Power Systems

IBM

IBM POWER9 Introduction Summit Training Workshop

Brian Thompto

POWER Systems, IBM Systems

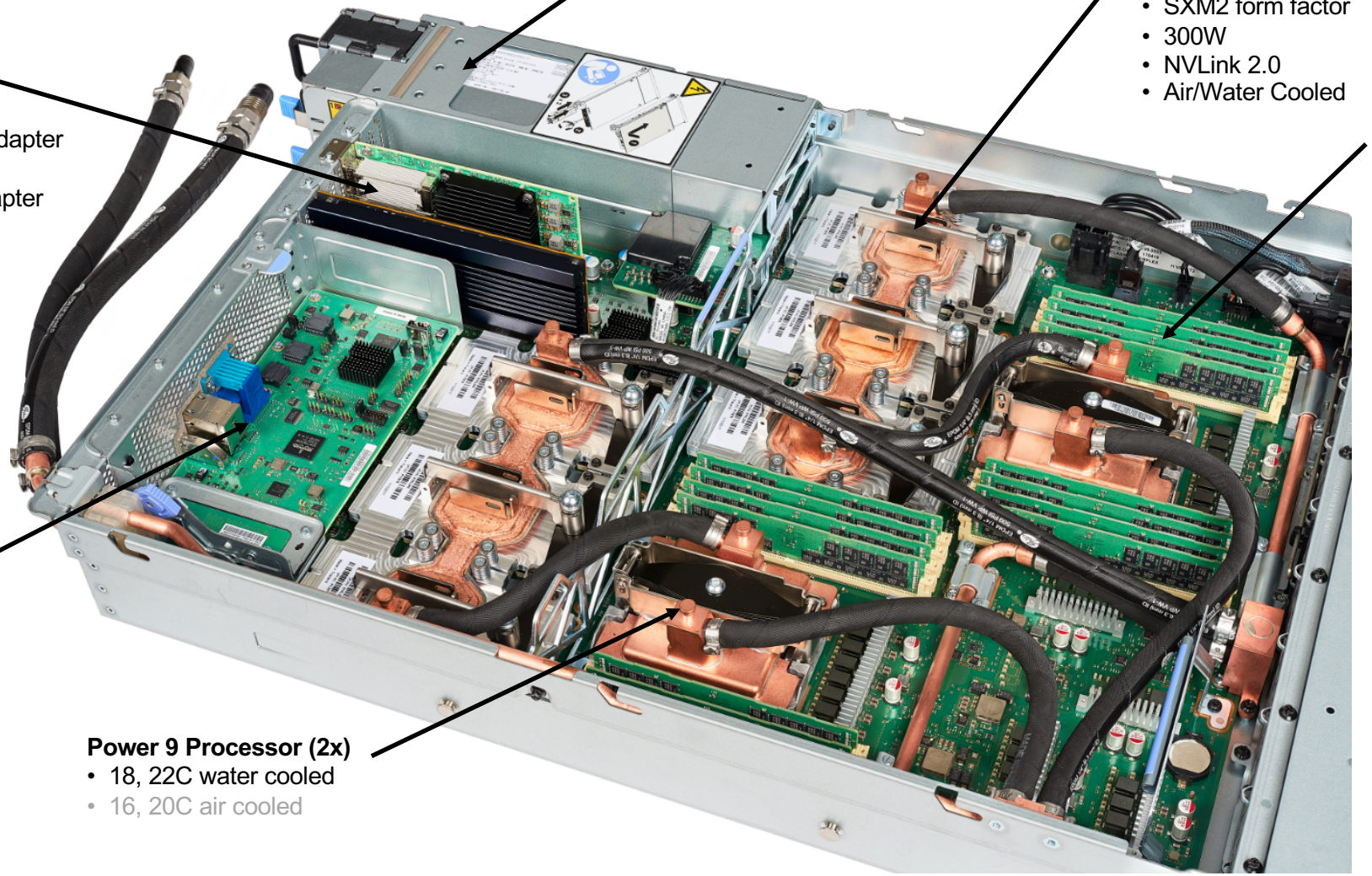
POWER AC922 Design – 6 GPU

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

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

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

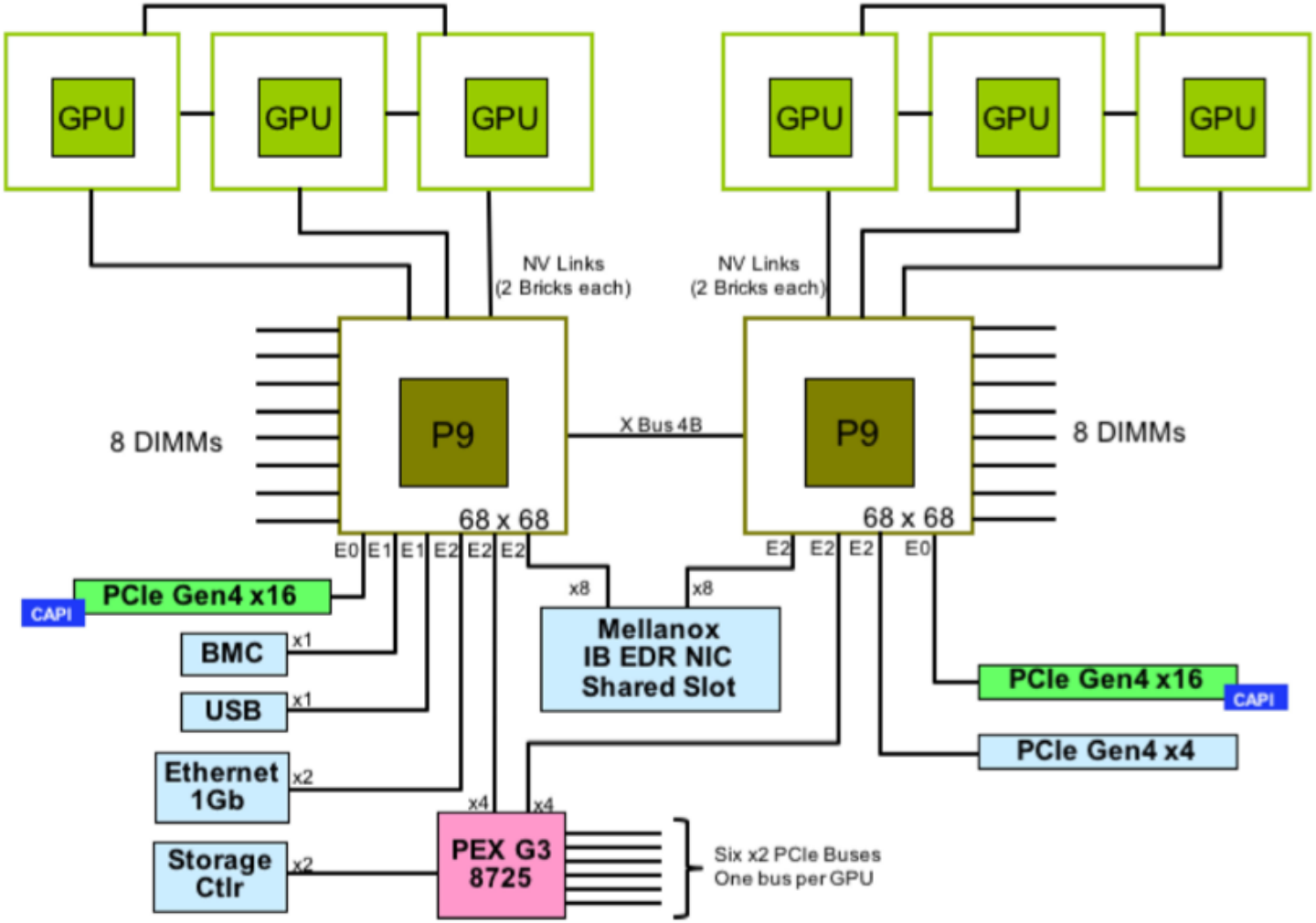
Memory DIMM's (16x)
- 8 DDR4 IS DIMMs per socket
- 8, 16, 32, 64, 128 GB DIMMs



