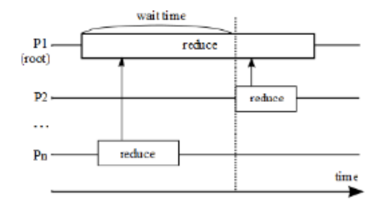


Problems with global collective operation performance:

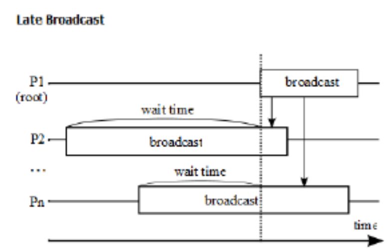
- Wait at Barrier



- Early Reduce



- Late Broadcast



Source:
https://software.intel.com/en-us/videos/quickly-discover-performance-issues-with-the-intel-trace-analyzer-and-collector-90-beta

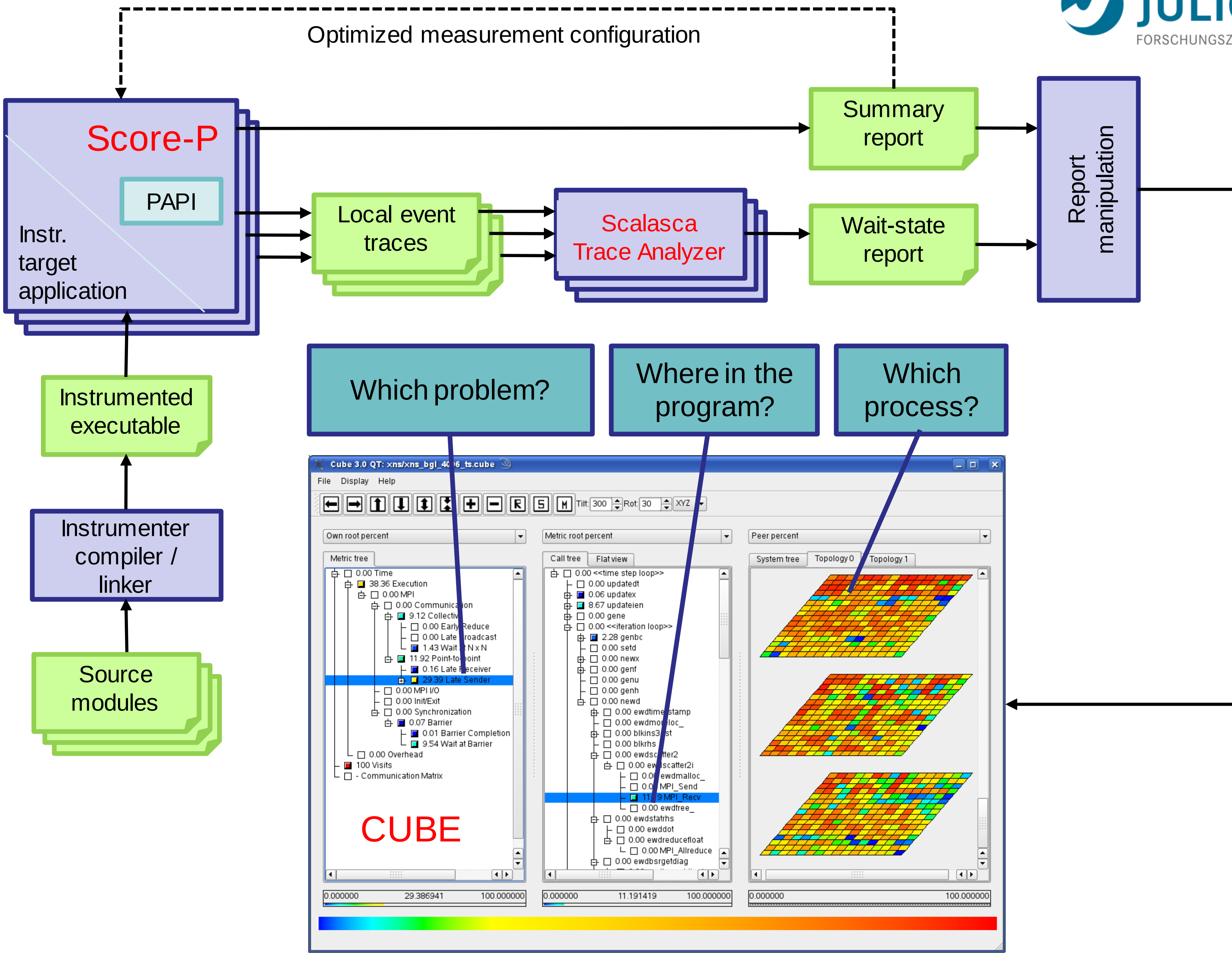
scalasca

Scalasca

- Scalable Analysis of Large Scale Applications
- Approach
  - **Instrument** C, C++, and Fortran parallel applications
    - Based on MPI, OpenMP, SHMEM, or hybrid
  - **Option 1: scalable call-path profiling**
  - **Option 2: scalable event trace analysis**
    - **Collect** event traces
    - **Search** trace for event patterns representing inefficiencies
    - **Categorize and rank** inefficiencies found
    - Supports MPI 2.2 (P2P, collectives, RMA, IO) and OpenMP 3.0 (exception: nesting)

http://www.scalasca.org/

TU Darmstadt


Optimized measurement configuration
Score-P
PAPI
Instr. target application
Summary report
Local event traces
Scalasca Trace Analyzer
Wait-state report
Report manipulation
Instrumented executable
Instrumenter compiler / linker
Source modules
Which problem?
Where in the program?
Which process?
CUBE

Scalasca Command

<table>
<tr><th></th><th>Scalasca 1</th><th>Scalasca 2</th></tr>
<tr><td>Prepare application objects and executable for measurement</td><td>1) scalasca –instrument <compile-or-link-command>
2) skin <compile-or-link-command></td><td>1) scalasca –instrument <compile-or-link-command>*
2) skin <compile-or-link-command>*
3) scorep <compile-or-link-command>**</td></tr>
<tr><td>Run application under control of measurement system</td><td colspan="2">1) scalasca –analyze <application-launch-command>
2) scan <application-launch-command>
3) set environment variables and run as usual</td></tr>
<tr><td>Interactively explore measurement analysis report</td><td colspan="2">1) scalasca –examine <experiment-archive|report>
2) square <experiment-archive|report></td></tr>
</table>

* command is deprecated and only provided for backwards compatibility with Scalasca 1.x.

