Development of server architecture and ecosystem for additional benefits.

—

Konstantin Gerberg
Power Technical specialist
Processor Technology Roadmap

Continued Investment in POWER

- **POWER8**
  - 22 nm
  - 12 Cores
  - SMT8
  - 2X DPFP
  - PCIE Gen 3
  - Coprocessor (CAPI)
  - Enhanced Prefetch

- **POWER8 with NVLink**
  - 22 nm
  - NVLink 1.0
  - 2X CAPI

- **POWER9**
  - 14 nm
  - 24 Cores
  - New µArchitecture
  - Direct-attach DDR4
  - Gen4 PCIe
  - CAPI 2.0
  - OpenCAPI 3.0
  - NVLink 2.0

- **POWER10**
  - 2020+
  - 48 Cores
  - New µArchitecture
  - Enhanced Memory
  - OpenCAPI 4.0
  - Future NVLink
  - Future OpenCAPI

- **POWER11**
  - Future
  - >48 Cores
  - New µArchitecture
  - 2x SIMD width
  - Future NVLINK
  - Future OpenCAPI

---

**Notes:**

openpowerfoundation.org

Inspur Power Commercial Systems Co., Ltd. – FP5180G2

Broadcom Limited – P225P – 2 x 25/10GbE PCIe NIC

Raptor Computing Systems, LLC – T2P9D01

Inspur Power Commercial Systems Co., Ltd. – Inspur Power System FP5466G2
POWER9 – Data Capacity & Throughput

Big Caches for Massively Parallel Compute and Heterogeneous Interaction

**L3 Cache: 120 MB Shared Capacity NUCA Cache**
- 10 MB Capacity + 512k L2 per SMT8 Core
- Enhanced Replacement with Reuse & Data-Type Awareness
  12 x 20 way associativity

Extreme Switching Bandwidth for the Most Demanding Compute and Accelerated Workloads

**High-Throughput On-Chip Fabric**
- Over 7 TB/s On-chip Switch
- Move Data in/out at 256 GB/s per SMT8 Core

POWER9

7 TB/s

- DDR
- PCIe
- CAPI
- NVLink 2
- OpenCAPI
- SMP

Memory
PCle Device
IBM & Partner Devices
Nvidia GPU
IBM & Partner Devices
POWER9

256 GB/s x 12

17 Layers of Metal
POWER Accelerator Attachment Roadmap

**Pascal**
Tesla P100

**Volta**

**Volta follow-on**

<table>
<thead>
<tr>
<th>Year</th>
<th>Architecture</th>
<th>Link</th>
<th>Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>POWER8 with NVLink</td>
<td>SXM2</td>
<td>NVLink 1.0 40+40 GB/s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>POWER9</td>
<td>SXM2</td>
<td>NVLink 2.0 75+75 GB/s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2020+</td>
<td>POWER10</td>
<td></td>
<td>NVLink 3.0 More Bandwidth</td>
</tr>
</tbody>
</table>

Note 1: Bandwidths are peak, per GPU, assuming two GPUs per socket.
Note 2: Improved performance comes from both faster signaling rate and wider interfaces.
Open Innovation Interface: OpenCAPI

Open Industry Coherent Attach
- Latency / Bandwidth Improvement
- Removes Overhead from Attach Silicon
- Eliminates “Von-Neumann Bottleneck”
- FPGA / Parallel Compute Optimized
- Network/Memory/Storage Innovation
# IBM CAPI Roadmap

<table>
<thead>
<tr>
<th>Feature</th>
<th>CAPI</th>
<th>CAPI 2.0</th>
<th>OpenCAPI 3.0</th>
<th>OpenCAPI 4.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU Generation</td>
<td>POWER8</td>
<td>POWER9</td>
<td>POWER9</td>
<td>Post-POWER9</td>
</tr>
<tr>
<td>PSL Location</td>
<td>FPGA/ASIC</td>
<td>FPGA/ASIC</td>
<td>Processor</td>
<td>Processor</td>
</tr>
<tr>
<td>Native DMA</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Native Atomic Ops</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Interface</td>
<td>PCIe (8 Gbps)</td>
<td>PCIe (16 Gbps)</td>
<td>25 Gbps</td>
<td>32 Gbps</td>
</tr>
<tr>
<td># Lanes per port</td>
<td>x8</td>
<td>x16</td>
<td>x 8*</td>
<td>x4, x8, x16, x32</td>
</tr>
<tr>
<td>BW per port</td>
<td>8+8 GB/s</td>
<td>32+32 GB/s</td>
<td>25+25 GB/s</td>
<td>Up to 128+128 GB/s</td>
</tr>
<tr>
<td>POWER Systems</td>
<td>All P8 models</td>
<td>All P9 models</td>
<td>Most P9 models</td>
<td>TBD</td>
</tr>
</tbody>
</table>
8 Direct DDR4 Ports
- Up to 120 GB/s of sustained bandwidth
- Low latency access
- Commodity packaging form factor
- Adaptive 64B / 128B reads

