Internal GPGPUs on Dedicated x16 Slots – Are They Needed for HPC?

Elliott Berger
HPC Solution Architect
Dell Research Computing
Elliott_Berger@Dell.com

Agenda

- "Gold" Standards in GPGPU Computing
  - Tesla S-series (S1060, S1070, S2050, S2070)
  - Generic 1U server with 2 GPUs
- C410x Design Goals and Description
  - Production-Ready
- C410x Benchmark Results
  - Internal vs. External Performance
  - GPU:Host Scaling
- Model for Enabling GPU Computing

GPGPU "Gold" Standards

NVIDIA Tesla S-Series

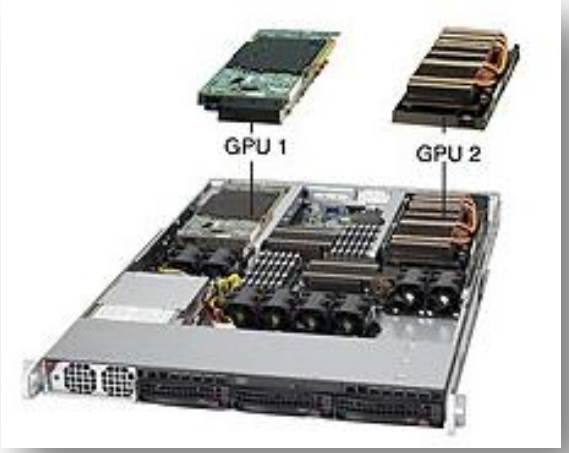
- 1U chassis (external)
- Up to 4 GPUs
- Connects to a "host" via a PCI-e HIC Card
  - Host is typically a 1U node/server
  - HIC = Host Interface Card
  - Two (2) PCI-e inputs per chassis
    - Each one addresses up to 2 GPUs

- 4 GPUs / 1 RackU (Tesla alone)
- 4 GPUs / 2 RackU (with one 1U host)
- 4 GPUs / 3 RackU (with two 1U hosts)

- **GPUs are *external* to the host**
- **2 GPUs / RackU or ~1.3 GPUs / RackU**
- **2 GPUs *sharing* a x16**
- **Single power supply**

SuperMicro SuperServer family

- 1U chassis (internal)
- Includes the host and GPUs
- Up to 2 GPUs
- Each GPU on a dedicated x16



- 2 GPUs / RackU

- **GPUs are *internal* to the host**
- **2 GPUs / RackU**
- **2 GPUs on *separate* x16's**

- **DDR performance for interconnect (x4 slot)**
- **Single power supply**

Dell PE C410x: Design Goals

Dell PowerEdge C410x Design Goals

- Increase density (more GPUs per RackU)
- Introduce "flexibility"
  - GPU/Host ratio = 1:1, 2:1, 3:1, 4:1, …, 8:1
- "Commercial Grade" – or – production-ready

"Today, most GPGPU buyers plan to use these processors for exploration and applications development rather than for production work. Assuming this experimentation period goes well, the transition to ***production computing*** will follow."

- IDC, February 19, 2010

Dell PE C410x: Description

Dell PowerEdge C410x:

- 3U chassis (external)
  - -“Room-and-Board” for PCIe Gen-2 x16 devices
  - Up to 8 hosts

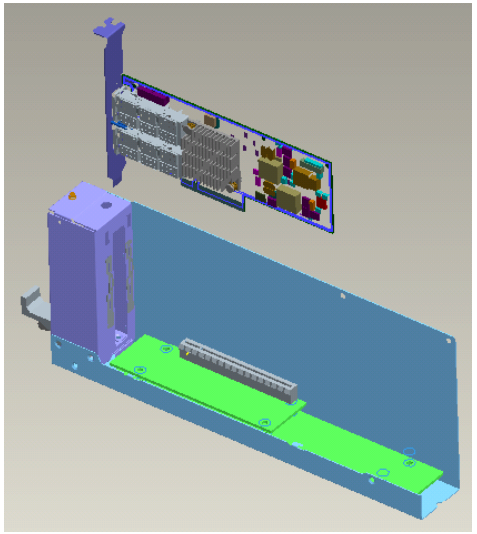
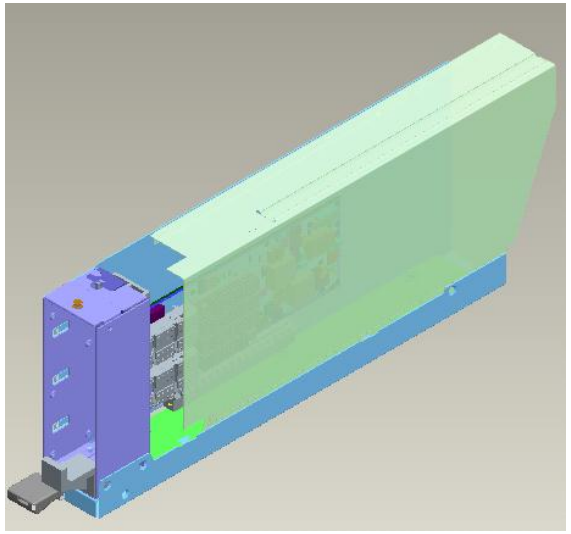
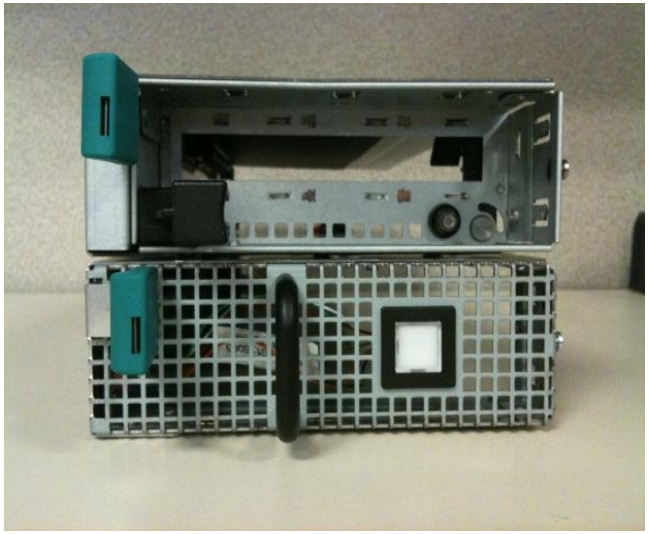
- Sixteen (16) x16 Gen-2 Devices
  - Initial Target = GPGPUs
  - Support for any FH/HL or HH/HL device
  - Each slot Double-Wide
  - Individually Serviceable

- N+1 Power (3+1)
  - Gold (90%)
- N+1 Cooling (7+1)

Dell PowerEdge C410x:

- Sixteen (16) x16 Gen-2 Modules
  - PCIe Gen-2 x16 compliant
  - Independently serviceable

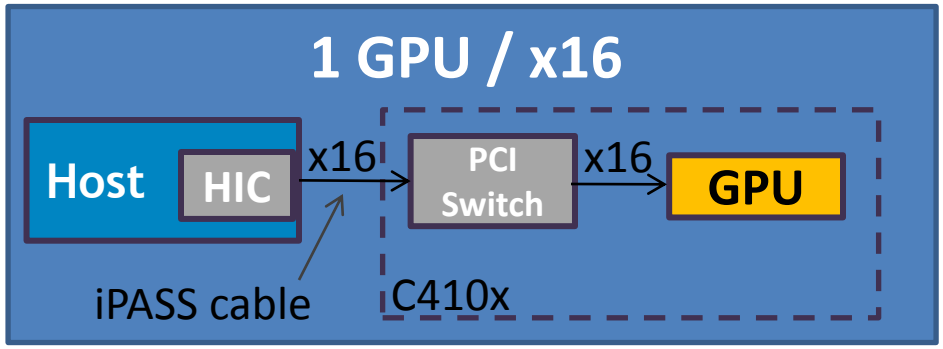
Dell PowerEdge C410x:

- Common Carrier supports low-profile, half-length, single-width, PCI Express cards with standard full-height bracket
- Allows external cabling (network fabric, etc.)

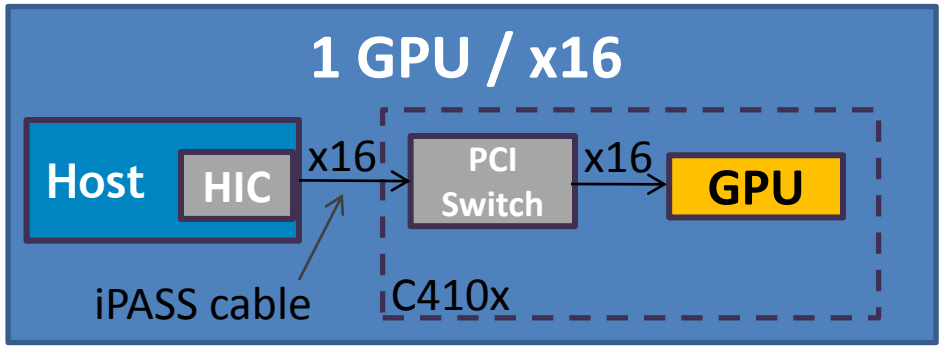
C410x Flexibility: One Host with a Single x16

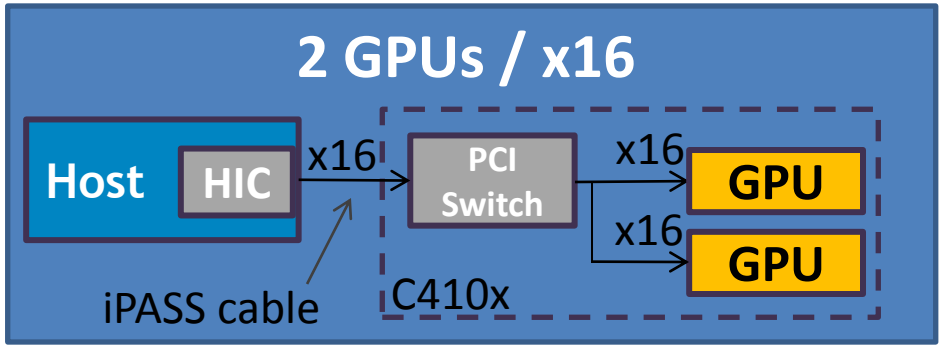
1:1, 2:1, 3:1, 4:1, 8:1



C410x Flexibility: One Host with a Single x16

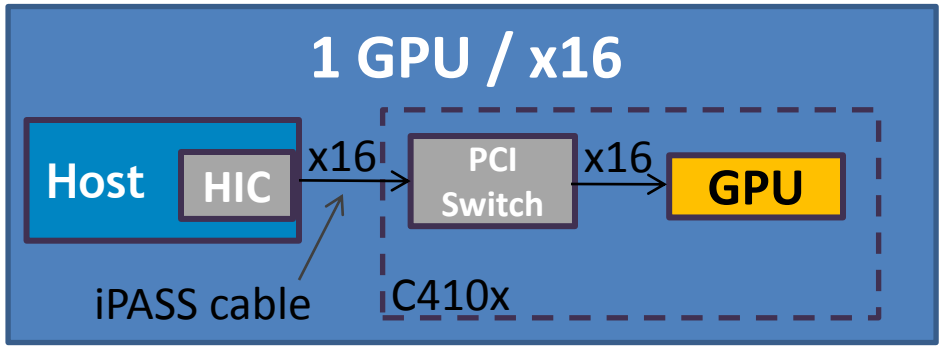
1:1, 2:1, 3:1, 4:1, 8:1

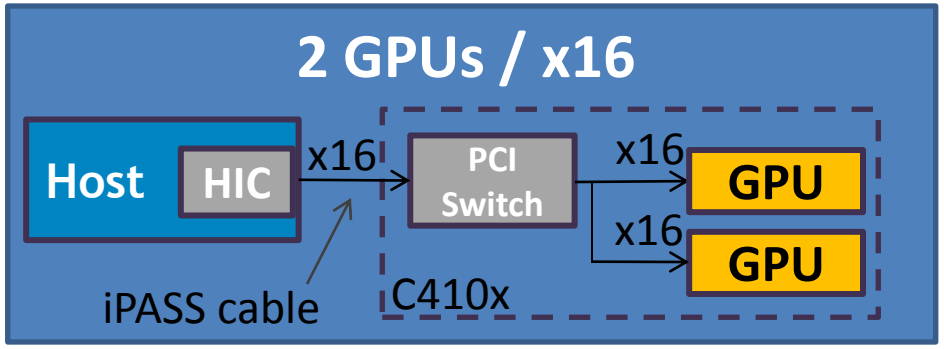


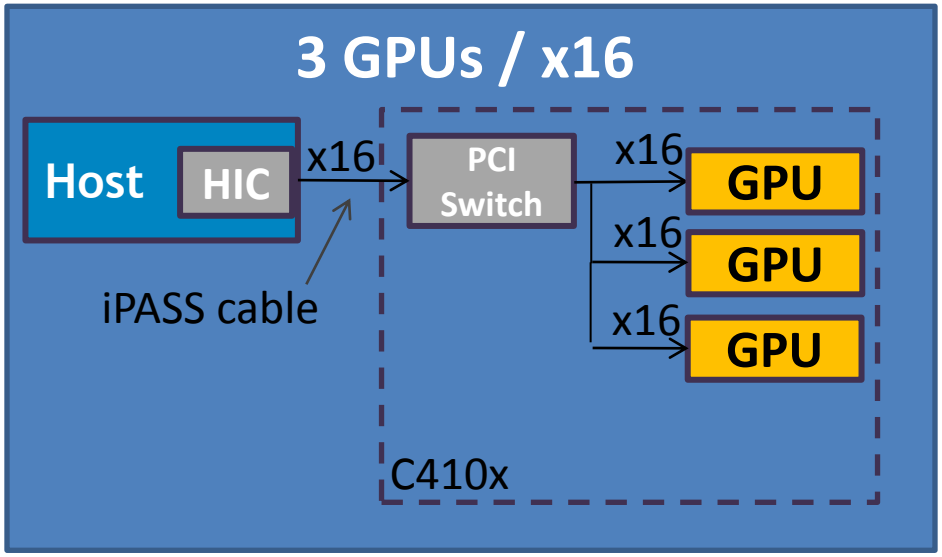


C410x Flexibility: One Host with a Single x16

1:1, 2:1, 3:1, 4:1, 8:1

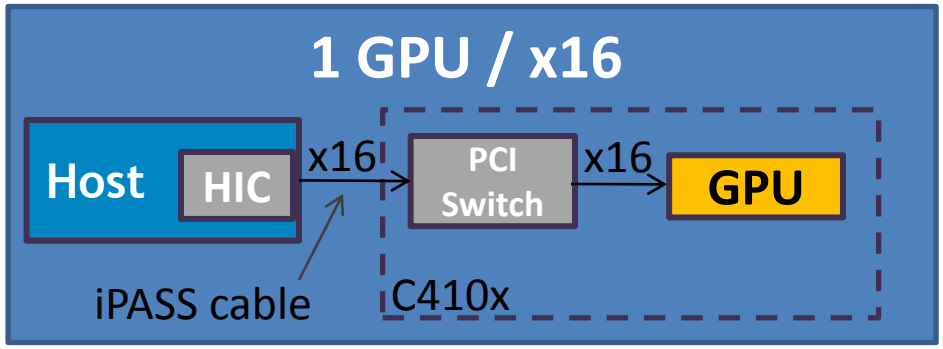


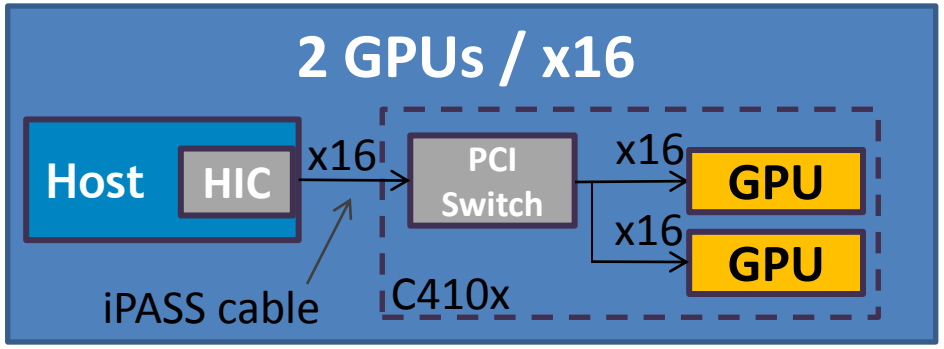


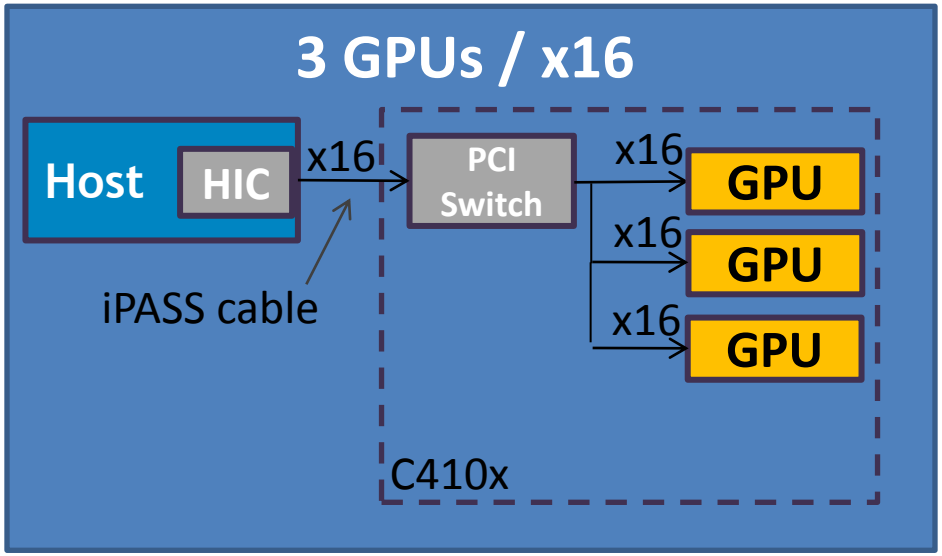


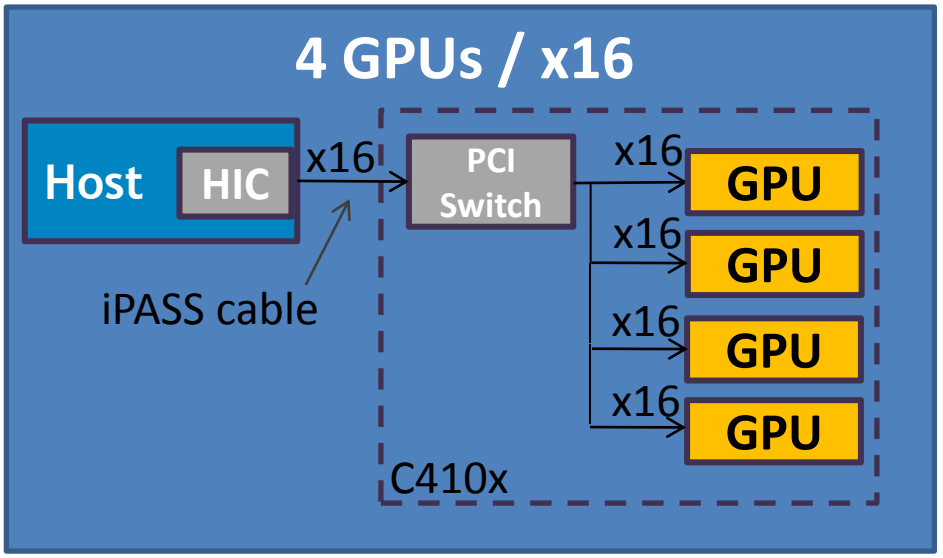
C410x Flexibility: One Host with a Single x16