** recommended option

Scalasca 1 vs Scalasca 2

	Scalasca 1	Scalasca 2
Instrumentation	EPIK	Score-P
Command line switches	different	
Manual instrumentation API	different	
Environmental variables	different	
Memory buffers	separate for each thread	memory pool on each process
Trace format	EPILOG	OTF2
Structure of the filterfile	different	
Scalable I/O	supports SIONlib	partially supports SIONlib
Report format	CUBE3	CUBE4
Experiment directory	epik_	scorep_
License	3-clause BSD	

For more information

- Zhukov, I. ; Feld, C. ; Geimer, M. ; Knobloch, M. ; Mohr, B. ; Saviankou, P.

  Scalasca v2: Back to the Future

  Niethammer, Christoph (Editor), ISBN: 978-3-319-16011-5

  Tools for High Performance Computing 2014, Stuttgart, Germany, 2015

  [doi:10.1007/978-3-319-16012-2_1]

Score-P

Scalable performance measurement infrastructure for parallel codes

Integration

- Need integrated tool (environment)
  for all levels of parallelization
  - Inter-node (MPI, PGAS, SHMEM)
  - Intra-node (OpenMP, multi-threading, multi-tasking)
  - Accelerators (CUDA, OpenCL)

- Integration with **performance modeling and prediction**

- No tool fits all requirements
  - **Interoperability of tools**
  - Integration via open interfaces

Score-P Functionality

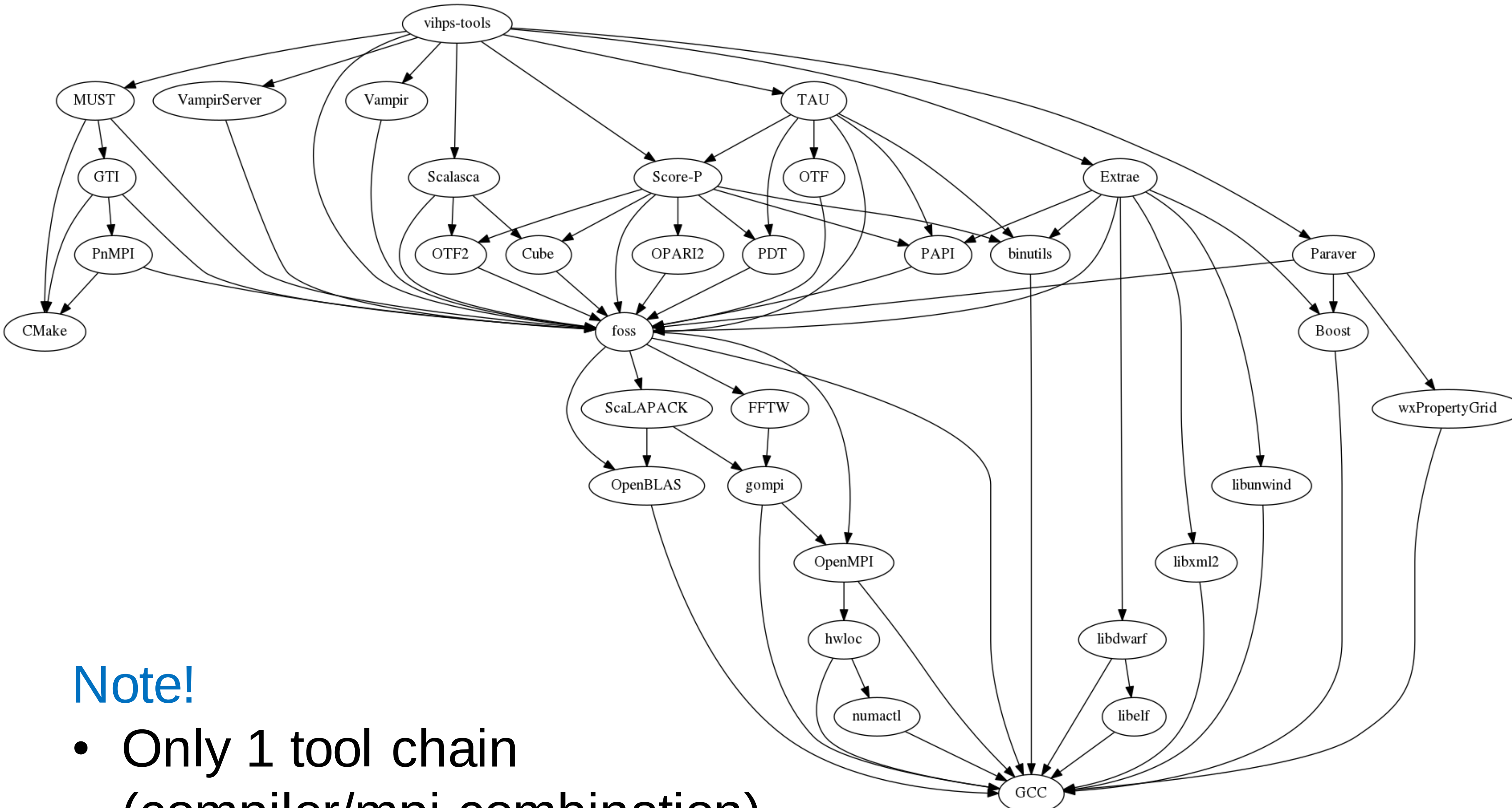
- Provide typical functionality for HPC performance tools
- Instrumentation (various methods)
  - Multi-process paradigms (MPI, SHMEM)
  - Thread-parallel paradigms (OpenMP, POSIX threads)
  - Accelerator-based paradigms (CUDA, OpenCL)
  - And their combination
- Flexible measurement without re-compilation:
  - Basic and advanced profile generation
  - Event trace recording
  - Online access to profiling data
- Highly scalable I/O functionality

- Support all fundamental concepts of partner’s tools

Non-functional Requirements

- **Portability:** support all major HPC platforms
  - IBM Blue Gene, Cray X*, Fujitsu K/FX10
  - x86, x86_64, PPC, Sparc, ARM clusters (Linux, AIX, Solaris)
- **Scalability**
  - Petascale, supporting platforms with more than 100K cores
- **Low measurement overhead**
  - Typically less than 5%
- **Robustness and QA**
  - Nightly Builds, Continuous Integration Testing Framework
- Easy and uniform installation through **EasyBuild**
- Open Source: New BSD License

