

# GPU Kernel Instrumentation

Working Group Outbrief

Scalable Tools Workshop 2025

# Attending members

Ronak Chauhan, Matin Raayai-Ardakani, Hsuan-heng Wu, Angus He, Yumeng Liu, David Kaeli, Sebastien Darche

# Instrumentation on MI300A

- Integrated memory - shared between device and host
- Memory management should be easier
- Packed wave id in a single register
- Flat scratch is read-only
- ALL the SGPRs are allocated for a single wave - no effect on occupancy
-

# Channels

- Rocprofiler-sdk provides a way to set callbacks for ATT (Advanced Trace Tracing)
- S\_TTRACEDATA to send data to the host

# Occupancy talk

- Instrumentation may affect occupancy
- Increased private memory & register usage
- Difficult to know how NVidia manages it (do registers affect it or is it just for shared memory ?)

# Luthier Demo

- Much faster than NVBit for instruction counting!