1:1, 2:1, 3:1, 4:1, 8:1



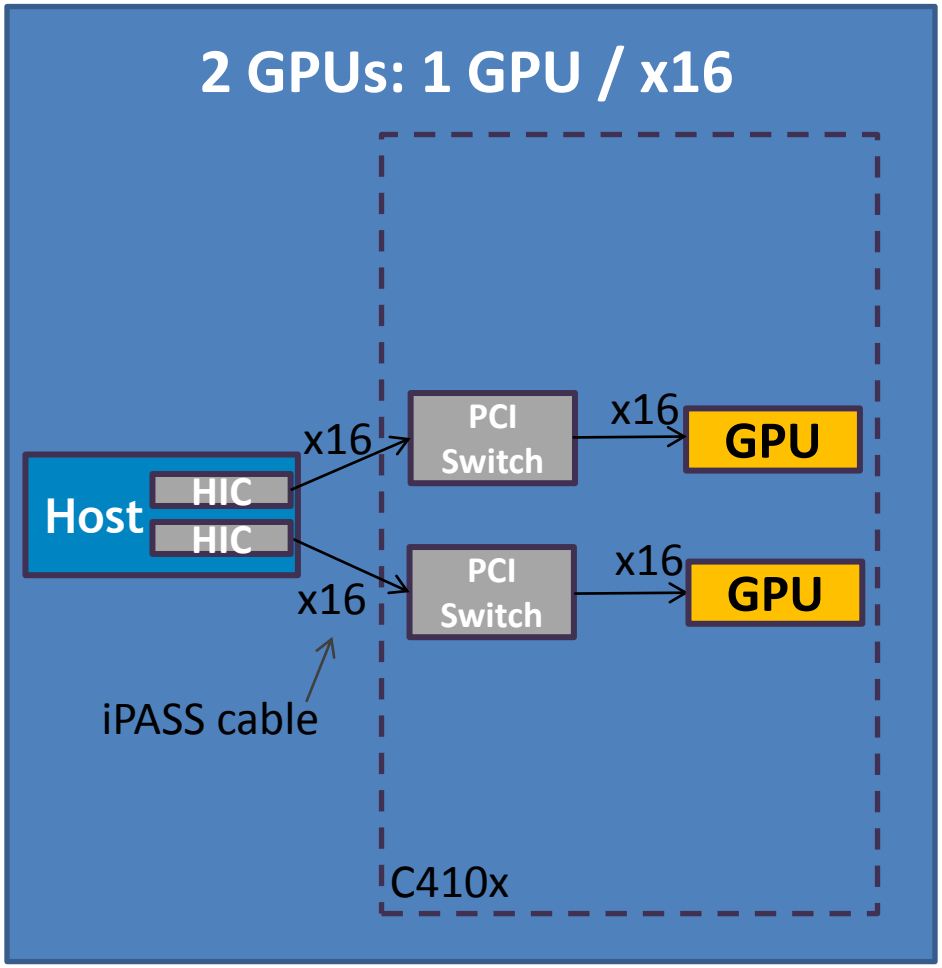






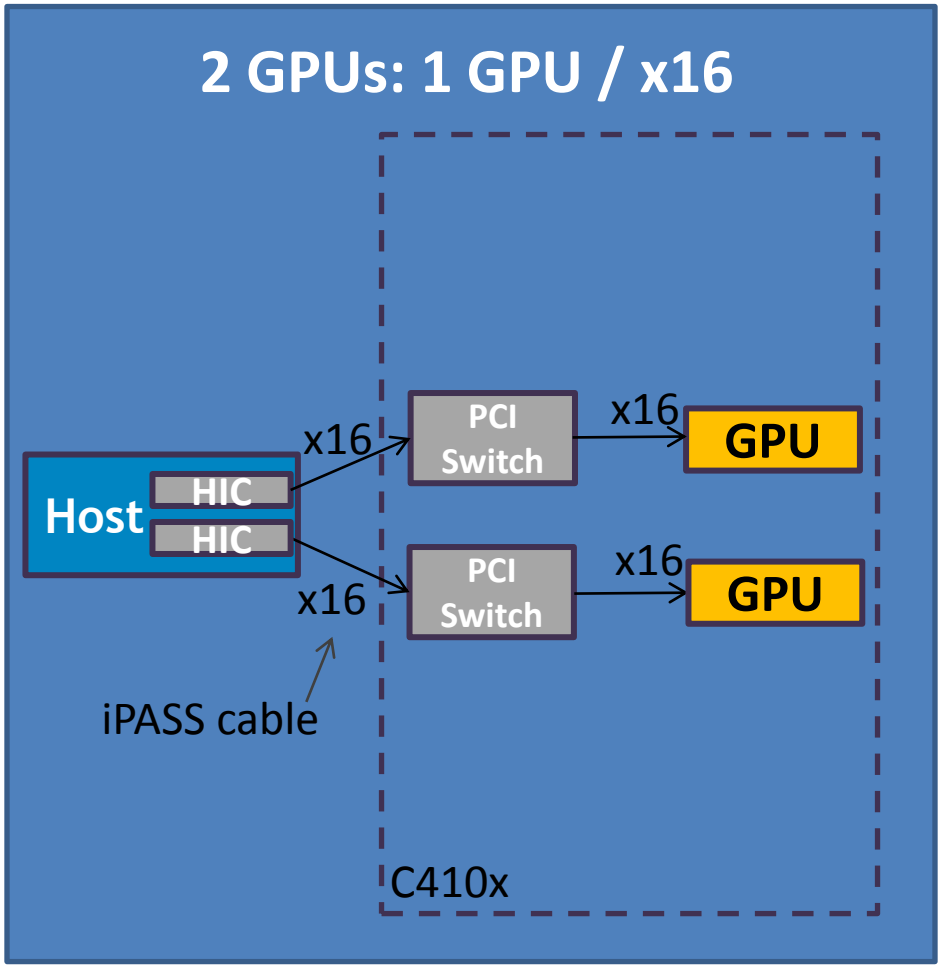
C410x Flexibility: One Host with Dual x16

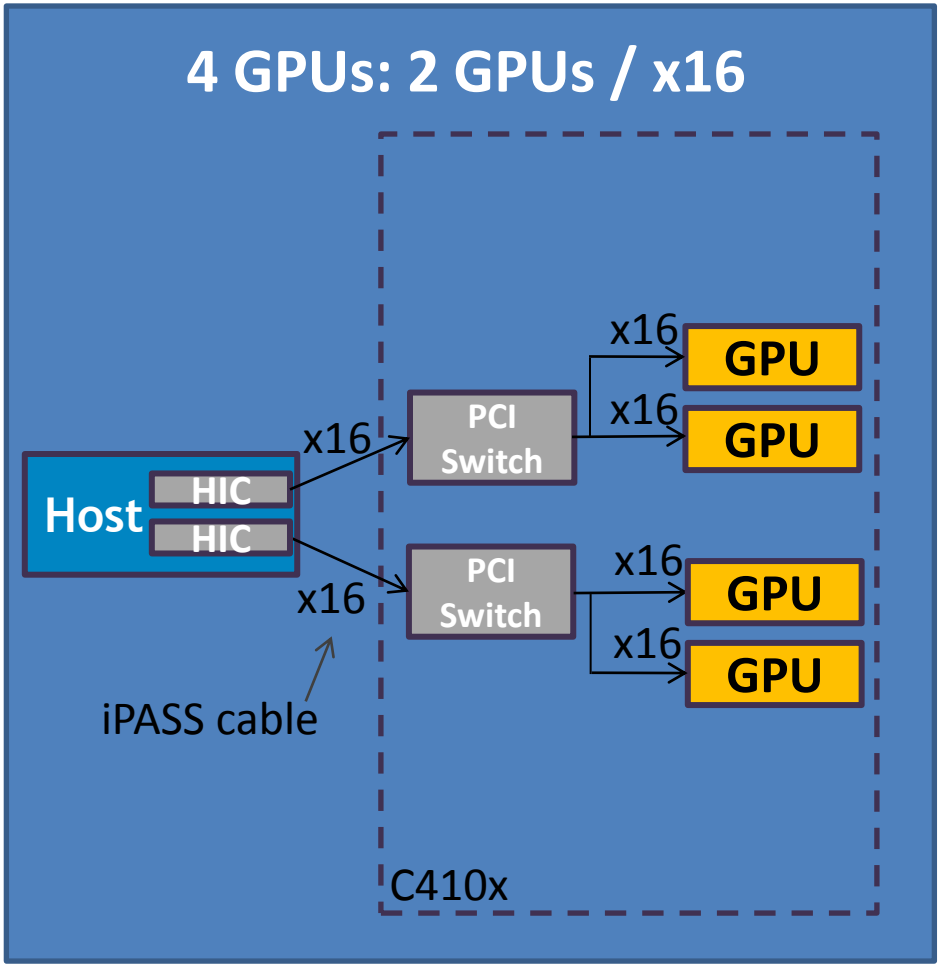
1:1, 2:1, 3:1, 4:1, 8:1



C410x Flexibility: One Host with Dual x16

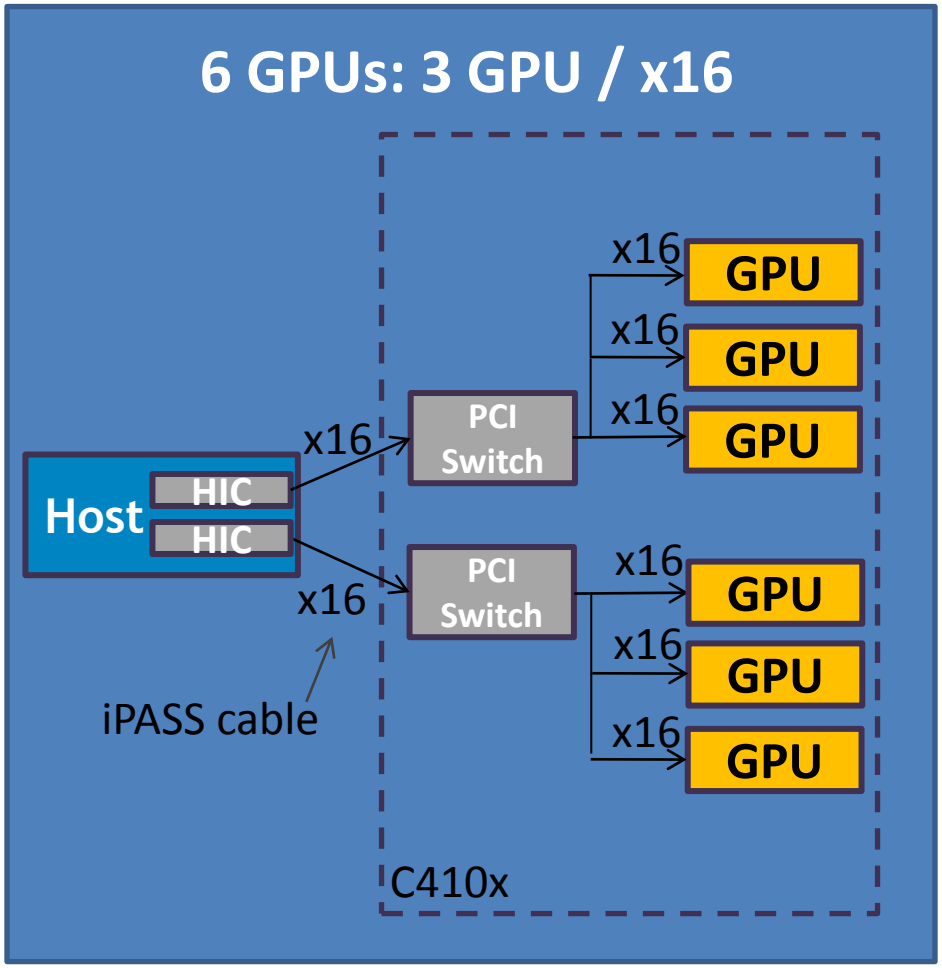
1:1, 2:1, 3:1, 4:1, 8:1





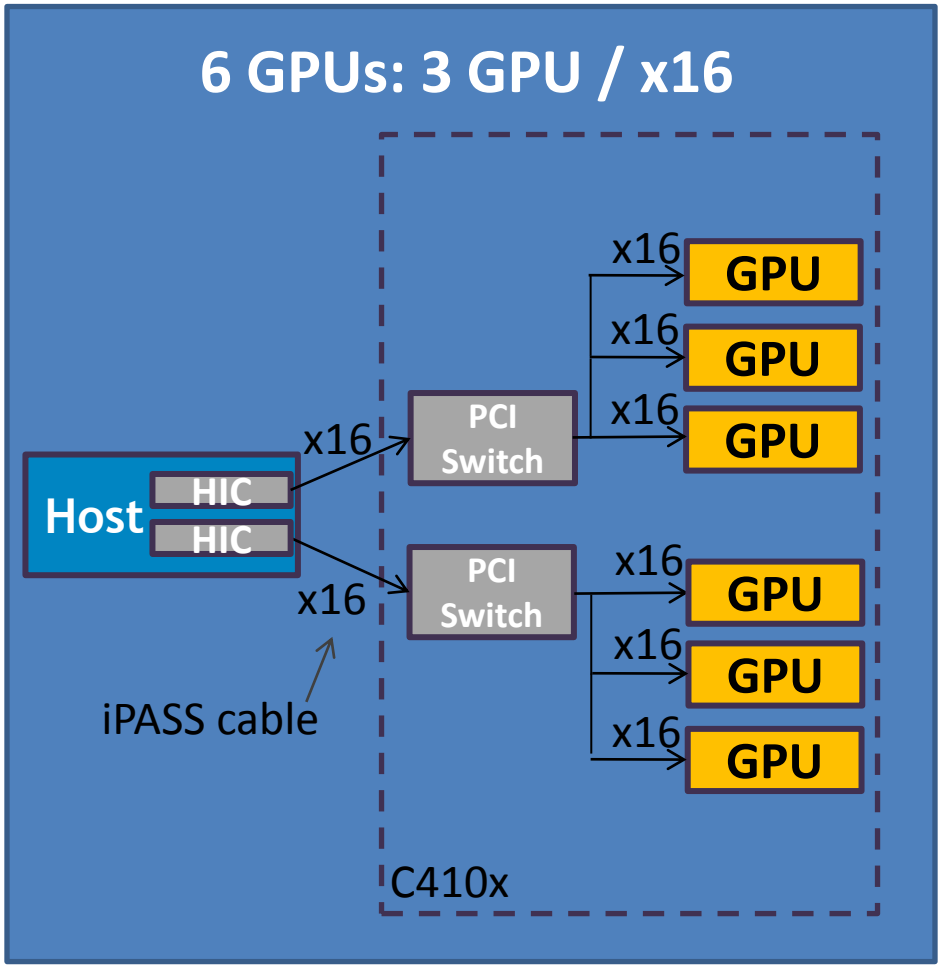
C410x Flexibility: One Host with Dual x16