Tool Dependencies



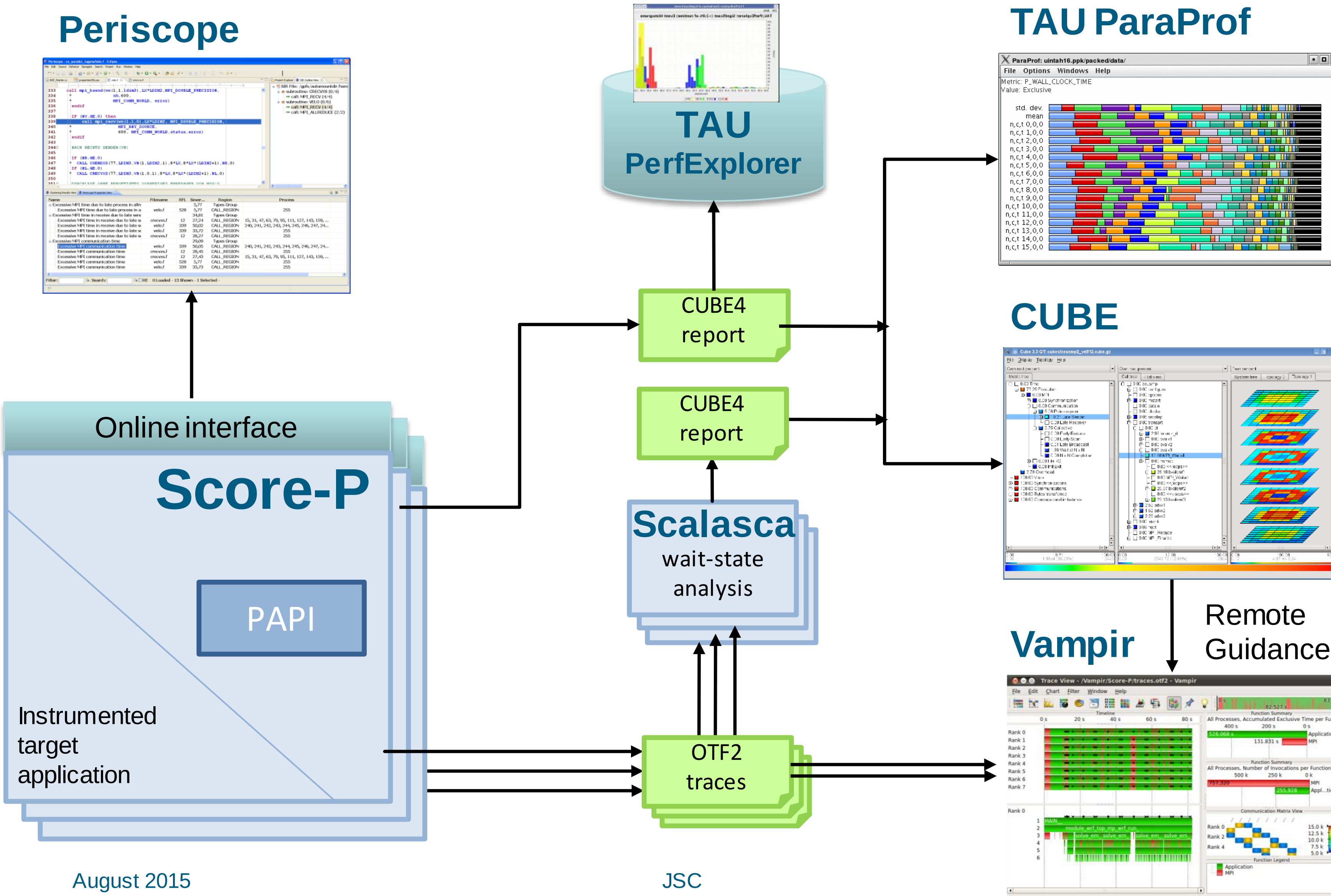
Note!

- Only 1 tool chain (compiler/mpi combination)
- Only 1 version

Score-P Partners

- Forschungszentrum Jülich, Germany
- German Research School for Simulation Sciences, Aachen, Germany
- Gesellschaft für numerische Simulation mbH Braunschweig, Germany
- RWTH Aachen, Germany
- Technische Universität Dresden, Germany
- Technische Universität München, Germany
- University of Oregon, Eugene, USA

The Score-P Tool Ecosystem



Past Funded Integration Projects

- **SILC (01/2009 to 12/2011)**
  - Unified measurement system (Score-P) for Vampir, Scalasca, Periscope
- **PRIMA (08/2009 to 10/2013)**
  - Integration of TAU and Scalasca
- **LMAC (08/2011 to 07/2013)**
  - Evolution of Score-P
  - Analysis of performance dynamics
- **H4H (10/2010 to 09/2013)**
  - Hybrid programming for heterogeneous platforms
- **HOPSA (02/2011 to 01/2013)**
  - Integration of system and application monitoring

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

ITEA 2
INFORMATION TECHNOLOGY FOR EUROPEAN ADVANCEMENT

SEVENTH FRAMEWORK PROGRAMME

MINISTRY OF EDUCATION AND SCIENCE OF THE RUSSIAN FEDERATION

Current Funded Integration Projects

- **Score-E (10/2013 to 09/2016)**
  - Analysis and Optimization of Energy Consumption
- **PRIMA-X (11/2014 to 10/2017)**
  - Extreme scale monitoring and analysis
- **RAPID (04/2014 to 03/2015)**
  - Enhanced support for node-level programming models
    - POSIX, ACE, Qt threads, MTAPI
  - Microsoft Windows support
- **Mont-Blanc-2 (10/2013 to 09/2016)**
  - OpenCL support
  - OmpSs support

