ROCm GPU profiling in Trace Compass

Arnaud Fiorini with Pr. Michel Dagenais

September 9, 2019

Polytechnique Montreal

DORSAL Laboratory

Agenda

I. Introduction to ROCm

II. TraceCompass plugin

III. Future work

Introduction to ROCm

- GPUs have a specific architecture that requires a modified computing model and tool chain to write programs for it.
- Code that runs on a GPU is executed in a small unit of code called a kernel. This « function » is executed by the GPU many times concurrently and its parameters have to be transferred to the target device.
- How this function is written depends mostly on the model used : OpenCL, CUDA, HIP, OpenMP …

Introduction to ROCm

Logical concepts are helpful for the developer to think about how a program runs on a GPU.

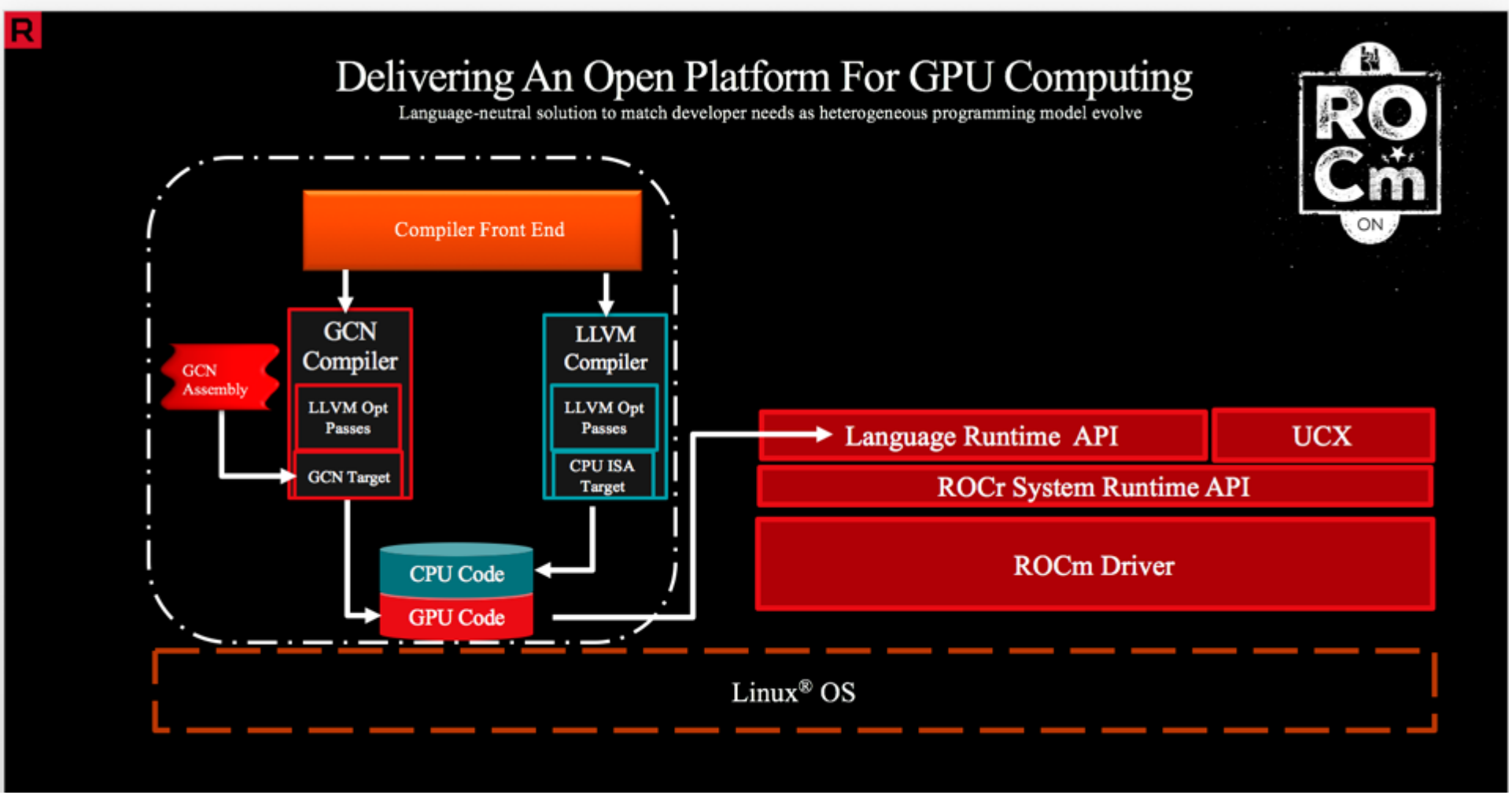
Architectural concepts

- Compute Unit
- SIMD Vector Unit
- ALU

Logical concepts

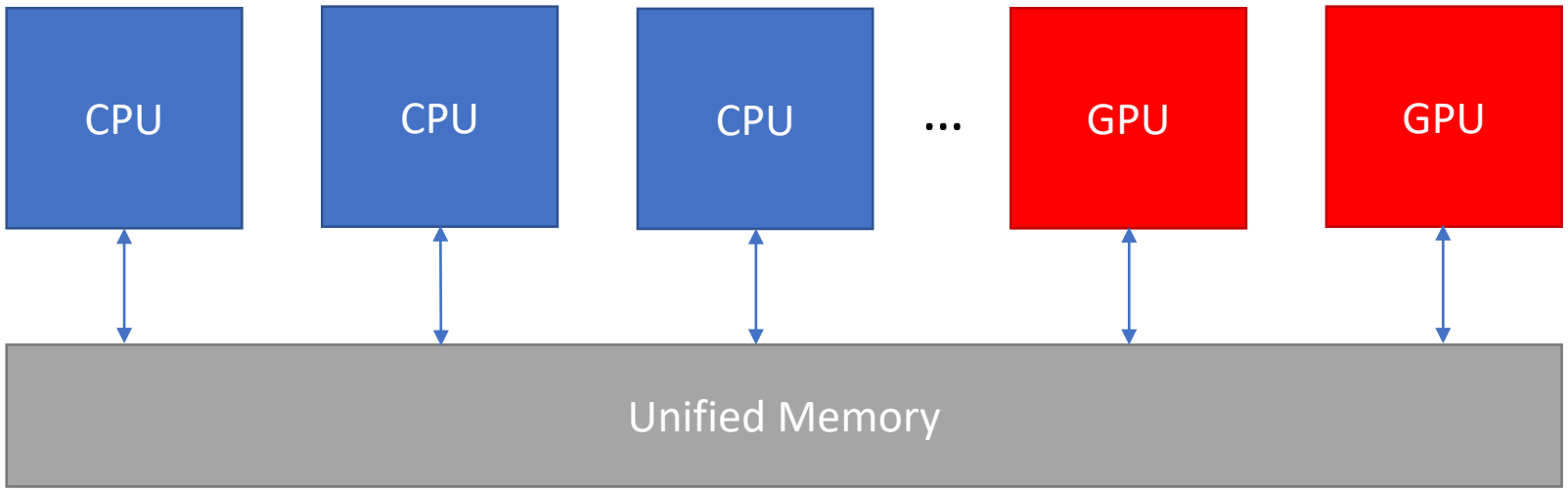
- Wavefront
- Thread
- Work-item

Introduction to ROCm



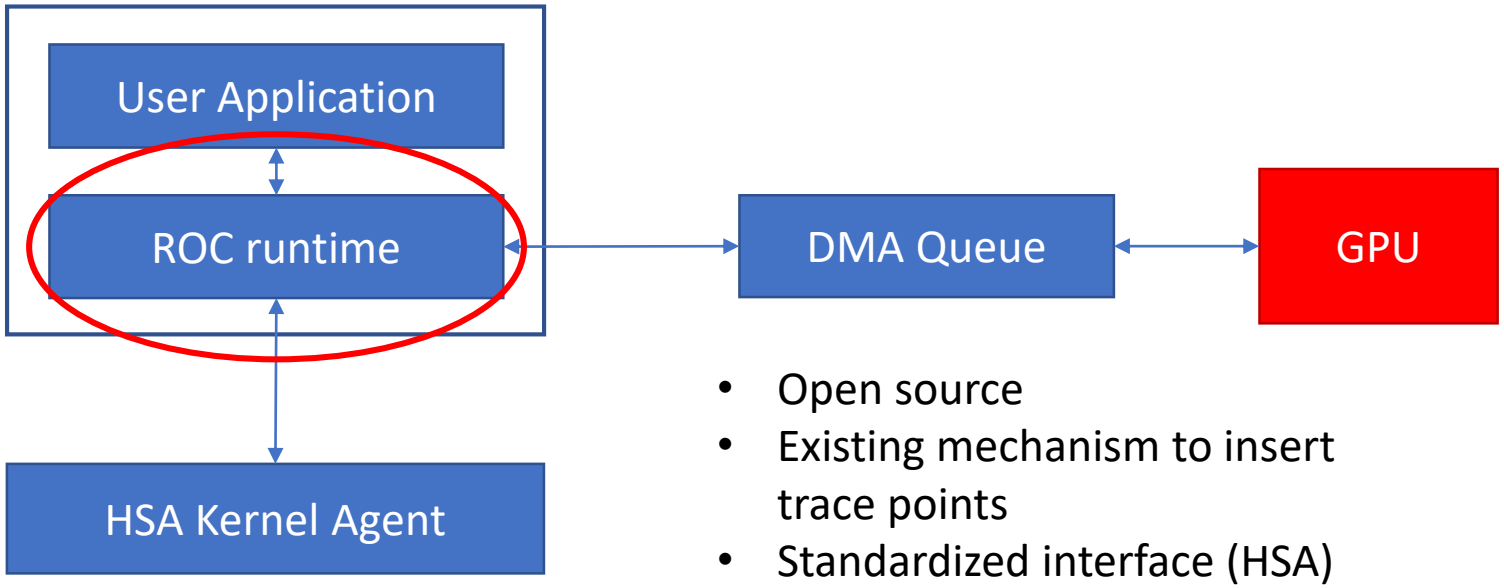
Introduction to ROCm

One of the key features of HSA is the heterogeneous Unified Memory Access :