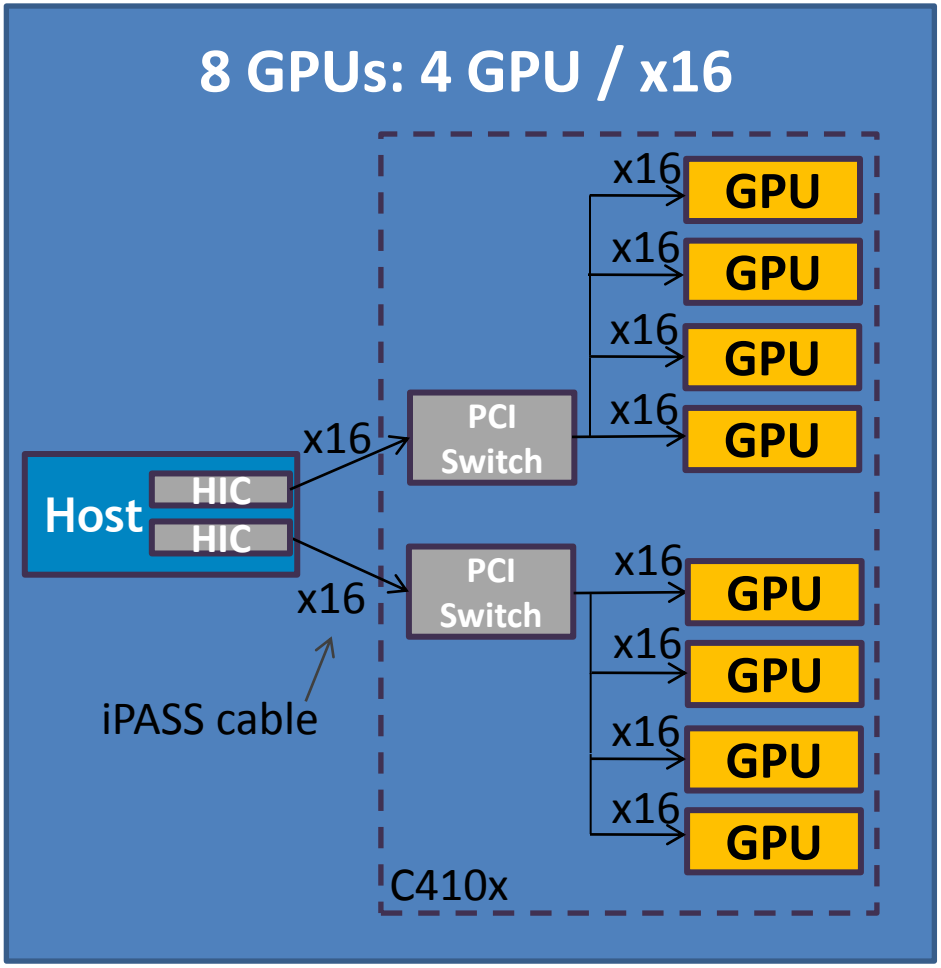
1:1, 2:1, 3:1, 4:1, 8:1



C410x Flexibility: One Host with Dual x16

1:1, 2:1, 3:1, 4:1, 8:1

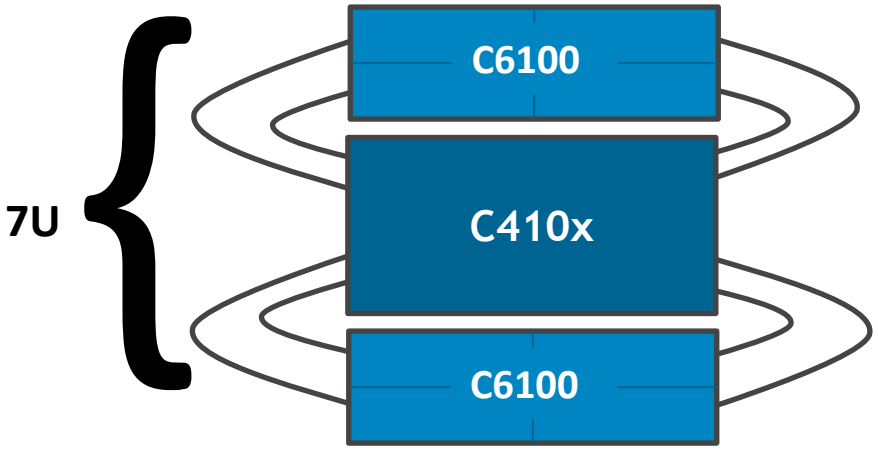


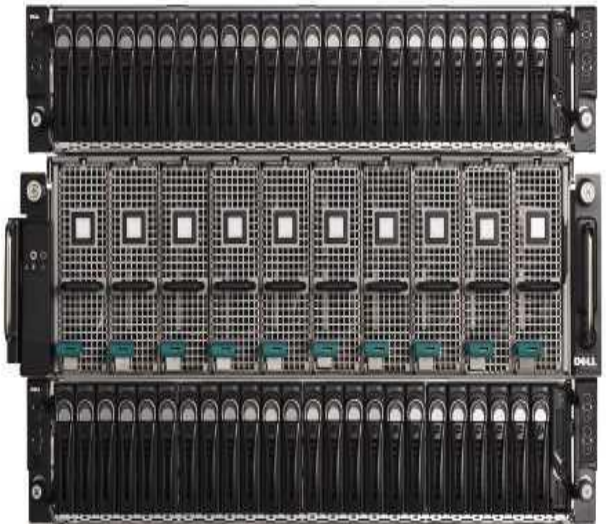


Recommended Host: Dell PowerEdge C6100

- **Four 2-Socket Nodes in 2U**
  - Intel Westmere-EP
- **Each Node:**
  - 12 DIMMs each
  - 2 GigE (Intel)
  - 1 Daughter Card (PCIe x8)
    - 10GigE
    - QDR IB
  - One PCIe x16 (half-length, half-height)
  - Optional SAS controller (in-place of IB)
- **Chassis Design:**
  - Hot Plug, Individually Serviceable System Boards / Nodes
  - Up to 12 x 3.5” drives (3 per node)
  - Up to 24 x 2.5” drives (6 per node)
- **N+1 Power supplies (1100W or 1400W)**
- **NVIDIA HIC certified**
- **DDR and QDR IB PCIe card certified**

C410x “Sandwich”





8-Card C410x Sandwich				16-Card C410x Sandwich
2 x C6100				2 x C6100
8 GPUs				16 GPUs
1 – QDR IB daughtercard				1 – QDR IB daughtercard
7U total				7U total
8 GPUs total				16 GPUs total
8 nodes total				8 nodes total
8/7 nodes / U				8/7 node / U
8/7 GPUs per U				16/7 GPUs per U
1 GPU per PCIe x16				2 GPUs per PCIe x16

Dell PowerEdge C410x:

- Increased density (more GPUs per RackU)
- Introduced “flexibility”
  - GPU/Host ratio = 1:1, 2:1, 3:1, 4:1, …, 8:1
- Purposely separate the Host from the GPUs
- Purpose-built to power, cool and manage PCI-e devices
  - (N+1) Power (3+1 “Gold” power supplies)
  - (N+1) Cooling (7+1 fans)
  - Onboard BMC Web interface to monitor, manage & configure
  - Each PCI-e Module is individually serviceable
    - -- no un-cabling
    - -- no un-racking
    - - - no opening of compute nodes
    - - - no bumped DIMMs
    - - - no disturbed dust
    - - - vertical insertion

Performance Testing

Performance Testing

- Internal vs. External Performance
  - Direct comparison between internal and external GPUs
  - Identical configurations
  - Identical applications

- GPU:Host Scaling
  - Examine GPU performance improvements (vs. CPUs)
  - Examine performance scaling with GPU count

Application Workloads

Workloads

- NAMD v2.7b2 (STMV benchmark)
  - a parallel, object-oriented molecular dynamics code designed for high-performance simulation of large biomolecular systems.
- LAMMPS (15 Jan 2010 Version) (LJ-Cut & Gay-Berne)
  - a classical molecular dynamics package written for parallel machines.
- CUDASW++ v2.0
  - Bio-informatics software for Smith-Waterman protein database searches
- GPU-HMMER - v0.92
  - Bioinformatics software that does protein sequence alignment using profile HMMs
  - Running a search of uniprot_sprot file for different profile HMM lengths
- HOOMD-blue – v0.8.2 Python-2.4
  - General purpose particle dynamics package
  - Running included LJ and Polymer benchmarks

System Configurations

- SuperMicro
  (Internal 2-x16)

  - CPU:
    - 2 x Intel Xeon
    - E5520 2.26 GHz
  - Memory:
    - 6 x 4GB 1333MHz DIMMs
  - GPU:
    - 2 x NVIDIA Tesla M1060

  - GPU Driver:
    - 3.0 linux-64 195.36.15
  - OS:
    - RHEL 5.3
  - CUDA
    - Toolkit: 3.0 linux-64 RHEL 5.3
    - SDK: GPU Computing 3.0 linux

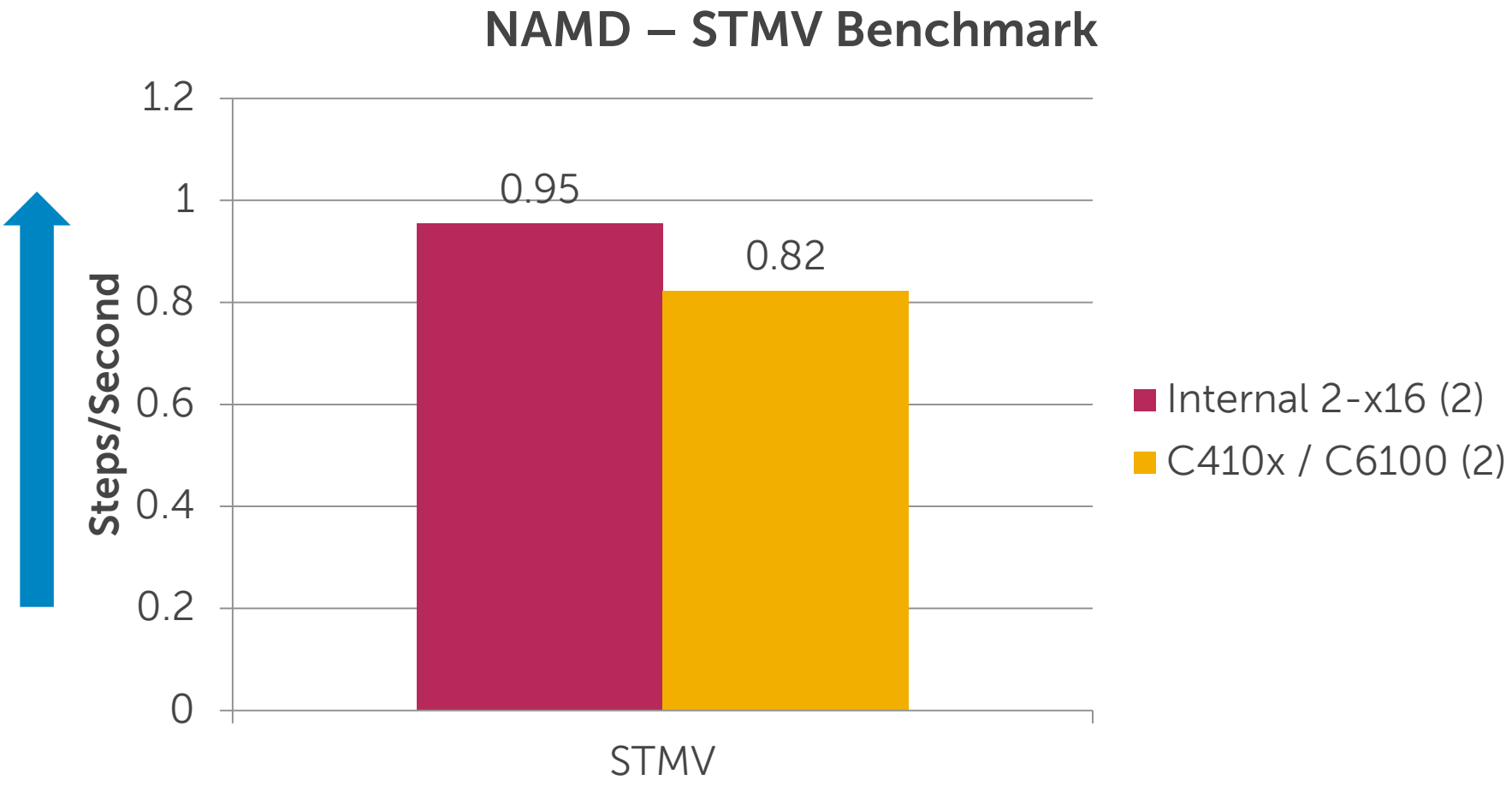
- C410x / C6100
  (External 1-x16)

  ✓ • CPU:
    - 2 x Intel Xeon
    - E5520 2.26 GHz
  ✓ • Memory:
    - 6 x 4GB 1333MHz DIMMs
  ✓ • GPU:
    - 2 x NVIDIA Tesla M1060

  ✓ • GPU Driver:
    - 3.0 linux-64 195.36.15
  ✓ • OS:
    - RHEL 5.5
  ✓ • CUDA
    - Toolkit: 3.0 linux-64 RHEL 5.3
    - SDK: GPU Computing 3.0 linux

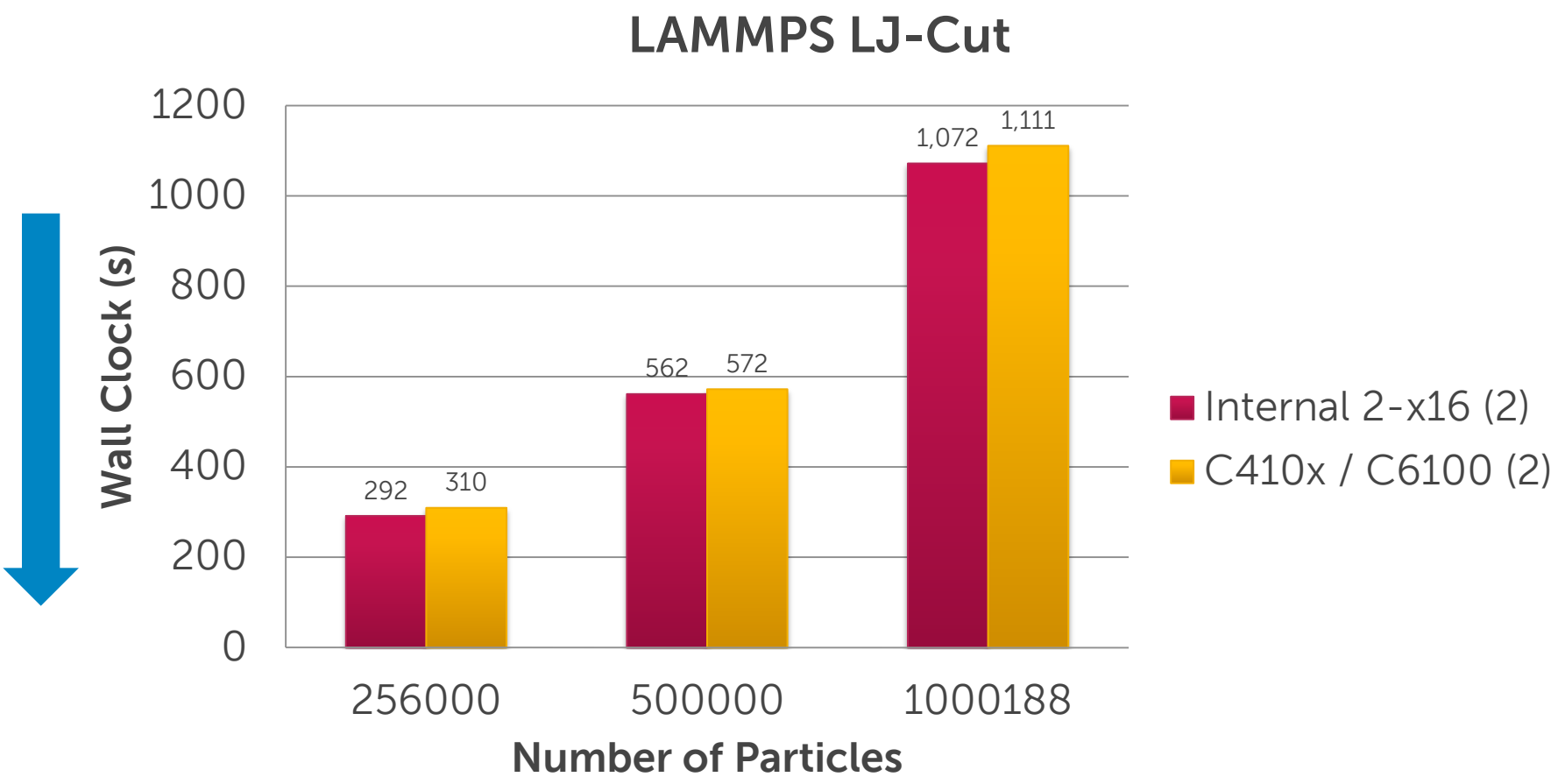
Internal vs. External x16 Slots

Internal vs. External: NAMD



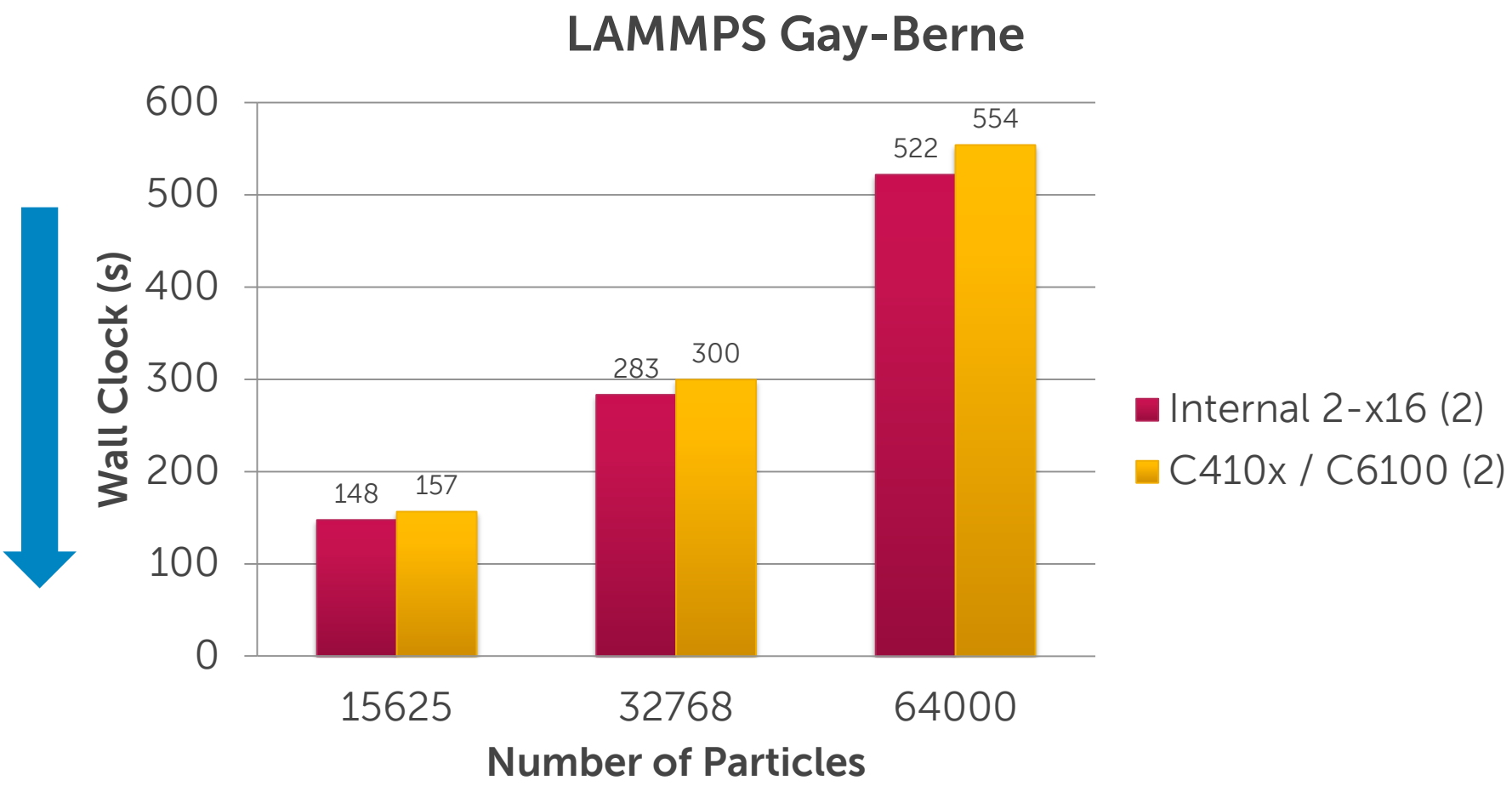
13.8% performance difference.

