nouveau: in the times of firmware

Dave Airlie
Distinguished Engineer
Red Hat Australia

NVIDIA hardware history

- 1999 - NV04
- 2004 - NV40
- 2006 - NV50 - Tesla - Introduction of per context VMA
- 2010 - GF1xx - NVC0 - Fermi
- 2012 - GK1xx - NVE0 - Kepler - Vulkan support
- 2014 - GM1xx - Maxwell
- 2014 - GM2xx - Maxwell 2 - Start of signed firmware
- 2016 - GP1xx - Pascal
- 2017 - GV1xx - Volta
- 2018 - TU1xx - Turing - GSP support
- 2020 - GA1xx - Ampere

Rise of signed firmware

- Started with Maxwell 2
- Firmware had to be signed
- Complicated firmware boot sequences
- Bespoke nouveau firmware
- No reclocking PMU firmware

GSP Firmware

- Introduced with R515 driver
- Initially only data-center GPUs support
- Limited to Turing and above
- Needs RISC-V GSP (GPU system processor)
- Most proprietary functionality on GSP

NVIDIA open kernel modules

- Fork of NVIDIA proprietary driver
- Announced May 2022
- Moved "secrets" to GSP
- Remaining kernel code isn't secret
- Allows MIT release of kernel drivers
- Not upstreamable

Nouveau Background

- Started in 2007
- Reverse Engineered driver for NVIDIA GPUs
- Supports NV04 - Ampere in various states
- Stagnating for a few reasons
  - Developers getting hired away
  - Only one fulltime developer
  - No reclocking makes it hard to justify driver efforts

Kernel driver

- Mostly HW enablement activity
- Secureboot since Maxwell2 is very time consuming
- Multiple firmwares diverged from what NVIDIA uses
- Only NVIDIA can really debug
- Started adding GSP support

OpenGL driver

- Is GL4.5 complete, but never submitted
- Broken multithreading/context - recently fixed
- Hasn't seen a lot of optimisation due to reclocking

Vulkan driver

- Started work on a Vulkan 1.0 for Kepler -> Ampere
- Passes 85% of CTS for 1.0
- Requires new uAPI to finish

New user APIs

- Split BO/VMA management
- syncobj handling
- VM_BIND/exec API

Upcoming problems

- Firmware size
- initramfs blowouts
- No stable fw ABI yet

Future plans/possibilities

- New GSP only driver - to move forward?
- Compute Stack?