8 Buffered Channels
- Up to 230 GB/s of sustained bandwidth
- Extreme capacity – up to 8TB / socket
- Superior RAS with chip kill and lane sparing
- Compatible with POWER8 system memory
- Agnostic interface for alternate memory innovations
IBM Centaur DIMM

OMI DDIMM

Ultra-scale

JEDEC DDR DIMM

- Technology agnostic
- Low cost
- Ultra-scale system density
- Enterprise reliability
- Low-latency
- High bandwidth

Approximate Scale
Final Addition to the POWER9 Processor Family

**Processor Chip Details**
- 728 mm² (25.3 x 28.8 mm)
- 8 Billion Transistors
- Up to 24 SMT4 Cores
- Up to 120 MB eDRAM L3 cache

**Semiconductor Technology**
- 14nm finFET
- Improved device performance
- Reduced energy
- eDRAM
- 17 layer metal stack

**High Bandwidth Signaling**
- 25 GT/s low energy differential
  - PowerAXON, OMI memory
- 16 GT/s low energy differential
  - Local SMP
- 16 GT/s PCIe Gen4

**The Bandwidth Beast**
**Advanced I/O (AIO)**

**Open Memory Interface (OMI)**
- 16 channels x8 at 25 GT/s
- 650 GB/s peak 1:1 r/w bandwidth
- Technology Agnostic
- Offered with Microchip DDR4 buffer (410 GB/s peak bandwidth)

**PowerAXON 25 GT/s Attach**
- Up to 16 socket glue-less SMP (4x24 SMP added to 3x30 local)
- Up to x48 NVIDIA NVLINK GPU attach
- Up to x48 OpenCAPI 4.0 coherent accelerator / memory attach

**Industry Standard I/O Attach**
- x48 PCIe Gen 4 at 16 GT/s
- Up to x16 CAPI 2.0 coherent accelerator / storage attach

**2 TB/s Raw Signaling Bandwidth**
**Shared by 6 Attach Protocols**
Servers line

E950
S922
AC922
LC921
L922
LC922
S924

E980
IBM Solutions

SAP HANA

H924/H924

ESS Elastic storage

HDP
(Hortonworks Data Platform)

AI Infrastructure Reference Architecture
Облачная инфраструктура

Традиционная виртуализация

Микросервисная архитектура
Интегрированная виртуализация

PowerVM Integrated Hypervisor

AIX
Linux
IBM i

Processor
Memory
I/O
Bus/Lane
PowerVM™ Hypervisor

Virtual I/O Server
AIX 6.1
IBM i 6.1

Red Hat Linux
SuSE Linux
Red Hat Linux
SuSE Linux

48 GB
20 GB
48 GB
12 GB

2 cores
8 cores
4 cores

2GB
64 GB
32 GB

16 Core (4 IFLs)
Shared Processor Pool

PowerVM™ Hypervisor

Storage
Ethernet

HMC

PowerVM 16 Core (4 IFLs)
Power Enterprise Pool

Sys A
64-core E880
4.35 GHz
Activations:
10 static
0 mobile
54 “dark”

Sys B
96-core 980
3.7 GHz
Activations:
30 static
55 mobile
11 “dark”

Sys C
96-core 880
3.7 GHz
Activations:
16 static
45 mobile
35 “dark”

Sys D
128-core 980
4.0 GHz
Activations:
40 static
60 mobile
28 “dark”

Pool Totals
Activations:
96 static
160 mobile
128 “dark”

Going back to initial starting point and moving the activations differently

Move 15 activations
Move 25 activations
Программная виртуализация

Server Hardware

Linux

KVM

Guest OS

Guest OS

Guest OS

App

App

App

bin
/\libs

bin
/\libs

bin
/\libs
Контейнерная виртуализация
OS on Power Systems

Full Enterprise Support

Community Supported distributions
**AIX 7.2 Standard Edition:** The AIX 7.2 Standard Edition is the edition that many people would think of as “AIX.” AIX 7.2 Standard Edition is eligible to run on any POWER7, POWER7+/TM and POWER8/9 server.

