Power management update on Nouveau

Karol Herbst

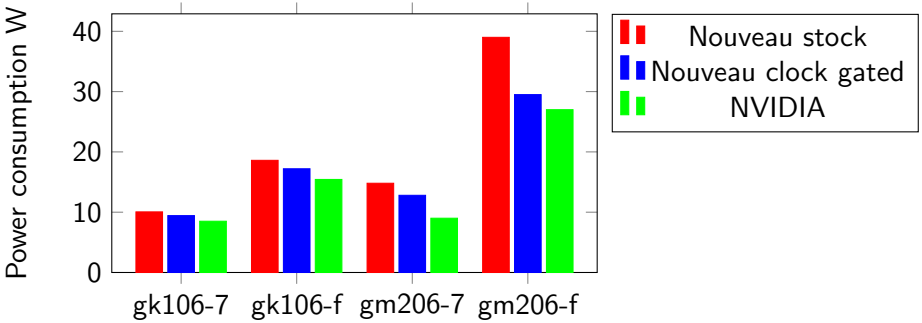
23rd September 2016

Overview

Power Sensors

- Initial support in 4.6
- Improved to be useful in 4.9
- Exposed through hwmon
- Based on Martins initial work
- Supported on mid+-end GPUs starting with Fermi
- Can hint at bad GPU utilization within mesa
- Should be moved into the PMU

Static Power Consumption - Clock Gating



- Up to 25% power savings
- Benefits some high loads
- WIP for Fermi, Kepler and Maxwell
- doesn't work without secboot

Static Power Consumption - Power Gating

- Might be able to close the gap
- Use counting of engines already in place
- Needs a lot of REing

GPU Boost - Introduction

- Advances DVFS
- 1.0: increase clocks until reaching power target
- 2.0: increase clocks until reaching temperature target
- Added officially with Kepler
- Already some bits on Fermi
- Mostly implemented in the kernel driver

GPU Boost - How it (really) works

- Limiting the voltage on high temperatures
- Information defined in a bunch of VBIOS tables

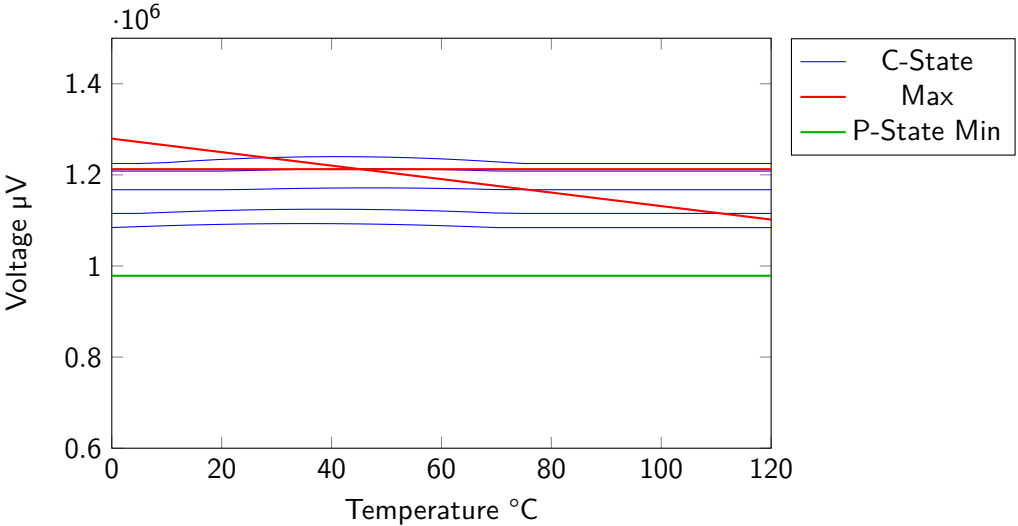
GPU Boost - Voltage

- VID table
  - How to voltage (GPIOs, PWM)
  - What voltages can be set (ranges, step sizes)
- VMAP table
  - Voltage calculations depending on
    - T: Temperature
    - S: Individual chip calibration factor (Speedo)
  - Defines "max" voltage entries
  - $volt_0 = \frac{c_0}{10} + \frac{c_1 * S}{10} + \frac{c_2 * S^2}{10^5}$
  - $volt_1 = \frac{c_0 * 5^6}{2^{18}} + \frac{c_1 * S * 5^6}{2^{18}} + \frac{c_2 * T * 5^6}{2^{10}} + \frac{c_3 * S * T * 5^6}{2^{18}} + \frac{c_4 * S^2 * 5^6}{2^{30}} + \frac{c_5 * T^2 * 5^6}{2^{18}}$
  - $volt_2 = min$
  - $volt_3 = \frac{min + max}{2}$
- CSTEP table
  - Voltage requirements for a C-State

