Zaius & Barreleye G2

48-volt Google / Rackspace OpenPOWER Platform

Aaron Sullivan  |  Distinguished Engineer  |  Rackspace
Adi Gangidi  |  Systems Design Engineer  |  Rackspace
Rob Lippert  |  Staff Engineer  |  Google
WE’RE TAKING A TOUR...
Introducing Zaius
Google is excited about POWER9 and our collaborative OCP platform codenamed Zaius!

- This platform is a result of a groundbreaking collaboration between the OpenPOWER Foundation, Open Compute Project, OpenCAPI, IBM, Google and Rackspace.
- POWER9 provides a significant upgrade in compute capability and memory capacity.
- Massive I/O connectivity with 80x gen4 PCIe lanes.
- Innovative OpenCAPI expansion ports for new class of accelerators.
- Openness in hardware, firmware, software helps us move faster.

2016 OCP Announcement Block Diagram

2017 OCP Working Prototype
Google Zaius Hardware Quick View

“Zaius” form factor sled – the Google version

- Designed for Google's datacenter environment
- Compatible with 48V open rack v2
- All I/O and service access in front
- Compact, simple, low cost sheet metal
- Short 1.5OU height requires slightly shorter than standard HH PCIe cards
- Shelf with horizontal bus bar allows flexibility in deployments, multiple expansion options
- Many small Google-specific touches to enhance connectivity, debug-ability
Rackspace Barreleye G2 Hardware Quick View

Zaius motherboard in a new and improved Barreleye chassis

- Built for full-depth 48V open rack v2
- Front IO & service access, hot swap fans
- Supports full height full length cards
- Low profile chassis option (1.250U)
- High capacity chassis option (2OU)
- User-friendly physical service
- High density storage bay
LOTS OF GROUND TO COVER
Topics We’re Covering

Zaius / Barreleye G2 Family Tree

PCIe Gen 4, OCP Mez on Gen4, Host & BMC IO Integration

Coherent Interconnects: 25G OpenCAPI and NVLink

Barreleye G2 Storage Design

48-volt Point of Load in Zaius, Barreleye G2

48-volt OpenRack v2 & Benchtop Solutions

Firmware (OpenPOWER / OpenBMC)
Simple Family Tree

- Zaius 1.5 OU Sled
- Zaius 1.5 OU Shelf
- Zaius Motherboard
  - 2 Storage Options
  - Barreleye G2 1.25 OU Chassis
  - 1 IO Cage
  - Barreleye G2 2 OU Chassis
  - 2 Storage Options
  - 3 IO Cage Options
Tour Map

Zaius Motherboard
*(most time, here)*

PCIe Gen4
OpenCAPI & NVLink

Storage Expander
(SAS/SATA/NVMe)

48v Open Rack
Bus Bar & Fan Adapter

Storage Canister Backplane
PCIe G4 and OCP Mez 2.0
IO Slots: Standard G4, G4 Mez 2.0, LOM+VGA

5 – 6 PCIe G4 slots
2 x8 slots
3 x16 slots
1 x8 / x16 OCP Mez 2.0 slot
(x8 if using LOM+VGA mux card)
## PCIe G4 Slots & Bandwidth

<table>
<thead>
<tr>
<th>Slot Type</th>
<th>QTY</th>
<th>TOTAL THROUGHPUT (UNI-DIR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>x16 Gen4</td>
<td>3</td>
<td>31.5 Gbytes / sec (252 Gbps)</td>
</tr>
<tr>
<td>x8 Gen4</td>
<td>2</td>
<td>15.75 Gbytes / sec (126 Gbps)</td>
</tr>
<tr>
<td>x16 Mez A/B Gen4</td>
<td></td>
<td>31.5 Gbytes / sec (252 Gbps)</td>
</tr>
<tr>
<td>or x8 Mez A Gen4*</td>
<td>1</td>
<td>15.75 Gbytes / sec (126 Gbps)</td>
</tr>
</tbody>
</table>

* Required if using LOM+VGA pass-through, as Mez B x8 connector is used by interposer card.
PCle G4 Mez 2.0 & LOM+VGA Pass-through

BergStak™
16Gb/s Gen4 support
Gen3 OCP Mez 2.0 Compatible
5 / 8 / 12 / 16 mm stack heights

