Open. Together.
The Journey of OCP NIC3.0
Solution Integration and Demo

Presenter:
Ben Wei, Software Engineer, Facebook
Damien Chong, Hardware Engineer, Facebook

Guest speaker:
Gady Rosenfeld, Broadcom
Yuval Itkin, Mellanox
NIC3.0 integration into Yosemite V2
Design Journey: Analysis
Design Journey: Game Plan

- Investigate and experiments with NIC sideband
- Influence NIC3.0 spec accuracy & completeness
- Multi-host platform for NIC vendor’s hardware & firmware readiness
- Fast availability
- NIC3.0 LFF (Large Form Factor) functionality
Take note:

- PVPERL in NIC3.0 is >1s instead of the common PCIe CEM >100ms.

One of the use-cases for longer duration allows NIC ASIC to attest for secure firmware before operational.
Signal Description

The USB interface shall be based on a $V_{BUS} = 3.3 \, \text{V}$. Both the baseboard and NIC device shall be capable of driving signals using 3.3 V logic. The OCP NIC 3.0 card may implement protection diodes and is up to the adapter vendor for placement.

To prevent leakage paths, a baseboard shall not use USB pull up resistors on the USB_DATp/n lines to indicate the bus data transmission rate. If used, pull up resistors shall only exist on the NIC side.
Yosemite V2 NIC3.0 adapter card
NIC3.0 supplier community partnership

* Pictures supplied by named companies
The Journey of OCP NIC3.0
Solution Integration and Demo

Guest speaker:
Gady Rosenfeld, Broadcom
Thor – Industry’s High Performance Ethernet NIC

200Gbps Ethernet Controller
8x 56G PAM4 SERDES
Native PCIe Gen4 Host Interface
< 800ns Latency
100 Million Packets Per Second
RoCEv2 with Smart Congestion Control
**TruFlow™** Virtualization Offloads, Up to 1K VFs
**BroadSafe™** Security – Secure Boot/Erasure, Si RoT
**Quad-Port** 200G Controller
Cloud, Storage, Web 2.0, Machine Learning & more
### New Broadcom OCP3.0 Portfolio - NetXtreme E-Series

<table>
<thead>
<tr>
<th>Name</th>
<th>Port Speed</th>
<th>I/O</th>
<th>Host I/F</th>
<th>Multi-Host</th>
<th>Chip Family</th>
<th>Part Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>N41GBT</td>
<td>4x 1G</td>
<td>RJ45</td>
<td>PCIe 2.0 x4</td>
<td></td>
<td>NX1</td>
<td>BCM95719-N1905C</td>
</tr>
<tr>
<td>N210P</td>
<td>2x 10G</td>
<td>SFP+</td>
<td>PCIe 3.0 x8</td>
<td></td>
<td>Whitney+</td>
<td>BCM957412-N4120C</td>
</tr>
<tr>
<td>N210GBT</td>
<td>2x 10GBase-T</td>
<td>RJ45</td>
<td>PCIe 3.0 x8</td>
<td></td>
<td>Whitney+</td>
<td>BCM957416-N4160C</td>
</tr>
<tr>
<td>N410G</td>
<td>4x 10G</td>
<td>SFP+</td>
<td>PCIe 3.0 x8</td>
<td>Y</td>
<td>Stratus</td>
<td>BCM957452-N410SG</td>
</tr>
<tr>
<td>N410GBT</td>
<td>4x 10GBase-T</td>
<td>RJ45</td>
<td>PCIe 3.0 x8</td>
<td>Y</td>
<td>Stratus</td>
<td>BCM957452-N410SGBT</td>
</tr>
<tr>
<td>N225P</td>
<td>2x 25G</td>
<td>SFP28</td>
<td>PCIe 3.0 x8</td>
<td></td>
<td>Whitney+</td>
<td>BCM957414-N4140C</td>
</tr>
<tr>
<td>N425G</td>
<td>4x 25G</td>
<td>SFP28</td>
<td>PCIe 3.0/4.0 x16</td>
<td>Y</td>
<td>Thor</td>
<td>BCM957504-N425G</td>
</tr>
<tr>
<td>N250G</td>
<td>2x 50G</td>
<td>SFP56</td>
<td>PCIe 3.0/4.0 x16</td>
<td>Y</td>
<td>Thor</td>
<td>BCM957504-N250G</td>
</tr>
<tr>
<td>N1100G</td>
<td>1x 100G</td>
<td>QSFP56</td>
<td>PCIe 3.0/4.0 x16</td>
<td>Y</td>
<td>Thor</td>
<td>BCM957504-N1100G</td>
</tr>
<tr>
<td>N2100G</td>
<td>2x 100G</td>
<td>QSFP56</td>
<td>PCIe 3.0/4.0 x16</td>
<td>Y</td>
<td>Thor</td>
<td>BCM957508-N2100G</td>
</tr>
<tr>
<td>N2200G</td>
<td>2x 200G</td>
<td>QSFP56</td>
<td>PCIe 3.0/4.0 x16</td>
<td>Y</td>
<td>Thor</td>
<td>BCM957508-N2200G</td>
</tr>
</tbody>
</table>

