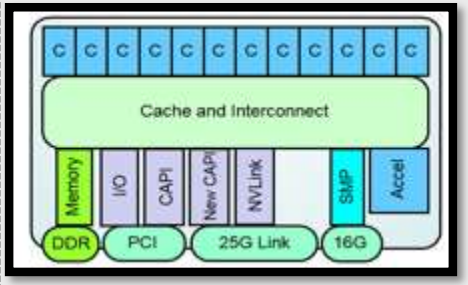
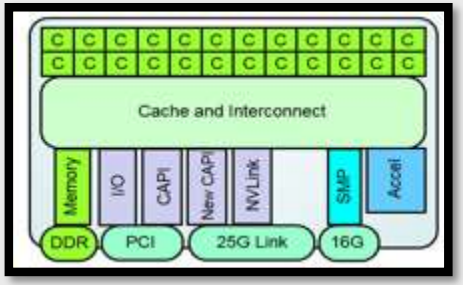
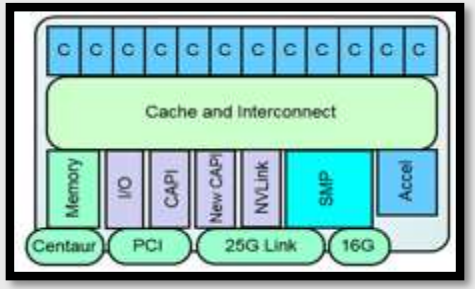
IBM Power Systems AC922

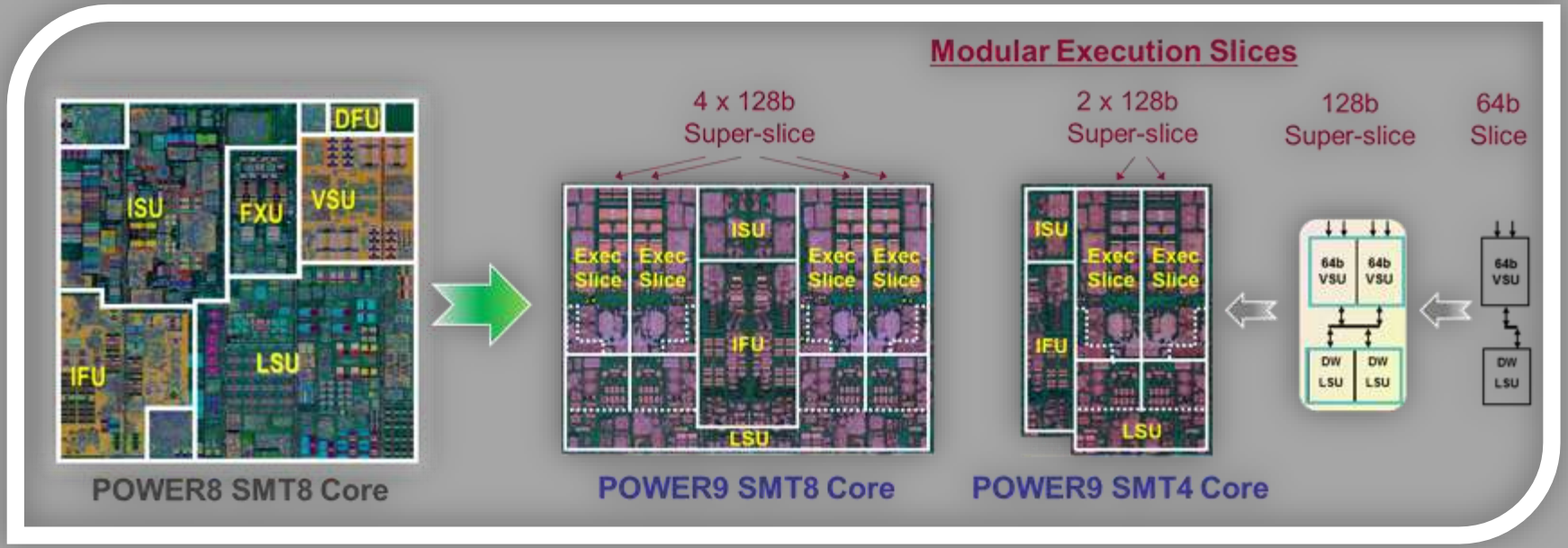
for HPC & Enterprise AI

POWER9 Processor Family Processors

Core Count / Size

SMP scalability Memory subsystem	SMT8 Core 12 SMT8 Cores / Chip PowerVM Ecosystem Continuity	SMT4 Core 24 SMT4 Cores / Chip Linux Ecosystem Optimized
Scale-Out – 2 Socket Optimized **Robust 2 socket SMP system** **Direct Memory Attach** • Up to 8 DDR4 ports • Up to 170 GB/s memory bandwidth • Commodity packaging form factor		
Scale-Up – 4+-Socket Optimized **Scalable System Topology / Capacity** • Large multi-socket **Buffered Memory Attach** • 8 Buffered channels • Up to 230 GB/s memory bandwidth		