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

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

POWER9 – AC922 with 6 GPU's – Block Diagram



Images / diagrams modified from:
"IBM POWER9 systems designed for commercial cognitive and cloud", *IBM J. Res. & Dev.*, vol. 62, no. 4/5, 2018.

POWER9 Processor – Common Features

New Core Microarchitecture

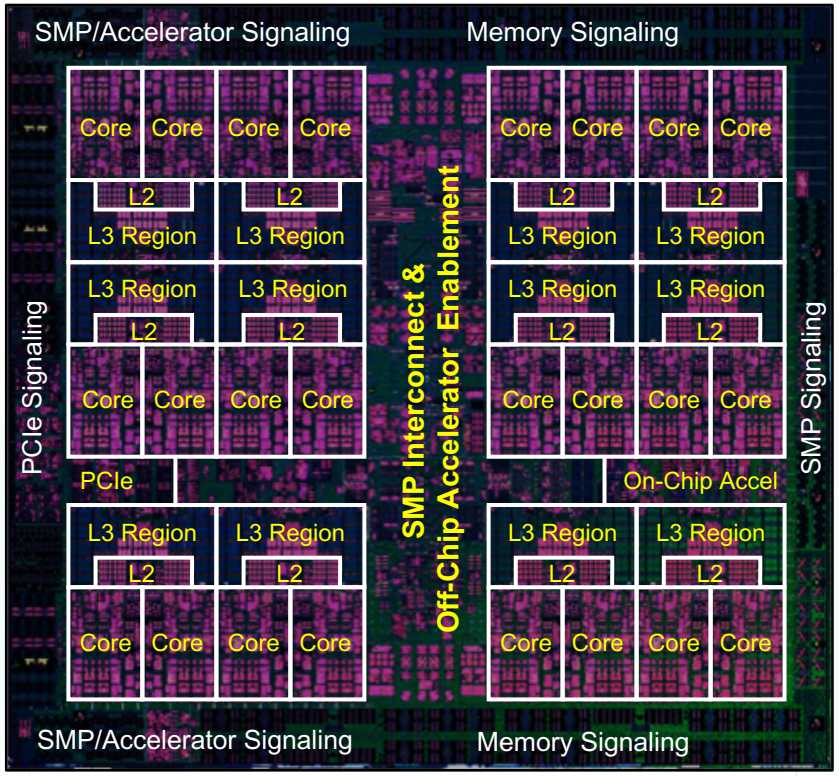
- 24 cores / die; 22 active for Summit
- Stronger thread performance
- Efficient agile pipeline
- POWER ISA v3.0

