Contribution ID: **119** Type: **not specified**

Open Source GPU compute stack - Not dancing the CUDA dance

Wednesday, 14 November 2018 09:45 (45 minutes)

Using graphics cards for compute acceleration has been a major shift in technology lately, especially around AI/ML and HPC.

Until now the clear market leader has been the CUDA stack from NVIDIA, which is a closed source solution that runs on Linux. Open source applications like tensorflow (AI/ML) rely on this closed stack to utilise GPUs for acceleration.

Vendor aligned stacks such as AMD's ROCm and Intel's OpenCL NEO are emerging that try to fill the gap for their specific hardware platforms. These stacks are very large, and don't share much if any code. There are also efforts inside groups like Khronos with their OpenCL, SPIR-V and SYCL standards being made to produce something that can work as a useful standardised alternative.

This talk will discuss the possibility of creating a vendor neutral reference compute stack based around open source technologies and open source development models that could execute compute tasks across multiple vendor GPUs. Using SYCL/OpenCL/Vulkan and the open-source Mesa stack, as the basis for a future task that development of tools and features on top of as part of a desktop OS.

This talk doesn't have all the answers, but it wants to get people considering what we can produce in the area.

I agree to abide by the anti-harassment policy

Yes

Primary author: AIRLIE, David

Presenter: AIRLIE, David

Session Classification: LPC Main Track

Track Classification: Refereed talk