POWER9 Processor



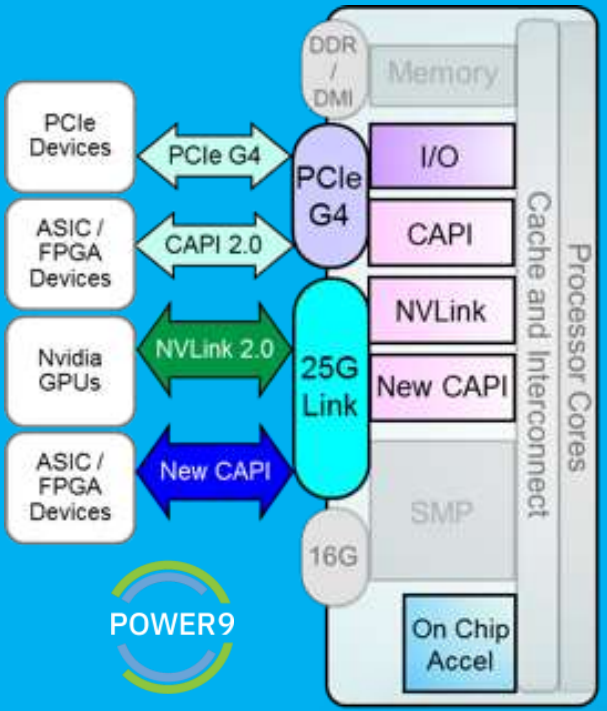
redesigned core provides improved efficiency and workload alignment with market needs

State of the Art I/O and Acceleration Attachment Signaling

- PCIe Gen 4 x 48 lanes – 192 GB/s duplex bandwidth
- 25G Link x 48 lanes – 300 GB/s duplex bandwidth

Robust Accelerated Compute Options with OPEN standards

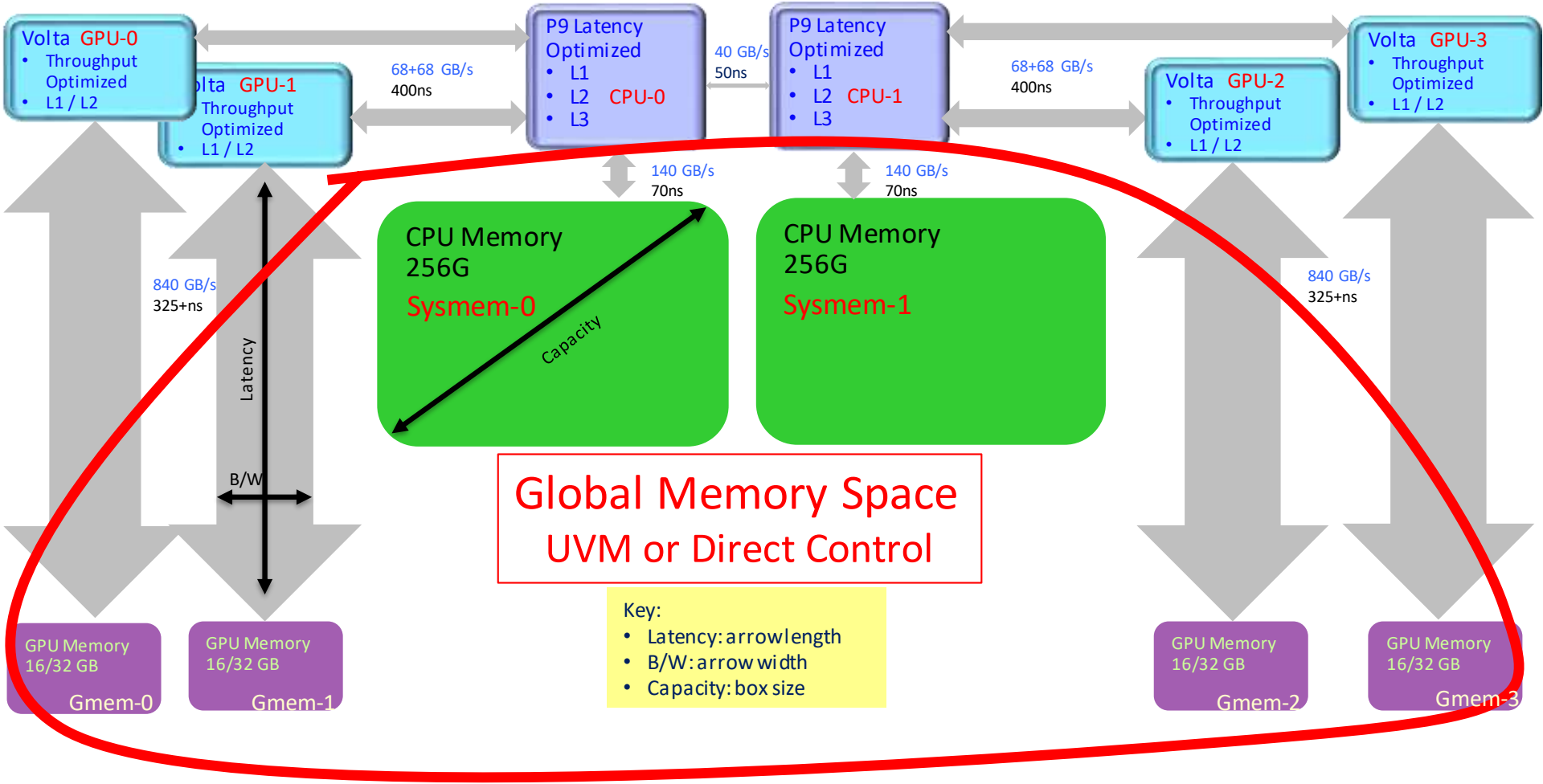
- On-Chip Acceleration – GZip x1, 842 Compression x2, AES/SHA x2
- CAPI 2.0 – 4x bandwidth of POWER8 using *PCIe Gen 4*
- OpenCAPI – High bandwidth, low latency and open interface using *25G Link*
- NVLink 2.0 – Next generation GPU←→CPU bandwidth and integration