Internal vs. External: LAMMPS LJ-Cut



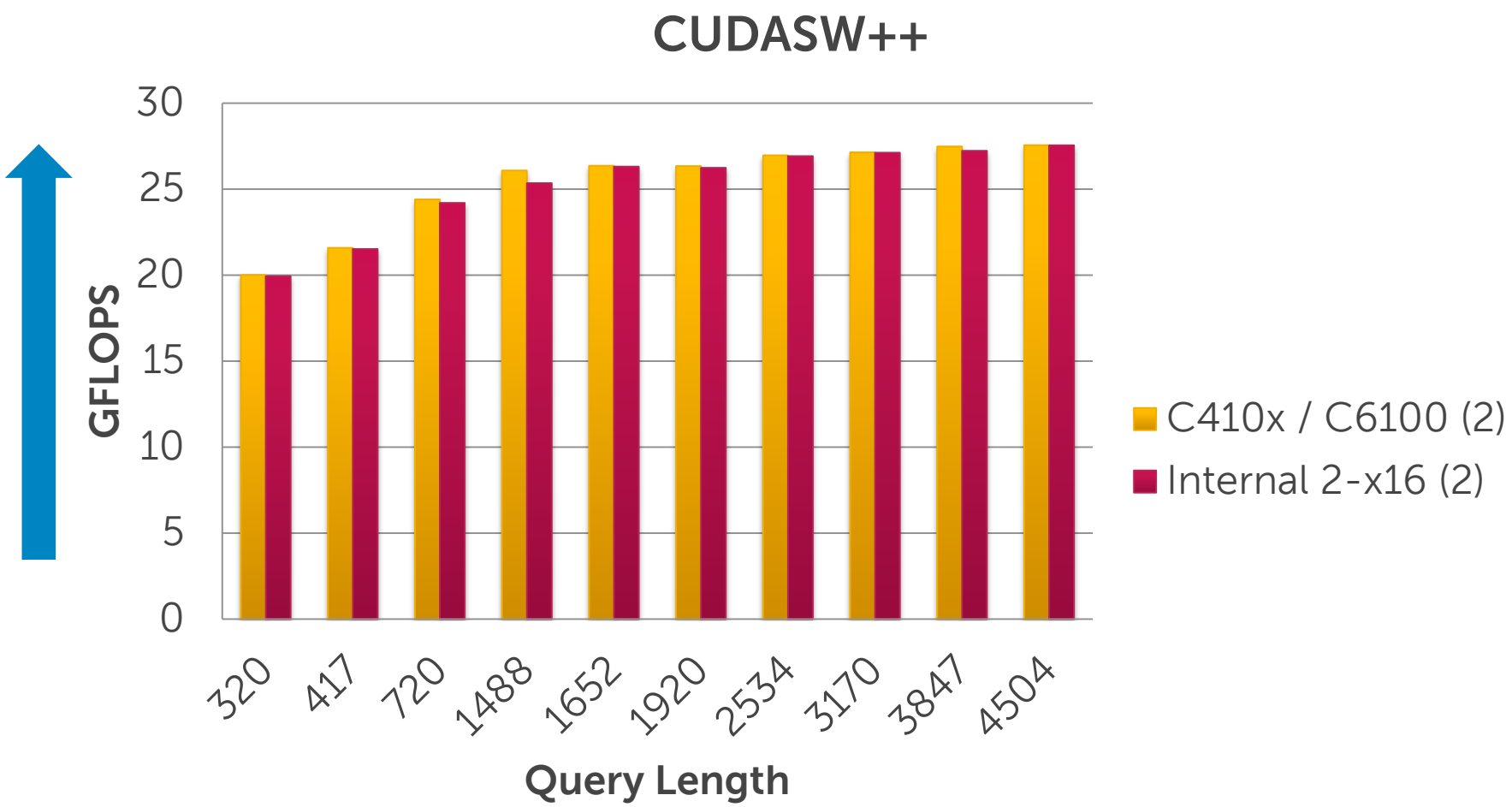
1.9 to 6.1% performance difference.

Internal vs. External: LAMMPS Gay-Berne



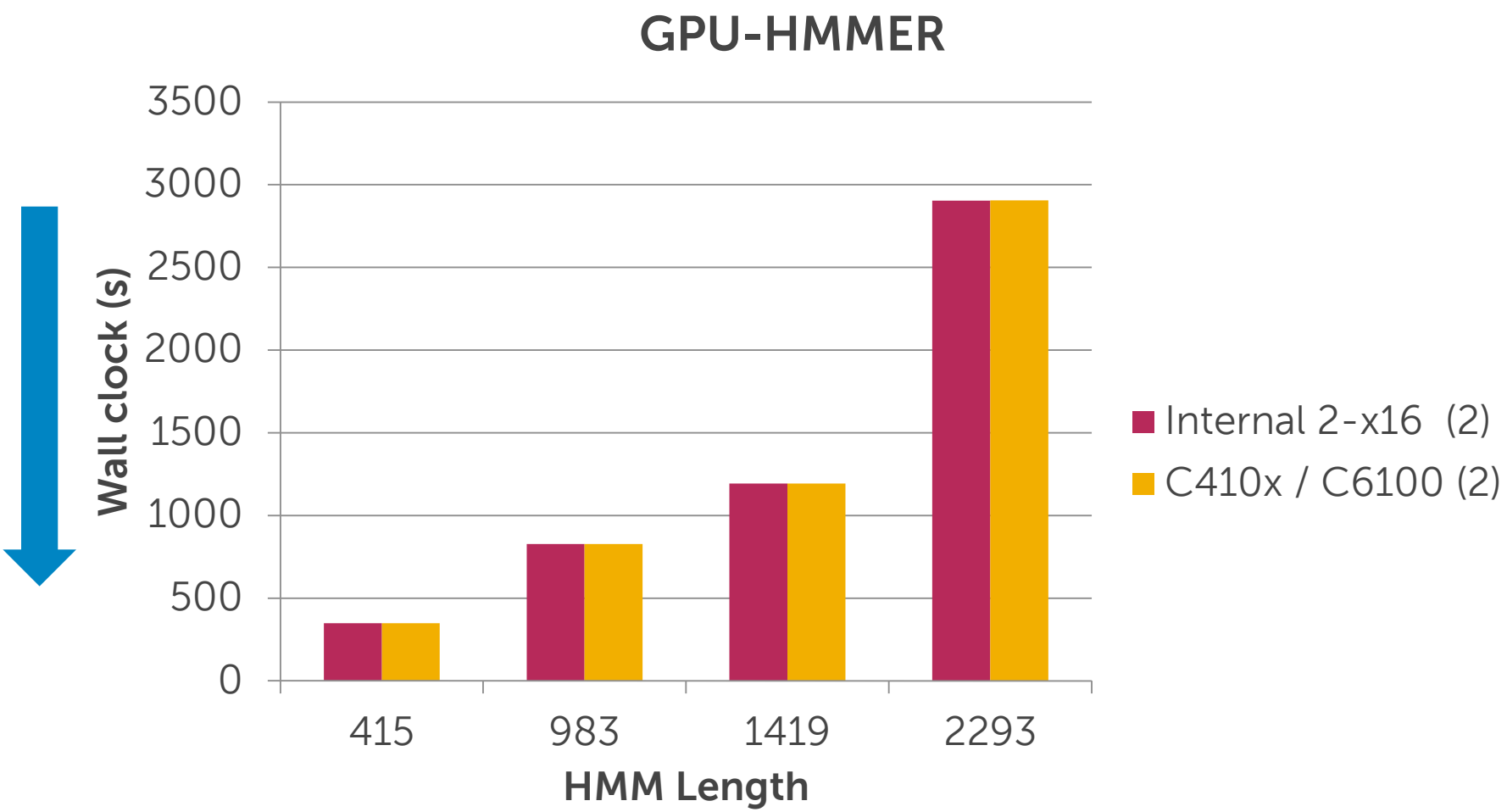
5.7 to 6.2% performance difference.

Internal vs. External: CUDASW++



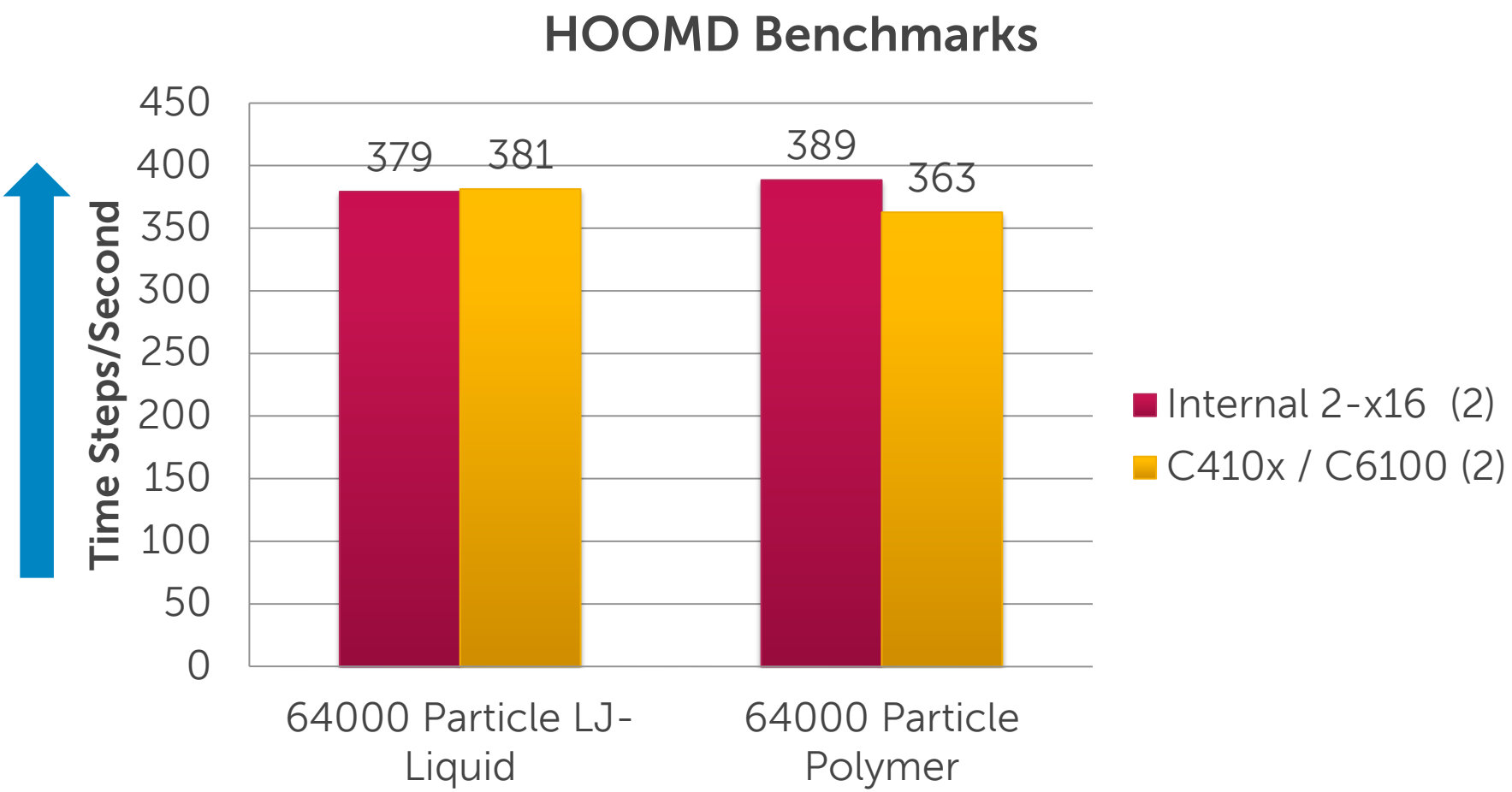
0.03 to 2.9% performance difference.

Internal vs. External: GPU-HMMER



0.04 to 0.23% performance difference.

Internal vs. External: HOOMD



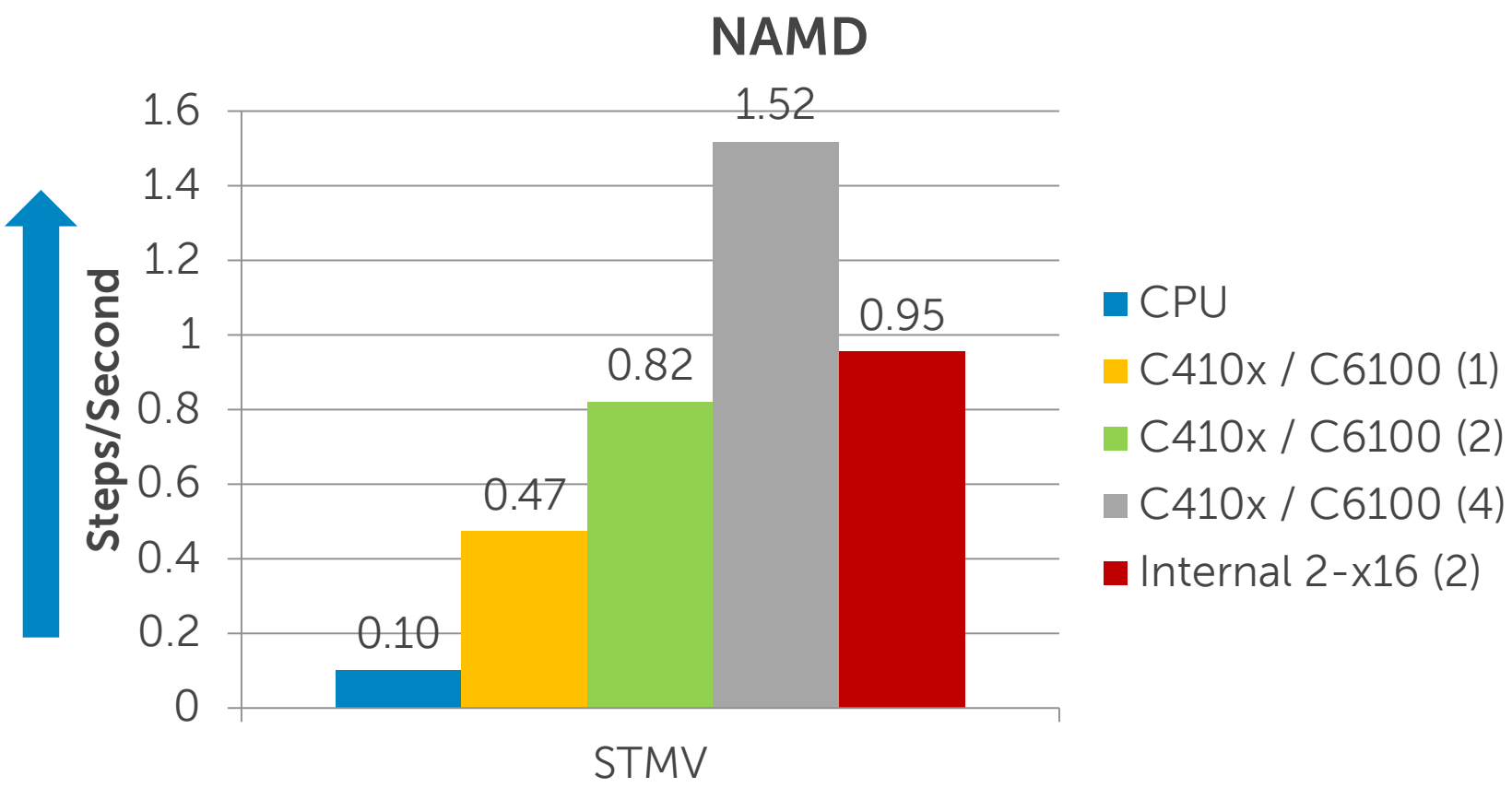
0.5% to 6.6% performance difference.