GEFÖRDERT VOM

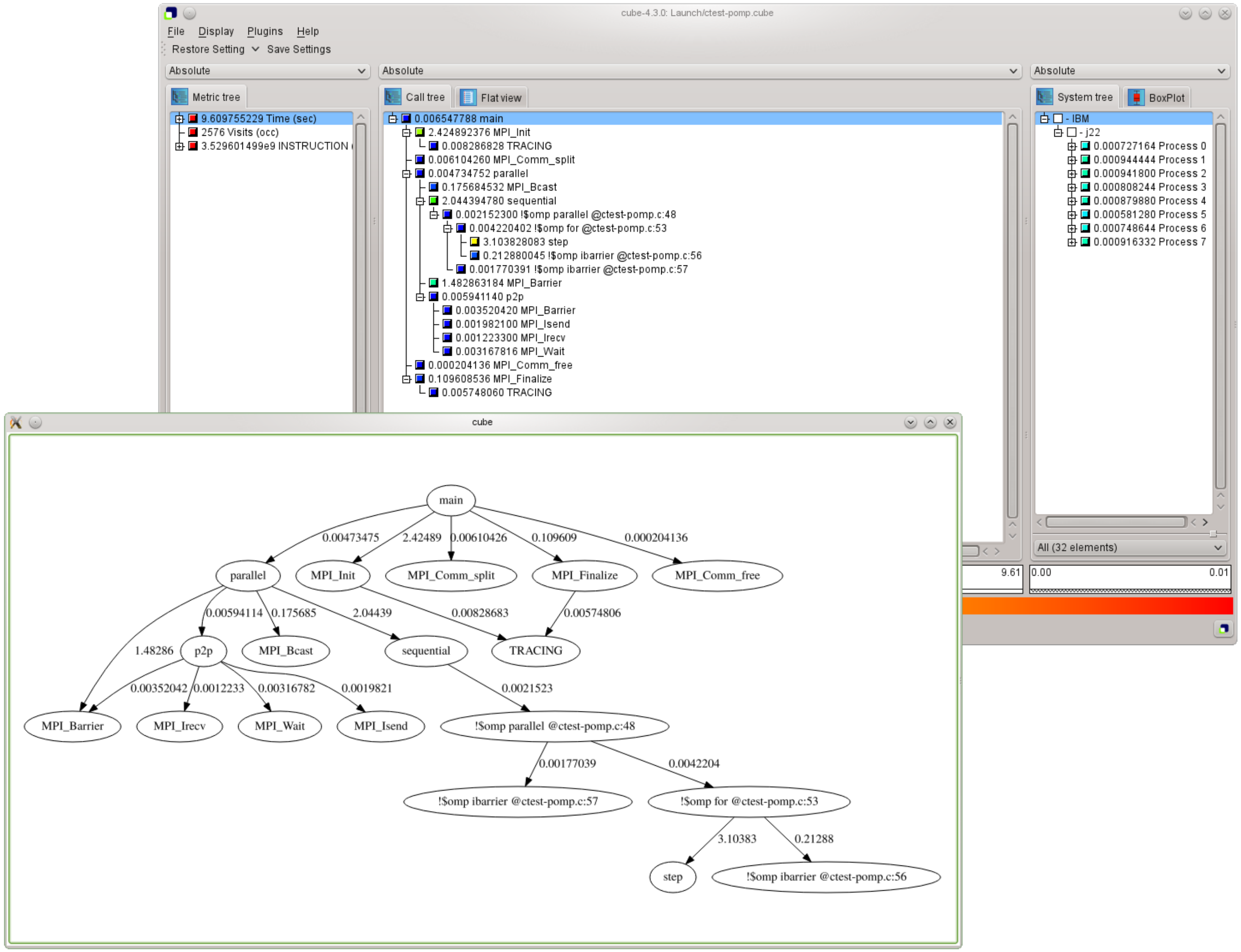
Bundesministerium für Bildung und Forschung