- Large processor/accelerator bandwidth with very low latency
- Coherent memory and virtual addressing capability for all accelerators
- OpenPOWER community enablement robust accelerated compute options

POWER9 Premier Acceleration Platform

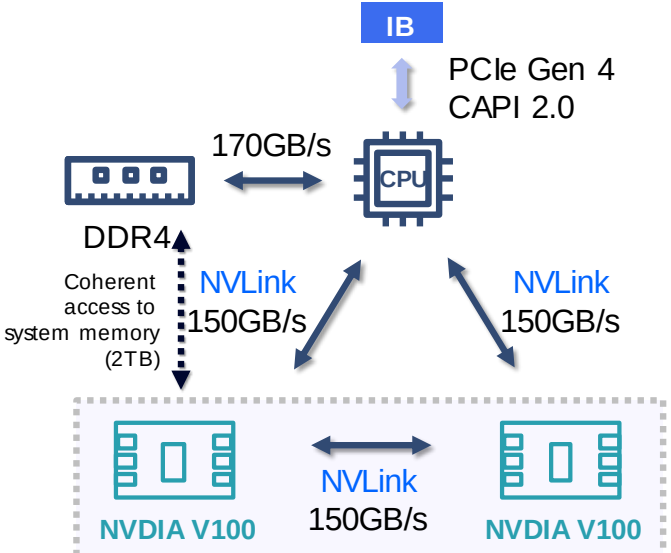
Open Power Server Memory Model



4 GPUs @150GB/s

CPU ←→ GPU bandwidth

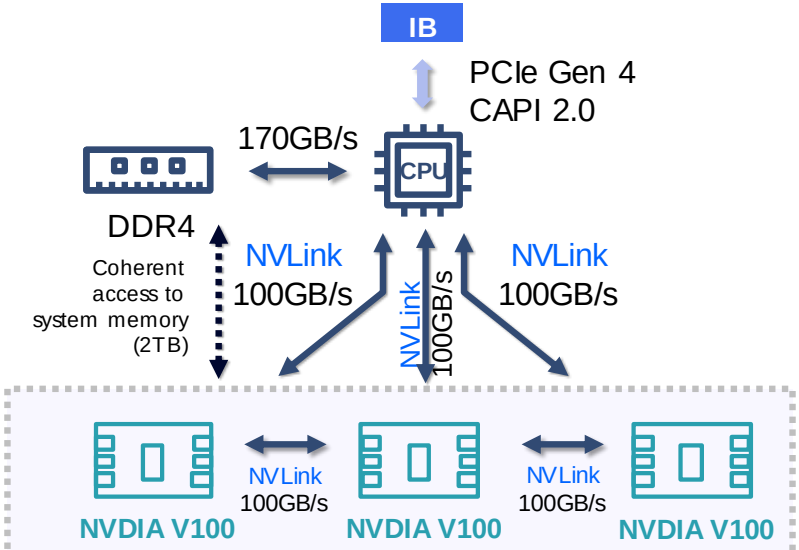
Coherent access to system memory
PCIe Gen 4 and CAPI 2.0 to InfiniBand
Air and Water cooled options



