Data Center Server Agnostic Power Management

Bryan Kelly – Principal Firmware Eng Manager
Project Olympus System Overview
Project Olympus

- System and Rack design

- Manufacturing collateral; schematics and board files open sourced in Feb 2017

- Open Sourced Firmware: Open EDKII, OpenBMC and Open PDU and Rack Manager.
## Power Management Distribution Unit

### Modules | Features
--- | ---
**PMDU** | • 48U and 42U support  
• AC Power for rack devices (FCI and Cable)  
• Enables blade management and power control  
• Dual Feed with Circuit Breakers  
• Hot Plug Rack Manager and AC/DC Converter  
• No active single point of repair

**Rack Manager + AC/DC** | • Hot pluggable  
• 1GbE Management  
• Power Meter for PMDU  
• UARTs for switch and Digi  
• Blade power control

**Management Switch** | • COTS 1GbE managed switch  
• Cold aisle cabling

**Standalone Rack Manager** | • 1U Standalone unit  
• Non-WCS rack management  
• Row management  
• Colo management  
• Reuse of Rack Manager Board Assy
Rack Level Power Control Mechanism

Universal PMDU
- Dual-feed, three-phase Power
- Integrated Rack Management
- Adapter AC whips for Data Centers
  - 208V-50A, 415V-30A, 400V-32A
- Power monitoring and throttle
  Servers blind-mate – no rear access

Three Phase Power Supply
- Three x 340W PSUs integrated in one
  - Dual feed, three phase AC power in
  - Auto selection of feed (IVS)
- Fault Mode Resiliency
  - AC feed failure, automatic fail over
  - PSU failure throttles if necessary
  - Double fault will be extremely rare

- Normal mode
  - 1020W
  - 12v: 340W
  - 12v: 340W
  - 12v: 340W

- Phase fault
  - 680W
  - 12v: 340W
  - 12v: 340W
  - 12v: 0W

- Double fault
  - 340W
  - 12v: 340W
  - 12v: 0W
  - 12v: 0W

Throttles & triggers repair
Rack Level Power Control Mechanism

- PMDU fans-out gpio logic from Rack Manager to servers
- Servers are configured with a Power Policy
  - Activated upon alert signal
- Rack Manager contains rack level power meter, that runs in a dedicated real-time coprocessor.
- Rack Manager assigned power limit for the rack
- If limit is reached, signal asserts causing servers to activate their policy (instantly throttle, depending on policy)
- Rack Manager has option for auxiliary in-puts (row, colo)
Intel Baseboard Power Control Logic

• Alert logic can immediately throttle CPU power (<1ms)
• BMC intercepts signals and applies power control policy to chipset, then deasserts logical alert
• Chipset implements power/frequency limits on the CPU
• Chipset perpetually polls HSC for platform power, and adjusts CPU limit
CPU Agnostic Baseboard Power Control Logic

- Alert logic can immediately throttles CPU power (<1ms)
- BMC intercepts signals and BMC co-processor applies power policy to CPUs, then deasserts logical alert
- BMC implements power/frequency limits on the CPU
- BMC perpetually polls HSC for platform power, and adjust CPU limit
Power Control – Rack Level Interactions

1. Detect OC power events
2. Hardware interrupt (1ms response)
3. Emergency Hard Power Cap (1ms of power event signal, 21ms including OC fault detection time)
4. Performance-aware Soft Power Cap (within ms of power event)
5. Read Alert events
   - Under voltage
   - Phase loss
   - Over current
   - Feed failure
   - HW Int
Dynamic Rack Level Power Control

• Static server level limit can impact system performance.
• Rack level permits servers run as high as they like, but cumulatively below the rack limit
• Tighter power provisioning made possible due to internal evitable workloads
• Fabric orchestrates policy, priority given to critical workloads
Row Management

- Standalone Rack Manager located in MOR
- Controls up to 24 Rack Managers
  - Remote Power Control
  - Bootstrap
  - Rack Level Throttle
- RJ45 Cables
- RM Throttle bypass

Middle of Row Manager, same hardware and firmware, different role.

[Diagram showing the connection of Rack Managers to a Middle of Row Manager (MOR)]
Colo Power Management

- Redundant Colo Managers at Power Source (same device, firmware, different role).
- Standalone Rack Manager located in Row distribution panels.
- Hierarchical design.
- Alert signal distribution propagates from data center infrastructure to server infrastructure.
Further Information

• Project Olympus:
  • [https://github.com/opencomputeproject/Project_Olympus](https://github.com/opencomputeproject/Project_Olympus)

• Rack Manager
  • [https://github.com/Project-Olympus/rackmanager](https://github.com/Project-Olympus/rackmanager)

• OpenUEFI
  • [https://github.com/tianocore/edk2-platforms/tree/devel-MinPlatform/Platform/Intel/PurleyOpenBoardPkg/BoardMtOlympus](https://github.com/tianocore/edk2-platforms/tree/devel-MinPlatform/Platform/Intel/PurleyOpenBoardPkg/BoardMtOlympus)
OPEN.
FOR BUSINESS.

OCP SUMMIT