Performance Analysis of OpenMP Target Offloading in Score-P

Authors: Jan André Reuter[1], William R. Williams[2], Bernd Mohr[1], **Contact:** j.reuter@fz-juelich.de

[1]Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich GmbH, Jülich, Germany

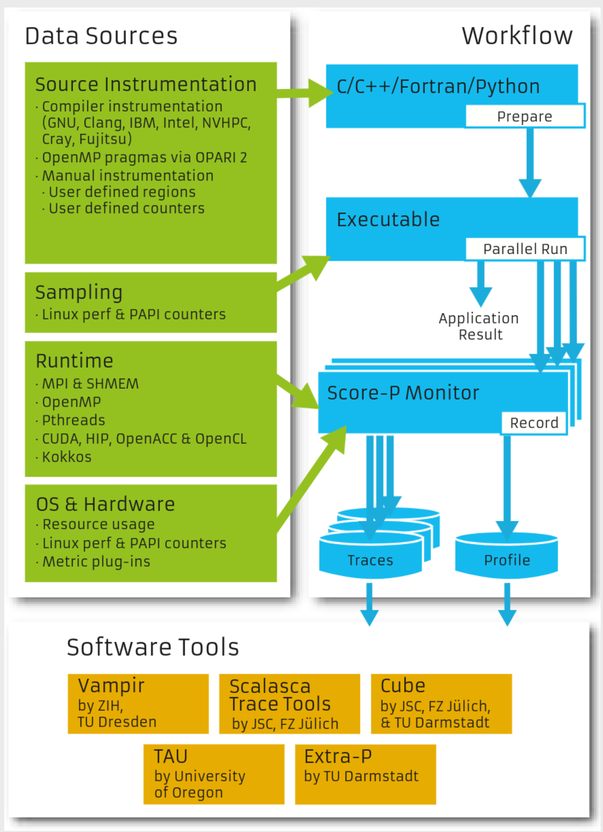
[2]Center for Interdisciplinary Digital Sciences (CIDS), Department Information Services and High Performance Computing (ZIH), TUD Dresden University of Technology, Dresden, Germany

Download the latest development version and try yourself

Introduction

- OpenMP supports offloading to accelerators via target directives, with the option to target different architectures, selecting them at runtime, interoperability with low-level APIs, and more. In recent years, OpenMP offloading gained a growing user base with frameworks offering support as well.
- With the OpenMP Tools Interface, performance analysis of OpenMP target regions is possible. However, the current support in performance tools is limited. Here, we present a prototype for performance instrumentation and measurement of applications using OpenMP offloading with Score-P.

Score-P Overview



Key features:

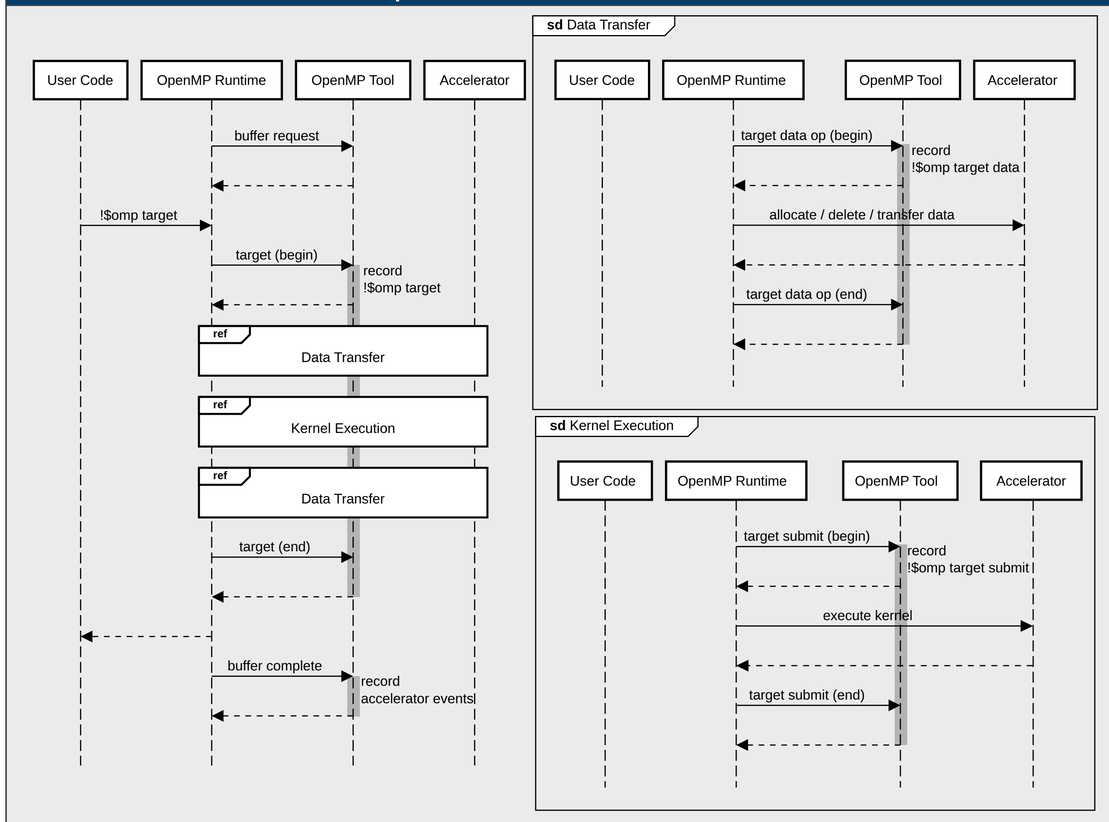
- Highly scalable, supporting platforms $\geq$ 500k cores
- Supported platforms: Linux clusters, HPE Cray, Fujitsu & more (using x86_64, Power & ARM)
- Supported parallelization models: MPI, SHMEM, OpenMP, Pthreads, CUDA, HIP, OpenACC, OpenCL & Kokkos
- Recording of I/O operations
- Profiling (CUBE4) & event tracing (OTF2)
- User library wrapping for C/C++ libraries
- Available as open source under a 3-clause BSD license