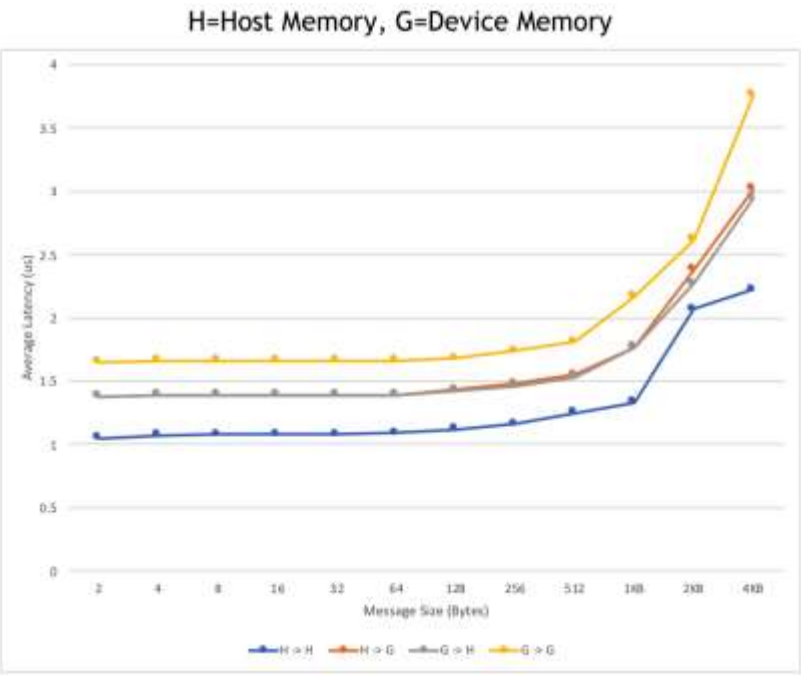
6 GPUs @100GB/s

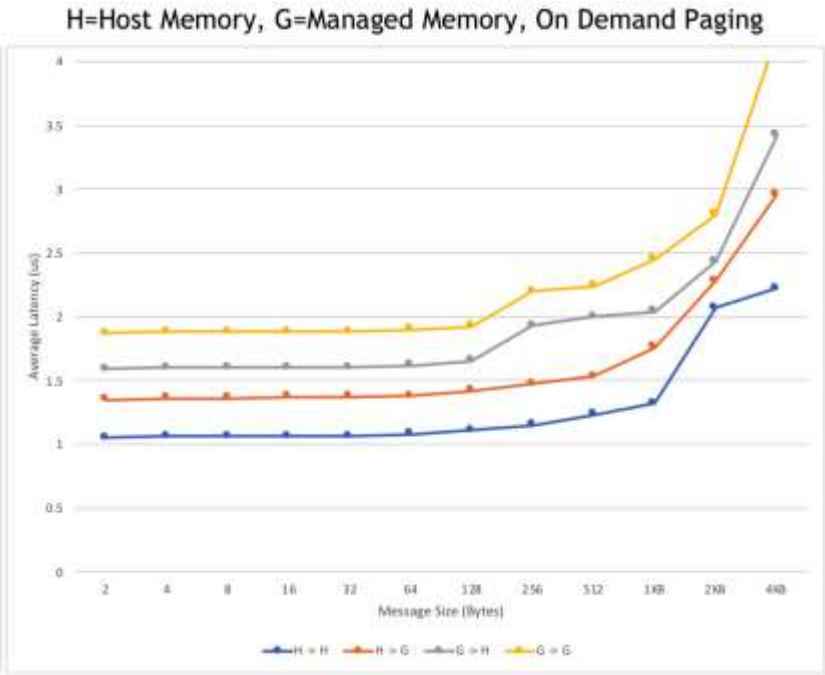
CPU ←→ GPU bandwidth

Coherent access to system memory
PCIe Gen 4 and CAPI 2.0 to InfiniBand
Water cooled only



