

NVIDIA tools discussion

Luthier team request - NVdisasm access API

Need actual access via NVbit for all kinds of deep analysis and SASS level work. They can brute force it out of NVdisasm, please make it easier. Note that AMD tools already do publish theirs in public docs. Just needs to know how the instructions work so that he can build his representations on top of it. (See Luthier presentation for details)

Specifically wants

- Information about memory allocations
- Instruction encoding decoding support (at least symbolically)
- Instruction semantics
- Any ELF relocations they need to know

Questions

- Do you need a library or is offline sufficient? Offline sufficient?
- Can you do without some instructions? As long as all the control flow maps to PTX it is good.

Matt has questions

- Would like to get flops counts from monitoring tools. Want to be able to talk about how much the GPU was saturated, not just when it was doing anything.
- Would like to be able to run for a very long time (some is there, some is coming)
- Need to be better able to parse the debug information.
- Need a better way to chop out the data for a smaller range
- Would like Top-Down Analysis on the GPU side.

PAPI questions

CUPTI Changes

- Workflow for get counter availability is more expensive
- Would prefer to destroy the context
- Why move away from the chip name?\
- When is it going to be enforce for future chips
 - Tested with rtx 6000 - it was available
- Used to get metrics w/o suffix roll-up, now have to post process
 - Newer versions of CUPTI will not get available metrics from just the hardware id w/o the time hit of “get”ing the set from CUPTI (time hit)

PAPI native avail

- Initialize map based on unique devices on the system
- Expensive to get description but not to just get metrics
 - Previous was 18 minutes, now cached