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



14nm finFET Semiconductor Process

- Improved device performance and reduced energy
- 17 layer metal stack and eDRAM
- 8.0 billion transistors

Leadership Hardware Acceleration Platform

- Enhanced on-chip acceleration
- Nvidia NVLink 2.0: High bandwidth, advanced new features
- CAPI 2.0: Coherent accelerator and storage attach (PCIe G4)
- OpenCAPI 3.0: Improved latency and bandwidth, open interface

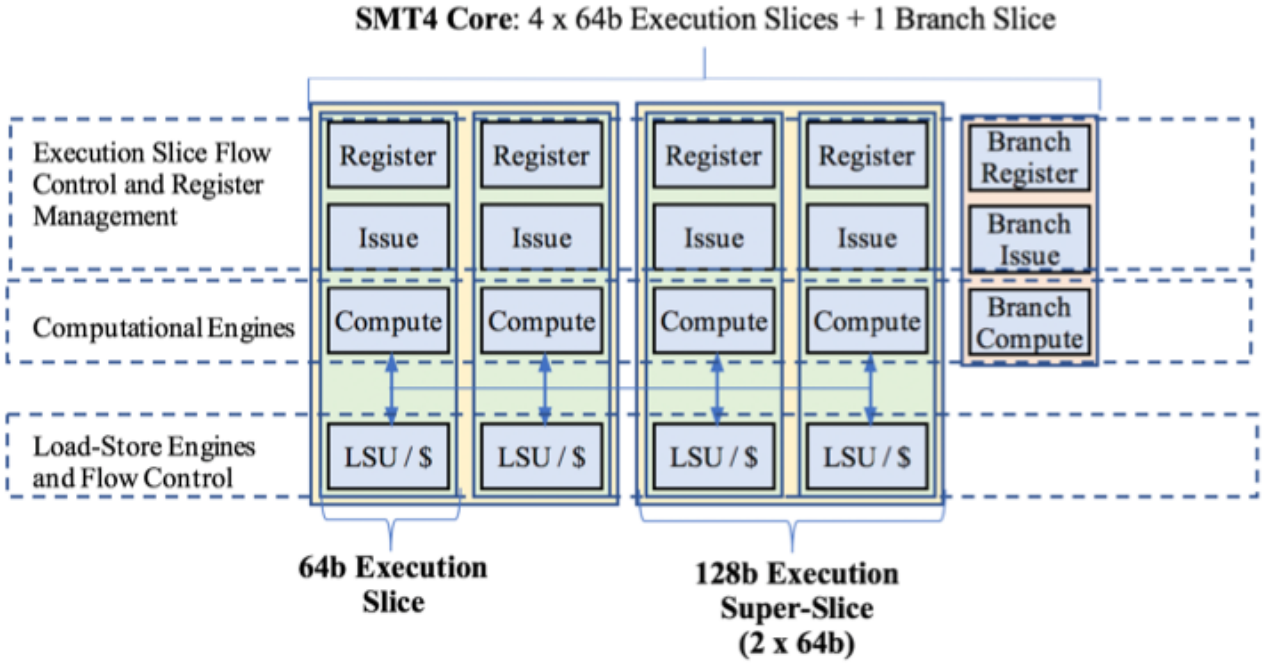
State of the Art I/O Subsystem

- PCIe Gen4 – 48 lanes

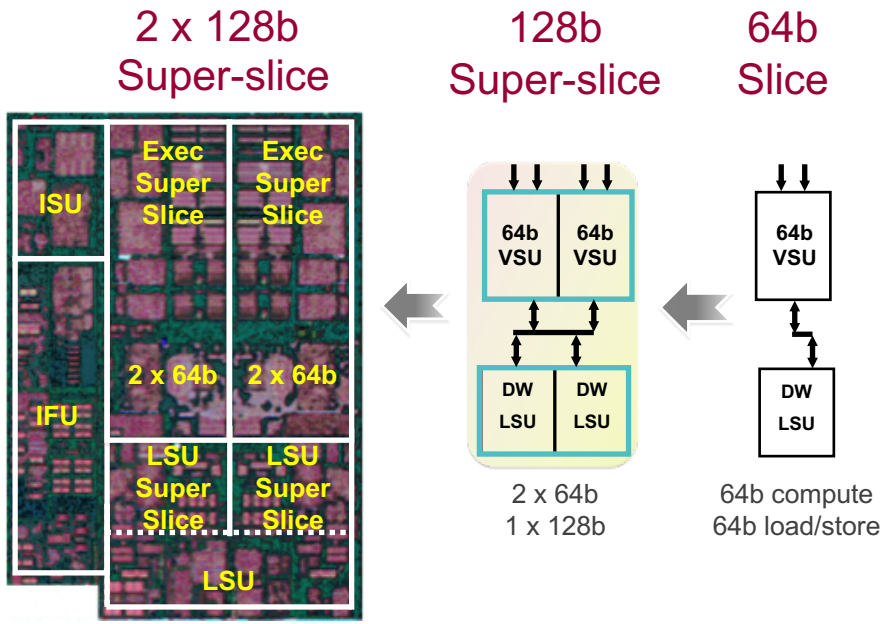
High Bandwidth Signaling Technology

- 16 Gb/s interface
  - – Local SMP
- 25 Gb/s interface – 25G Link
  - – Accelerator, remote SMP

POWER9 SMT4-Core Pipeline



POWER9 SMT4 Core – Sliced Micro-arch



POWER9 SMT4 Core

Images / diagrams modified from:
"POWER9: Processor for the cognitive era", Proc. Hot Chips 28 Symp., pp. 1-19, Aug. 2016.
"IBM POWER9 processor core", IBM Journal of Research and Development, vol. 62, no. 4/5, pp. 2:1-2:12, 2018.

