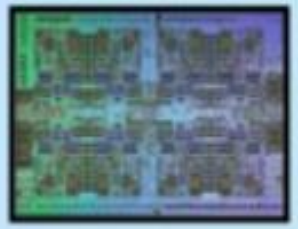
NVIDIA V100 Performance Characteristics on IBM AC922 (POWER9) System and HPC Application Performance

Pidad D'Souza (pidsouza@in.ibm.com)
Aditya Nitsure (anitsure@in.ibm.com)
ISDL, IBM India Pvt Limited

Agenda

- IBM POWER Systems
- NVIDIA Tesla V100
- AC922 System Configuration
- AC922 System & GPU Characteristics
- HPC Application Characteristics
  - NAMD
  - GROMACS
  - CPMD
  - HPCG
  - 3D-FFT

POWER Processor Technology Roadmap

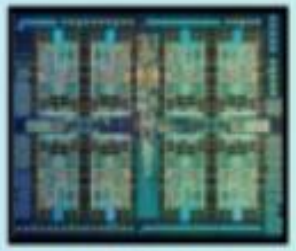


POWER7
45 nm

Enterprise

- 8 Cores
- SMT4
- eDRAM L3 Cache

1H10

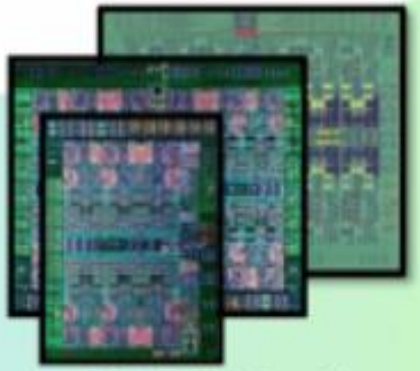


POWER7+
32 nm

Enterprise

- 2.5x Larger L3 cache
- On-die acceleration
- Zero-power core idle state

2H12

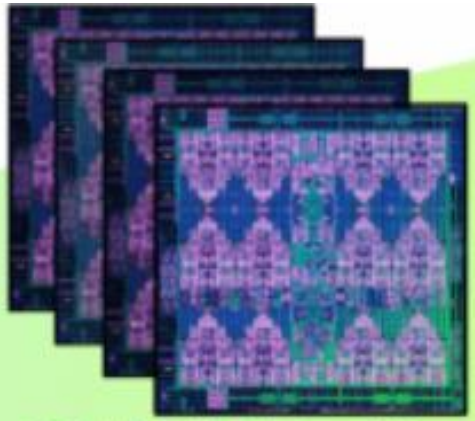


POWER8 Family
22nm

Enterprise &
Big Data Optimized

- Up to 12 Cores
- SMT8
- CAPI Acceleration
- High Bandwidth GPU Attach

1H14 – 2H16



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

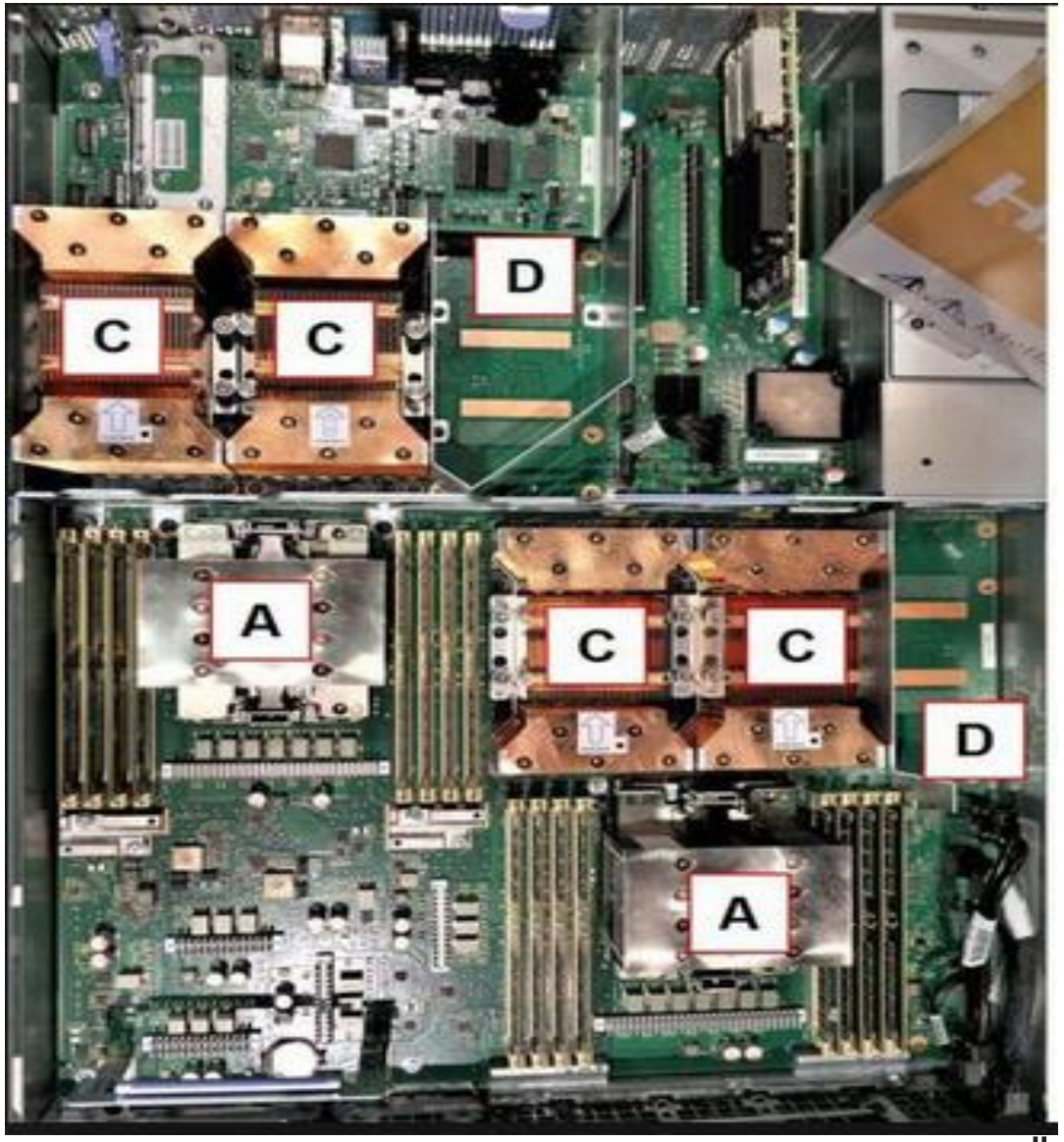
2H17 – 2H18+

IBM AC922 (POWER9)

IBM® Power System™ Accelerated Compute Server (AC922)

- Delivers unprecedented performance for modern HPC, analytics, and artificial intelligence (AI)
- Designed to fully exploit the capabilities of CPU and GPU accelerators
  - Eliminates I/O bottlenecks and allows sharing memory across GPUs and CPUs
  - Extraordinary POWER9 CPUs
  - 2-4 NVIDIA® Tesla® V100 GPUs with NVLink

- Co-optimized hardware and software for deep learning and AI
- Supports up to 5.6x more I/O bandwidth than competitive servers
- Combines the cutting edge AI innovation Data Scientists desire with the dependability IT
- Next Gen PCIe - PCIe Gen4 2x faster