Results Summary

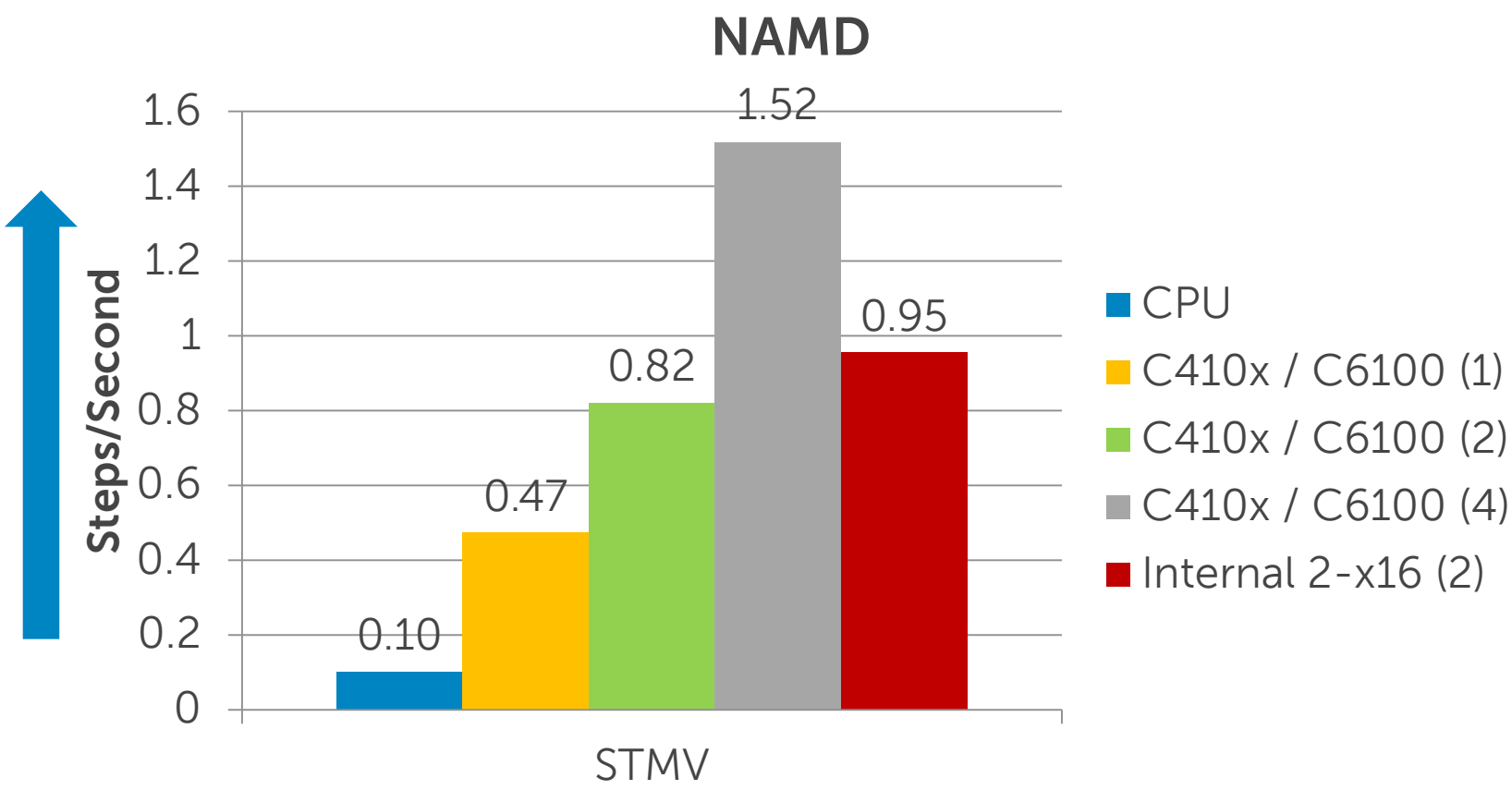
Application	Negatively Affected by External GPU's and/or Shared x16's	Scales Beyond 2 GPU's (4:2)
NAMD	Y (13.8%)	
LAMMPS LJ-CUT	Y (6.1%)	
LAMMPS GB	Y (6.2%)	
CUDASW++	N (0.03 to 2.9%)	
GPU-HAMMER	N (0.04 to 0.23%)	
HOOMD	Y (6.6%)	

