The IBM POWER9 Scale Up Processor

Jeff Stuecheli
William Starke

POWER Systems, IBM Systems
POWER Processor Technology Roadmap

POWER7+ 32 nm
- 2.5x Larger L3 cache
- On-die acceleration
- Zero-power core idle state
- Up to 12 Cores
- SMT8
- CAPI Acceleration
- High Bandwidth GPU Attach

POWER8 Family 22nm

POWER9 Family 14nm
Built for the Cognitive Era
- Enhanced Core and Chip Architecture Optimized for Emerging Workloads
- Processor Family with Scale-Up and Scale-Out Optimized Silicon
- Premier Platform for Accelerated Computing

POWER7 45 nm
Enterprise
- 8 Cores
- SMT4
- eDRAM L3 Cache

1H10
2H12
1H14 - 2H16
2H17 - 2H18+
POWER9 Processor – Common Features

New Core Microarchitecture
- Stronger thread performance
- Efficient agile pipeline
- POWER ISA v3.0

Enhanced Cache Hierarchy
- 120MB NUCA L3 architecture
- 12 x 20-way associative regions
- Advanced replacement policies
- Fed by 7 TB/s on-chip bandwidth

Cloud + Virtualization Innovation
- Quality of service assists
- New interrupt architecture
- Workload optimized frequency
- Hardware enforced trusted execution

Leadership Hardware Acceleration Platform
- Enhanced on-chip acceleration
- Nvidia NVLink 2.0: High bandwidth and advanced new features (25G)
- CAPI 2.0: Coherent accelerator and storage attach (PCIe G4)
- OpenCAPI: Improved latency and bandwidth, open interface (25G)

State of the Art I/O Subsystem
- PCIe Gen4 – 48 lanes

High Bandwidth Signaling Technology
- 16 Gb/s interface
  - Local SMP
- PowerAXON 25 GT/sec Link interface
  - Accelerator, remote SMP

14nm finFET Semiconductor Process
- Improved device performance and reduced energy
- 17 layer metal stack and eDRAM
- 8.0 billion transistors
POWER9 – Dual Memory Subsystems

**Scale Out**

**Direct Attach Memory**
- 8 Direct DDR4 Ports
  - Up to 150 GB/s of sustained bandwidth
  - Low latency access
  - Commodity packaging form factor
  - Adaptive 64B / 128B reads

**Scale Up**

**Buffered Memory**
- 8 Buffered Channels
  - Up to 230GB/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
PowerAXON → High-speed 25 GT/s Signaling

Utilize Best-of-Breed 25 GT/s Optical-Style Signaling Technology

Flexible & Modular Packaging Infrastructure
Scale Out
Direct Attach Memory
2 Socket SMP

Scale Up
Buffered Memory
16 Socket SMP
Scale Up Chipset Block Diagram

- **PCI G4 I/O Attach (16 GHz)**
- **Processor Cores**
- **Accel**
- **Local SMP Links (16 GHz)**
- **Remote SMP Link Control**
- **Cache and Interconnect**

**Memory Control**
- 8 Channels per Processor
- 4 ops Bypass
- 32B
- 2 KB
- 4 KB
- Active Operations 64 + 12
- Framing / Parsing

**Framing / Parsing**
- Command, 1B write, 2B read @ 9.6 GHz

**Centaur Memory Buffer**
- 16 MB
- 10B DDR4
- DDR4 Sched
- Operation and Data Control
- 16 MB Cache

**PowerAXON 25 GHz Signaling**
- 4x8 6x8 4x24

**Connect to other POWER9, Scale-up chips, GPU's, and OpenCAPI Devices**

© 2018 IBM Corporation
Four Socket Server Topology

16 GHz x32

POWER9 Scale-Up

POWER9 Scale-Up

POWER9 Scale-Up

POWER9 Scale-Up

1 TB Buffered Memory
Sixteen Socket Server Topology
Power E980 Server

- Modular, Scalable POWER9 server
  - 1 to 4 x 5U CEC drawers + 2U Control Unit
- POWER9 Enterprise processor
- Up to 192 cores in a single system
- Up to 64 TB DDR4 memory
- Up to 32 PCIe Gen4 slots
- PowerAXON 25Gb/s ports
  - Used for SMP cabling between nodes - 4x bandwidth improvement
  - Enabled for OpenCAPI accelerators
Demonstration of POWER9 capability, does not imply specific system plans.
Demonstration of POWER9 capability, does not imply specific system plans
Future Evolution of System Architecture

OpenCAPI Northbound

Yesterday’s Plumbing
Tomorrow’s Differentiation

OpenCAPI & PCI

Yesterday’s Plumbing
Tomorrow’s Differentiation

CPU/Accelerator Bandwidth

System bottleneck
OpenCAPI Memory

• Signaling → AXON @25.6GHz vs DDR4 @ 3200 MHz
  – 4 times the bandwidth per IO
• Idle latency over traditional DDR
  – POWER8/9 Centaur design ~10 ns
  – OpenCAPI target of ~5 ns

• Centaur → One proprietary design
• OpenCAPI → Open
## Proposed POWER Processor Technology and I/O Roadmap

### POWER7 Architecture
- **2010**
  - POWER7
  - 8 cores
  - 45nm
  - New Micro-Architecture
  - New Process Technology

### POWER8 Architecture
- **2012**
  - POWER7+
  - 8 cores
  - 32nm

- **2014**
  - POWER8
  - 12 cores
  - 22nm
  - New Micro-Architecture
  - New Process Technology

- **2016**
  - POWER8 w/ NVLink
  - 12 cores
  - 14nm
  - Enhanced Micro-Architecture
  - New Process Technology

### POWER9 Architecture
- **2017**
  - P9 SO
  - 12/24 cores
  - 14nm
  - New Micro-Architecture
  - Direct attach memory
  - New Process Technology

- **2018**
  - P9 SU
  - 12/24 cores
  - 14nm
  - Enhanced Micro-Architecture
  - Buffered Memory
  - New Memory Subsystem

- **2019**
  - P9 w/ Adv. I/O
  - 12/24 cores
  - 14nm
  - Enhanced Micro-Architecture
  - New Technology

- **2020+**
  - P10
  - TBA cores
  - New Micro-Architecture
  - New Technology

### Sustained Memory Bandwidth
- **POWER7**
  - 85 GB/s
  - PCIe Gen2
  - N/A

- **POWER8**
  - 160 GB/s
  - PCIe Gen3
  - CAPI 1.0

- **POWER9**
  - 350+ GB/s
  - PCIe Gen4 x48
  - CAPI 2.0, OpenCAPI2.0

- **2020+**
  - 435+ GB/s
  - PCIe Gen5
  - CAPI 2.0, OpenCAPI4.0

### Standard I/O Interconnect
- **POWER7**
  - PCIe Gen2

- **POWER8**
  - PCIe Gen3

- **POWER9**
  - PCIe Gen4 x48

- **2020+**
  - PCIe Gen5

### Advanced I/O Signaling
- **POWER7**
  - N/A

- **POWER8**
  - 160 GB/s
  - CAPI 1.0, NVLink 1.0

- **POWER9**
  - 300 GB/s
  - CAPI 2.0, OpenCAPI2.0

- **2020+**
  - 300 GB/s
  - CAPI 2.0, OpenCAPI4.0

### Advanced I/O Architecture
- **POWER7**
  - N/A

- **POWER8**
  - CAPI 1.0

- **POWER9**
  - CAPI 2.0, OpenCAPI2.0

- **2020+**
  - CAPI 2.0, OpenCAPI4.0

---

Statement of Direction, Subject to Change
Thanks

Questions?