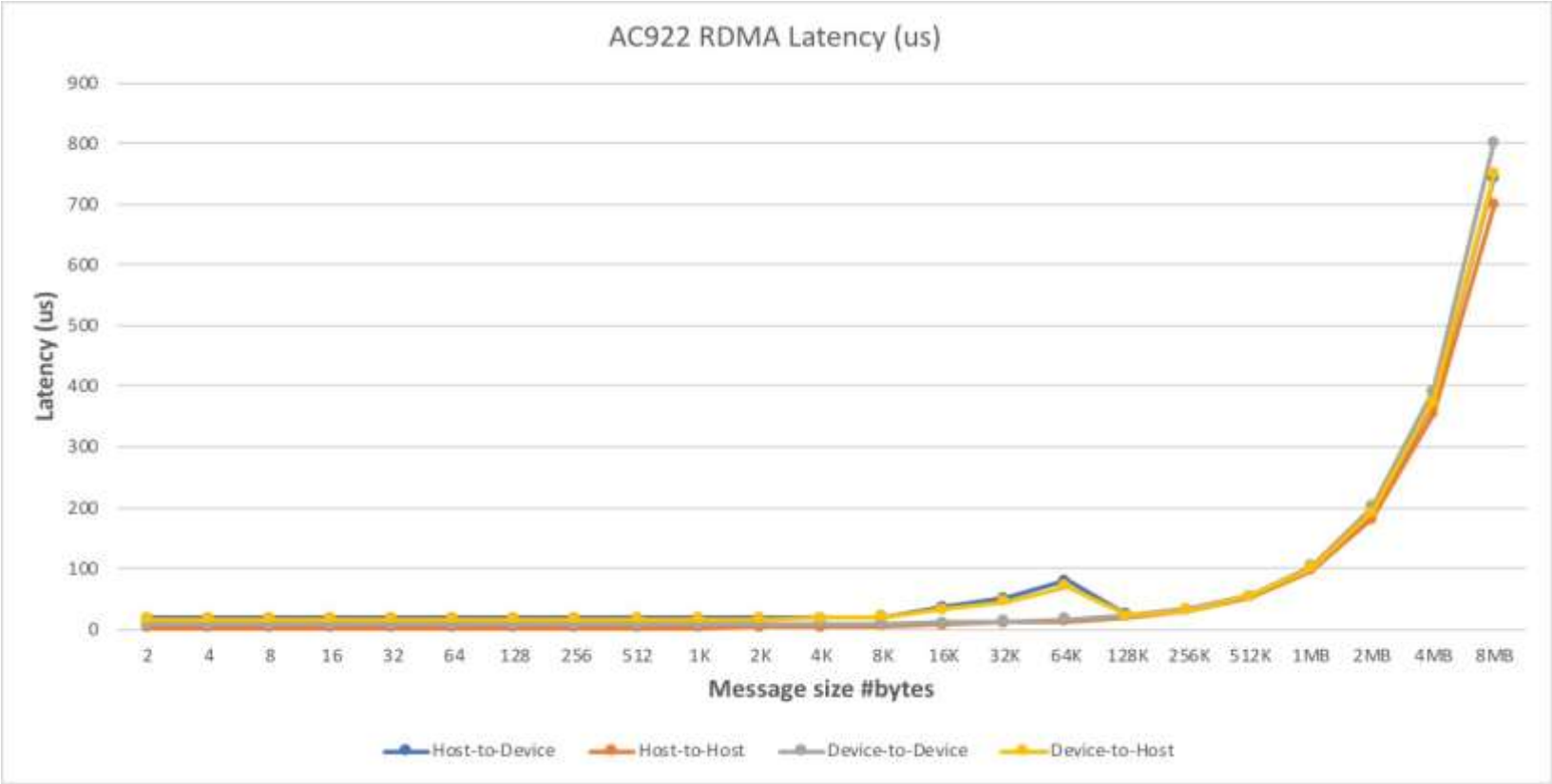
AC922 Intra-node RDMA latency





ib_send_lat, 4 Tesla V100-SXM2-16GB POWER9 server, half-roundtrip latency 1000 iterations, max inline data 0B, RC protocol,CX-5 CAPI mode enabled, tx:mlx5_0/GPU0 rx:mlx5_0/GPU1, loopback

AC922 RDMA Inter-node Latency



384 hours (16 days)
to train a model built on ImageNet-22K using ResNet-101 on a server with 8 GPUs.