GPU:Host Ratio Performance

GPU:Host Scaling: NAMD

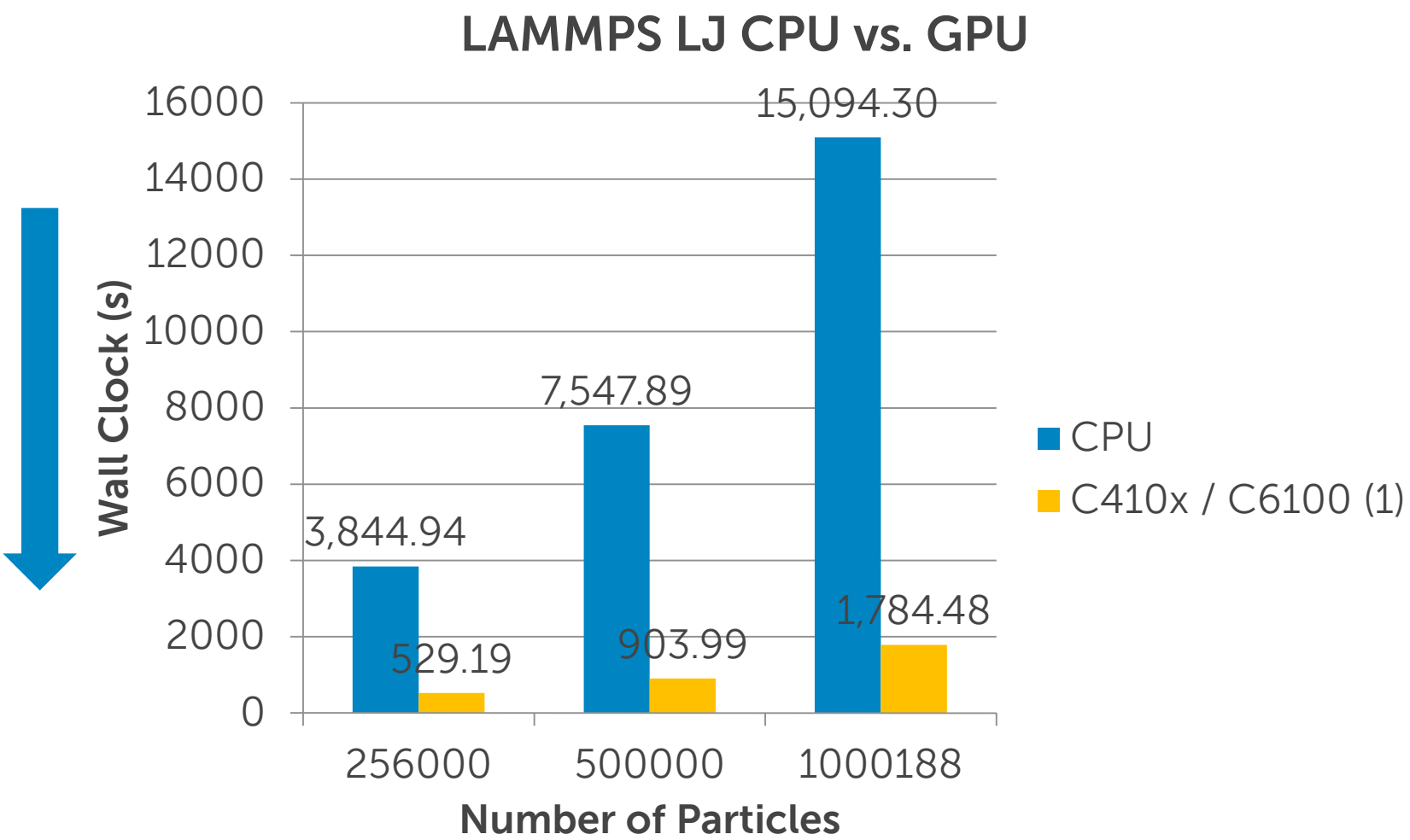


GPU:Host Scaling: NAMD

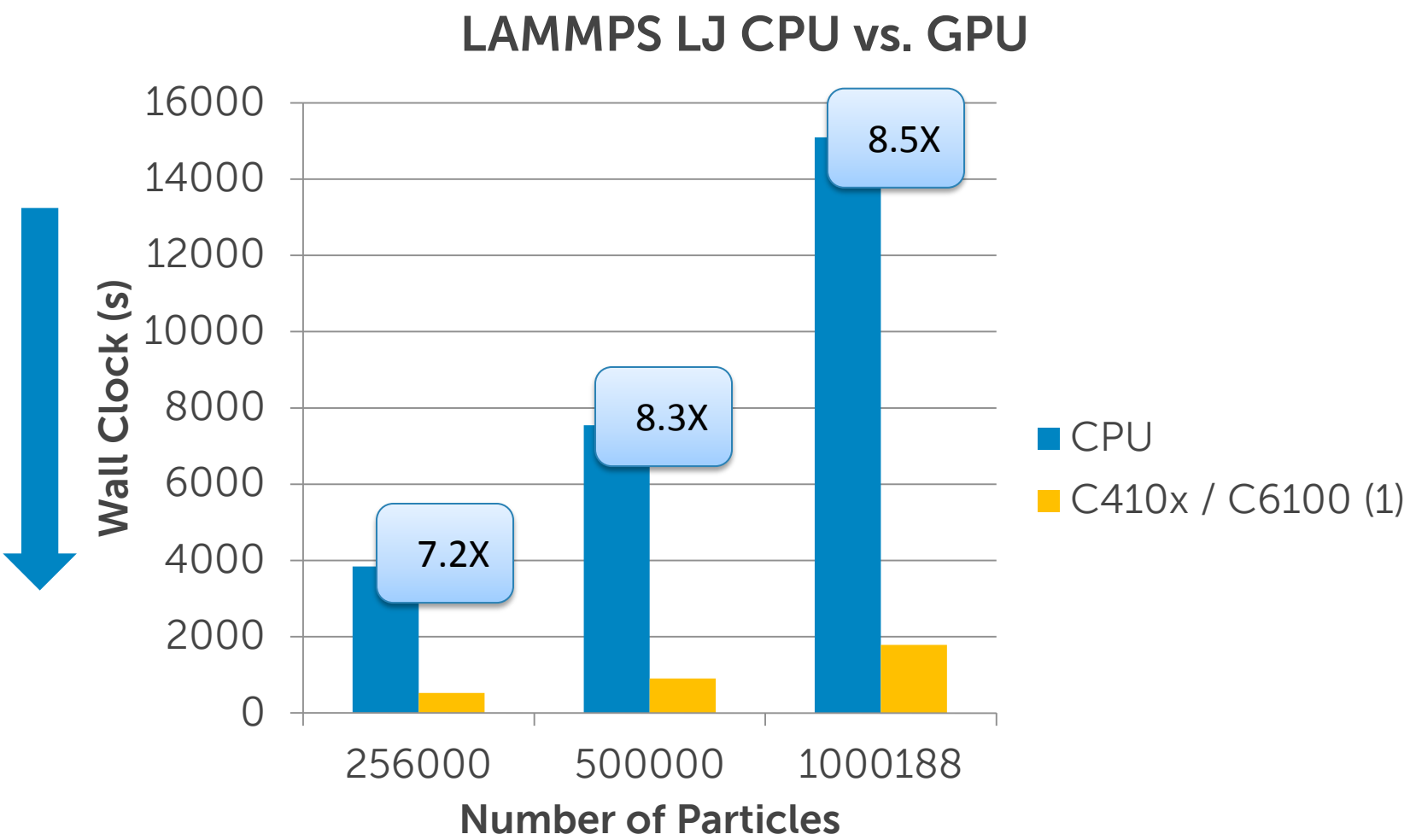


Speedup
4.7X
8.2X
15.2X
9.5X

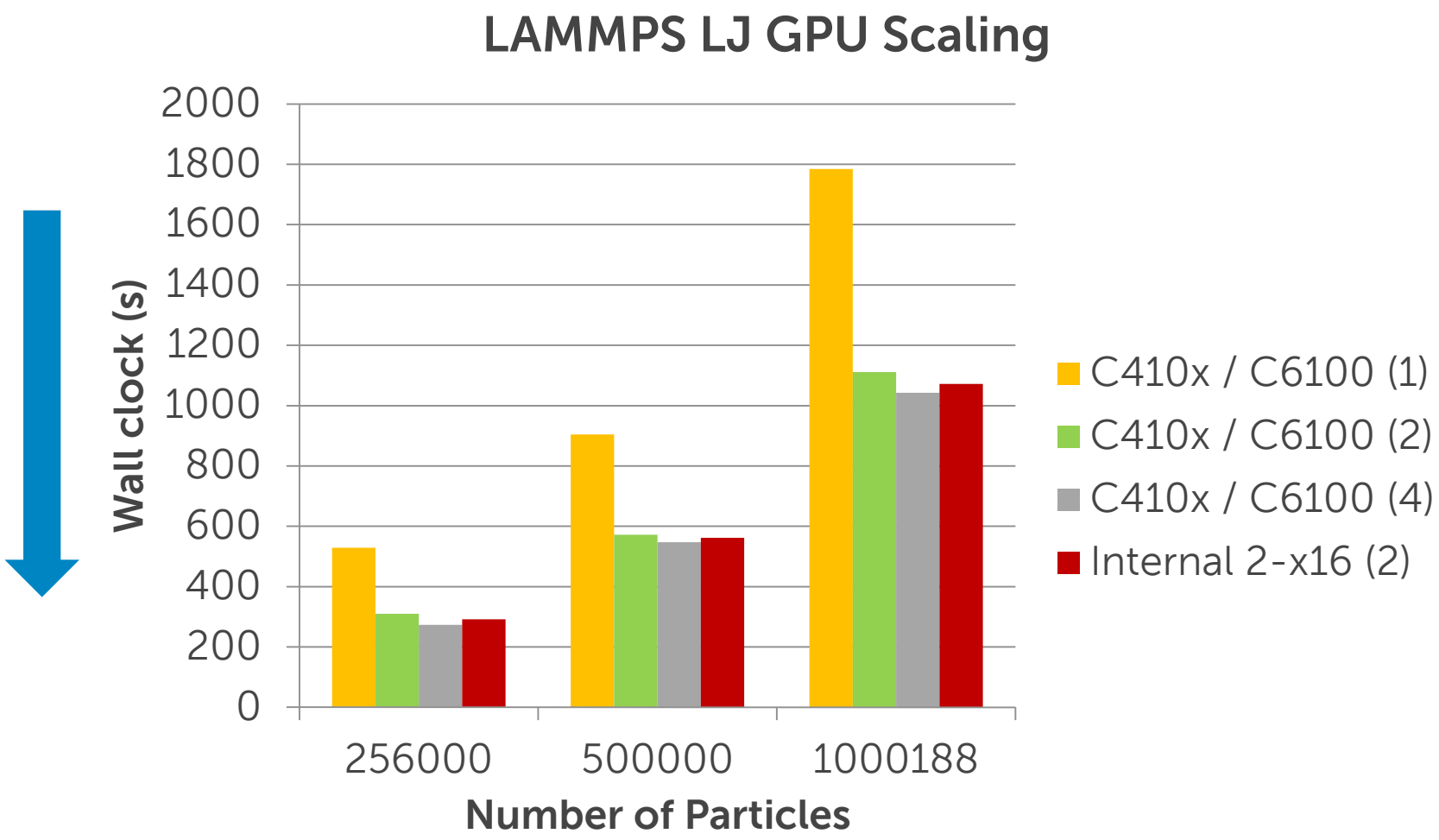
GPU:Host Scaling : LAMMPS LJ-Cut



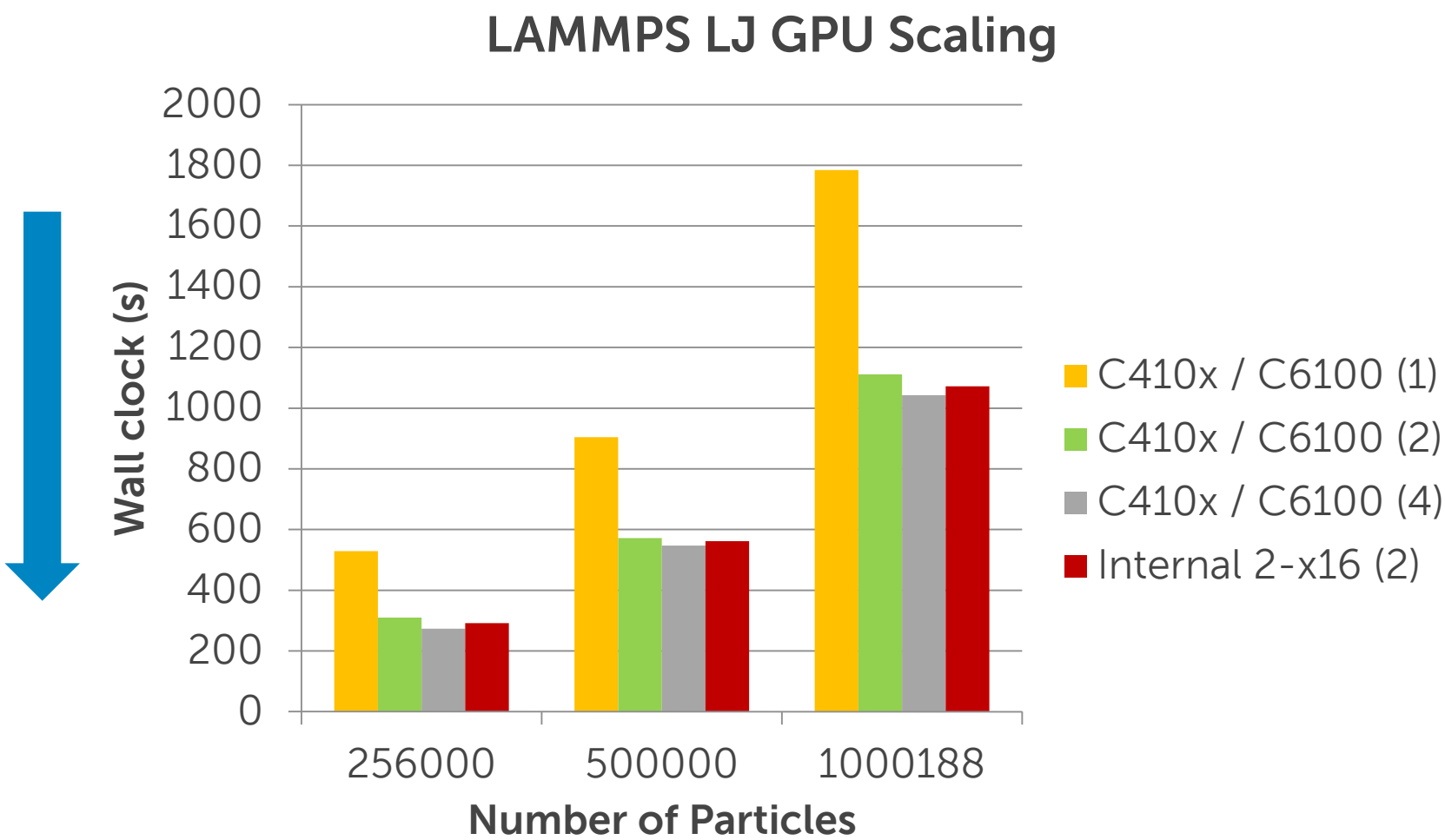
GPU:Host Scaling : LAMMPS LJ-Cut



GPU:Host Scaling : LAMMPS LJ-Cut

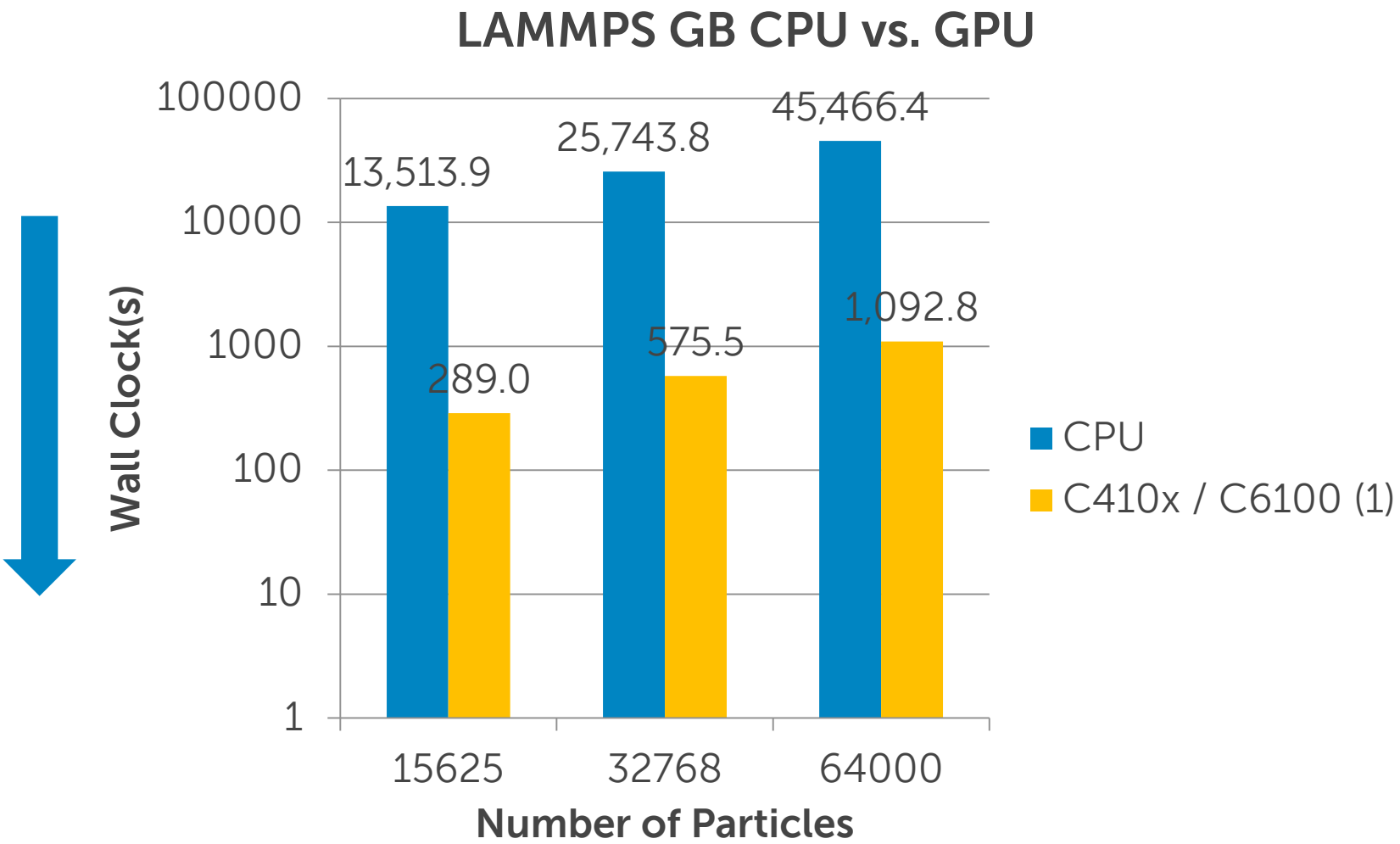


GPU:Host Scaling : LAMMPS LJ-Cut

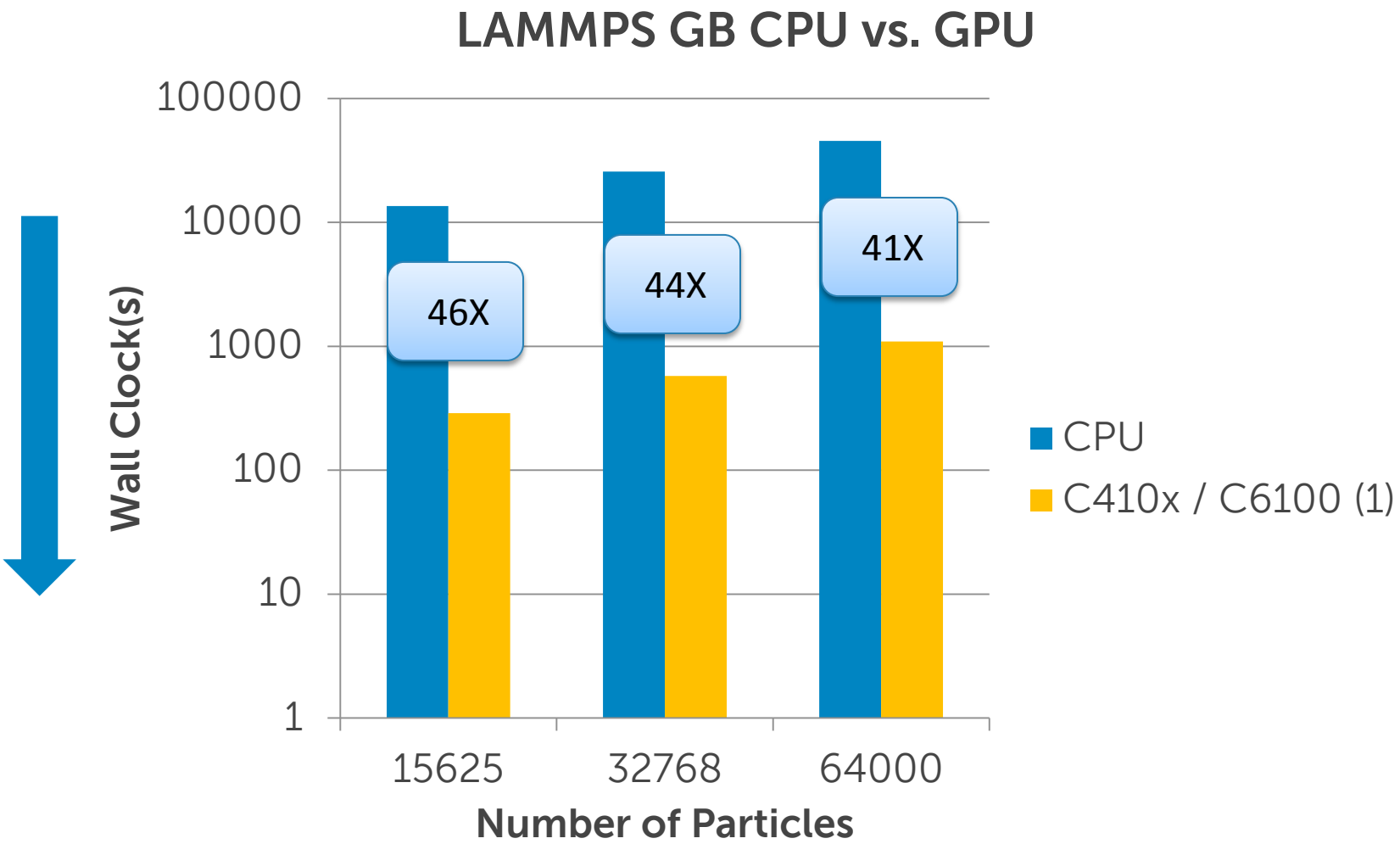


Speedup
8.5X
13.5X
14.4X
14.0X

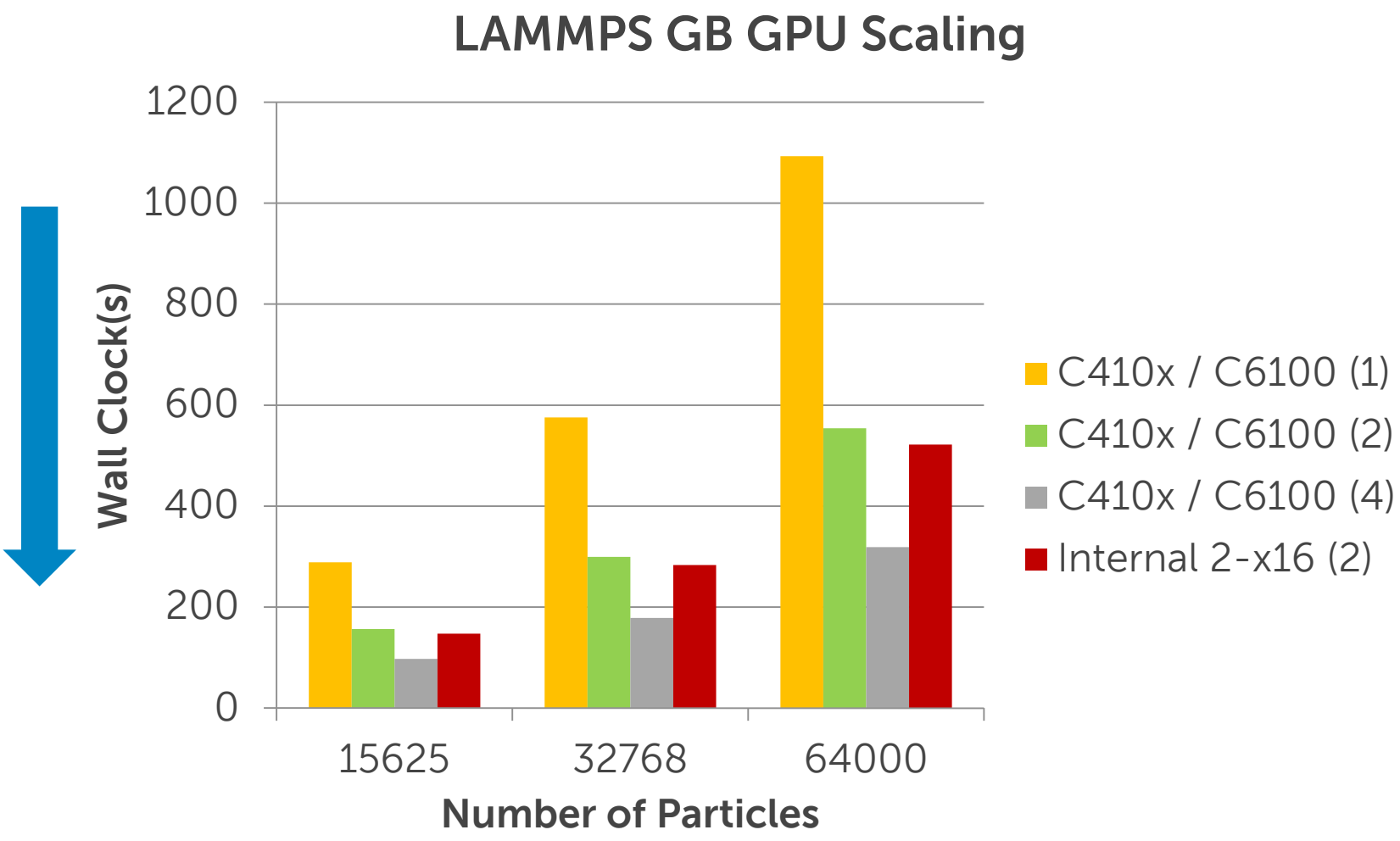
GPU:Host Scaling : LAMMPS Gay-Berne



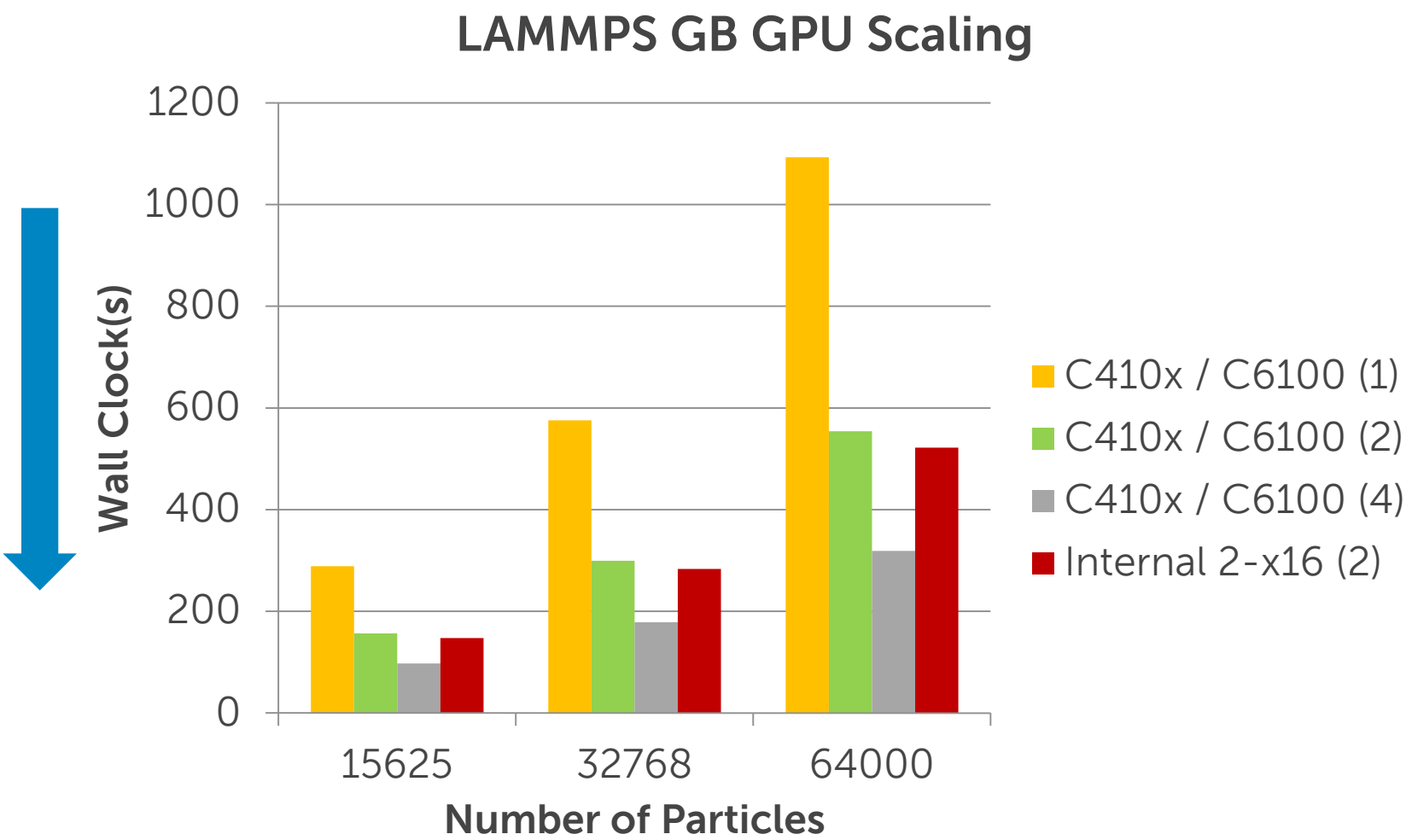
GPU:Host Scaling : LAMMPS Gay-Berne



GPU:Host Scaling : LAMMPS Gay-Berne

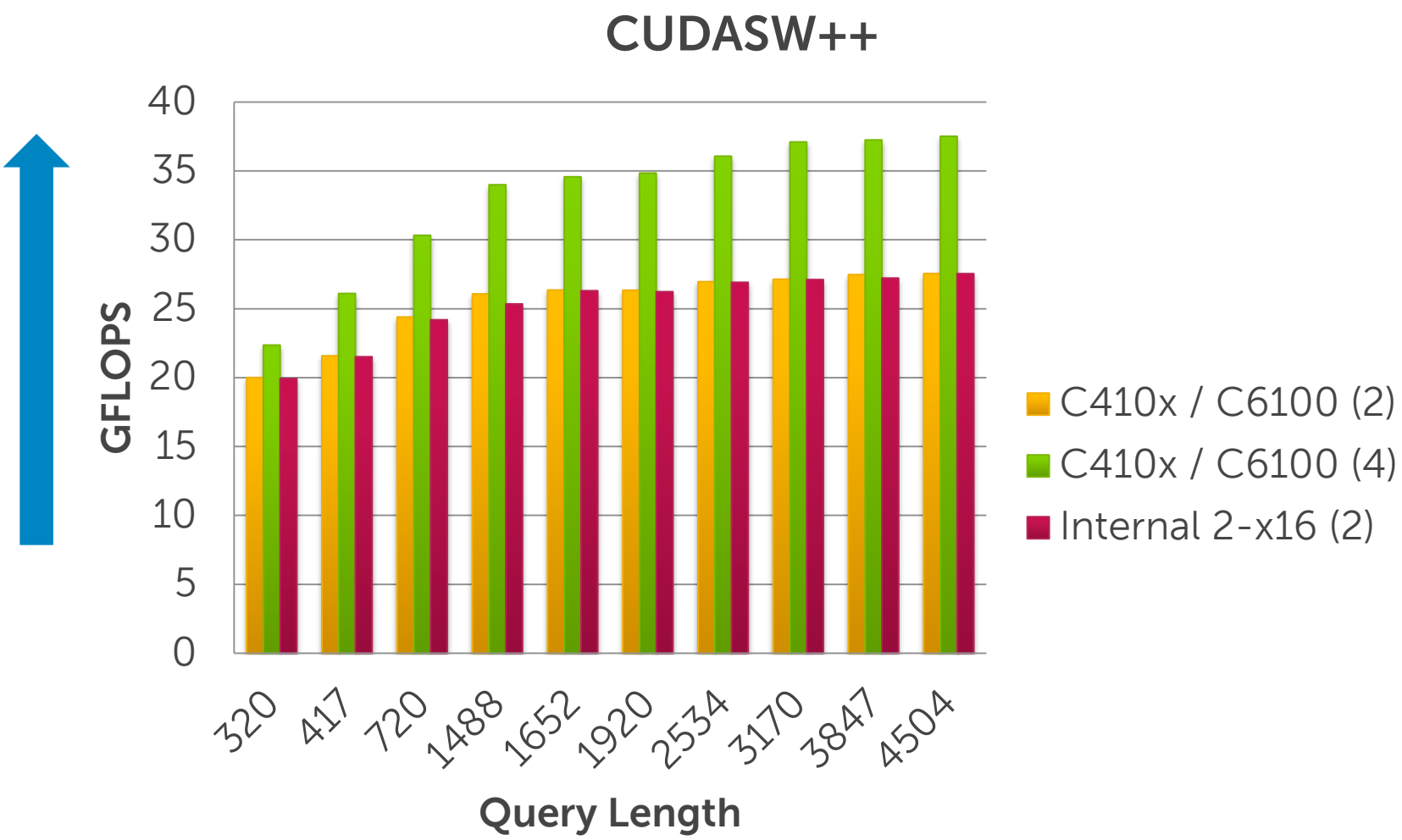


GPU:Host Scaling : LAMMPS Gay-Berne

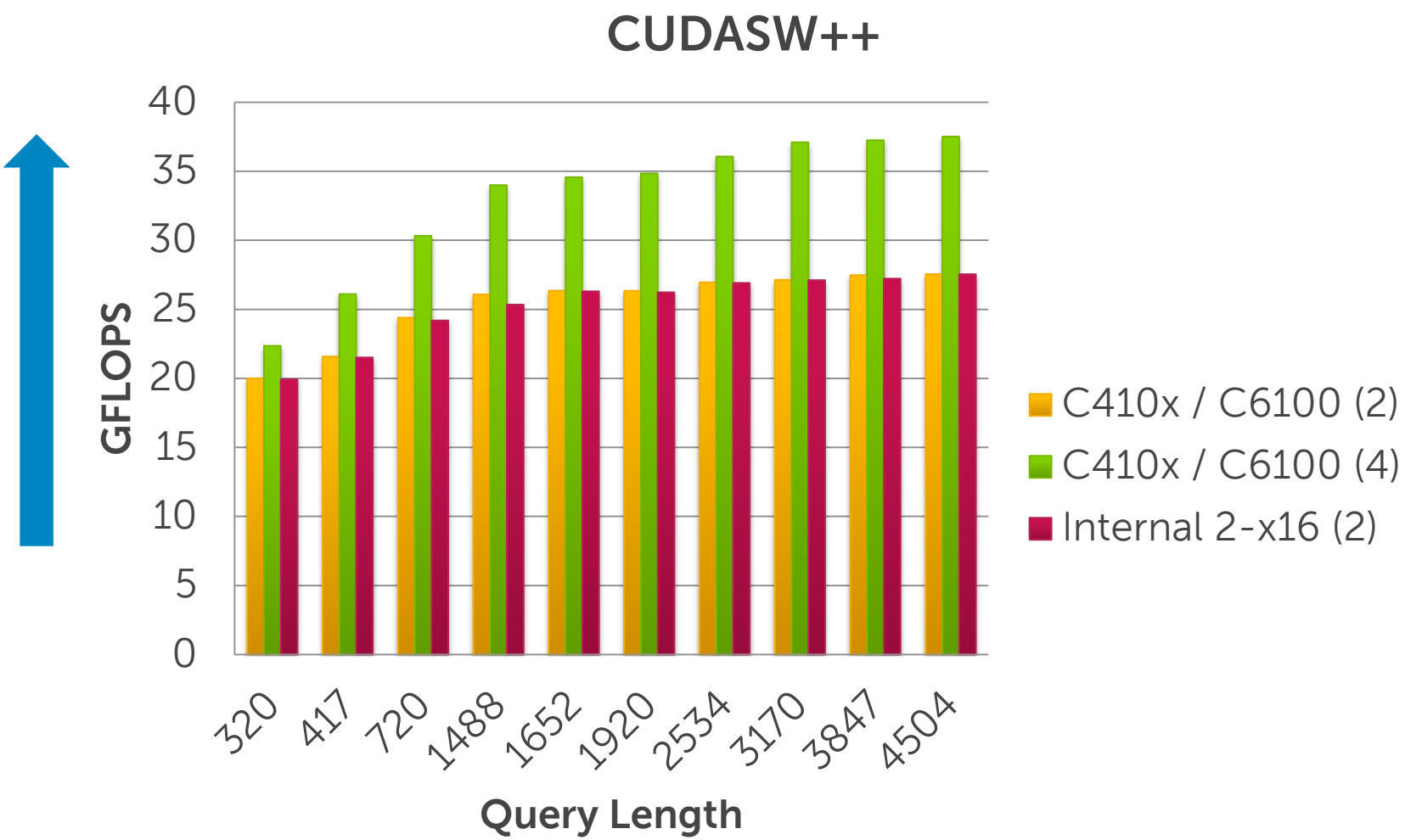


Speedup
41X
82X
142X
87X

GPU:Host Scaling : CUDASW++

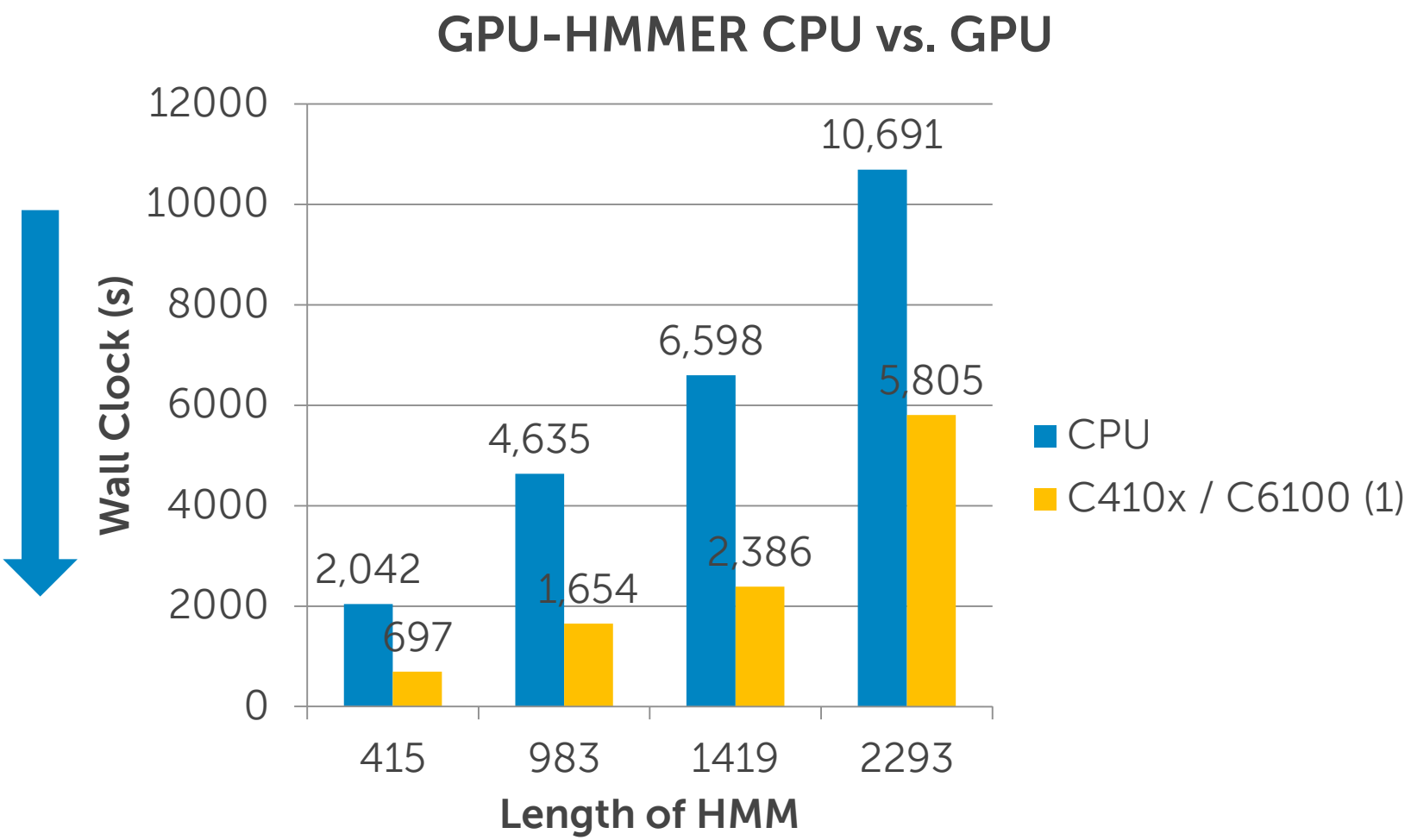


GPU:Host Scaling : CUDASW++

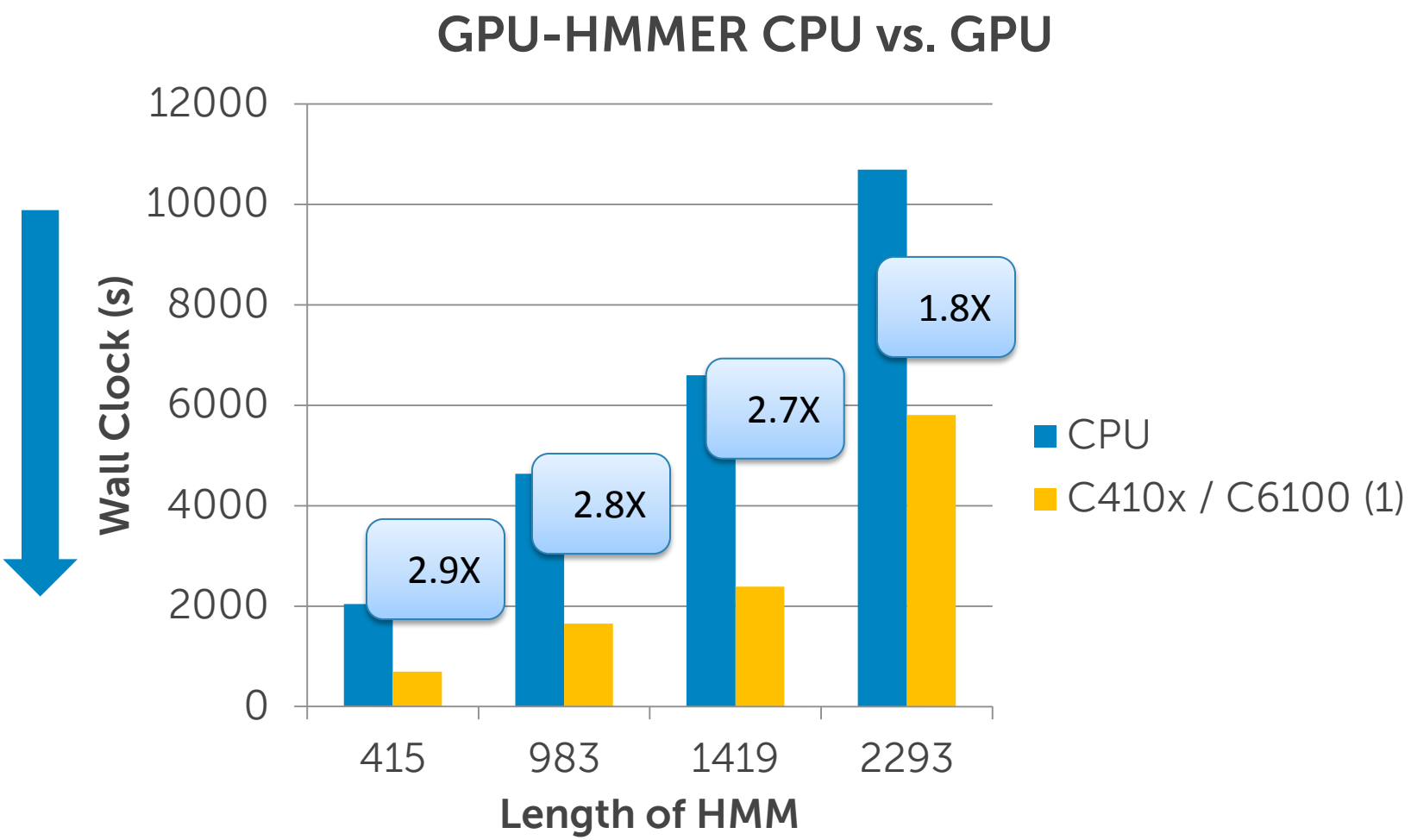


Speedup
1.8X
2.4X
1.8X

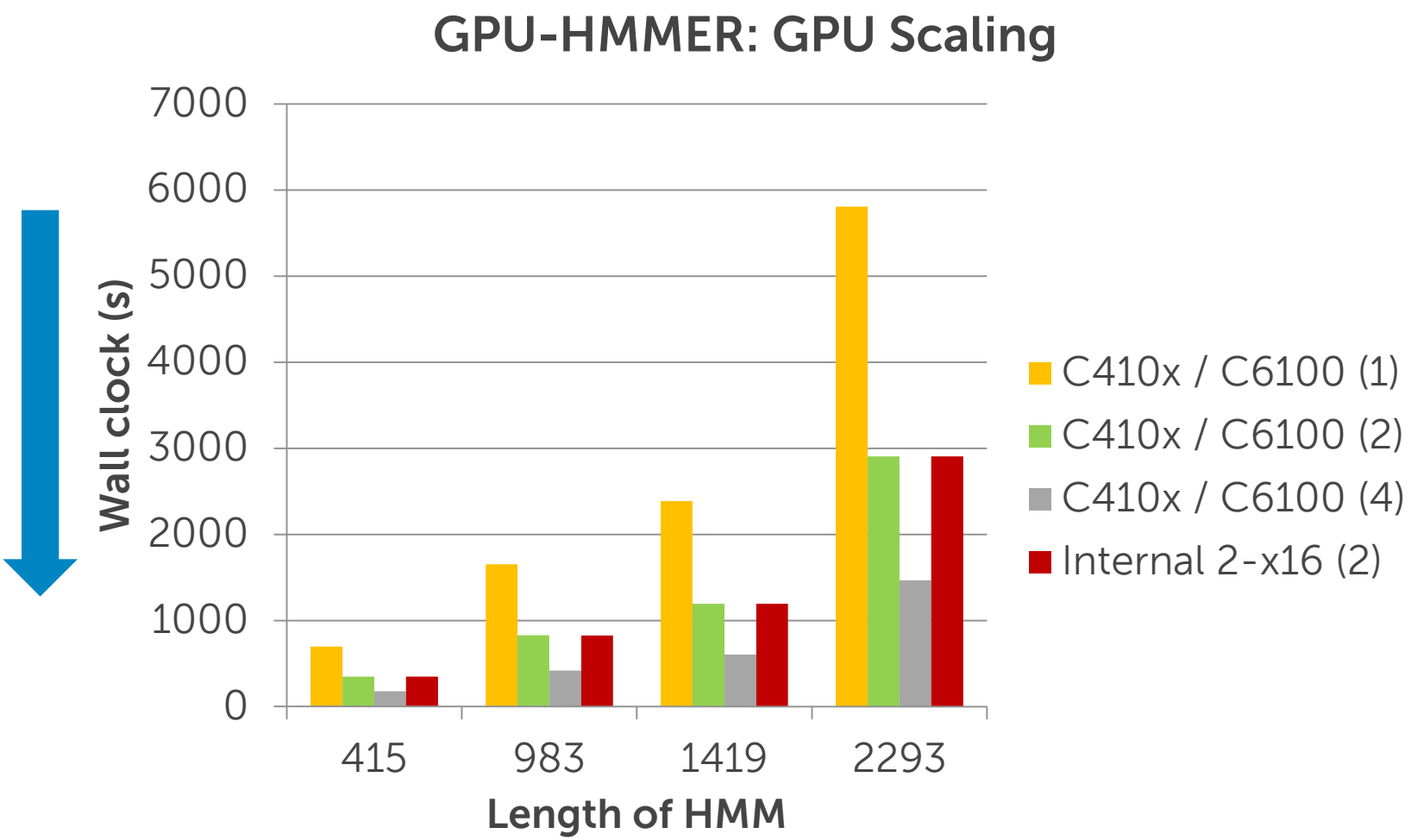
GPU:Host Scaling : GPU-HMMER



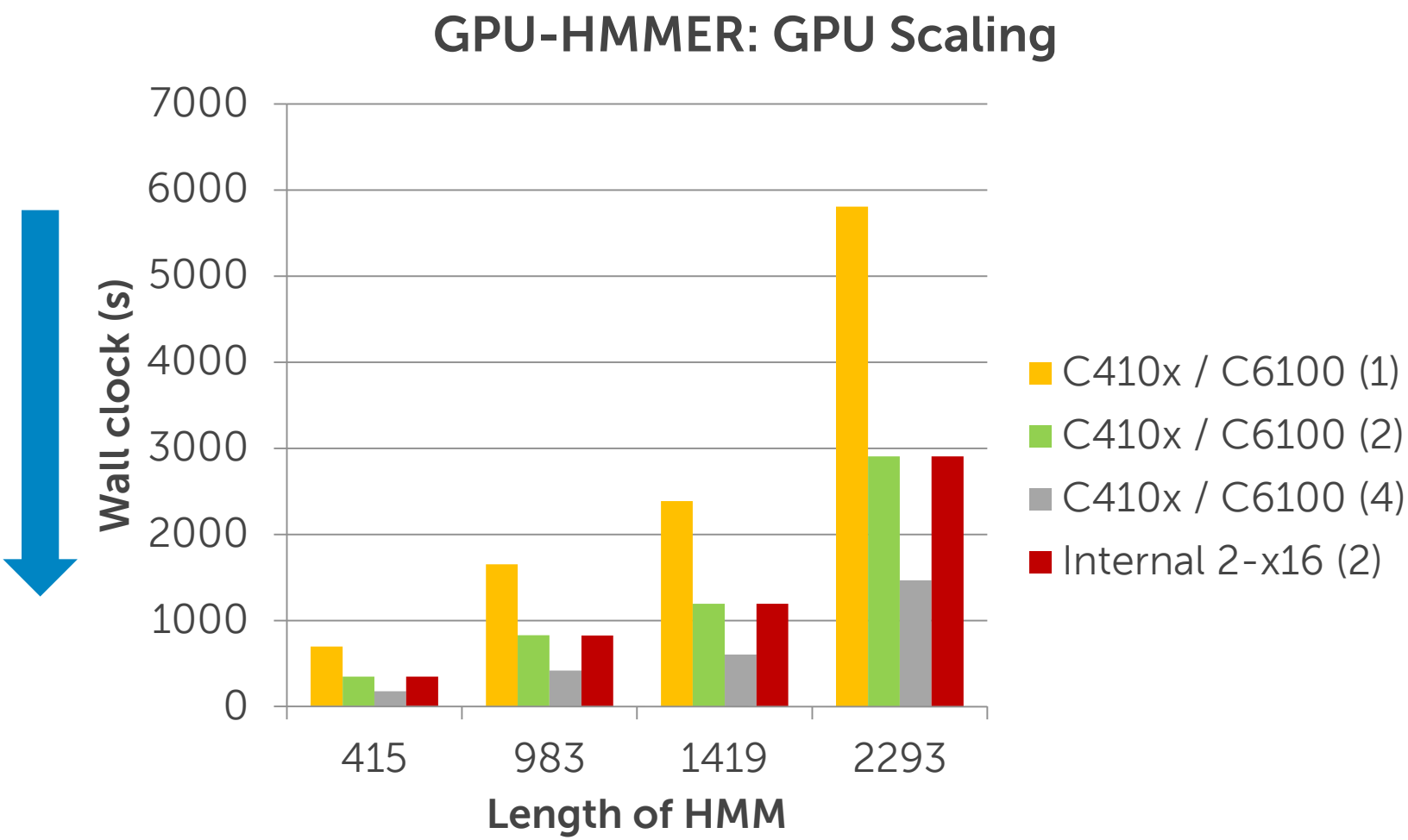
GPU:Host Scaling : GPU-HMMER



GPU:Host Scaling : GPU-HMMER

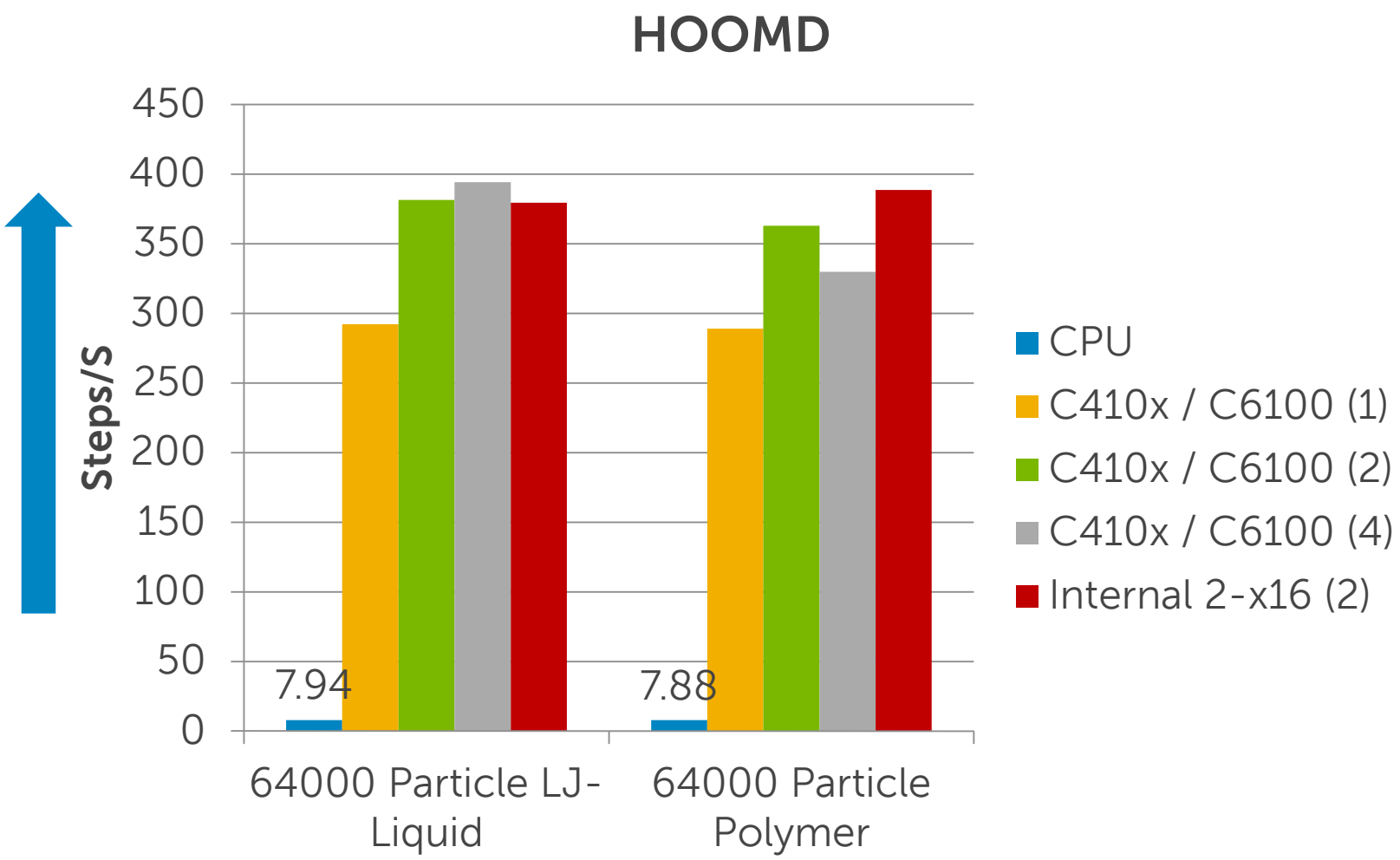


GPU:Host Scaling : GPU-HMMER

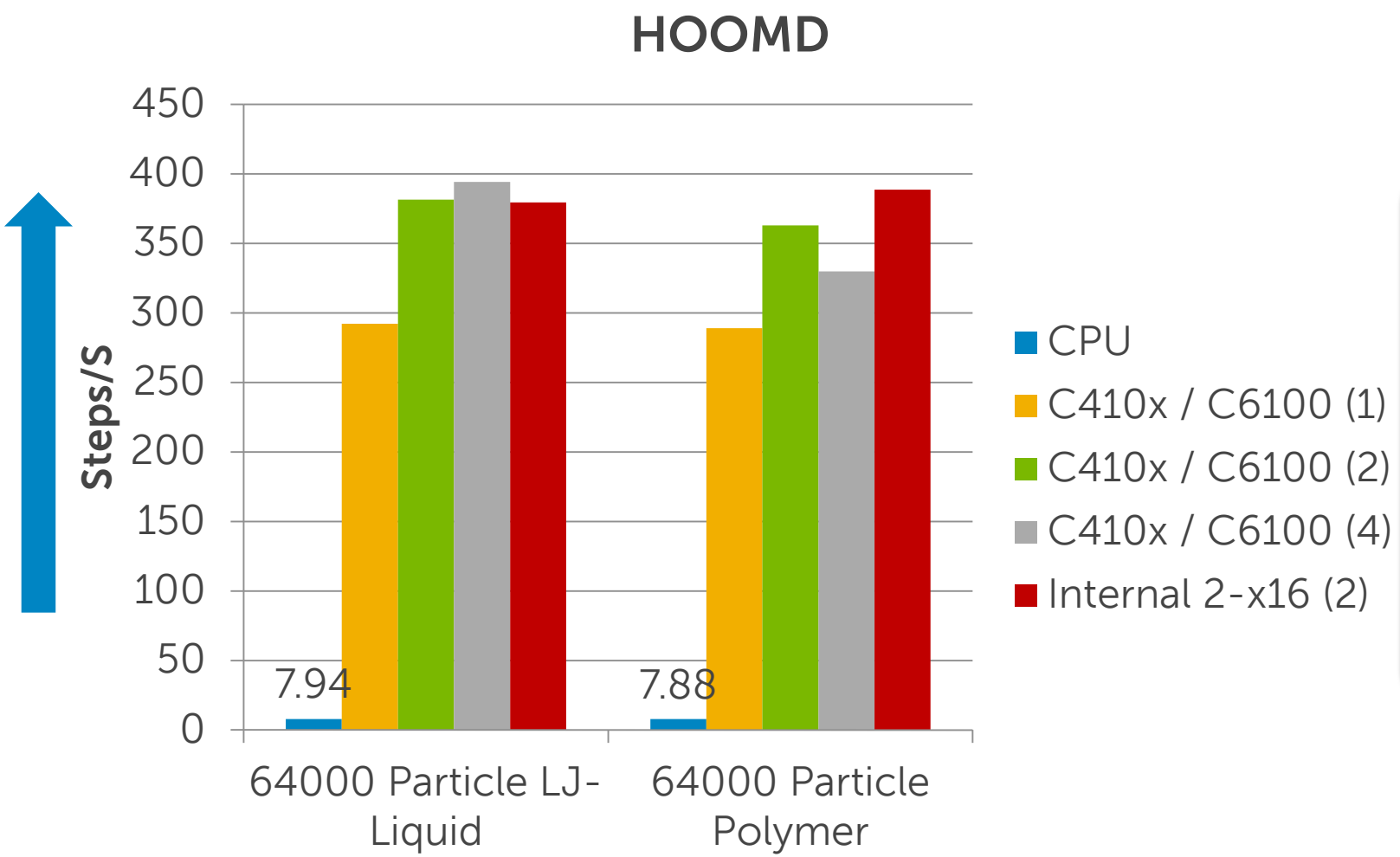


Speedup
1.8X
3.6X
7.2X
3.6X

GPU:Host Scaling : HOOMD



GPU:Host Scaling : HOOMD



Speedup
1.0X
36X
46X
41X
49X

Summary

Application	Negatively Affected by External GPU's and/or Shared x16's	Scales Beyond 2 GPU's (4:2)
NAMD	Y (13.8%)	Y (1.8X)
LAMMPS LJ-CUT	Y (6.1%)	N
LAMMPS GB	Y (6.2%)	Y (1.6X)
CUDASW++	N (0.03 to 2.9%)	Y (1.3X)
GPU-HAMMER	N (0.04 to 0.23%)	Y (1.9X)
HOOMD	Y (6.6%)	N

Conclusions

- **GPU:Host Ratio** is critically important for performance