URL: https://www.score-p.org – **Contact:** support@score-p.org

OpenMP Tools Interface (OMPT)

- Portable interface for analysis tools. Supported by performance tools like Score-P, HPCToolkit, TAU and Nsight Systems, and correctness tools like LLVM Archer.
- Tools register OpenMP runtime callbacks, which get invoked on directives and runtime function calls.
- Support for heterogeneous accelerators through device tracing interface, where events are written to buffers, which are flushed during program execution (similar to CUPTI).
- Event correlation via two IDs: One per target region, one for data transfers and submitted kernels.

OMPT Workflow Example



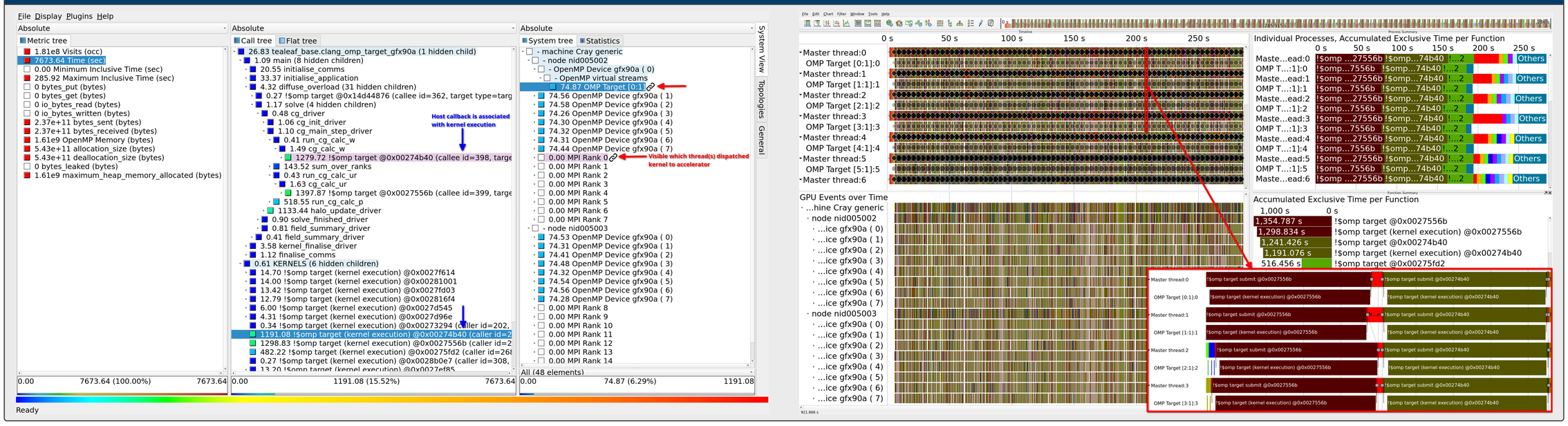
Simplified diagram for callbacks during an OpenMP program run. Accelerator events are recorded to buffer, dispatched to the OpenMP tool via `buffer complete` asynchronously. With this, we can write accelerator events correlated to host callbacks.

Current OMPT Compiler Vendor Support for Score-P

Compiler	Host Events		Accelerator Events
	Host	Target	Device Tracing
AOMP 18.0-1	Full	Full	Full
CCE 17.0.0	None	Full	None
Clang 18.1.6	Full	Partial	None
GCC 14.1	None	None	None
NVHPC 24.5	Full	Full	None
oneAPI 2024.1	Full	Partial	None
ROCm 6.1	Full	Full	Partial

- Full: Required parts of the interface are implemented, but may have runtime bugs.
- Partial: Missing features for full support (e.g. multiple accelerators).
- None: No or insufficient support.
- Accelerator events require all three event types.

First Profile and Trace Results – SPEC HPC 618.tealeaf_s on LUMI-G with AOMP 18.0-1



References & Acknowledgements

- Knüpfer, Andreas, et al. "Score-P: A joint performance measurement run-time infrastructure for Periscope, Scalasca, TAU, and Vampir." Tools for High Performance Computing 2011
- Feld, Christian, et al. "Score-P and OMPT: Navigating the Perils of Callback-Driven Parallel Runtime Introspection" International Workshop on OpenMP 2019
- OpenMP Application Programming Interface, Version 5.2: https://www.openmp.org/wp-content/uploads/OpenMP-API-Specification-5-2.pdf

- This poster is based upon work supported by the BMBF as part of the ENSIMA project (Grant No. 16ME0630)
- We acknowledge the EuroHPC Joint Undertaking for awarding this project access to the EuroHPC supercomputer LUMI, hosted by CSC (Finland) and the LUMI consortium through a EuroHPC Regular Access call.