CUMU
userspace CUDA emulator

May 13, 2023

why

current situation

CUDA is the de-facto API for GPU accelerated application
however proprietary and deeply tied to NVIDIA hardware

alternatives

HIP/SYCL/OpenCL/Vulkan Compute
mostly requires code change
with weaker ecosystems

what if

a CUDA implementation that interoperates with existing codes
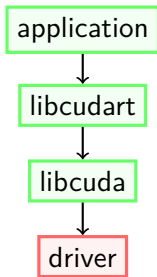a perfect chance to upset Jensen Huang

me

experiment NCCL without CUDA capable hardware

linux distributions

ease the testing of CUDA based softwares

researchers

a deep dive into CUDA internals

```
application
    ↓
 libcudart
    ↓
  libcuda
    ↓
  driver
```

libcudart

CUDA Runtime Library

libcuda

CUDA Driver Library

The runtime API eases device code management by providing implicit initialization, context management, and module management.
And by reimplementing libcuda, everything atop just works.

were

CUDA deviceemu

the official CUDA emulator at CUDA 3.0

https://github.com/vosen/ZLUDA

a drop-in replacement for CUDA on Intel GPU

https://github.com/gtcasl/gpuocelot

a dynamic compilation framework for PTX

- varying degrees of success
- mostly discontinued

Jensen Huang tries his best to stop us

- `dlsym` $\to$ `cuGetProcAddress`
- `cuGetExportTable` symbol name $\to$ uuid
- …