<table>
<thead>
<tr>
<th>HBA &amp; SHARED LOM+VGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>x8 HBA</td>
</tr>
<tr>
<td>LOM+VGA Connector</td>
</tr>
<tr>
<td>VGA &amp; LOM</td>
</tr>
<tr>
<td>Host + BMC Access</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>x16 HBA (INDEPENDENT LOM, NO VGA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>x16 HBA</td>
</tr>
<tr>
<td>Host LOM</td>
</tr>
<tr>
<td>BMC LOM</td>
</tr>
<tr>
<td>Host Access Only</td>
</tr>
<tr>
<td>BMC Access Only</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>x8 HBA (INDEPENDENT LOM, NO VGA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>x8 HBA</td>
</tr>
<tr>
<td>Host LOM</td>
</tr>
<tr>
<td>BMC LOM</td>
</tr>
<tr>
<td>Host Access Only</td>
</tr>
<tr>
<td>BMC Access Only</td>
</tr>
</tbody>
</table>
LOM + VGA Pin Assignments

Connect to LOM BCM 5719 via PCIex4

Connect to BMC AST2500 via PCIex1
Why This Matters

• Gen4 adapters can be fit into OCP mez 2.0 form factor

• This is possible, but not necessarily optimal
  ▪ May not optimal for low profile chassis solutions
  ▪ Little space for NICs with extra co-processor & memory footprint

• Now is a good time to revisit OCP mezzanine spec
25G OpenCAPI & NVLink
OpenCAPI / NVLink Connector Location & Type

- 2 “Bricks” per CPU Socket
- 24G Slimline SAS Connectors (25GByte per Brick)
- Follows SFF-8654 Standard
Zaius / BG2 OpenCAPI & NVLink Interface

13mm (23.5 mated)

23.5mm (25.95 mated)

7.1mm (9.24 mated)
Zaius / BG2 OpenCAPI & NVLink Cable

<table>
<thead>
<tr>
<th>DESCRIPTION</th>
<th>QTY</th>
<th>TOTAL THROUGHPUT (UNI-DIR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVLink 2.0 / OpenCAPI 3.0 Brick</td>
<td>-</td>
<td>25 Gbytes / sec (200 Gbps)</td>
</tr>
<tr>
<td>Bricks per Socket</td>
<td>2</td>
<td>50 Gbytes / sec (400 Gbps)</td>
</tr>
<tr>
<td>Bricks per Server</td>
<td>4</td>
<td>100 Gbytes / sec (800 Gbps)</td>
</tr>
</tbody>
</table>

### SPECs

<table>
<thead>
<tr>
<th>MEASUREMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>RECEPTACLE WIDTH:</td>
</tr>
<tr>
<td>PLUG WIDTH:</td>
</tr>
<tr>
<td>MATED HEIGHT:</td>
</tr>
<tr>
<td>MATED LENGTH:</td>
</tr>
<tr>
<td>30AWG DIAMETER:</td>
</tr>
<tr>
<td>PCB THICKNESS:</td>
</tr>
</tbody>
</table>

### SPECS MEASUREMENT

- Receptacle Width: 23.5mm
- Plug Width: 25.95mm
- Mated Height: 22.4mm
- Mated Length: 9.24mm
- 30AWG Diameter: 0.255mm
- PCB Thickness: 1.0mm
Example Card and Cable Mount Options*

**Perpendicular Mount**
- Accelerator Card w/PCIe card edge & 25G Connector
- 25G Cable
- PCIe slot
- Sideband signals (low freq)

**Parallel Mount w/ Mez**
- Accelerator card (e.g. GPU, OpenCAPI, etc)
- Supplier native connector
- PCIe to Native mez adapter
- PCIe slot
- Sideband signals (low freq)

* Possible approach to co-processor card, mount, and cable designs.
** Does not refer to OCP Mez standard.
Why This Matters

- Coherent attach is powerful – soon to be common
- Too many wire & connector standards
- Gratuitous differentiation at work here
- Could fragment motherboard & chassis specs
- Rackspace considers SFF-8654 ("Slimline SAS") to be a good choice for OCP
- Let’s work together with processor & co-processor companies to unify on a common standard
  - Include PCI-SIG, CCIX, OpenCAPI
Storage Design
Barreleye G2 Storage Overview

Storage Expander Board

Storage Canister Backplane

Host Bus Adapter
BG2 EVT Storage Architecture (Today)

24 HIGH SPEED LANES IN DRIVE BACK PLANE

Drive Back-Plane

x1 SAS / SATA wired to 24 Drives

x8 SAS [7:0]

SlimSAS Conn

MiniSAS Conn

X8 Phy [23:16]

X8 Phy [15:8]

X8 Phy [7:0]

X8 PCIe Gen4

BCM 9361-8i

SAS 35x48

Storage Expander Board
Takeaways

• Barreleye G2 2 O.U will support SAS / SATA / NVMe in each slot

• Drive backplane design is getting complicated
  ▪ Need for next generation connector and protocol standards to ease transition

• Tri-mode Transition
  ▪ U.3 connector proposed to follow U.2
  ▪ Tri-mode Expander
  ▪ U.3 Drives
48 Volt & Point of Load VR
48V Barreleye G2 Voltage Distribution