Distributed Deep Learning trained this model in 7 hours
58x faster by scaling the workload across 64 servers and 256 GPUs. Now iterate!

POWER9 scales with 95% efficiency.

Limited memory on a GPU ~~is~~ ***was*** a problem for deep neural network training

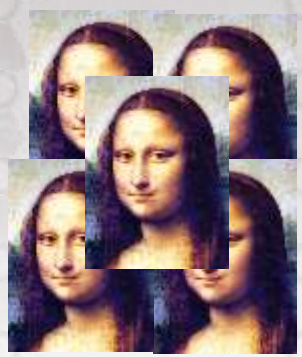
POWER9

IBM PowerAI

OpenCAPI

Traditional Model Support (Competitors)

Limited memory on GPU forces trade-off in model size / data resolution which leads to **less complex, shallower neural nets that don't perform**



Large Model Support (IBM Power)

Use system memory and GPU coherency with NVLink 2.0 to train deep neural nets with higher resolution data and **develop more accurate models for better inference capability**

Caffe 3.7x

Chainer 3.8x

TensorFlow 2.3x

train more
build more
know more

POWER9

POWER9 delivers 3.8x reduction in AI training with same NVIDIA GPU

train more | build more | know more

Critical capabilities (regression, nearest neighbor, recommendation systems, +++) operate on more than just the GPU memory

Use Server and GPU memory to support higher resolution data by moving large amounts of data between the CPU and GPU

PowerAI automatically enables seamless use of Server and GPU memory

NVLink 2.0 and POWER9 significantly cuts training times and boosts performance (accuracy) of the model with higher resolution data