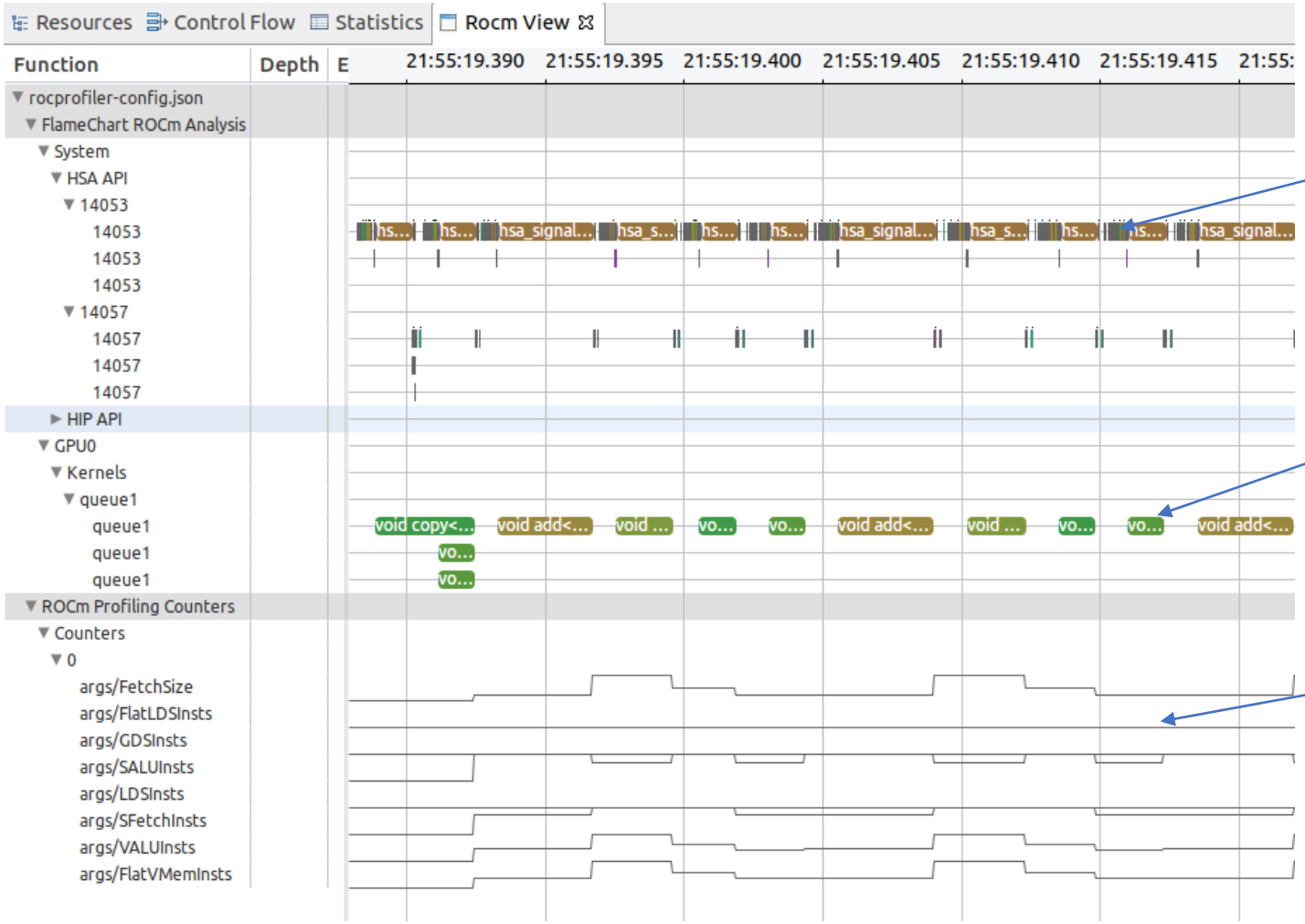
TraceCompass ROCm plugin

rocprofiler and roctracer
https://github.com/ROCm-Developer-Tools/rocprofiler
https://github.com/ROCm-Developer-Tools/roctracer



- Open source
- Existing mechanism to insert trace points
- Standardized interface (HSA)

