Power Systems AC922 Overview

Chris Mann
IBM Distinguished Engineer
Chief System Architect, Power HPC Systems
December 11, 2017
IBM POWER HPC Platform Strategy

- High-performance computer and high-performance analytics drive common platform design
- Servers will be predominately 2-socket designs
- Developing deeper relationships with technology partners – ref OpenPOWER
- Majority of floating-point performance will come from GPUs
- Utilize Industry-standard compliant 19” racks and electronics enclosures
  - Air and water cooling options
- Platforms will be based on a common enclosure form factor
  - Enclosure provides working envelope that we will continue to enhance with the latest technology from IBM, NVIDIA, Mellanox and other OpenPOWER partners
  - Enclosure provides a platform with sufficient power, cooling capability to support these enhancements
IBM POWER GPU Intensive Roadmap

POWER S822LC
- 2 POWER8 Processors
- 190 Turismo module
- 2 x16 Gen 3 FHFL PCIe slots
- Supports 2 NVidia K80 GPU’s
- Supports 2 PCIe adapters
- 1 x8 Gen 3 HHHL PCIe, CAPI
- 1 x16 Gen 3 HHHL PCIe, CAPI
- 32 DDR3 IS DIMM’s
- 4, 8, 16, 32GB DIMMs
- 32 – 1024GB Memory Capacity
- 2 SATA SFF HDD / SSD
- 2 1300W Power Supplies
- 200VAC Input
- BMC support structure
- IPMI, USB, EN, VGA
- Air cooled

POWER S822LC for HPC
- 2 POWER8 w/ NVLink Processors
- 190 module
- 1, 2, 4 NVidia “Pascal” GPU’s
- 300W, SXM2 Form Factor, NVLink 1.0
- 2 x16 Gen 3 HHHL PCIe, CAPI enabled
- 1 x8 Gen3 HHHL PCIe, CAPI enabled
- 32 DDR4 IS DIMM’s
- 4, 8, 16, 32GB DIMMs
- 2 SATA SFF HDD / SSD
- Pluggable NVMe storage adapter
- 1.6, 3.2TB Capacity
- 2 1300W power supplies
- 200VAC Input
- BMC Support Structure
- IPMI, USB, EN, VGA
- Air and water cooled options

POWER AC922
- 2 POWER9 Processors
- 190, 250W modules
- 4-6 NVidia “Volta” GPU’s
- 300W, SXM2 Form Factor, NVLink 2.0
- 6 GPU configuration, water cooled
- 4 GPU configuration, air or water cooled
- 2 Gen4 x16 HHHL PCIe, CAPI enabled
- 1 Gen4 x4 HHHL PCIe
- 1 Gen4 Shared x8 PCIe adapter
- 16 IS DIMM’s
- 8, 16, 32, 64, 128GB DIMMs
- 2 SATA SFF HDD / SSD
- 2 2200W power supplies
- 200 VAC, 277VAC, 400VDC input
- N+1 Redundant
- Second generation BMC Support Structure
- Pluggable NVMe storage adapter option

2015 2016 2017 - 2018
Introducing Power AC922 2U HPC Server – Premier Accelerator Platform
Leveraging POWER9 Innovation
High-performance core, bandwidth, accelerator differentiation

For compute-intensive workloads, accelerators are key to Cognitive/AI economics
**Power AC922 - POWER9 with increased GPU and IO bandwidth for differentiation**

Realize unprecedented performance and application gains with POWER9 and NVLink 2.0
- 2 POWER9 CPUs and up to 4 “Volta” NVLink 2.0 GPUs in a versatile 2U Linux server
- PCIe Gen4 bus has double I/O Bandwidth vs. PCIe Gen3
- CPU (Turbo)/GPU (Boost) enabled for improved data center efficiency and performance to be maintained at high levels

**High level System Overview**
- 2-Socket, 2U Packaging
- 40 P9 Processor cores
- 4 NVIDIA Volta 2.0 GPUs
- 1 TB Memory (16x - 64GB DIMMs)
- 4 PCIe Gen4 Slots
- 2x SFF (HDD/SSD), SATA, Up to 7.7 TB storage
- Supports 1.6TB and 3.2TB NVMe Adapters
- Default 3 year 9x5 warranty, 100% CRU
POWER9 HPC Product Description
POWER9 Processor – Common Features

- 14HP Technology
- 24 POWER9 Cores with new SMC uArchitecture
  - 4 threads per core
  - 2 cores per chiplet
  - LINUX Radix Page Table support
- Large, Low-latency Cache
  - 512k private L2, 10MB NUCA L3 per chiplet
- Direct Attach Memory Support
  - 8 DDR4 channels
  - 2666 MHz DIMM support
  - 140 GB/s streaming bandwidth
- SMP 2 Socket Support via 4B, 16 Gb/s X-Bus
- Leadership Hardware Acceleration Platform
  - Enhanced on-chip acceleration
  - NVIDIA NVLink 2.0
  - CAPI 2.0, Coherent accelerator and storage attach via PCIe G4
  - OpenCAPI 3.0, Improved latency and bandwidth, open interface
- Network Interconnect
  - CAPI 2.0 attached Mellanox CX-5 support
  - Address translation support
  - Atomics
  - NVIDIA GPU direct
- Gen4 PCIe, 48 Lanes @ 16Gb/s
  - 6 PHBs
  - PCIe Bus 0 turbo-charged to support next generation HDR IB
- Self Boot Capability
- Instant ON/OFF
- Cloud Management QoS
POWER AC922 Design – 4 GPU