Chainer

GoogLeNet – 1000 epochs

LOWER IS BETTER

IBM PowerAI

4xTesla V100 GP NVLink 2.0

POWER9

[2622] seconds

IBM 3.8x faster

[9709] seconds

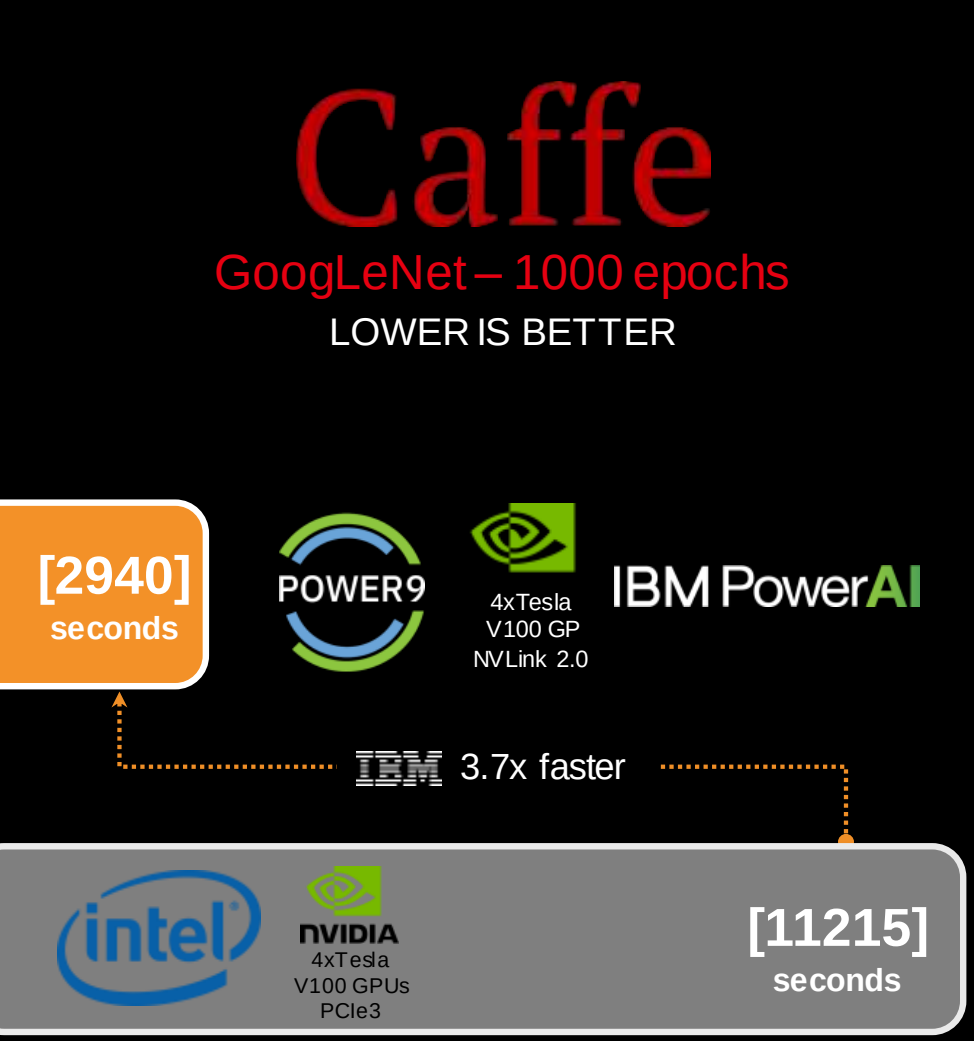
NVIDIA 4xTesla V100 GPUs PCIe3

intel

Benchmark details in speaker notes.

POWER9 delivers 3.7x reduction in AI training with same NVIDIA GPU

train more | build more | know more



Benchmark details in speaker notes.

Critical capabilities (regression, nearest neighbor, recommendation systems, +++) operate on more than just the GPU memory

Use Server and GPU memory to support higher resolution data by moving large amounts of data between the CPU and GPU

PowerAI automatically enables seamless use of Server and GPU memory

NVLink 2.0 and POWER9 significantly cuts training times and boosts performance (accuracy) of the model with higher resolution data

POWER9 delivers 2.3x more images processed per second vs tested x86 systems

train more | build more | know more

Critical capabilities (regression, nearest neighbor, recommendation systems, +++) operate on more than just the GPU memory

Use Server and GPU memory to support higher resolution data by moving large amounts of data between the CPU and GPU

PowerAI automatically enables seamless use of Server and GPU memory

NVLINK 2.0 and POWER9 significantly cuts training times and boosts performance (accuracy) of the model with higher resolution data

TensorFlow

GoogLeNet – 1000 epochs

HIGHER IS BETTER

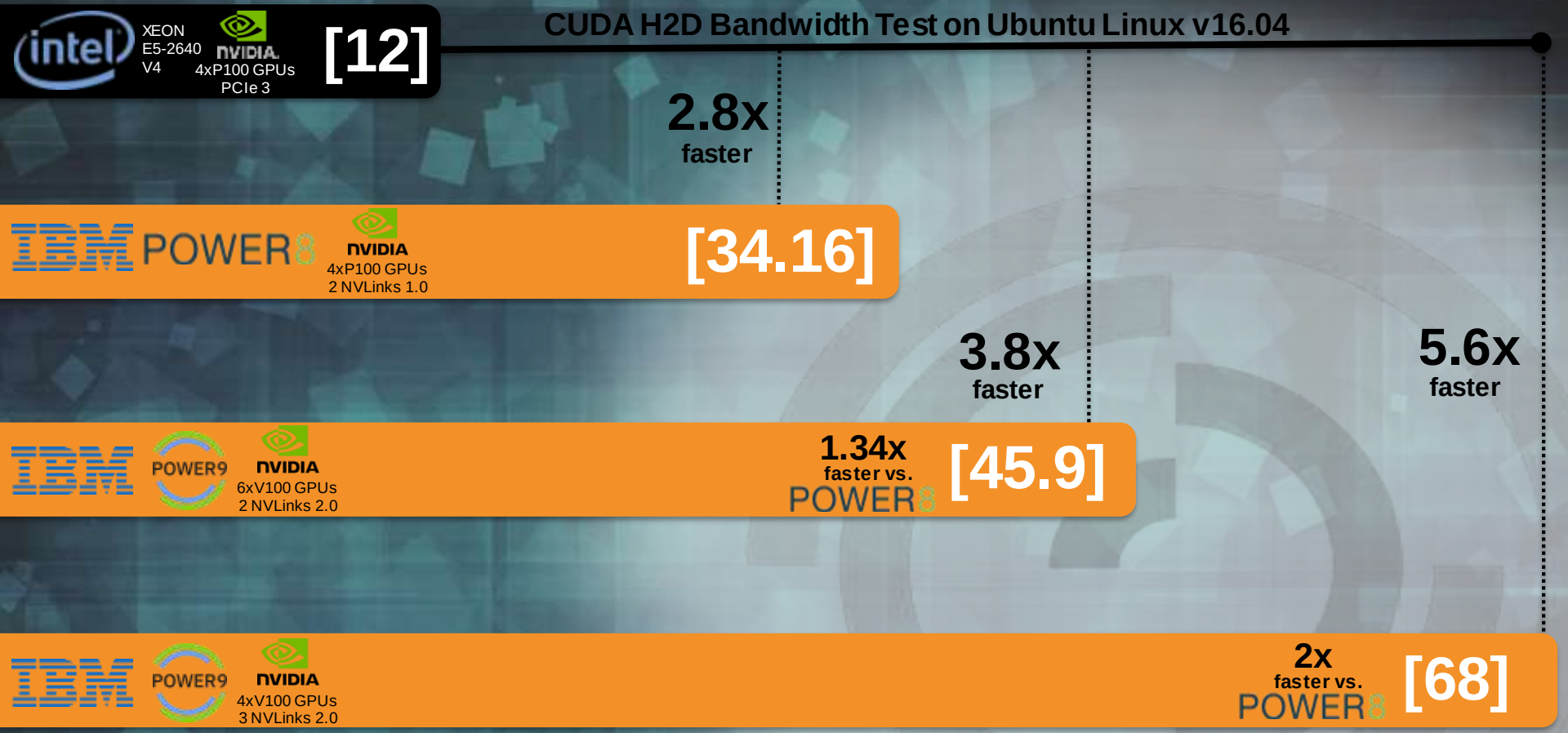
IBM POWER9 NVIDIA 4xTesla V100 GP NVLINK 2.0 **[4763] images / second**