TraceCompass ROCm plugin

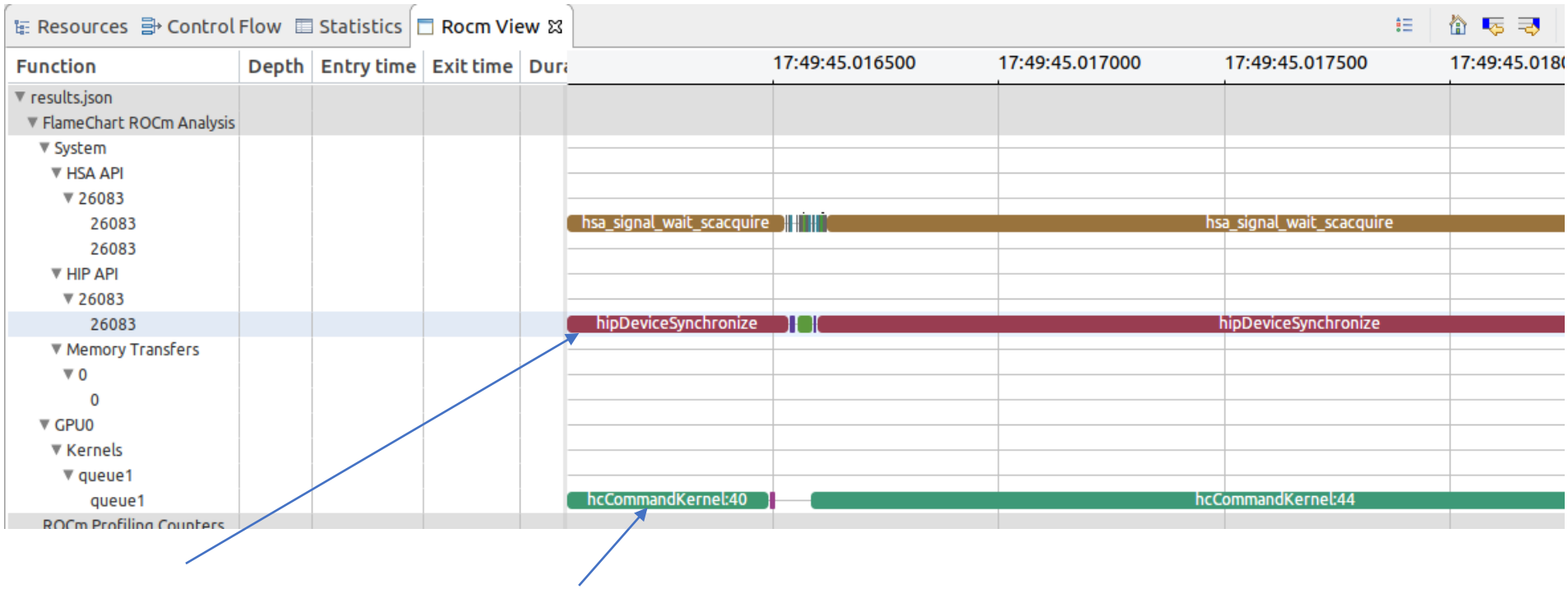


HSA function calls separated by thread

Kernel executions

Performance counters

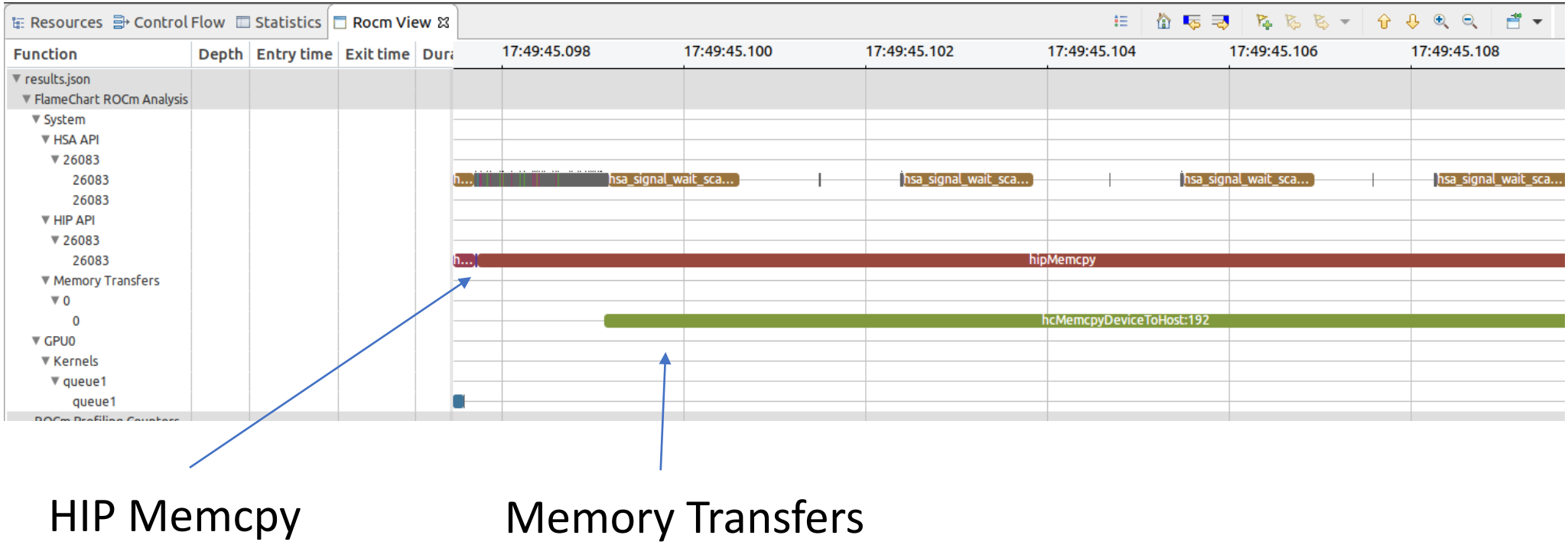
TraceCompass ROCm plugin



HIP function calls separated by thread

Kernel executions

TraceCompass ROCm plugin

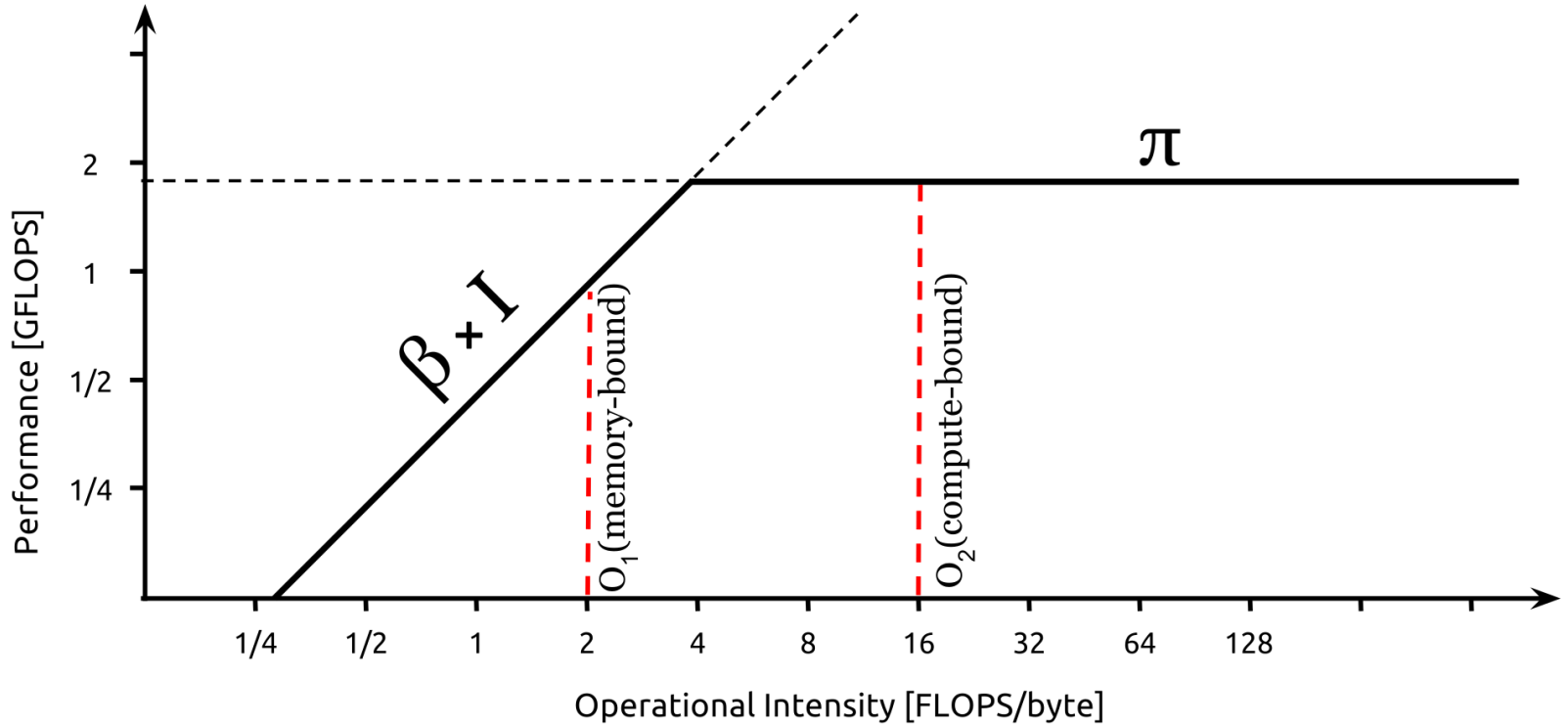


HIP Memcpy

Memory Transfers

Future Work

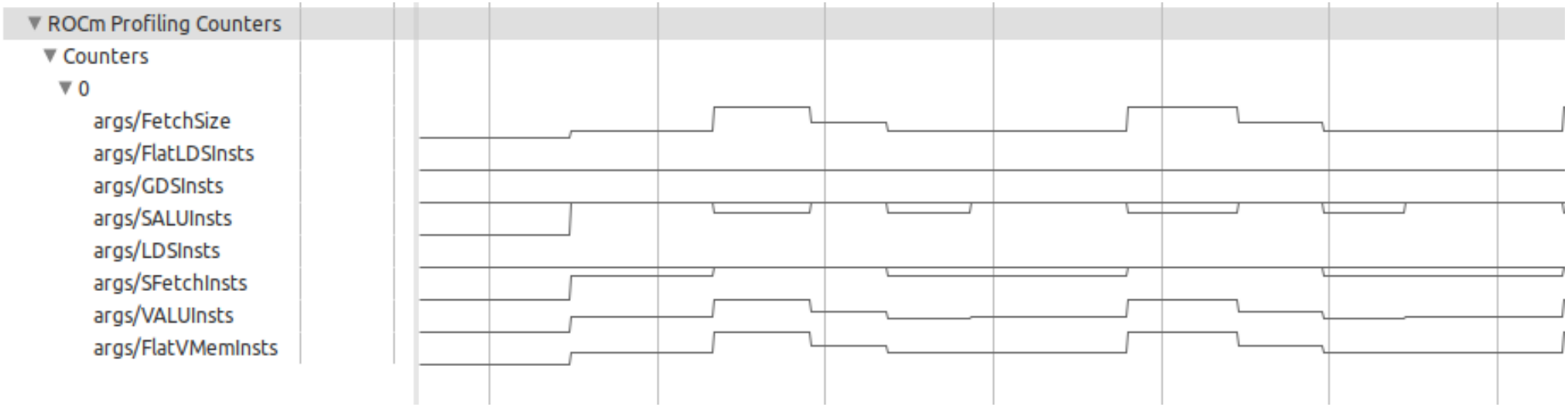
- Roofline Model coming soon



β : peak bandwidth
I : arithmetic intensity
π : peak performance

Future Work

- Top-down analysis



Future Work

- Synchronizing with kernel events will help to compare different parallel programming models (Kernel programming – Cuda, Hip – compared to OpenMP)

Thank you for listening !

Questions ?

References

- https://github.com/RadeonOpenCompute/ROCm
- https://rocm-documentation.readthedocs.io/en/latest/
- http://www.hsafoundation.com/
- HSA Runtime Programmer’s Reference Manual, Version 1.2
- HSA Programmer's Reference Manual, Version 1.2
- HSA Platform System Architecture Specification, Version 1.2
- https://github.com/ucb-bar/opencl-kernels/blob/master/saxpy/kernel.cl
- https://medium.com/@smallfishbigsea/basic-concepts-in-gpu-computing-3388710e9239
- https://www.techpowerup.com/gpu-specs/docs/amd-gcn1-architecture.pdf