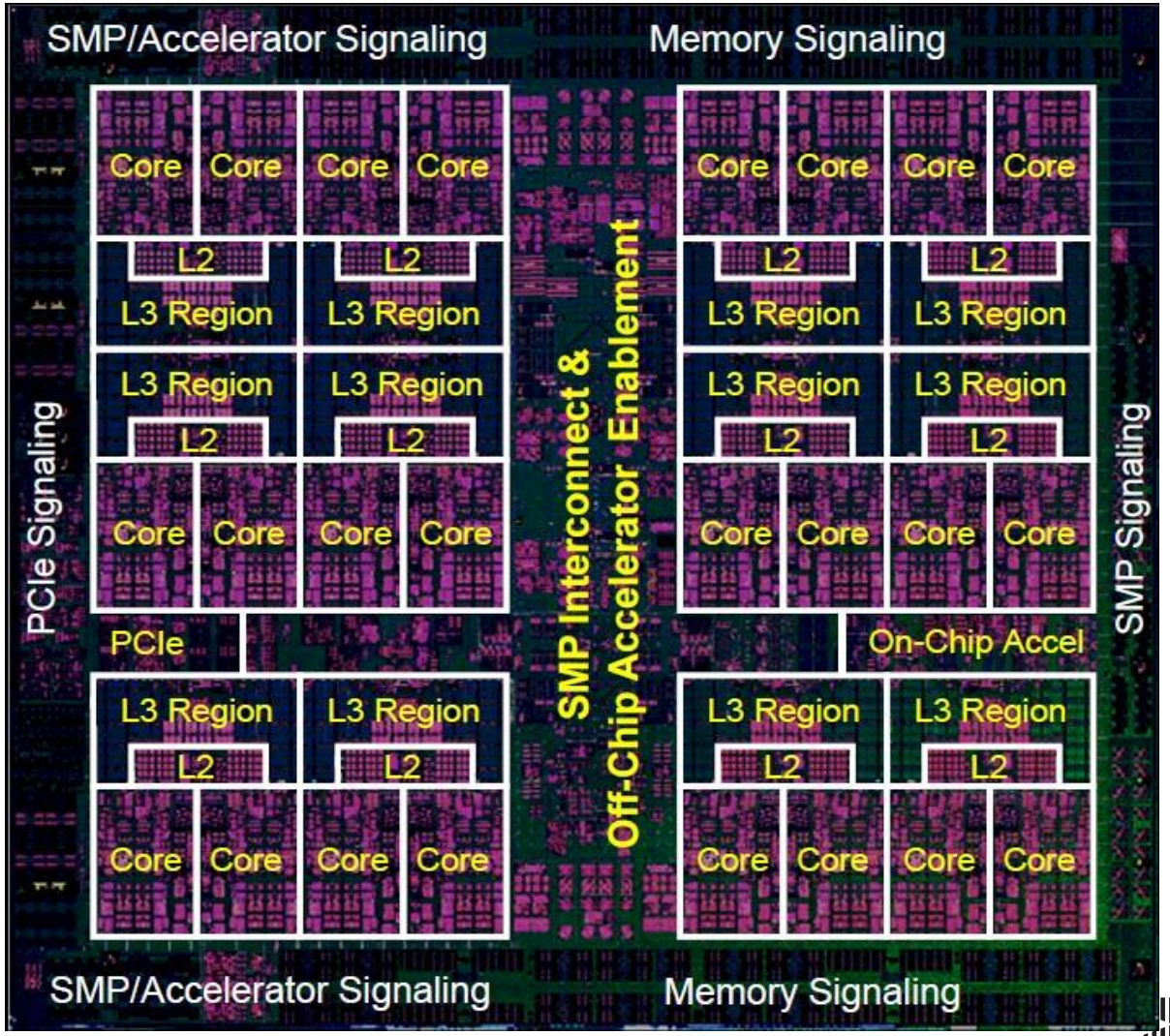


Air Cooled

A. POWER9POWER9 processor with air-cooling heat sink
C. NVIDIA V100 NVLink module with air-cooling heat sink
D. Empty NVLink socket

IBM POWER9 Processor – General features

- 14nm finFET semiconductor
- Stronger thread performance
- POWER ISA 3.0
- Enhanced Cache Hierarchy
- NVIDIA NVLink 2.0 high speed link with 25GB/s per link bandwidth
- I/O System – PCIe Gen4
- Improved device performance and reduced energy
- Nominal & Peak freq: 2.2GHz & 2.8GHz



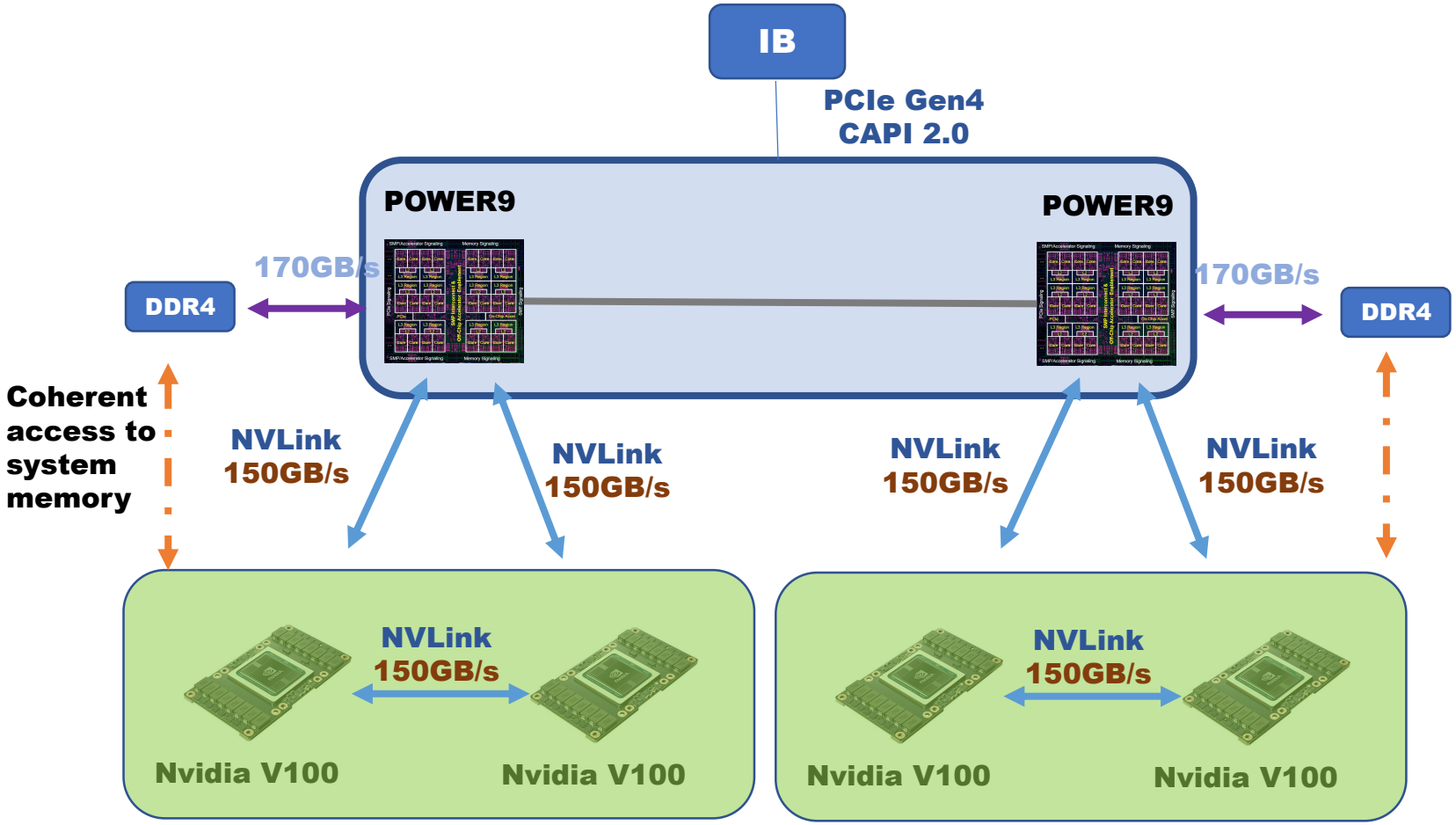
NVIDIA Tesla V100 GPU

- Designed for AI Computing and HPC
- Second-Generation NVLink™
- HBM2 Memory: Faster, Higher Efficiency
- Enhanced Unified Memory and Address Translation Services
- Maximum Performance and Maximum Efficiency Modes

- Number of SM/cores : 80/5120
- Double Precision Performance : 7.5 TFLOPS
- Single Precision Performance : 15 TFLOPS
- GPU Memory : 16GB
- Memory bandwidth : 900 GB/s

https://devblogs.nvidia.com/inside-volta/

AC922 System Configuration – 4 GPU



High bandwidth interconnect NVLink 2.0

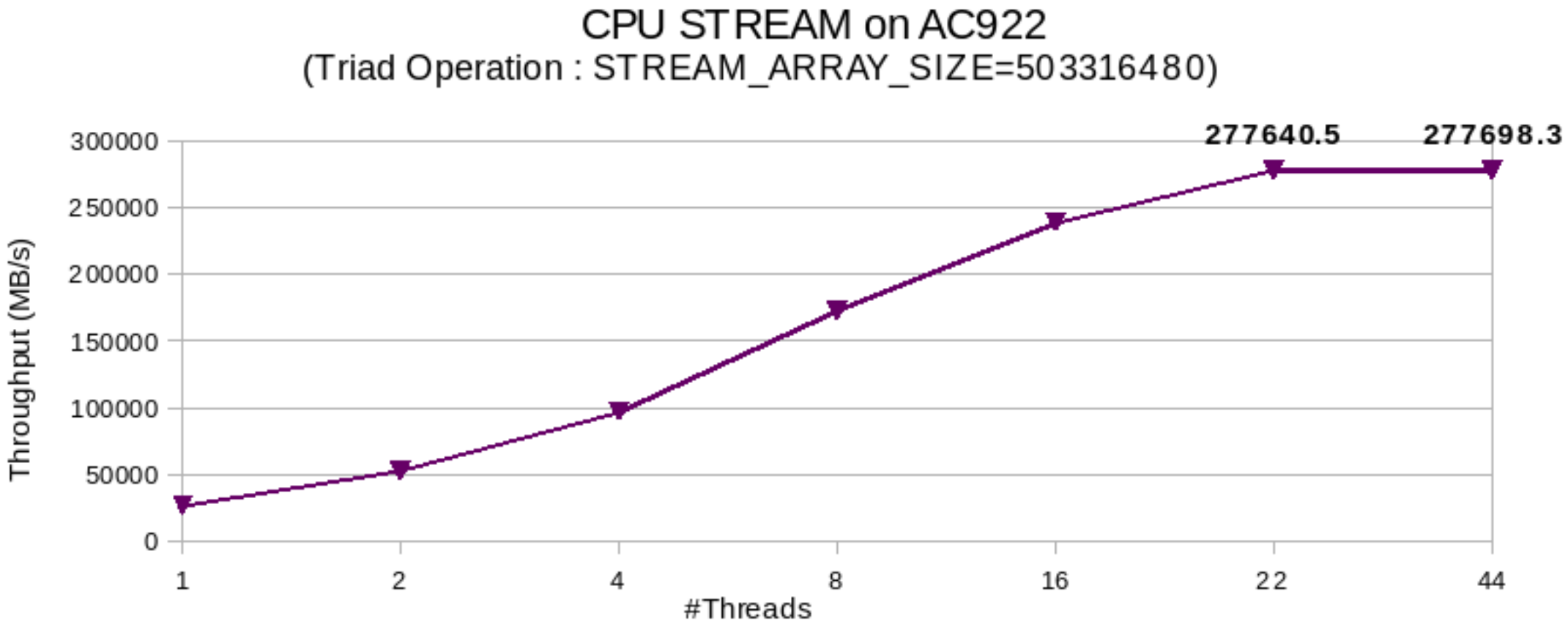
150GB/s bi-directional bandwidth between CPU-GPU and GPU-GPU