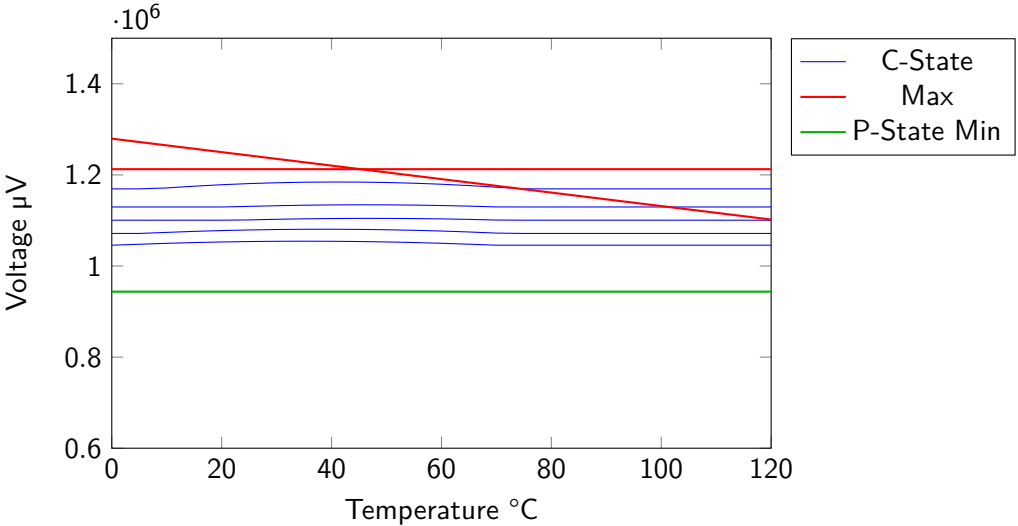
GPU Boost - Clocking

- CSTEP table
  - List of GPC clocks (C-States)
  - Defines max C-State of P-States
- PMode table
  - "Default" clocks for each engine
- BOOST table
  - Clock adjustments for Engines (eg. 90% of GPC clock)
  - Min/Max clocks for each P-State
- VPSTATE table
  - Actually (partly) documented!
  - Specifies Base clocks (confirmed)
  - ... and Boost clocks

GPU Boost - Example: GTX 780 Ti - speedo 0x610



GPU Boost - Example: GTX 780 Ti - speedo 0x691



GPU Boost - nv_cmp_volt

- Reverse engineering tool
- Reads out hardware state
- Doesn't interact with the NVIDIA driver
- Verifies Nouveau voltage algorithm against NVIDIA
- Demo later

GPU Boost - Solved problems

- Didn't respect Voltage limits
- Tried to perfectly match voltage
- Missing voltage calculation
- No updates on temperature changes (WIP)

State of reclocking - What is in a good state?

- Hardware thermal protection up to GM10x
- Fan management up to GM10x
- Engine reclocking for most cards up to Maxwell
- Memory reclocking for Tesla (G92+), Kepler and Maxwell
- PCIe link configuring
- GPU Boost (partly in 4.9)

State of reclocking - Work in Progess

- PMU engine load counters for GT215+
- Memory reclocking on Fermi (Roy Spliet)
- Software thermal protection
- Power budgets

State of reclocking - What is missing?

- Clock read out while hardware thermal protected
- Maxwell2 fan control (signed PMU firmware required)
- Advanced GK110+ engine reclocking
- Hi-PLL Fermi+ engine reclocking
- Linebuffer/isohub? (flicker free memory reclocking)
- New fan control VBIOS table for Kepler+
- Memory reclocking on G80-G86
- Pascal is completly different

State of reclocking - DVFS

- Policies
  - What is usable in general?
  - How to deal with spiky loads?
- PMU notifies host about high/low loads
- Various benchmarks needed
  - Performance
  - Interactive (Desktop environment)
  - Web browsers
- Prototype!

Demos

- nv_cmp_volt
- DVFS