SIEMENS

SEVENTH FRAMEWORK PROGRAMME

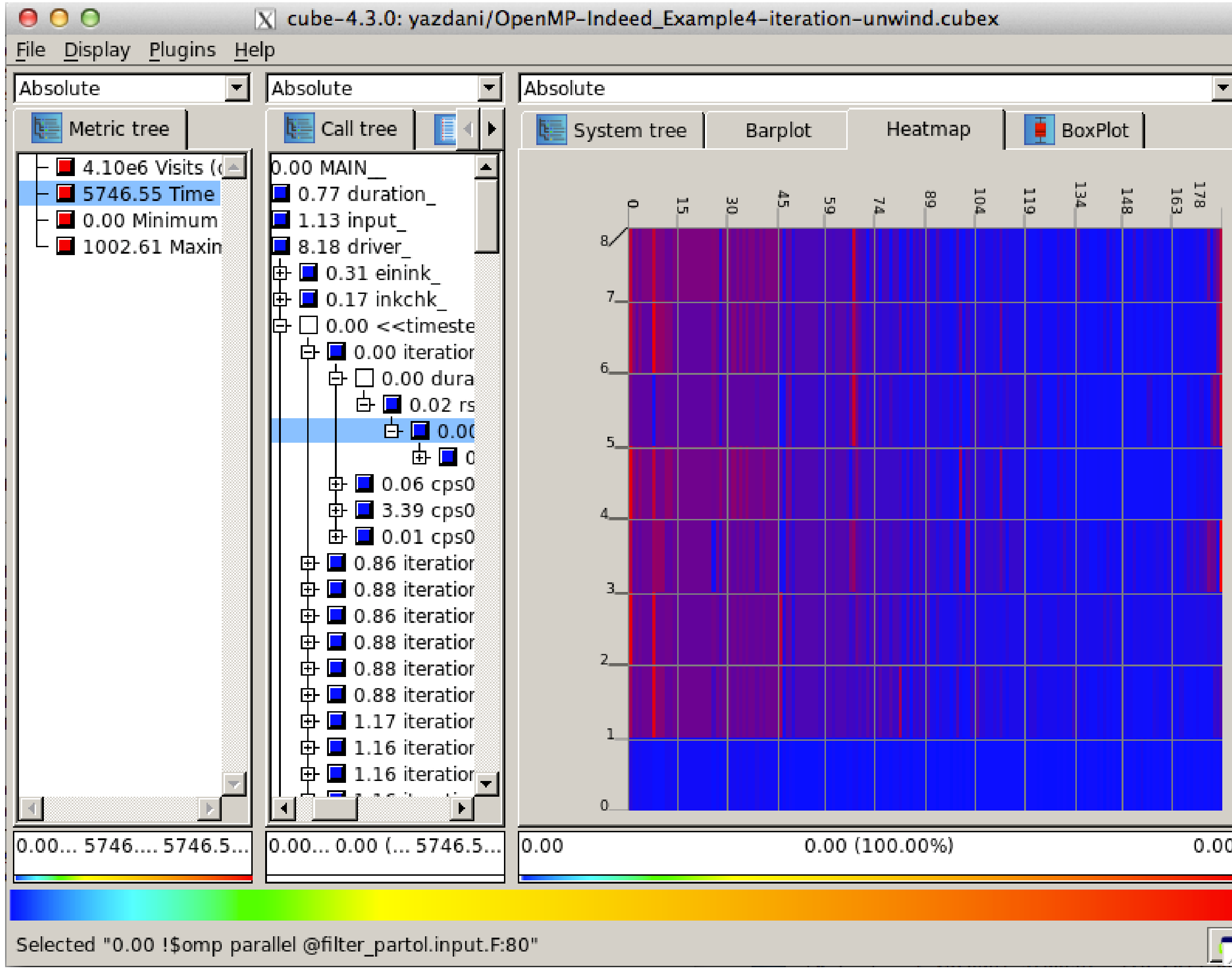
CUBE V4 PLUGIN INTERFACE

GUI Plugin: CallGraph



Cube Viz Plugins: Phase Heatmap

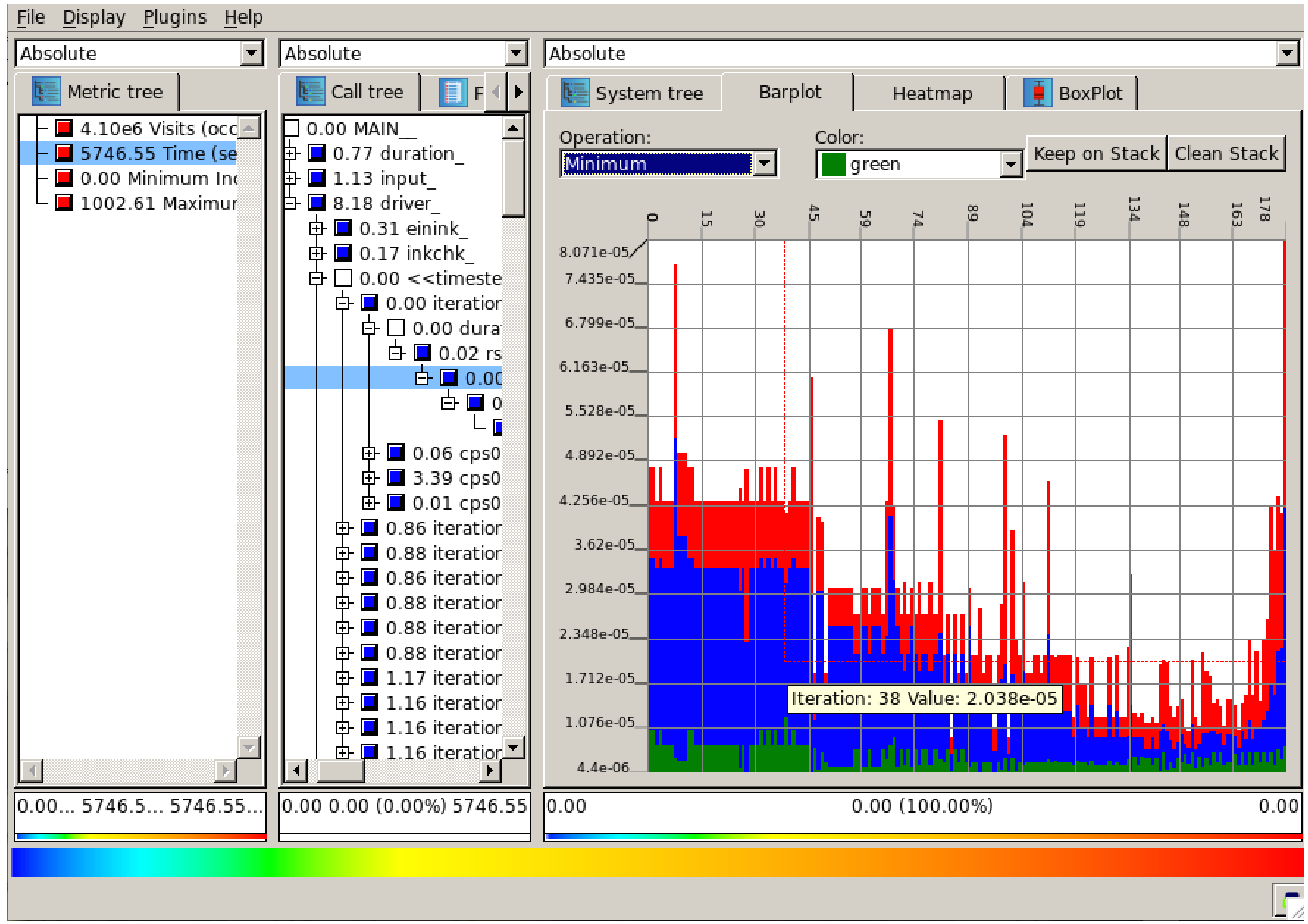
- **Phase profiling**
- Collects data for each instance of phases marked in program instead of aggregating it
- Shows data over “time” (phase instances) for each rank/thread