Coherent access to CPU memory

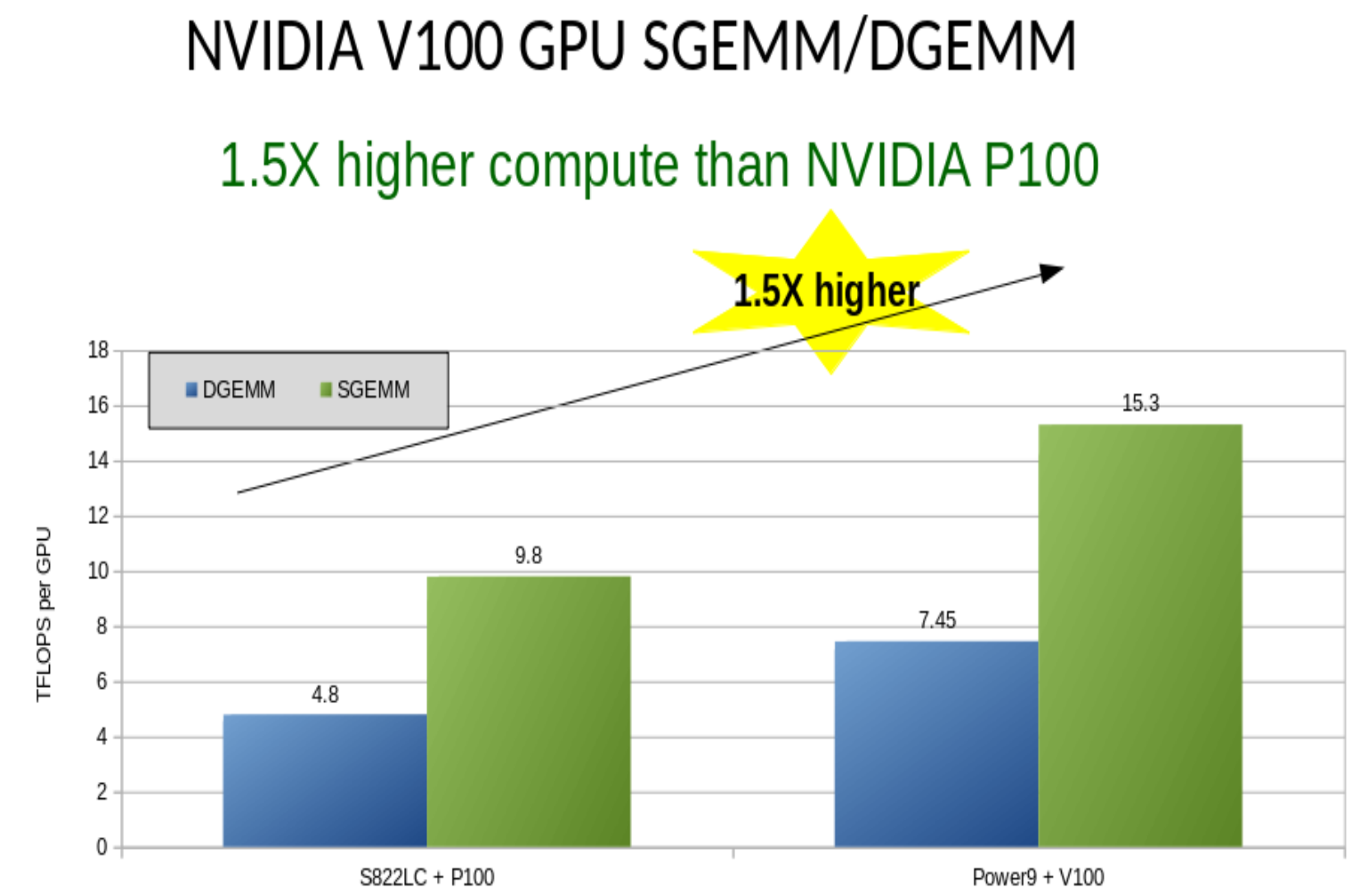
AC922 System Characteristics

CPU STREAM Bandwidth(preliminary, not submitted)

STREAM benchmark : Measure system memory bandwidth.
(https://www.cs.virginia.edu/stream/)



NVIDIA Volta 100 Compute – Single and Double Precision

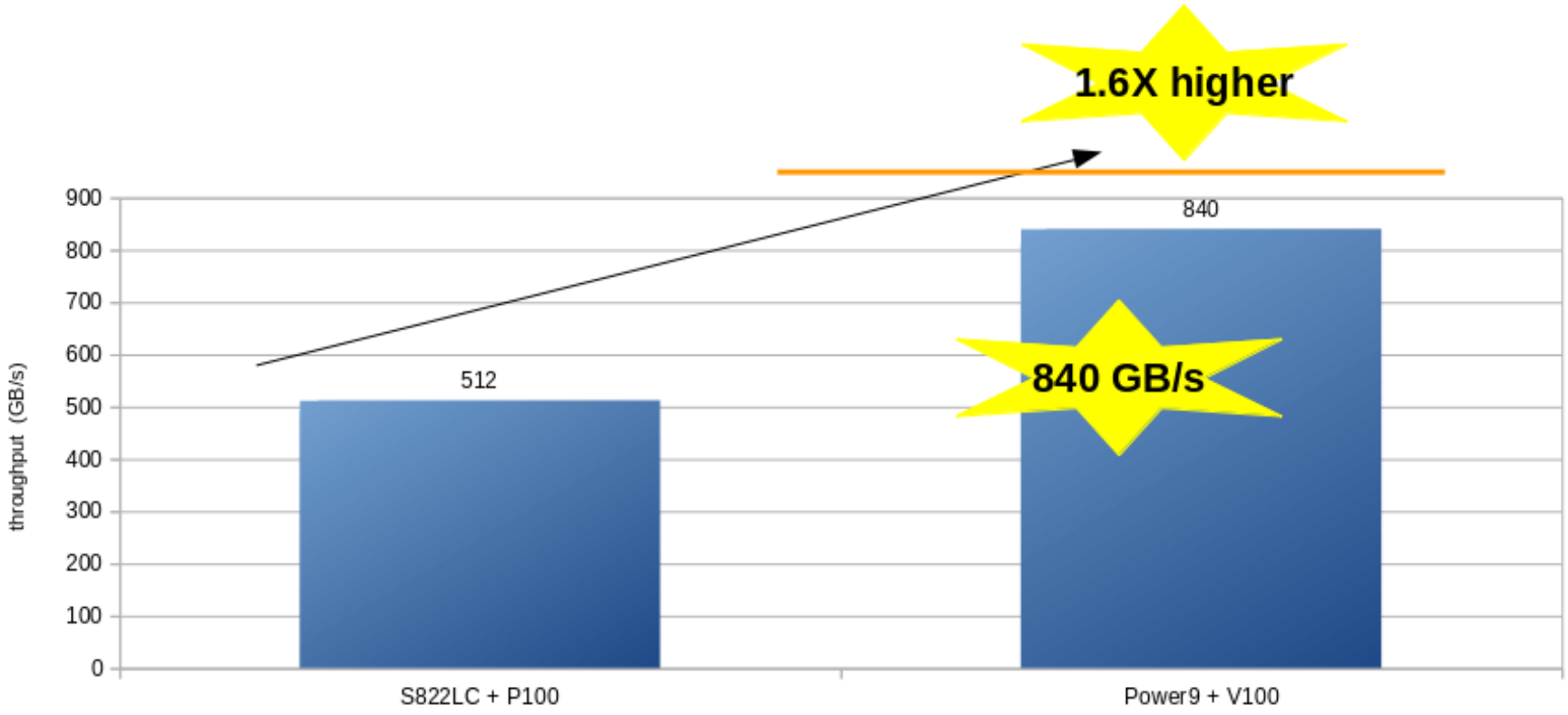


Note : cuBLAS APIs used for measurements

- Applications to have more compute power
- Shorten time to completion
- Accomplish more simulation/experiment