  - application dependent
  - applications exist which scale beyond 1 or 2 GPUs
  - flexibility is important in unknown situations
- **Dedicated x16's** may not be important for performance
- **External Connections** minimally affect performance

- **C410x**
  - Introduces flexibility while increasing density
  - Purpose-built to power, cool & manage PCI-e devices
  - Separates the Host/GPU ; addresses technology curves

Model for Enabling GPU Computing

Model for Enabling GPU Computing

- Develop anywhere
- Test in the office
- Deploy in the data center

Dell Precision M4600

GPU development ***anywhere***

Mobile 15” workstation starting at 6.15lbs

Maximized Performance

- **Latest technology**
- Up to Intel Core i7 Extreme Edition
- DDR3 1333MHz or 1600MHz

- **Workstation graphics**
- NVIDIA Quadro 1000M
- NVIDIA Quadro 2000M
- AMD FirePro M5950
- UltraSharp Display

Application Empowering

- **ISV certified**
  - Avid/DigiDesign Tier 1 Certification
  - 95 applications from 35 key industry vendors
- **GPU support**
  - Quadro 4000M = 192 CUDA cores with 2GB
- **OS support**
  - Red Hat Linux WS 6
  - Windows 7

Mobile & Scalable

- **Expandable**
  - Up to 32GB RAM at 1333MHz or 16GB RAM at 1600MHz
  - 2 Storage drives
- **Flexible**
  - e-SATA interface for additional storage
  - SSD Mini card option
  - Dual integrated high quality speakers and dual integrated noise cancelling digital array microphones

Total Control of Ownership

- **Manageability**
  - Transition periods
  - Intel vPro support
  - Remote systems management
- **Security**
  - Dell ControlVault
  - FIPS FP reader
  - Contactless SC reader
  - FIPS Encrypted drives
- **Dell Services**
  - Lifecycle Services
  - ImageDirect

Dell Precision M6600