POWER9 – Core Compute

SMT4 Core Resources

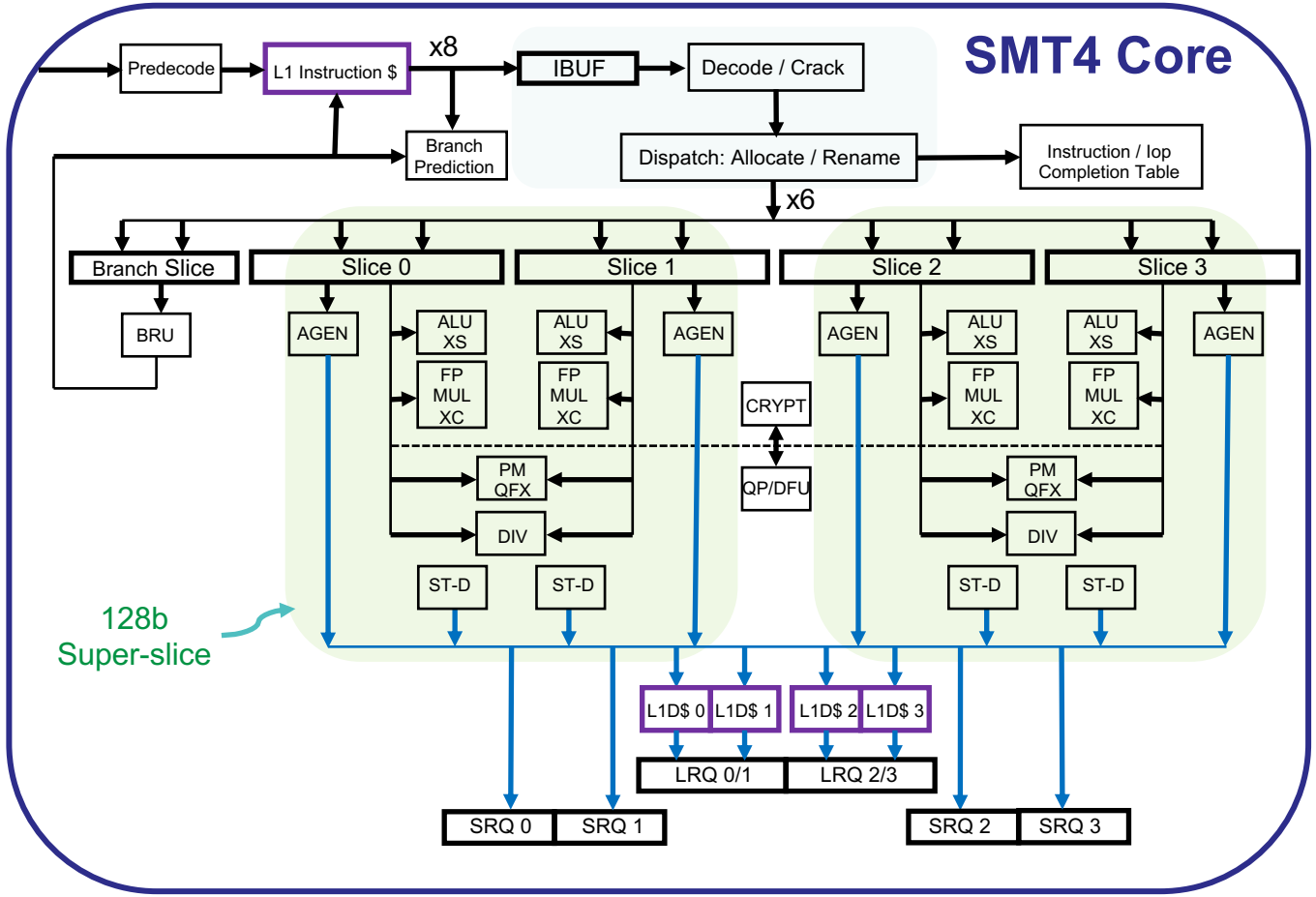
Fetch / Branch
- 32kB, 8-way Instruction Cache
- 8 fetch, 6 decode
- 1x branch execution

Slices issue VSU and AGEN
- 4x scalar-64b / 2x vector-128b
- 4x load/store AGEN

Vector Scalar Unit (VSU) Pipes
- 4x ALU + Simple (64b)
- 4x FP + FX-MUL + Complex (64b)
- 2x Permute (128b)
- 2x Quad Fixed (128b)
- 2x Fixed Divide (64b)
- 1x Quad FP & Decimal FP
- 1x Cryptography