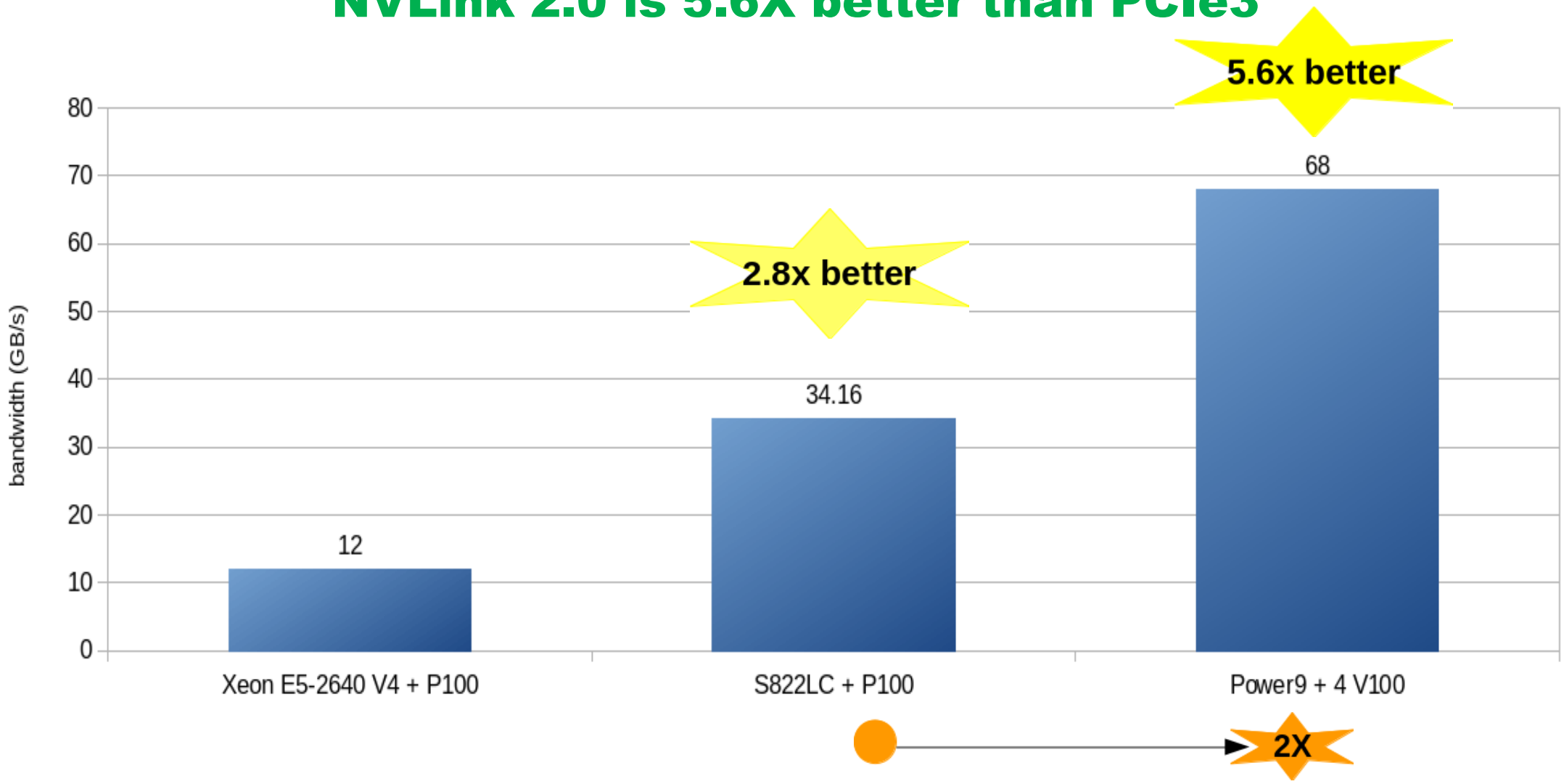
NVIDIA V100 GPU memory bandwidth (GPU STREAM)

1.6X higher bandwidth than NVIDIA P100



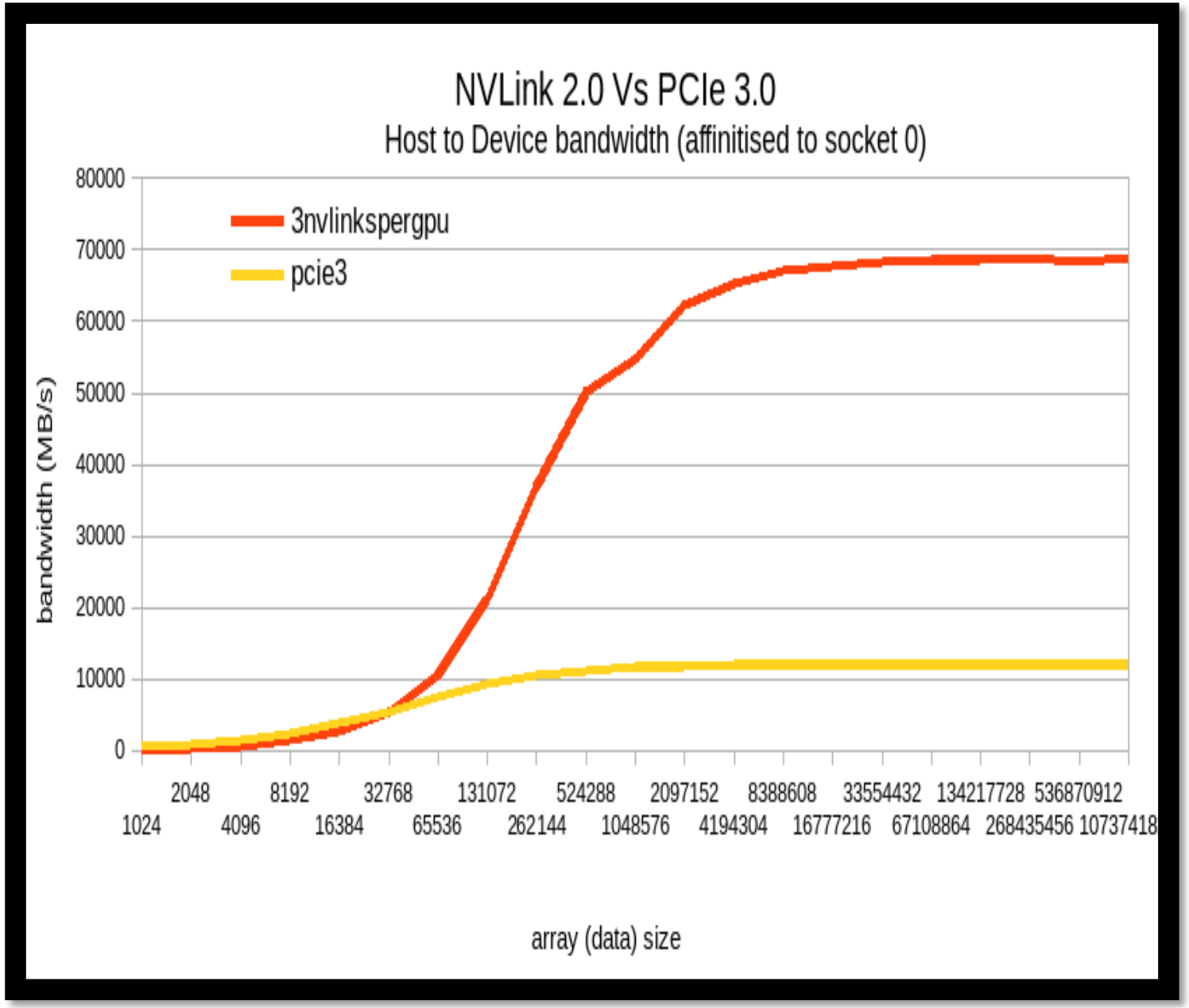
CPU to GPU NVLink Vs PCIe3 bandwidth

NVLink 2.0 is 5.6X better than PCIe3



Note: NVIDIA bandwidth test used for measurement

NVLINK Bandwidth with varied data sizes



- Minimize communication latencies
- Unlock PCIe bottlenecks
- Transfer larger data at high speed
- Ideal for out-of-core computations

HPC Application Performance

System Components Impacting HPC performance

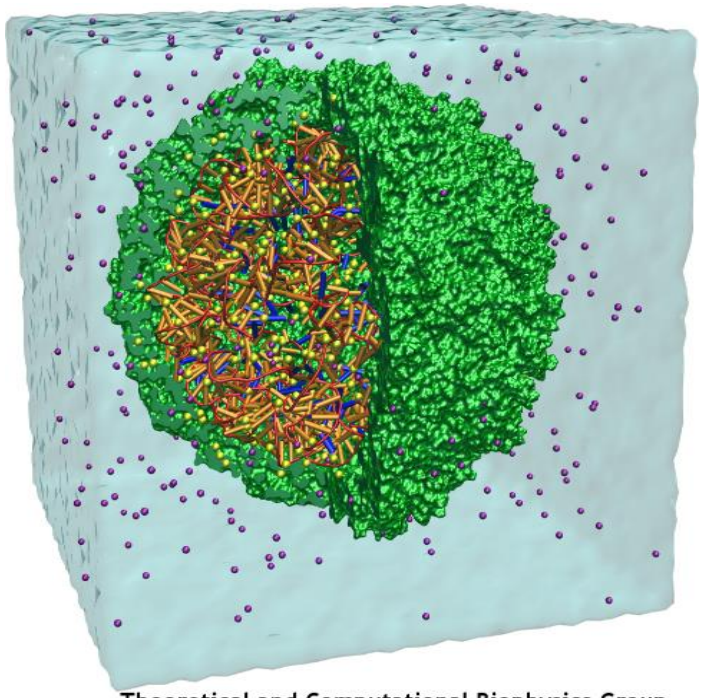
- Compute (CPU & Accelerators)
- Memory capacity
- Memory bandwidth
- Interconnect Latency and Bandwidth
- IO performance
- IBM HPC Software Stack