Cube Viz Plugins: Phase Barplot

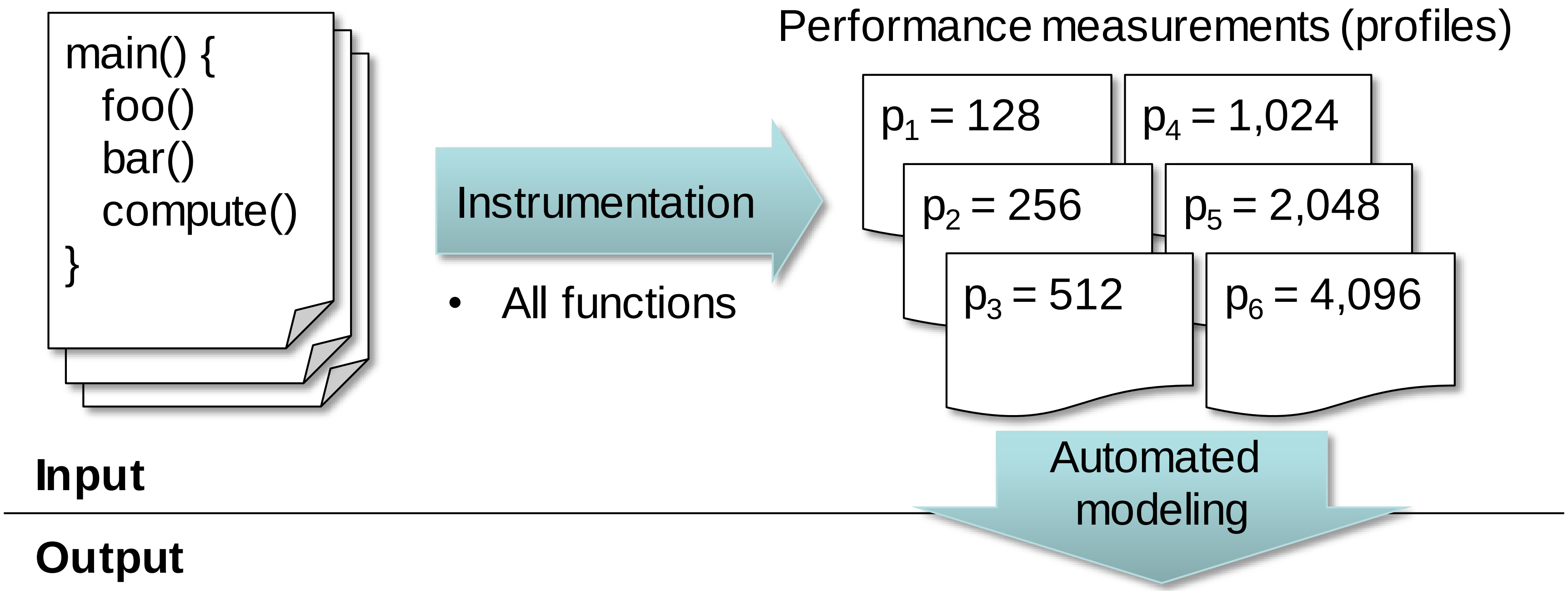
JÜLICH FORSCHUNGSZENTRUM

- **Phase profiling**
- Collects data for each instance of phases marked in program instead of aggregating it
- Shows min/max/avg metric value over “time” (phase instances)



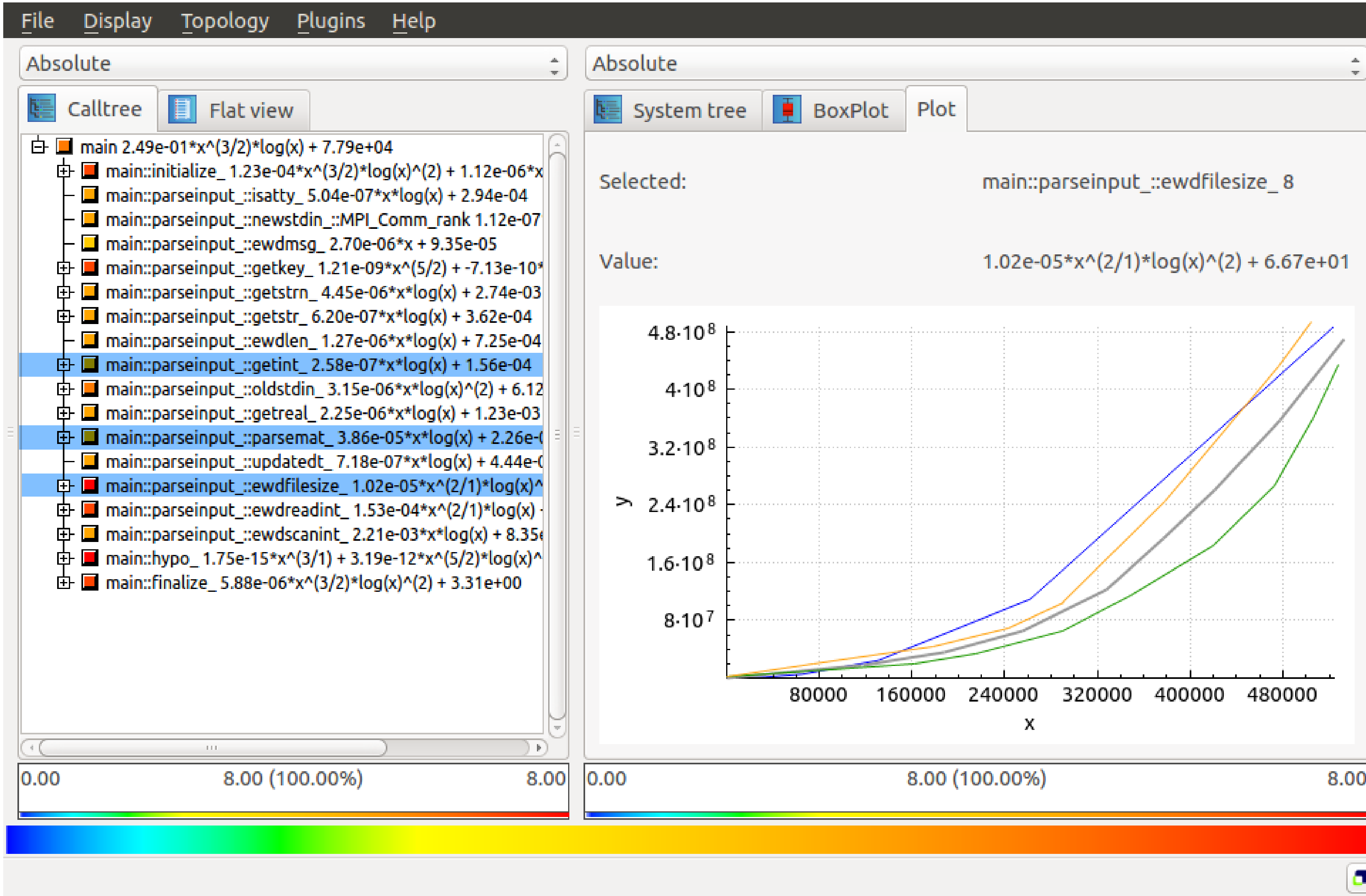
Integration of Measurement and Modelling

- Example: DFG SPPEXA Catwalk Project



Rank	Function	Model [s]
1	bar()	4.0 * p + 0.1*log(p)
2	compute()	0.5 * log(p)
3	foo()	65.7