Load Store Unit (LSU) Slices
- 32kB, 8-way Data Cache
- Up to 4 DW load or store

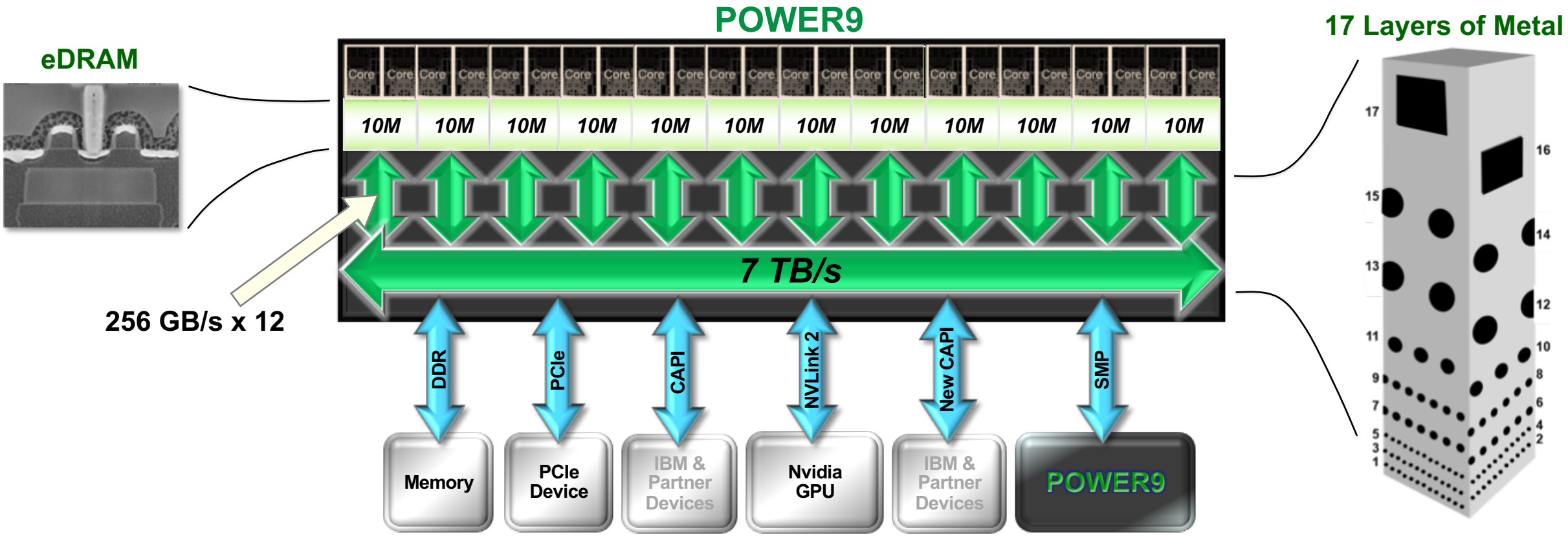
SMT4 Core x 22 per Socket for Summit Systems



POWER9: Cache Capacity

Caches per pair of SMT4 cores (up to 1-8 threads)

- **L2: 512k, 8-way**
- **L3: 10 MB, 20-way**
  - Enhanced L3 Cache Effectiveness with enhanced Replacement
  - Aggregate 110 MB, 11 x 20 way associativity when 22 cores active (out of 24) on Summit



POWER ISA v3.0

New Instruction Set Architecture Implemented on POWER9 vs. POWER8

Broader data type support

- 128-bit IEEE 754 Quad-Precision Float – Full width quad-precision for financial and security applications
- Expanded BCD and 128b Decimal Integer – For database and native analytics
- Half-Precision Float Conversion – Optimized for accelerator bandwidth and data exchange

Support Emerging Algorithms

- Enhanced Arithmetic and SIMD
- Random Number Generation Instruction

Accelerate Emerging Workloads

- Memory Atomics – For high scale data-centric applications

Cloud Optimization

- Enhanced Translation Architecture – Optimized for Linux
- New Interrupt Architecture – Automated partition routing for extreme virtualization
- Enhanced Accelerator Virtualization
- Hardware Enforced Trusted Execution

Energy & Frequency Management

- POWER9 Workload Optimized Frequency – Manage energy between threads and cores with reduced wakeup latency
  - Enables boost of frequency beyond the 3.1 Ghz base; Linux governors can also restrict / lower frequency to save power or boost other cores