NAnoscale Molecular Dynamics (NAMD)

The Molecular Dynamics software from the University of Illinois at Urbana-Champaign
http://www.ks.uiuc.edu/Research/namd/

Compute atomic trajectories by solving equations of motion numerically using empirical force fields

Calculate the force as the negative gradient of a scalar potential energy function



Theoretical and Computational Biophysics Group
Beckman Institute
University of Illinois at Urbana-Champaign

$$\vec{F}(\vec{r}) = -\nabla U(\vec{r}),$$

$\ldots\vec{r}$ *is position vector*

For systems of biomolecules, this potential function involves the summing

$$U(\vec{r}) = \sum U_{bonded}(\vec{r}) + \sum U_{nonbonded}(\vec{r})$$

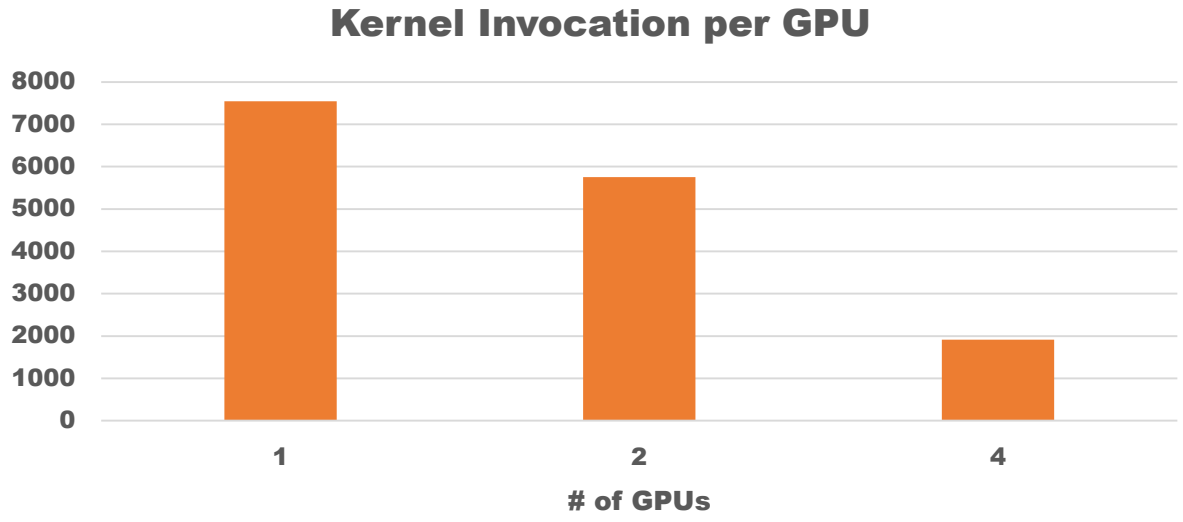
Research areas

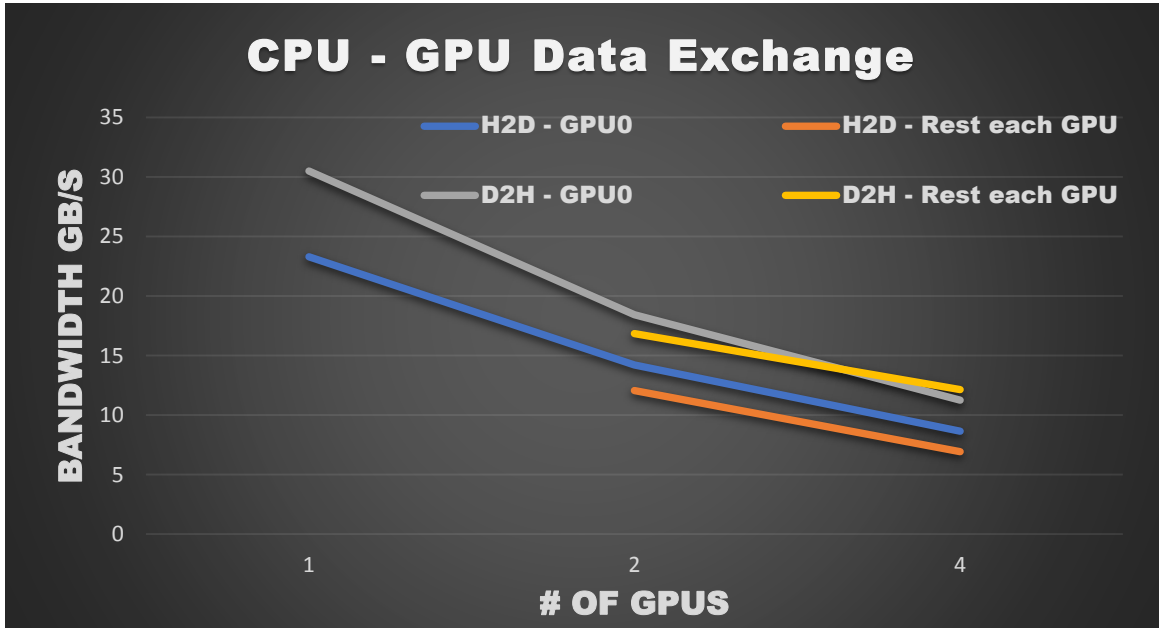
- Protein folding
- Viruses
- Symbiont bacteria
- Molecular motors
- Nanoengineering
- Bioenergetics
- Quantum biology
- Neurobiology

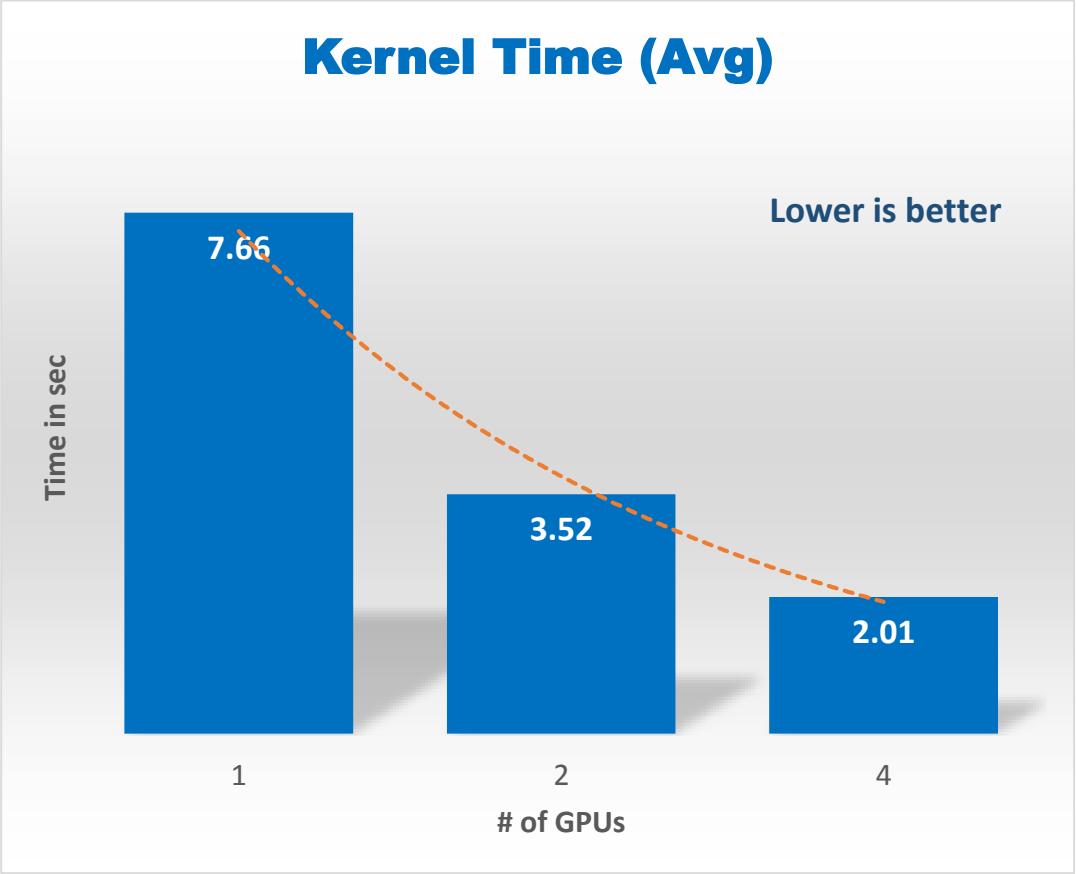
NAMD version : 2.13 (sandbox)
Object Oriented, Charm++, MPI & CUDA parallelization