Catwalk: Result Visualization



CUBE Derived Metrics

- Cube v4 now also supports definition of derived metrics
  - Based on CubePL DSL
  - PreDerived and PostDerived metrics
- List of selected features:
  - Support for various arithmetic calls
  - Support of arrays and variables
  - Automatic data type conversion
  - Lambda-function definitions
  - Predefined variables
  - Redefinition of aggregation operation

Saviankou, P. ; Knobloch, M. ; Visser, A. ; Mohr, B.
Cube v4: From Performance Report Explorer to Performance Analysis Tool
International Conference On Computational Science (ICCS 2015)
Procedia computer science 51, 1343 - 1352 (2015) [doi:10.1016/j.procs.2015.05.320]

SUCCESS STORIES

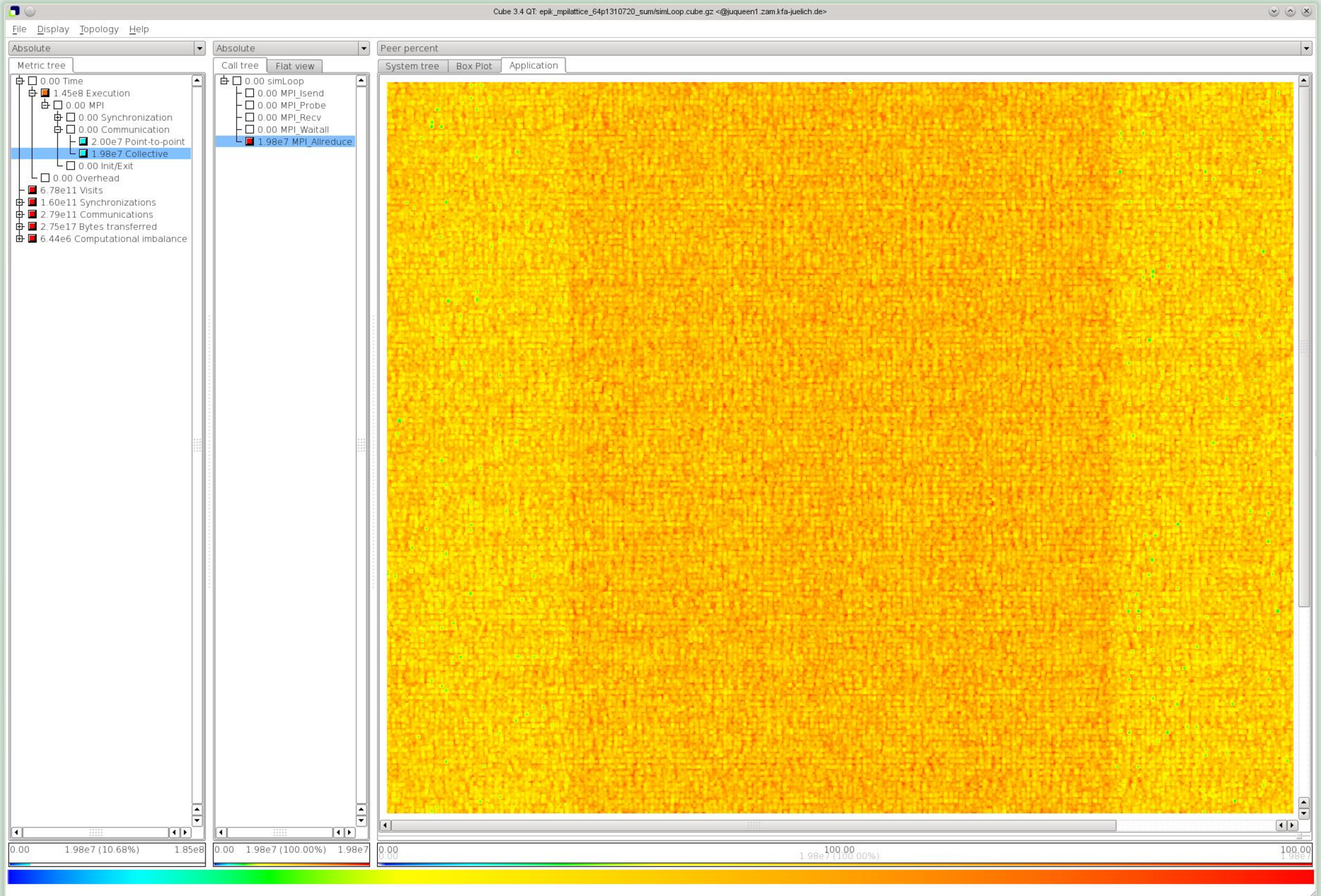
Performance Tool Scaling: Scalasca

- **Latest test case**
  - Granular Dynamics Simulation
  - Based on Physics Engine (PE) Framework (Erlangen)
  - PRACE @ ISC Award winner
  - MPI only

- **Scalasca 1.x Experiments on JUQUEEN**
  - Full machine experiment: 28,672 nodes x 32 MPI ranks
    - **917,504 processes** [Limit: Memory / System metadata]
  - Largest no. of threads: 20,480 nodes x 64 MPI ranks
    - **1,310,720 processes** [Limit: Memory / System metadata]

- **Scalasca 2.x / Score-P 1.4.1** NAS BT-MZ on JUQUEEN
    - Profiles: 16,384 x 64 = **1,048,576 threads** [Limit: BT-MZ]
    - Traces: 10,240 x 64 = 655,360 thread [Limit: OTF2]