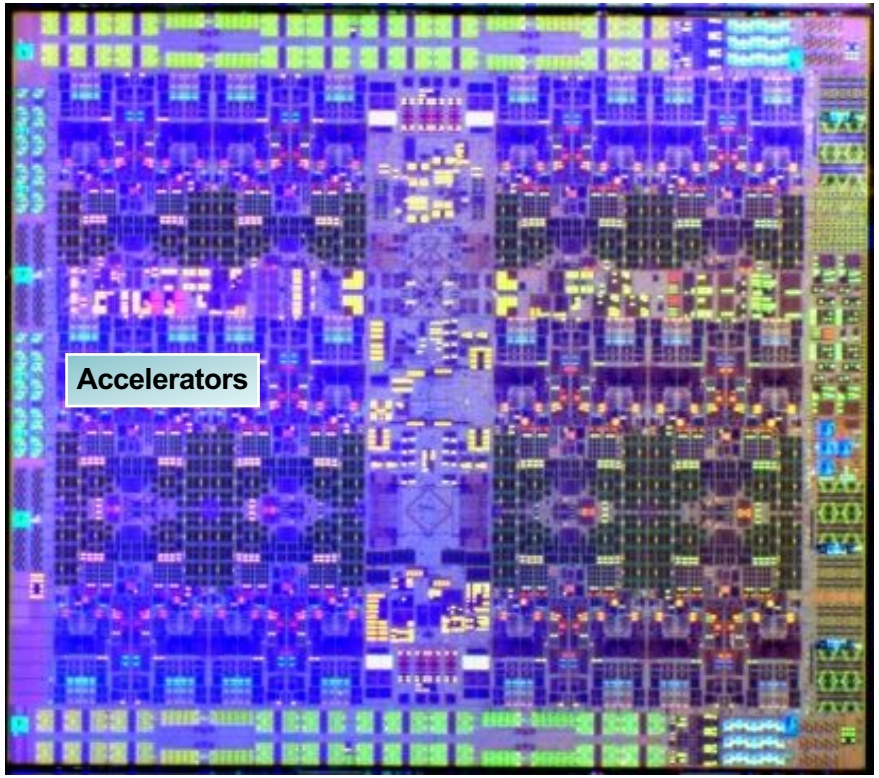
POWER9: On-Processor Accelerators

Virtualized: User mode invocation (No Hypervisor Calls)

Shared accelerators, accessible from each Thread

Accelerator Types

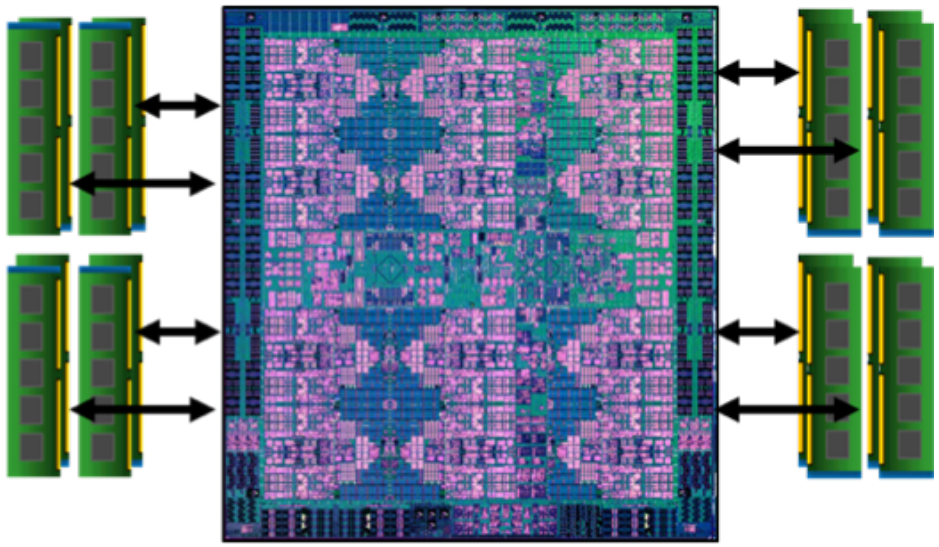
- Industry Standard GZIP Compression / Decompression
  - Up to 16GB/s of gzip / gunzip
- AES / SHA Cryptography Support
  - AES 128b
  - AES 256b
  - SHA 256
  - SHA 512
- Memory compression engine
- True Random Number Generation
- Data Mover