NAMD(2.13 Dev Sandbox) Observations

Work distribution on GPUs

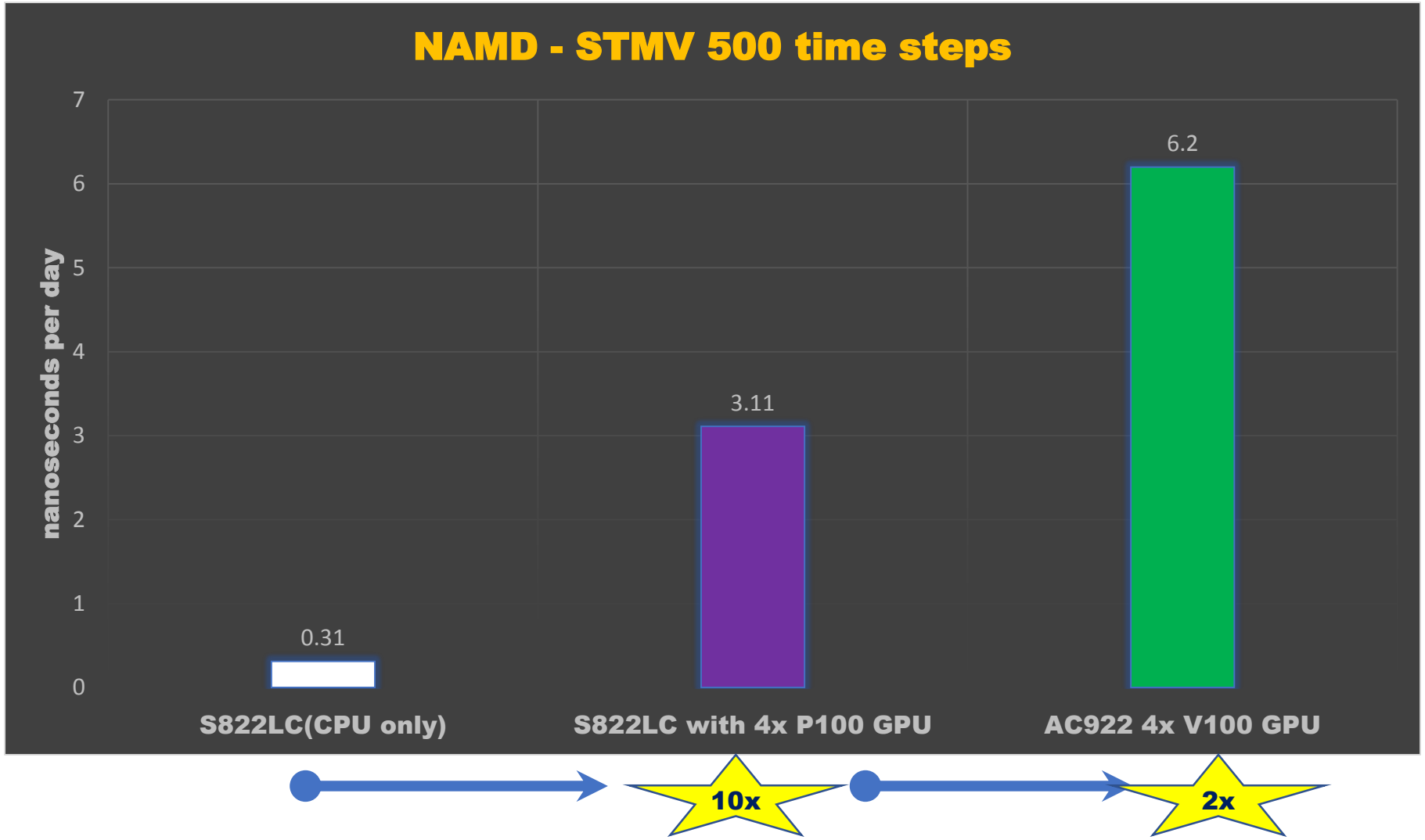






- Less amount of data exchange with increased number of GPUs
- Large amount of data transfer – Potential NVLink benefit

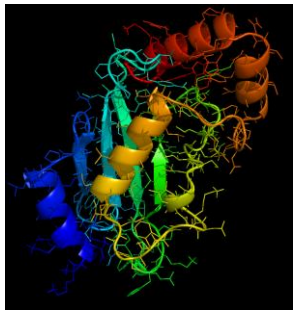
NAMD on IBM POWER Systems



Higher performance with fewer systems

GROMACS

GROningen MAchine for Chemical Simulations



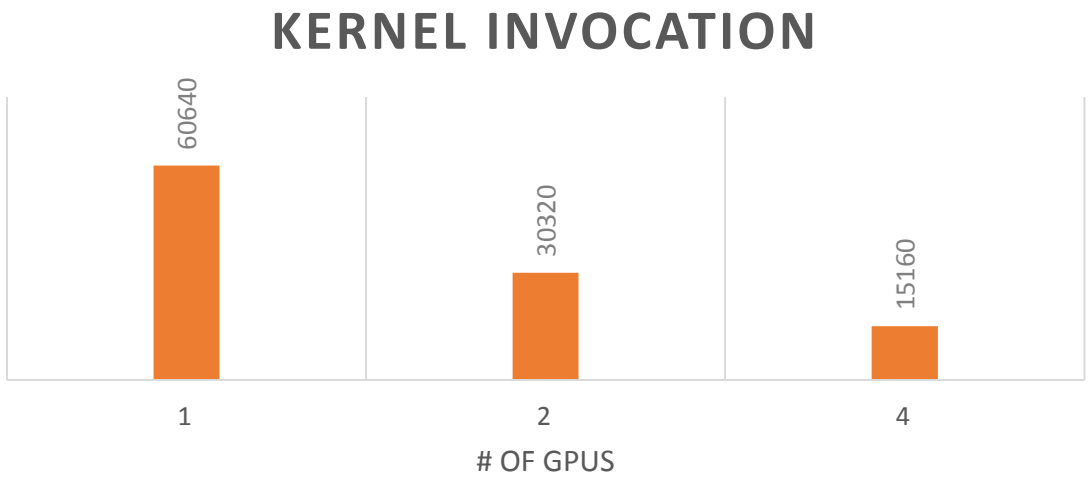
- A versatile package to perform molecular dynamics simulation with the Newtonian equations of motion for systems with hundreds to millions of particles

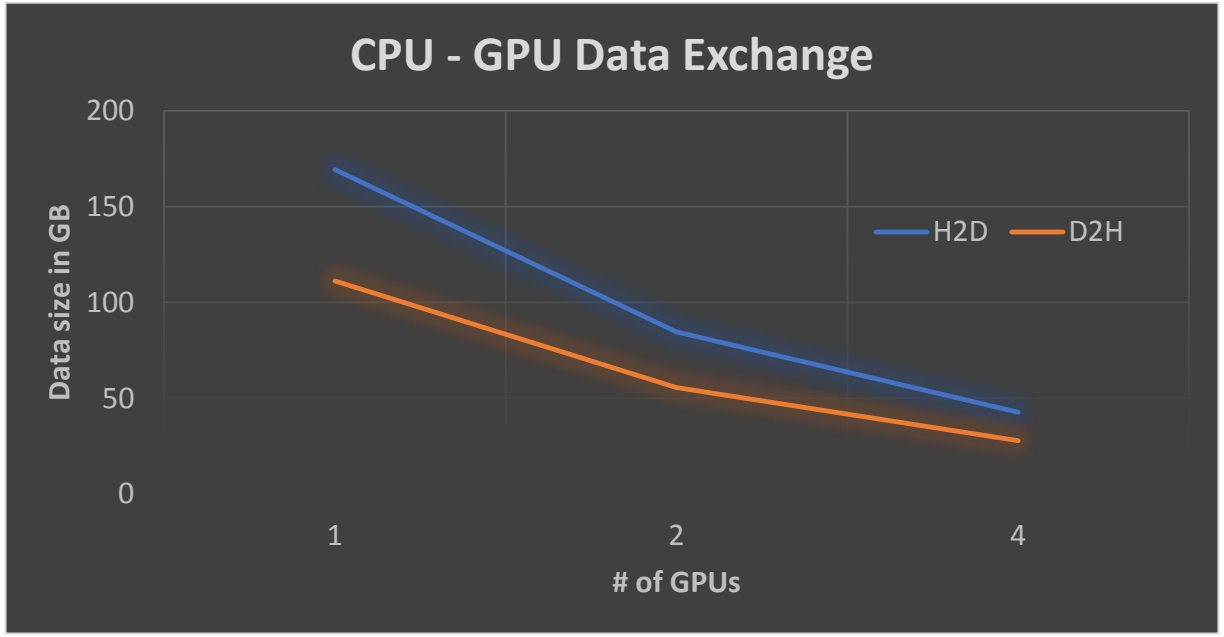
- GPU acceleration is core part of GROMACS that works in combination with GROMACS domain decomposition and load balancing code