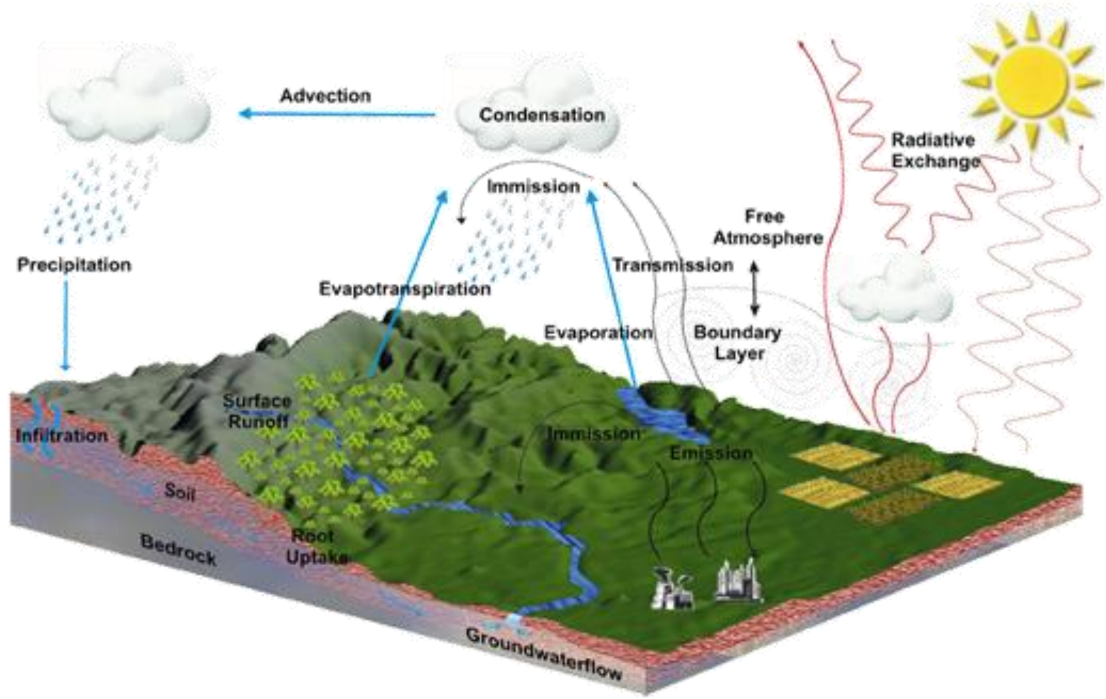
Scalasca: 1,310,720 process test case



Showcase: TerrSysMP

JÜLICH FORSCHUNGSZENTRUM

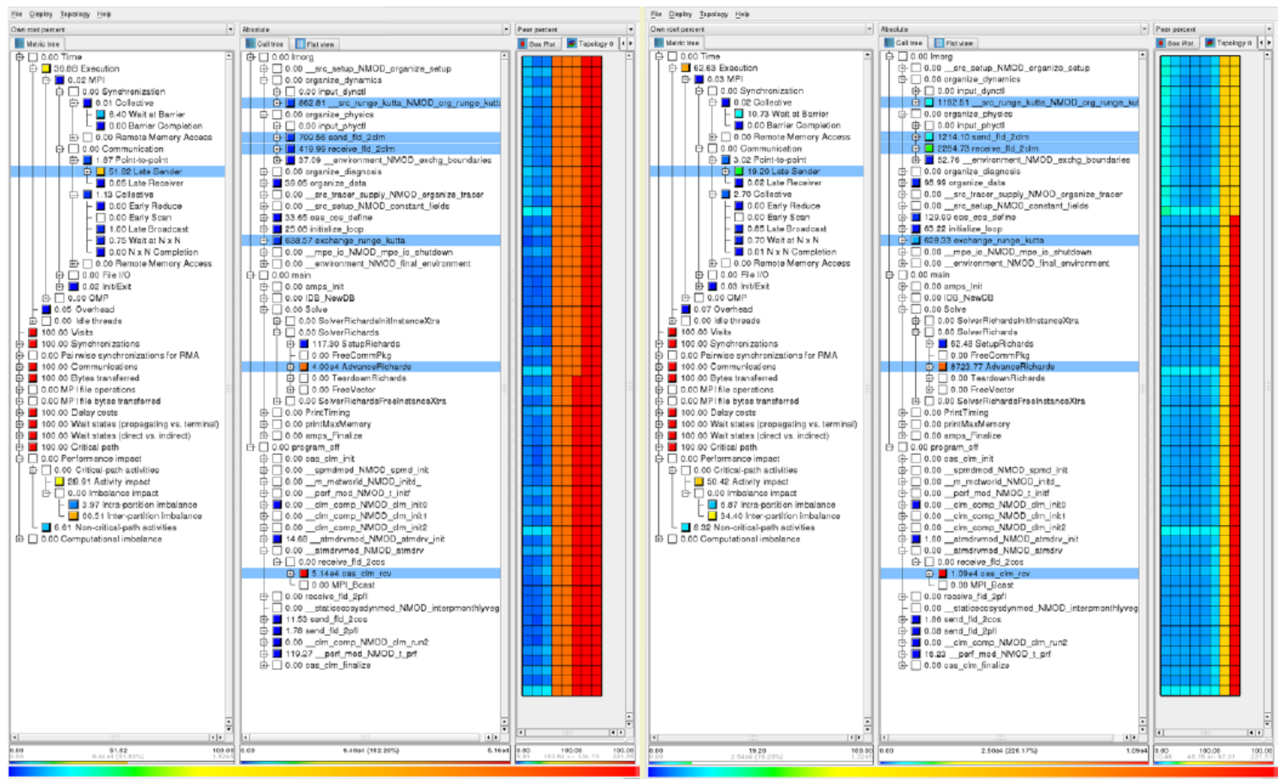
- Scale-consistent highly modular integrated multi-physics sub-surface/surface hydrology-vegetation atmosphere modelling system



- Fully-coupled MPMD simulation consisting of
  - COSMO (Weather prediction)
  - CLM (Community Land Model)
  - ParFlow (Parallel Watershed Flow)
  - OASIS coupler

Success Story: TerrSysMP

- Identified several sub-components bottlenecks:
  - Inefficient communication patterns
  - Unnecessary/inefficient code blocks
  - Inefficient data structures
- Performance of sub-components improved by factor of 2!
- Scaling improved from 512 to 32768 cores!



JÜLICH
FORSCHUNGSZENTRUM

The Team


Markus Geimer


Michael Knobloch


Bernd Mohr


Christian Rössel


Marc Schlütter


Pavel Saviankou


Alexandre Strube


Brian Wylie


Anke Visser


Ilja Zhukov

Sponsors

HELMHOLTZ ASSOCIATION

ITEA 2
INFORMATION TECHNOLOGY FOR EUROPEAN ADVANCEMENT

SEVENTH FRAMEWORK PROGRAMME

Deutsche Forschungsgemeinschaft
DFG

Bundesministerium für Bildung und Forschung

DEPARTMENT OF ENERGY • UNITED STATES OF AMERICA

Questions?

- Check out **http://www.scalasca.org**
- Or contact us at **scalasca@fz-juelich.de**