POWER9 – Memory Architecture

Scale Out

Direct Attach Memory



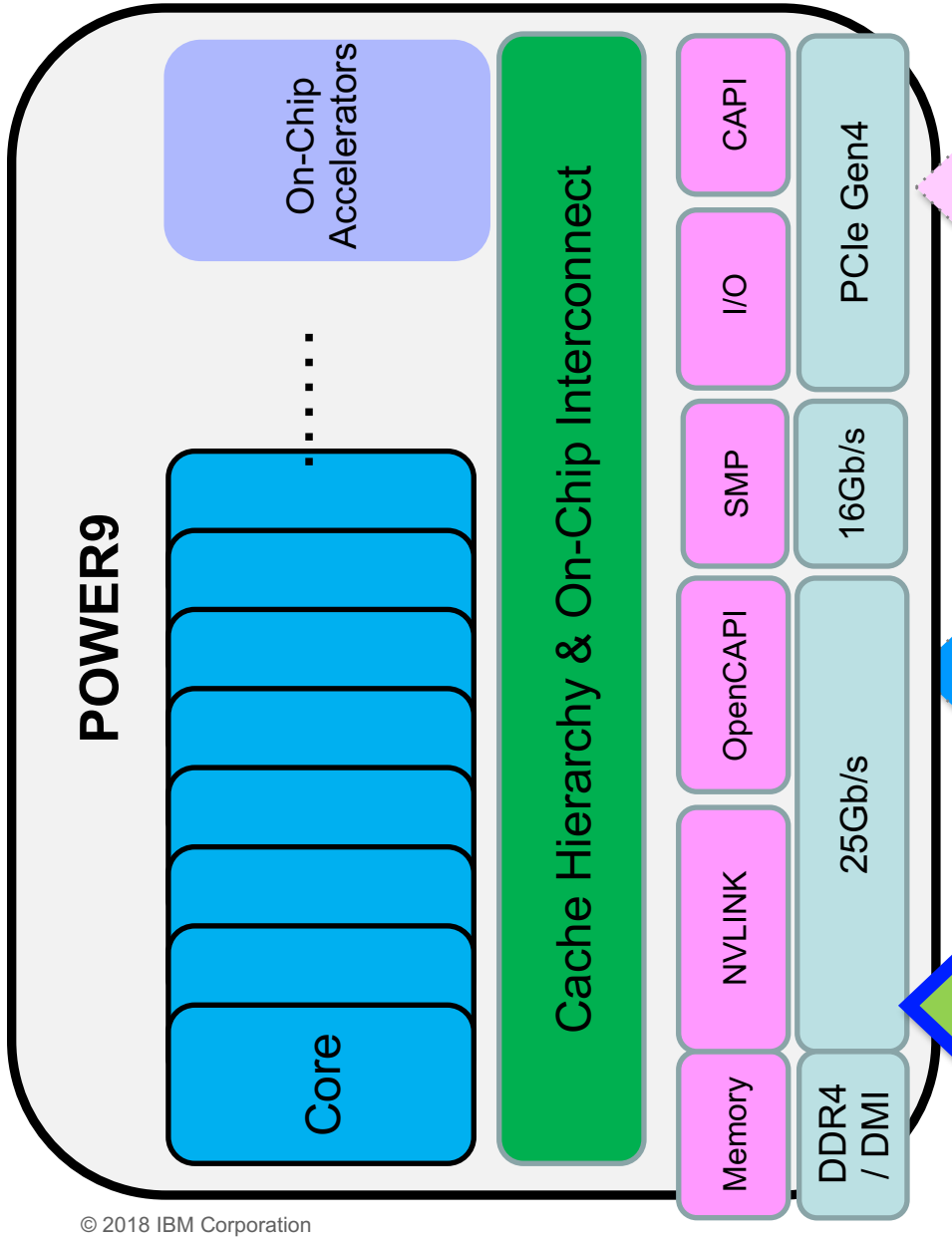
8 Direct DDR4 Ports Per Socket

- 140 GB/s streaming, 170 GB/s of bandwidth peak
- Up to 4TB memory capacity
- Low latency access
- Commodity packaging form factor
- Adaptive 64B / 128B reads

Memory AC922 Summit Systems

- 16 direct attach industry standard DDR4 DIMMs
  - **32 GB DIMM, 2666 MHZ**
- 512 GB Memory Capacity per System

POWER9 Premier Acceleration Platform



CAPI 2.0 (128 GB/s)

Coherent, Open Attach over PCIeG4 Physical Connections

- Attach storage class memory (SCM), network adapters (NIC), FPGA/GPU accelerators, or storage controllers as coherent peer to processor core
- Coherent model enables simpler programming, reduced overhead, and new applications
- Up to 4x attach bandwidth of CAPI 1.0 (POWER8)

OpenCAPI 3.0 (200 GB/s)

Coherent, Open, Host-Agnostic Attach over 25Gb/s Physical Connections

- Higher BW, lower latency physical connection
- Host agnostic protocol enables the same device to attach to multiple CPU architectures

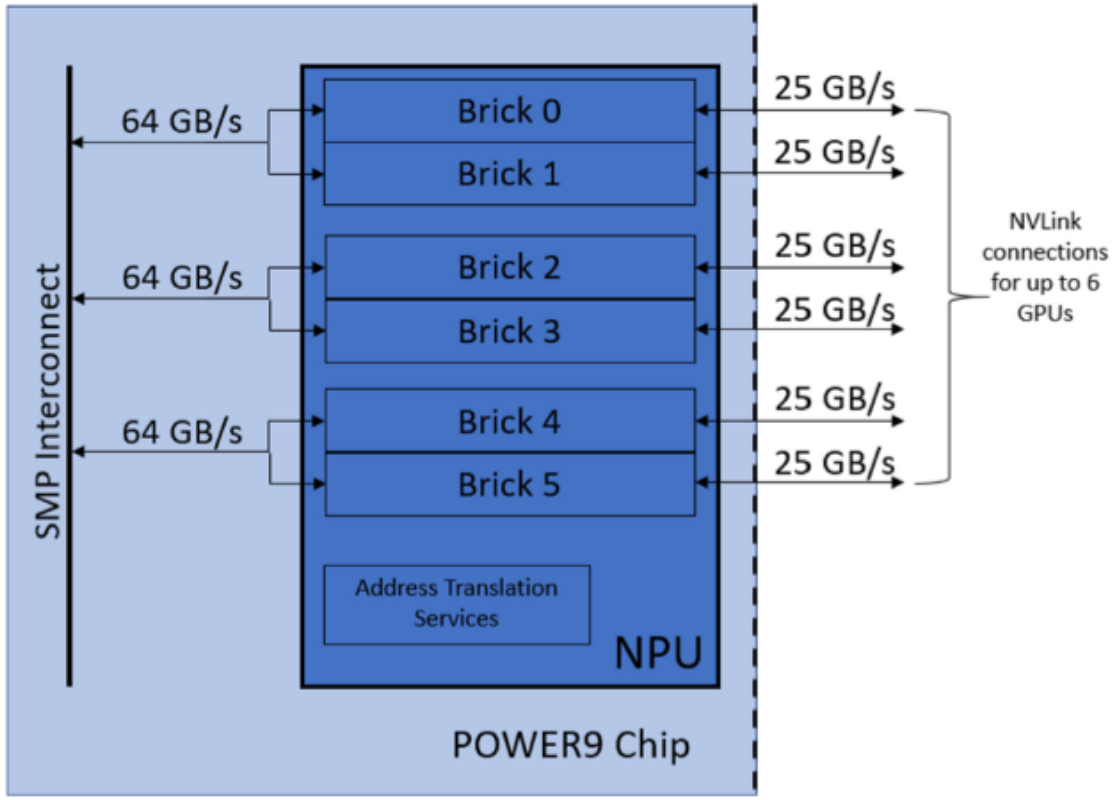
NVLINK 2.0 (300 GB/s)