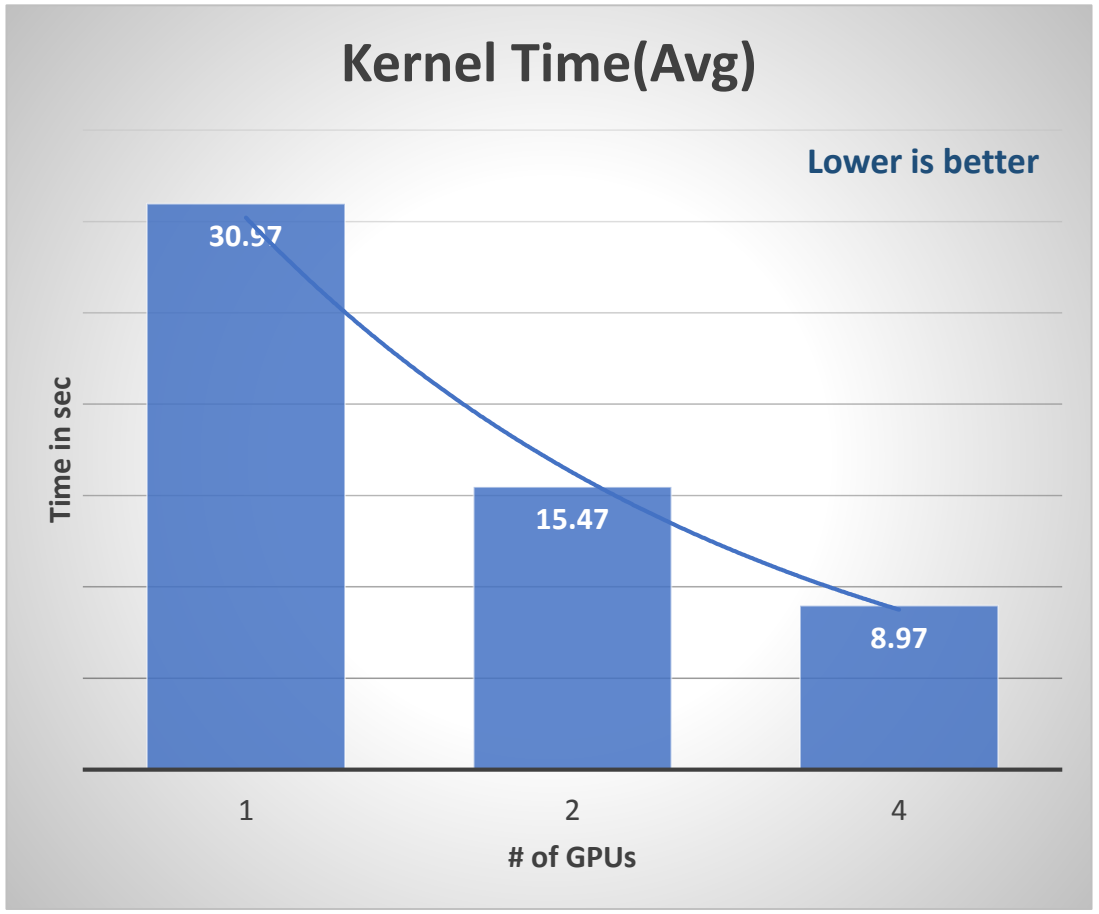
- Gromacs team claims –
  - GPU accelerated code performances up to 5x better compared to CPU-only processing.

Gromacs 2016.3/RF Observations

Work distribution on GPUs







- Less amount of data exchange with increased number of GPUs
- Large amount of data transfer – Potential NVLink benefit

Car-Parrinello Molecular Dynamics (CPMD)

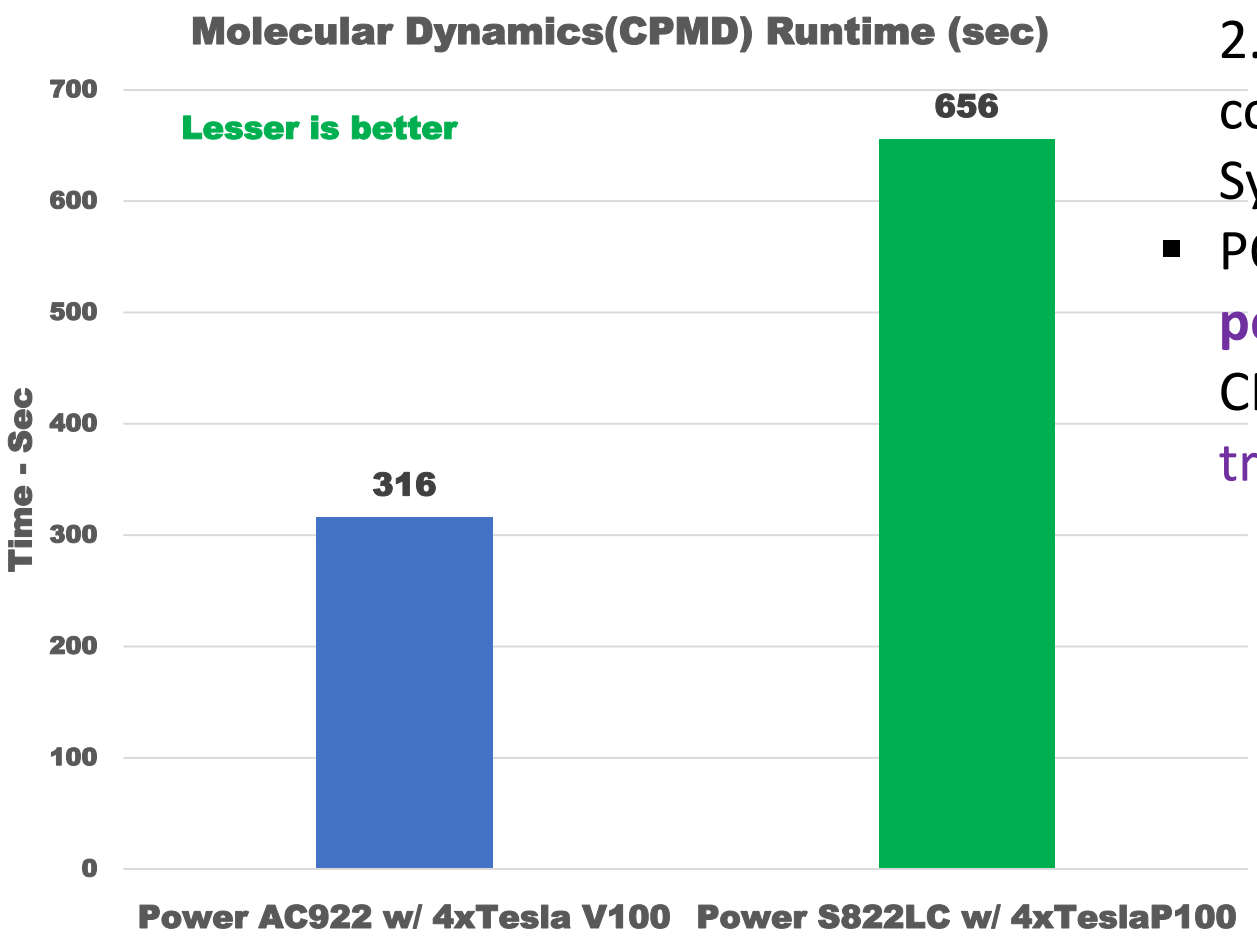
The Molecular Dynamics software from the University of Illinois at Urbana-Champaign
http://www.ks.uiuc.edu/Research/namd/

- A parallelized plane wave / pseudopotential implementation of density functional theory
- Designed for ab initio molecular dynamics
- Most suitable for dynamical simulations of condensed systems or large molecules
  - less for small molecules or bulk properties of small crystals
- Computationally highly optimized for vector and scalar supercomputers, shared and distributed memory machines and combination thereof (SMP)

Capabilities

- Solution of electronic and ionic structure
- XC-functional: LDA, GGA, meta-GGA, hybrid
- Molecular dynamics in NVE, NVT, NPT ensembles
- General constraints (MD and geometry optimization)
- Metadynamics, Free energy functional
- Path integrals, QM/MM, Wannier functions
- Response properties: TDDFT, NMR, Raman, IR, . . .
- Norm conserving and ultra-soft pseudo potentials

Improved computational chemistry(CPMD) simulation CPMD on IBM POWER9™ with NVLink



- CPMD on IBM POWER9™ AC922 with NVLink 2.0 delivers 2.0X reduction in execution time compared to prior generation IBM POWER System S822LC for HPC
- POWER9 with NVLink 2.0 **unlocks the performance of GPU-accelerated version** of CPMD by enabling lightning fast CPU-GPU data transfers
  - 3.3 TB of data movement required between CPU and GPU
  - 70 seconds for NVLink 2.0 transfer time vs 300+ seconds for traditional PCIe bus transfer time