**AIX 7.2 Enterprise Edition:** The AIX 7.2 Enterprise Edition includes all the UNIX capabilities of AIX Standard Edition, but also includes significant enhancements that come with the inclusion of IBM Cloud PowerVC Manager, PowerSC, IBM Tivoli® Monitoring and IBM BigFix Lifecycle. AIX Enterprise Edition includes all of these products under a single ordering and support structure.

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Product</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rapidly deploy and easily manage private clouds</td>
<td>PowerVC Cloud</td>
<td>Self-service infrastructure provisioning at scale</td>
</tr>
<tr>
<td></td>
<td>Spectrum Scale</td>
<td>Provides software defined capabilities to enable SAN less power clouds</td>
</tr>
<tr>
<td></td>
<td>Cloud Management Console (SaaS)</td>
<td>Cross-datacenter monitoring of all Power Systems infrastructure</td>
</tr>
<tr>
<td></td>
<td>Cloud App Management</td>
<td>Application-Aware Infrastructure Monitoring</td>
</tr>
<tr>
<td>Simplify security and compliance management</td>
<td>PowerSC Std. Ed.</td>
<td>Security and Compliance management of Power environments</td>
</tr>
<tr>
<td></td>
<td>PowerSC MFA</td>
<td>Raise the overall assurance level with multi-factor authentication</td>
</tr>
<tr>
<td></td>
<td>Big Fix Lifecycle</td>
<td>Discover, secure and manage endpoints on 90 different OS versions</td>
</tr>
<tr>
<td>Simplify High Availability Management</td>
<td>VM Recovery Manager HA</td>
<td>Simplified, OS agnostic HA solution for cloud deployments</td>
</tr>
<tr>
<td>Accelerate file transfer in cloud environments</td>
<td>Aspera</td>
<td>High Speed large-file exchange between (multi-/hybrid-) cloud environments</td>
</tr>
</tbody>
</table>
PowerVC

Упрощение управления виртуализацией
Легкость тиражирования виртуальных машин для быстрого развертывания. Автоматизированная конфигурация ресурсов ввода/вывода обеспечивает мобильность и отказоустойчивые конфигурации

PowerVP

Понимание распределения нагрузок на виртуальном и аппаратном уровне
Простое графическое представление о работе в реальном времени

PowerSC
Средство повышение уровня безопасности систем
Cognitive Systems HA/DR solution family

PowerHA for AIX
- Platinum level HA/DR, for dedicated mission critical environments
- Covers planned and unplanned outages for both software and hardware
- Advanced capabilities such as HyperSwap (AIX and IBM i) and three site configurations
- Operating system based technology
- best case RTO
- N+1 licensing

PowerHA for Linux

PowerHA for IBM i

PowerHA System Mirror
- Gold/Silver level HA/DR, ideal for cloud environments
- Primarily for planned and unplanned hardware outages
- Manage and monitor large numbers of VMs (LPARS)
- Relatively easy to implement and manage
- Operating system independent (supports AIX, i and Linux)
- RTO not as fast as PowerHA (VM reboot)
- N+0 licensing

VM Recovery Manager HA

VM Recovery Manager DR

active/passive HA/DR clustering

active/inactive VM restart clustering
## PowerHA for AIX

### PowerHA SystemMirror for AIX **Standard Edition**
- Shared storage HA clustering
- High Availability within a Datacenter

### PowerHA SystemMirror for AIX **Enterprise Edition**
- Enables long distance HA/DR clustering
- Low cost host based mirroring support (GLVM)
- Extensive support for storage array replication
- IBM DS8K, SVC, XIV, A9000, EMC, Hitachi, HP
PowerHA for Linux:

**SAP HANA System Replication**

SAP HANA management:
- SAP HANA system replication Active standby configuration support
- PowerHA automates failover to standby node
- Re-integration of recovered LPAR into cluster as standby
- RHEL and SUSE support

HA deployment Wizard
Graphical management (common with PowerHA for AIX)

**SAP HANA Cold Standby**

Admin can deploy cold restart HA for SAP HANA
- Disks are shared across nodes
- PowerHA automates the start of the SAP HANA during failover
- Recovery time is longer for SAP HANA
VM Recovery Manager

- Server, VM, and workload-level HA
- GUI management experience
- Capacity adjustment policy
- Co-location and anti-colocation policy support
- Monitoring agents for DB2, Oracle, and SAP HANA
Thank you

Konstantin Gerberg