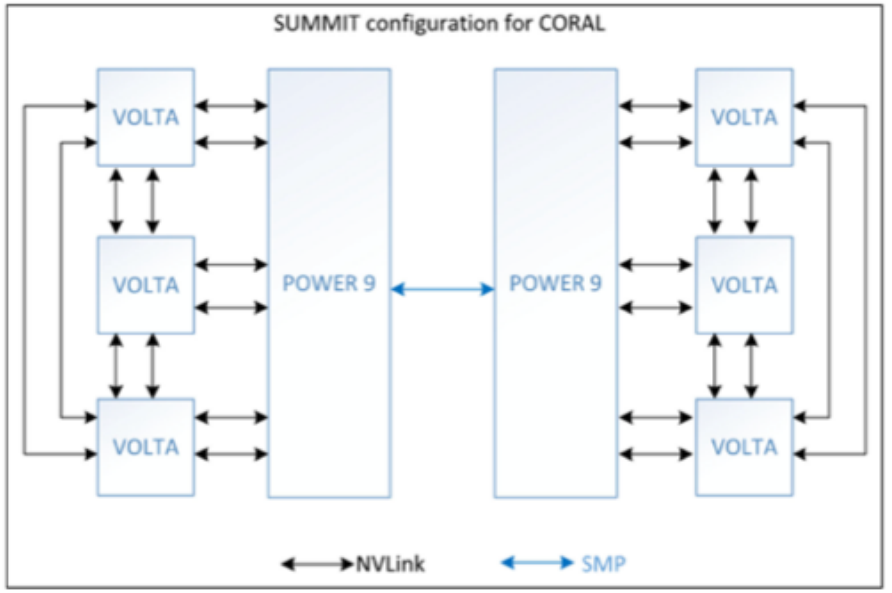
NVIDIA GPU Attach (x3 GPU)

- 10x industry standard PCIe attach BW
- Reduced overhead + simpler programming w/ virtual address support + coherent memory sharing

POWER9 – AC922 NVLINK 2.0

- **Extreme Processor / Accelerator Bandwidth and Reduced Latency**
- **300 GB/s duplex between each POWER9 socket and 3 Volta GPU's**
- **Coherent Memory and Virtual Addressing Capability**





Images / diagrams modified from:
"Functionality and performance of NVLink with IBM POWER9 processors", *IBM J. Res. & Dev.*, vol. 62, no. 4/5, pp. 9:1-9:10, 2018.

Thank you

Special notices

This document was developed for IBM offerings in the United States as of the date of publication. IBM may not make these offerings available in other countries, and the information is subject to change without notice. Consult your local IBM business contact for information on the IBM offerings available in your area.

Information in this document concerning non-IBM products was obtained from the suppliers of these products or other public sources. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. Send license inquiries, in writing, to IBM Director of Licensing, IBM Corporation, New Castle Drive, Armonk, NY 10504-1785 USA.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

The information contained in this document has not been submitted to any formal IBM test and is provided "AS IS" with no warranties or guarantees either expressed or implied.

All examples cited or described in this document are presented as illustrations of the manner in which some IBM products can be used and the results that may be achieved. Actual environmental costs and performance characteristics will vary depending on individual client configurations and conditions.

IBM Global Financing offerings are provided through IBM Credit Corporation in the United States and other IBM subsidiaries and divisions worldwide to qualified commercial and government clients. Rates are based on a client's credit rating, financing terms, offering type, equipment type and options, and may vary by country. Other restrictions may apply. Rates and offerings are subject to change, extension or withdrawal without notice.

IBM is not responsible for printing errors in this document that result in pricing or information inaccuracies.

All prices shown are IBM's United States suggested list prices and are subject to change without notice; reseller prices may vary.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

Any performance data contained in this document was determined in a controlled environment. Actual results may vary significantly and are dependent on many factors including system hardware configuration and software design and configuration. Some measurements quoted in this document may have been made on development-level systems. There is no guarantee these measurements will be the same on generally-available systems. Some measurements quoted in this document may have been estimated through extrapolation. Users of this document should verify the applicable data for their specific environment.

Revised September 26, 2006

Special notices (continued)