- 12V from Expander Board to Drive plane
- 48V Input to MB
- 12V to Fans
- 48V Input from Bus Bar
- 12V to Expander Board
- 12V from Expander Board to Drive plane

* Zaius sled & shelf use only motherboard. Additional PCBAs not used.
** Barreleye G2 chassis uses Zaius motherboard and additional PCBAs shown here.
# Zaius Chassis / Motherboard 48V Step Down

## 48V Motherboard @ 41.185A

- 1x Vicor CPU 0 VR 1
  - CPU 0.7V
  - 0.7V @ 74.4 A
  - 90% Efficiency
- 3x Vicor CPU 0 VR 2
  - CPU 0.8V
  - 0.8V @ 297 A
  - 90% Efficiency
- 2x Vicor DIMM VR 1
  - 32X DIMM 1.2V
  - 1.2V @ 95.54A each
  - 90% Efficiency

## 48V Hot-Swap Controller ADM1272

- 1x CPU VR1
  - CPU 0.7V
  - 0.7V @ 74.4 A
  - 90% Efficiency
- 2x CPU VR2
  - CPU 1.2V
  - 1.2V @ 95.54A each
  - 90% Efficiency
- 3x CPU VR2
  - CPU 1.2V
  - 1.2V @ 297 A
  - 90% Efficiency

## AVSBus

- Finer-grained CPU power

## Delta Power Brick 48V -> 12V

- System Fans @12V
- OCP Mezz / PCIe @12V
- Misc Infineon VR @5V
- Misc Infineon VR @3.3V
- Misc Infineon / Intersil VR < 3.3V
- USB/VGA/HDD/MEZZ AUX @5V
- SATA /LOM/ BMC / I2C / PCIe AUX / Mezz AUX / On-board storage @ 3.3V
- CPU/1 0.96 – 3.3V
- DDR4CPU 0/1 0.6 – 2.5V
Takeaways

• 48V POL Regulators & AVSBus
  ▪ Compact solution with high efficiency

• Barreleye G2 Add-on power board
  ▪ Provides choice of high-wattage sub-systems (GPU / Storage) without penalty of routing power through Motherboard
48 Volt Rack & Power Shelf
48-Volt Server-to-Bus-Bar Clip

6* Fan(60x60x56 )

Bus bar Clip
48-Volt Open Rack v2 for Barreleye G2

OCP 48V Rack
+48VDC bus bar
(one power zone)
30U 30kW power shelf
(36kW without redundancy)
41 - 420U total
38 - 390U with power shelf
48-Volt Rack Power System

30 kW Power Shelf
- 3 OU height, 2x – 4x 60A AC input
- 4x – 6x C14 socket 10A PDU (rear)
- 12x 3kW PSU Modules
- 6x ATS Modules (1 per 2 PSUs)
- 10+2 Redundancy (if ATS fails)
- Rack Management Controller

12 PCS 3kW PSU
6 PCS 6kW ATS

RMC
48v Lunchbox (~1700w)

48v Output and Cable

110 – 240v Input and Cables
OpenPOWER / OpenBMC
Zaius is part of the OpenPOWER community and is planned to be a fully compliant/tested OpenPOWER system.

- Committed to providing a fully open-source firmware and management stack to go along with server – no black boxes.
- External contributors welcome and encouraged to contribute!

OpenPOWER boot firmware (aka BIOS) fully supported and booting now:

- `git clone --recursive https://github.com/open-power/op-build`
- `op-build zaius_defconfig && op-build`
- Zaius work in progress:
  - Stability, error logging/reporting, slot bifurcation, ...

OpenBMC on Zaius AST2500/2520 working to initialize hardware and start booting the machine:

- `git clone https://github.com/openbmc/openbmc` (Advances Facebook BMC with new kernel and drivers)
- `export TEMPLATECONF=meta-openbmc-machines/meta-openpower/meta-ingrasys/meta-zaius/conf; . openbmc-env`
- `bitbake obmc-phosphor-image`
- Zaius work in progress:
  - Sensors, fan control, NC-SI muxing, LEDs, POST card, boot time, ...
# It Takes a Village: Partner Appreciation