Portable high-performance GPU development Mobile 17” workstation starting at 8.1lbs

Unparalleled Performance

- **Latest technology**
  - Up to Intel Core i7 Extreme Edition
  - DDR3 1600MHz
  - RAID 0, 1, or RAID 5 SSD based storage options
- **Workstation graphics**
  - NVIDIA Quadro 3000M
  - NVIDIA Quadro 4000M
  - AMD ATI FirePro M9800
  - Dell UltraSharp Display

Application Empowering

- **ISV certified**
  - Avid/DigiDesign Tier 1 Certification
  - 95 applications from 35 key industry vendors
- **GPU support**
  - 4000M = 336 CUDA cores with 2GB
  - COMING SOON: Quadro 5010M = 384 CUDA cores with 4GB
- **OS support**
  - Red Hat Linux WS 6
  - Windows 7

Massive Scalability

- **Expandable**
  - Up to 32GB RAM
  - 4 DIMM slots
  - 2 Storage drives
  - Up to over 1.5TB storage capacity
- **Flexible**
  - e-SATA interface for additional storage
  - USB 3.0 option
  - SSD Mini card option
  - Dual integrated high quality speakers and dual integrated noise cancelling digital array microphones

Total Control of Ownership

- **Manageability**
  - Transition periods
  - Global Standard Platforms
  - Remote systems management
- **Security**
  - Dell ControlVault
  - FIPS FP reader
  - Contactless SC reader
  - FIPS Encrypted drives
- **Dell Services**
  - Lifecycle Services
  - ImageDirect

Delivering a personal supercomputer workstation

Desktop GPU Computing enables rapid development and at-speed testing in the office

NVIDIA® Tesla™
C2050
with 448 cores

Dell Precision
T7500 workstation

NVIDIA® Tesla™
C1060
with 240 cores

Dell Precision
T5500 or
T7500 workstation

More information?
http://www.dellhpcsolutions.com

Thank You!

Elliott Berger
HPC Solution Architect
Dell Research Computing
Elliott_Berger@Dell.com

The power to do more