Source: Manish Modani, ISDL, IBM

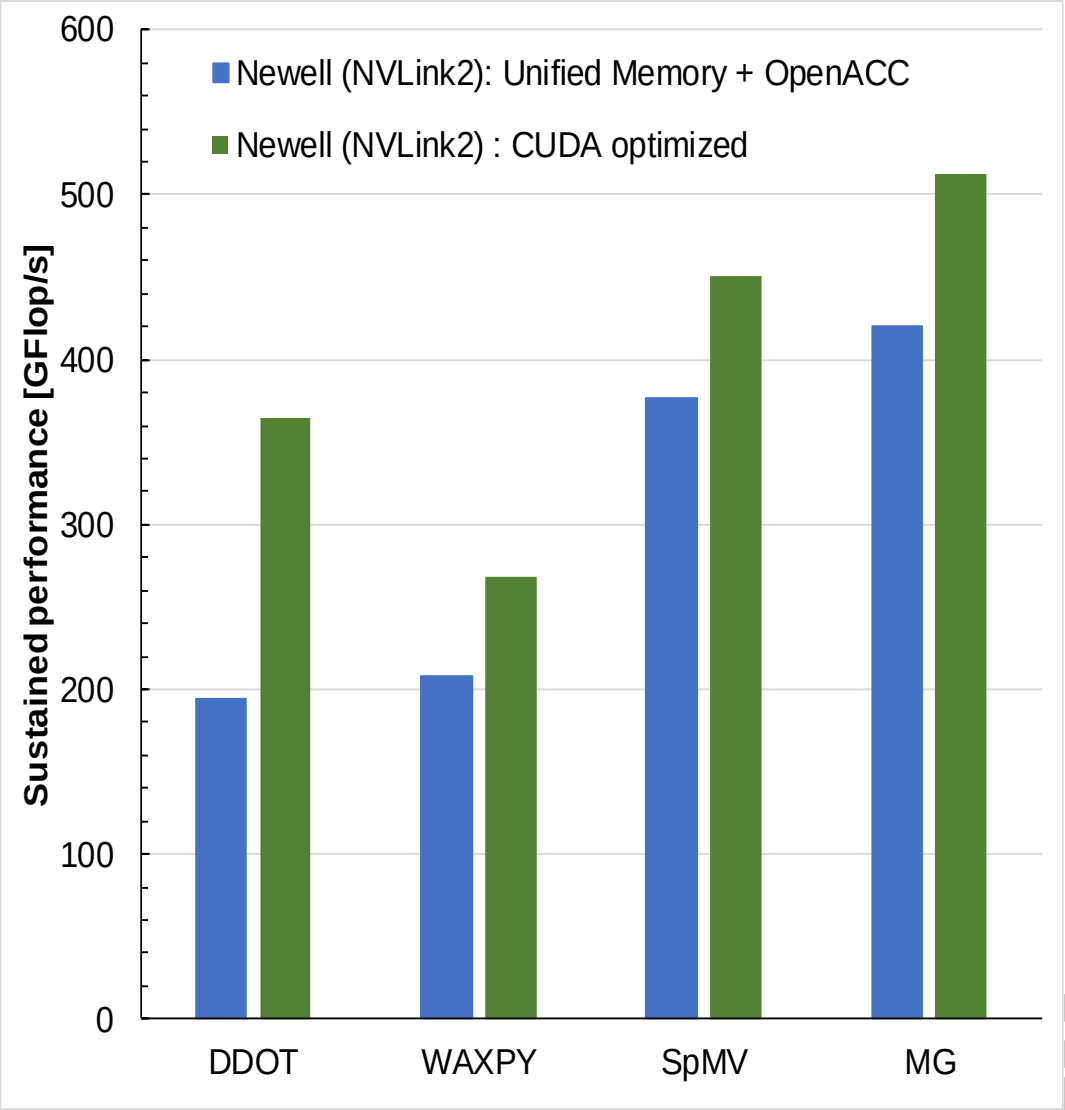
HPCG: Unified Memory + OpenACC Example

OpenACC

- Directive based programming to offload workload to GPUs
- With Unified Memory no data transfer directives are needed

HPCG Benchmark

- Solves linear equation of sparse problem with CG method
- Added 25 lines of OpenACC directives to local SMP codes, achieved 80% of CUDA optimized implementation



Source: Jun Doi, Tokyo Research Labs, IBM

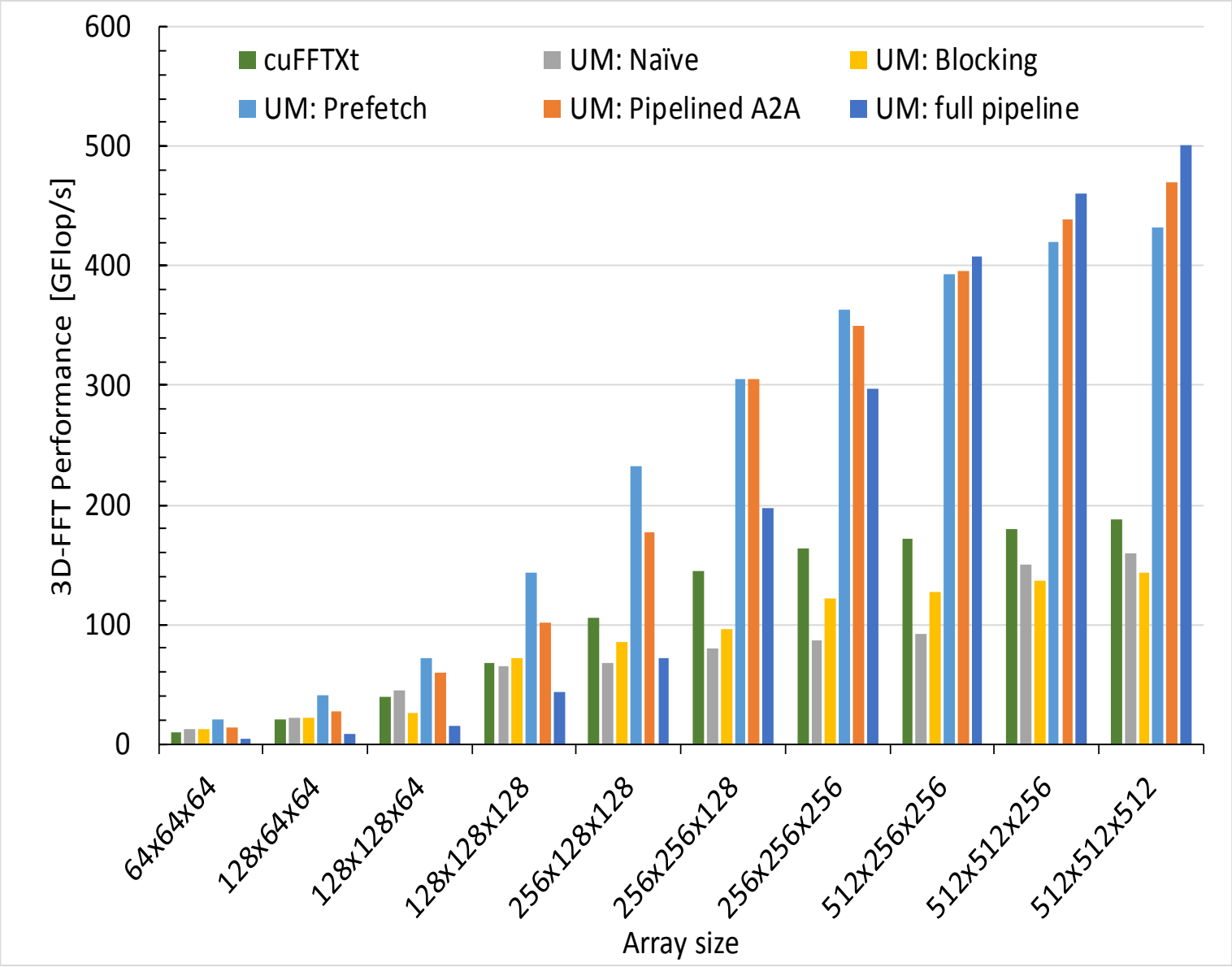
HPCG: Unified Memory + OpenACC Example

```
#ifdef _OPENACC
#pragma acc data deviceptr(pColIdx,pMatVal,rv,xv)
#else
#pragma omp target data map(to:nrow,nnz)
use_device_ptr(pColIdx,pMatVal,rv,xv)
#endif
      {

            for(c=0;c<pOpt->nColor;c++){
                  is = pOpt->colorOffset[c];
                  ie = pOpt->colorOffset[c + 1];
```

```
#ifdef _OPENACC
#pragma acc kernels async(1)
#pragma acc loop independent private(irow,sum,icol,j)
#else
#pragma omp target teams distribute parallel for
private(irow,sum,icol,j)
#endif
                  for(i=is;i<ie;i++){
                        irow = i;

                        sum = rv[irow];
```

Source: Jun Doi, Tokyo Research Labs, IBM

3D-FFT on AC922 with 4 Tesla V100



- FFT kernels are very slow if data is not on that GPU, because of overheads and latencies of implicit data copy (data migration)
- Prefetch version achieves double of the performance of cuFFTXt because we only exploit 1 transpose
- Calculation time is much smaller than all-to-all time, overlapped time by pipelining is very small
- With prefetch, large performance improvement

Source: Jun Doi, Tokyo Research Labs, IBM

Conclusion

- AC922 Designed for HPC, Analytics, AI
- High speed interconnect NVLink between CPU & GPU
- Unified memory, OpenACC simplifies programing, with good performance

References

- https://www.ibm.com/us-en/marketplace/power-systems-ac922
- http://www.gromacs.org/
- http://www.ks.uiuc.edu/Research/namd/

Acknowledgements

We are grateful to Tom Heller, Jun Doi, Manish Modani and Pradeep Kumar Madhavan for their guidance and support.

Thank You