<table>
<thead>
<tr>
<th>Category</th>
<th>Brand 1</th>
<th>Brand 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bus Bar Solutions</td>
<td>TE</td>
<td>Amphenol</td>
</tr>
<tr>
<td>Networking</td>
<td>Mellanox</td>
<td>BROADCOM</td>
</tr>
<tr>
<td>Storage</td>
<td>BROADCOM</td>
<td>SAMSUNG</td>
</tr>
<tr>
<td>POWER9</td>
<td>IBM</td>
<td>OpenPOWER</td>
</tr>
<tr>
<td>OpenCAPI / NVLINK 25G Connectivity</td>
<td>Amphenol</td>
<td>OpenCAPI</td>
</tr>
<tr>
<td>Accelerators</td>
<td>NVIDIA</td>
<td>XILINX</td>
</tr>
<tr>
<td>Memory</td>
<td>SAMSUNG</td>
<td>Micron</td>
</tr>
<tr>
<td>Motherboard, PCBA, Chassis, Integration</td>
<td>Ingrasys</td>
<td>Ingrasys</td>
</tr>
<tr>
<td>Rack &amp; Power Shelf</td>
<td>Ingrasys</td>
<td>Delta</td>
</tr>
<tr>
<td>Voltage Regulators</td>
<td>VICOR</td>
<td>Delta</td>
</tr>
</tbody>
</table>
JOIN US
Want to Participate?

Inquire About Development Samples

OpenCAPI & NVLink Solution Developers Also Welcome to Join
Want to Know More?

OCP Community Preview
http://files.opencompute.org/oc/public.php?service=files&t=4302505c9d59160ec03aaf4ea67b146c&download

OCP Server Mailing List
http://lists.opencompute.org/mailman/listinfo/opencompute-server

OCP Server Mail Archives
http://lists.opencompute.org/pipermail/opencompute-server/
Design Package

Available on OCP Server Wiki and GitHub

Enhancements Coming Throughout 2017

GitHub
https://github.com/opencomputeproject/zaius-barreleye-g2 (don’t forget to install Git LFS)

OCP
http://www.opencompute.org/wiki/Server/Working#Open_Rack
THANK YOU
Archive Material
Zaius Motherboard

Zaius 1.5 OU Deployment
Shelf & Sled

IO Cage
- 5x PCIe + x16 OCP Mezz

Storage System
- 1x internal 2.5" or 3.5" SATA + 1x internal M.2 2280 SATA

Barreleye G2 2 OU Chassis

IO Cages
- 5x PCIe + x16 OCP Mezz

Storage Systems
- 24x 2.5" SAS / SATA / NVMe Drive Canister + 1x internal M.2 SATA (2280)
- Diskless (1x internal M.2 SATA 2280) – not shown in deck

2x FHFL GPU + 1 HHHL PCIe + x16 OCP Mezz

2 – 4 OpenCAPI / NVLink + 1 HHHL PCIe + x16 OCP Mezz

Diskless (1x internal M.2 SATA 2280) – not shown in deck

Barreleye G2 1.25 OU Chassis

IO Cage
- 2x FHFL + 1 HHHL PCIe + x16 OCP Mezz

Storage Systems
- 15x 2.5" SAS / SATA / NVMe Drive Canister + 1x internal M.2 SATA (2280)
- Diskless (1x internal M.2 SATA 2280) – not shown in deck
Other NVLink & OpenCAPI Interface Option*

15.8mm (25.3 mated)

23.5mm (25.95 mated)

6.6mm (8.5 mated)

* For illustrative purposes only. Could be used in cards, but is not mounted on motherboard
**Other NVLink & OpenCAPI Interface Option**

<table>
<thead>
<tr>
<th>DESCRIPTION</th>
<th>QTY</th>
<th>TOTAL THROUGHPUT (UNI-DIR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVLink 2.0 / OpenCAPI 3.0 Brick</td>
<td>-</td>
<td>25 Gbytes / sec (200 Gbps)</td>
</tr>
<tr>
<td>Bricks per Socket</td>
<td>2</td>
<td>50 Gbytes / sec (400 Gbps)</td>
</tr>
<tr>
<td>Bricks per Server</td>
<td>4</td>
<td>100 Gbytes / sec (800 Gbps)</td>
</tr>
</tbody>
</table>

* For illustrative purposes only. Could be used in cards, but is not mounted on motherboard

---

<table>
<thead>
<tr>
<th>SPECS</th>
<th>MEASUREMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>RECEPTACLE WIDTH:</td>
<td>23.5mm</td>
</tr>
<tr>
<td>PLUG WIDTH:</td>
<td>25.95mm</td>
</tr>
<tr>
<td>MATED HEIGHT:</td>
<td>8.55mm</td>
</tr>
<tr>
<td>MATED LENGTH:</td>
<td>25.3mm</td>
</tr>
<tr>
<td>30AWG DIAMETER:</td>
<td>0.255mm</td>
</tr>
<tr>
<td>PCB THICKNESS:</td>
<td>1.0mm</td>
</tr>
</tbody>
</table>

---

* For illustrative purposes only. Could be used in cards, but is not mounted on motherboard.
BG2 Storage Limitations: U.2 vs U.3