IBM 2.3x faster

intel NVIDIA 4xTesla V100 GPUs PCIe3 **[2042] images / second**

Benchmark details in speaker notes.

CUDA H2D Bandwidth Test on Ubuntu Linux v16.04



Benchmark details in speaker notes.

POWER9 delivers ~5x faster data movement CPU ←→ GPU when running CPMD simulations

HIGHER IS BETTER

[~50GB/s] IBM POWER9 NVIDIA NVLink 2.0

~5x faster

[~10GB/s] intel PCIe G3

train more | build more | know more

POWER9 delivers 2.6x faster CPMD simulation runs with same NVIDIA GPU

LOWER IS BETTER

IBM POWER9 NVIDIA 4xV100 GPUs NVLink 2.0 [351] seconds

2.6x faster

IBM NVIDIA 4xP100 GPUs NVLink 1.0 POWER8 [673] seconds

intel PCIe G3 NVIDIA 4xV100 GPUs NVLink 2.0 [917] seconds

Benchmark details in speaker notes.

intel POWER8

both use
PCIe G3

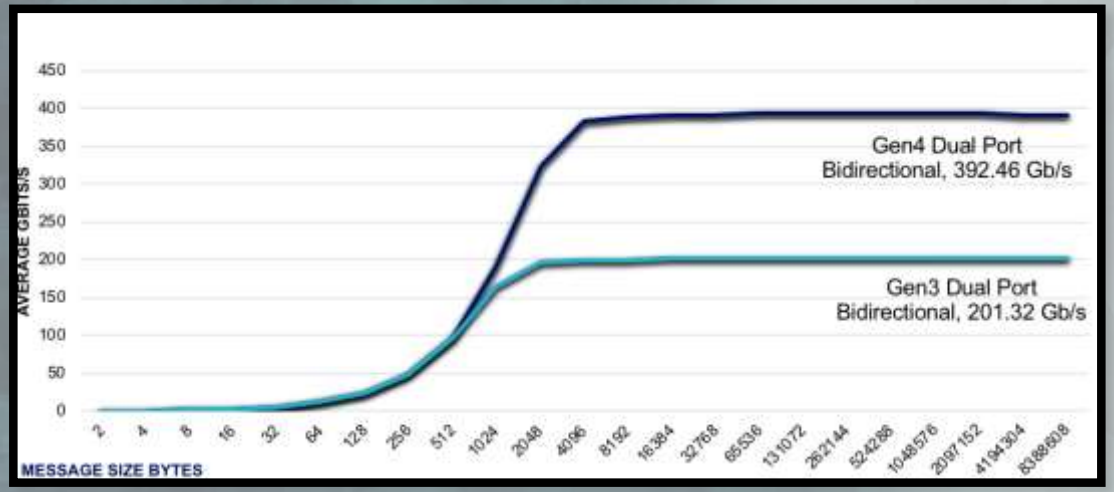
EDR InfiniBand 100GB/s Bandwidth

(Gen 3 vs Gen 4)

IBM POWER9 **PCIe G4**

PCIe G4 bus versus PCIe G3

2x faster



Benchmark details in speaker notes.