Power Supplies (2x)
- 2200W
- 200VAC, 277VAC, 400VDC input

NVidia Volta GPU
- 3 per socket
- SXM2 form factor
- 300W
- NVLink 2.0
- Air/Water Cooled

Memory DIMM’s (16x)
- 8 DDR4 IS DIMMs per socket
- 8, 16, 32, 64, 128GB DIMMs

Power Processor (2x)
- 18, 22C water cooled
- 16, 20C air cooled

PCIe slot (4x)
- Gen4 PCIe
- 2, x16 HHHL Adapter
- 1, Shared slot
- 1 x8 HHHL Adapter

BMC Card
- IPMI
- 1 Gb Ethernet
- VGA
- 1 USB 3.0
## NVIDIA Volta Specifications

<table>
<thead>
<tr>
<th>NVIDIA Volta GPU Features</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak double precision floating point performance</td>
<td>7.8 TFLOPS</td>
</tr>
<tr>
<td>Memory bandwidth</td>
<td>900 GB/sec</td>
</tr>
<tr>
<td>GPU Memory Size</td>
<td>16 GB</td>
</tr>
<tr>
<td>NVLink “Bricks” (8 lane interface)</td>
<td>6</td>
</tr>
<tr>
<td>NVLink Interconnect Bi-Directional</td>
<td>300GB/s</td>
</tr>
<tr>
<td>Maximum Power</td>
<td>300W</td>
</tr>
</tbody>
</table>

NVLink Evolution in POWER HPC

NVIDIA P100 GPU with NVLink 1.0

System Memory  POWER8

NVIDIA Volta GPU with NVLink 2.0

System Memory  POWER9

2016  2017-2018
GPU Attach Bandwidth Comparison, PCIe Gen3 verses NVLink

All values are **MEASURED** Single Direction Bandwidths
I/O Attachment Evolution in POWER HPC

First industry implementation of Gen4 PCIe
Multi-host attachment of POWER9 and the Mellanox EDR-IB adapter
InfiniBand EDR 100Gb/s – PCIe Gen 4 verses PCIe Gen 3

Comparing PCI Express Gen3 to Gen4 EDR Infiniband IB Dual Port Bidirectional Bandwidth Test

Gen4 Dual Port Bidirectional, 392.46 Gb/s

Gen3 Dual Port Bidirectional, 201.32 Gb/s

MESSAGE SIZE BYTES

AVERAGE GBITS/S

0 50 100 150 200 250 300 350 400 450
0 64 128 256 512 1024 2048 4096 8192

InfiniBand EDR 100Gb/s – PCIe Gen 4 verses PCIe Gen 3
Front and Rear Views

**Front**

80mm CR Cooling Fans (4x)

*Note: Front bezel is removed in this illustration*

**Rear**

Power Supplies (2X)

PCIe Slot 1
- Gen4 x4 (x8 Connector)
- HHHL Slot

PCIe Slot 2
- Gen4 Shared x8,x8
- HHHL Slot
- CAPI Enabled

PCIe Slot 3 & 4
- Gen4 x16
- HHHL Slot
- CAPI Enabled

IPMI

VGA

USB 3.0

1Gb Eth (2x)

Water lines (Option)

Service Indicators

Power Button

USB 3.0

SFF-4 Carrier (2X)
- SFF SATA HDD or SSD
Rear View – AC Input

Unique AC input connection

- Requires unique AC cord to PDU
- Enables high voltage DC input to power supply
- Enables 480V Input voltage to the rack (future direction)
  - 277VAC distributed to power supplies
POWER AC922 Memory

- 16 direct attach industry standard DDR4 DIMMs are supported in Witherspoon
  - 8 DIMMs per P9
  - Maximum memory capacity increased to 2TB
  - Maximum memory bandwidth of 170GB/s per socket, 340GB/s for the enclosure
    - Measured Stream is 140GB/s per socket
- All memory DIMMs are installed in order to maximize memory bandwidth

<table>
<thead>
<tr>
<th>DIMM Size</th>
<th>Maximum System Memory</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>16GB</td>
<td>256GB</td>
<td>DOE CORAL contract configuration LLNL</td>
</tr>
<tr>
<td>32GB</td>
<td>512GB</td>
<td>DOE CORAL contract configuration ORNL</td>
</tr>
<tr>
<td>64GB</td>
<td>1024GB</td>
<td></td>
</tr>
</tbody>
</table>
CORAL
IBM, Mellanox, and NVIDIA awarded $325M U.S. Department of Energy’s CORAL Supercomputers

Coral: Leadership Class Supercomputers

5x – 10x Higher App Perf than Current Systems

June 2018 System Acceptance
POWER AC922 Design – 6 GPU

Power Supplies (2x)
- 2200W
- 200VAC, 277VAC, 400VDC input

NVidia Volta GPU
- 3 per socket
- SXM2 form factor
- 300W
- NVLink 2.0
- Air/Water Cooled

Memory DIMM's (16x)
- 8 DDR4 IS DIMMs per socket
- 8, 16, 32, 64, 128GB DIMMs

Power 9 Processor (2x)
- 18, 22C water cooled
- 16, 20C air cooled

PCIe slot (4x)
- Gen4 PCIe
- 2, x16 HHHL Adapter
- 1, Shared slot
- 1 x8 HHHL Adapter

BMC Card
- IPMI
- 1 Gb Ethernet
- VGA
- 1 USB 3.0

NVidia Volta GPU

Memory DIMM's (16x)
CORAL Installation at LLNL
CORAL Installation at ORNL
Thank you!

ibm.com/systems/hpc