**Complete Portfolio of OCP NIC 3.0 Adapters**

1G to 200G

---

**Open. Together.**
The Journey of OCP NIC3.0
Solution Integration and Demo

Guest speaker:
Yuval Itkin, Mellanox
OCP NIC 3.0: Mellanox’s Journey

Shaping The Standard
Early participation | Active Contributor

Sharing Knowledge
Millions of OCP 0.5/2.0 | Hyperscale | Smart-NICs

Management Features
Auto-adjusting mode | PLDMoRMII FW Update

Hardware

Mechanical and Thermal
ConnectX-5 and Beyond

**2*100G, Gen4 x16, 150 MPPS**
Hardware-based XDP acceleration 100 MPPS

**Secured Firmware**

**Multi-Host and Management**
Up to 4 hosts | NCSI | MCTPoPCIe | PLDM

**Shipping in Volumes**
OCP | Major OEMs | Hyperscale

**Storage and Cloud Offloads**
NVMeF | VXLAN | Header Re-writes | Erasure Coding…

**Advanced Security**
Inline IPsec and TLS encryption, Secured Boot

**Advanced Networking**
Enhanced Multi-host buffer | congestion handling

**Advanced Offloads**
Flexible program’ pipeline | vSwitch/vRouter

**Sampling Soon**
PLDM Firmware Update over RMII

- **Platform Level Data Model**
- **OpenBMC** in OCP platforms perform:
  - Firmware updates through the BMC
  - Inventory management operations
  - Hardware monitoring: temperature, health, cable state, etc.
- Open source and standard environment
- Based on **DMTF** pre-standard draft (expected e/2019)
The Journey of OCP NIC3.0
Solution Integration and Demo

Presenter:
Ben Wei, Software Engineer, Facebook
Damien Chong, Hardware Engineer, Facebook
OpenBMC NIC Bring-Up journey: NIC3.0

- **OCP 3.0 NIC:**
  - Supports NC-SI v1.1 standard (firmware pending)
  - Allows BMC to power on/off NIC independent of sled platform power
  - Provides more flexibility for management controller, BMC may power cycle NIC as a last resort recovery mechanism if a NIC fatal error is detected or NIC becomes unresponsive

- We worked with Mellanox, Broadcom, & ODM partners to bring up NIC3.0 on Yosemite V2 platforms as a proof of concept
  - Yosemite V2 already supports OCP 2.0
  - To support OCP 3.0, OpenBMC implements auto NIC type detection (OCP v2.0 vs v3.0) via MEZZ_PRSNT pins.

- Upon detection of OCP NIC 3.0, OpenBMC powers up NIC via AUX_PWR_EN & MAIN_PWR_EN

- NIC Sensor monitoring – both through I2C bus and through PLDM over NC-SI
OpenBMC Initialization flow for OCP 3.0 NIC

- Read MEZZ_PRSNT pins to determine NIC type
  - use GPIOL0 (MEZZ_PRSNTA2_N) and GPIOL1 (MEZZ_PRSNTB2_N).
  - the supported OCP 2.0 NIC on Yosemite V2 all have 2 connectors, that is, PRSNT_A2 and PRSNT_B2 to baseboard
  - OCP3.0 cards only has PRSNT_A2 to baseboard

- If OCP 3.0 NIC is detected, turn on Aux and Main power
- Initialize NC-SI interface
- Read NIC firmware version to determine NIC vendor
- Continue with vendor-specific initialization process
  - Sending NC-SI OEM commands to continue NIC initialization
NIC Sensor Monitoring

- OCP 3.0 NIC contains onboard I2C temperature sensor
- Additionally, new NIC firmware supports PLDM Type 2 (PLDM For Platform Monitoring and Control) that allows BMC to access and monitor additional sensors
- As part of OCP 3.0 NIC Bring-Up, we added PLDM sensor monitoring to OpenBMC
  - Sensor are accessed through PLDM over NC-SI/RBT
  - BMC tracks sensor history and logs min/max value over time, in addition to display real time sensor values
  - Addition sensors will be added once they are available
PLDM Firmware Update

- As part of the bring-up process, we worked with our partners to prototyped PLDM over NC-SI/RMII
- In addition to reading sensors over PLDM, OpenBMC supports NIC firmware update through PLDM over NC-SI/RBT
- Firmware supports delayed-activation: at the end of firmware update process, new firmware is not activated until host power cycles or BMC resets NIC
Power-on journey: Expect the unexpected

10-Kohm pull-down

Low ~1.3v
High ~3.3v

100-ohm pull-down

Low ~0.1v
High ~3.0v

No change to NIC3.0 spec yet
Power-on journey: 4-hosts (multi-host) data transfer
Power-on journey: Out-of-band data transfer
THIS JOURNEY
1% FINISHED
Call to Action

Join OCP NIC3.0 community in defining and developing scalable solution.

OCP NIC3.0 Wiki with latest specification:
http://www.opencompute.org/wiki/Server/Mezz

Visit us at OCP Experience Lab for live demo

Mini plug-fest 3-4pm today!
Back-up