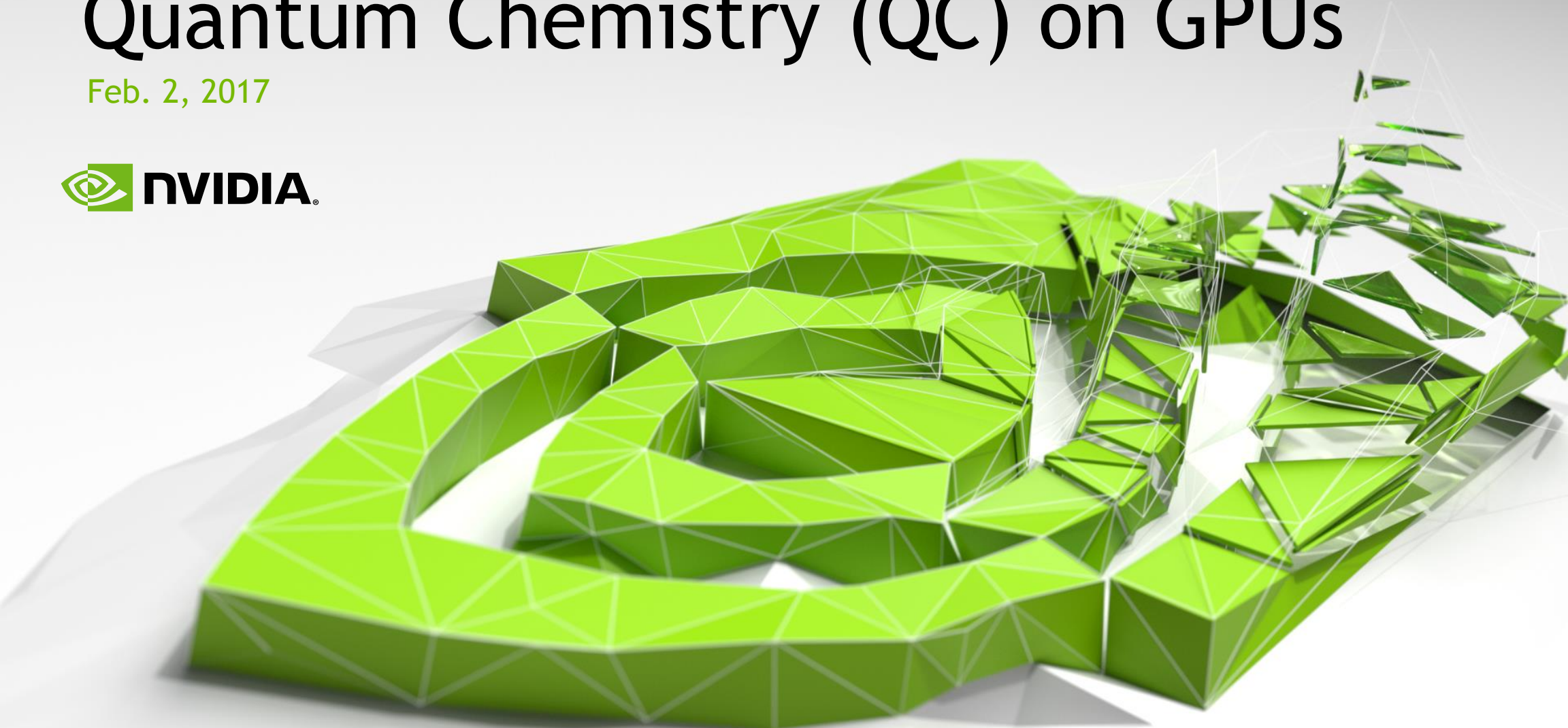
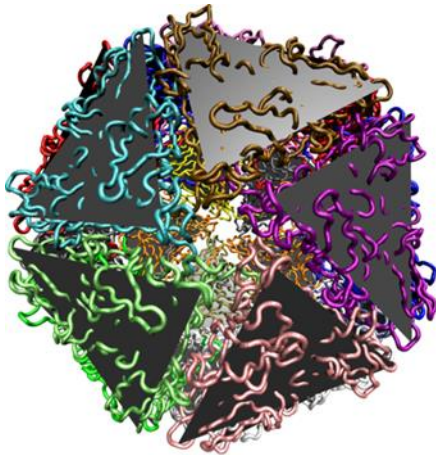


Quantum Chemistry (QC) on GPUs

Feb. 2, 2017



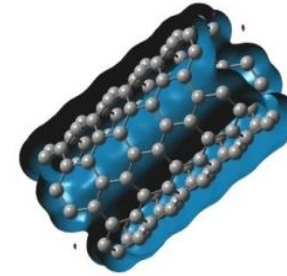
Overview of Life & Material Accelerated Apps



MD: All key codes are GPU-accelerated

- ▶ Great multi-GPU performance
- ▶ Focus on dense (up to 16) GPU nodes &/or large # of GPU nodes
- ▶ **ACEMD***, **AMBER (PMEMD)***, BAND, CHARMM, DESMOND, ESPRESSO, Folding@Home, GPUgrid.net, GROMACS, HALMD, **HOOMD-Blue***, LAMMPS, **Lattice Microbes***, mdcore, MELD, miniMD, NAMD, OpenMM, PolyFTS, **SOP-GPU*** & more

green* = application where >90% of the workload is on GPU



QC: All key codes are ported or optimizing

- ▶ Focus on using GPU-accelerated math libraries, OpenACC directives
- ▶ GPU-accelerated and available today:
 - ▶ ABINIT, ACES III, ADF, BigDFT, CP2K, GAMESS, GAMESS-UK, GPAW, LATTE, LSDalton, LSMS, MOLCAS, MOPAC2012, NWChem, **OCTOPUS***, PEtot, QUICK, Q-Chem, QMCPack, Quantum Espresso/PWscf, QUICK, **TeraChem***
- ▶ Active GPU acceleration projects:
 - ▶ CASTEP, GAMESS, Gaussian, ONETEP, **Quantum Supercharger Library***, VASP & more

MD vs. QC on GPUs

“Classical” Molecular Dynamics	Quantum Chemistry (MO, PW, DFT, Semi-Emp)
Simulates positions of atoms over time; chemical-biological or chemical-material behaviors	Calculates electronic properties; ground state, excited states, spectral properties, making/breaking bonds, physical properties
Forces calculated from simple empirical formulas (bond rearrangement generally forbidden)	Forces derived from electron wave function (bond rearrangement OK, e.g., bond energies)
Up to millions of atoms	Up to a few thousand atoms
Solvent included without difficulty	Generally in a vacuum but if needed, solvent treated classically (QM/MM) or using implicit methods
Single precision dominated	Double precision is important
Uses cuBLAS, cuFFT, CUDA	Uses cuBLAS, cuFFT, OpenACC
Geforce (Accademics), Tesla (Servers)	Tesla recommended
ECC off	ECC on

Accelerating Discoveries

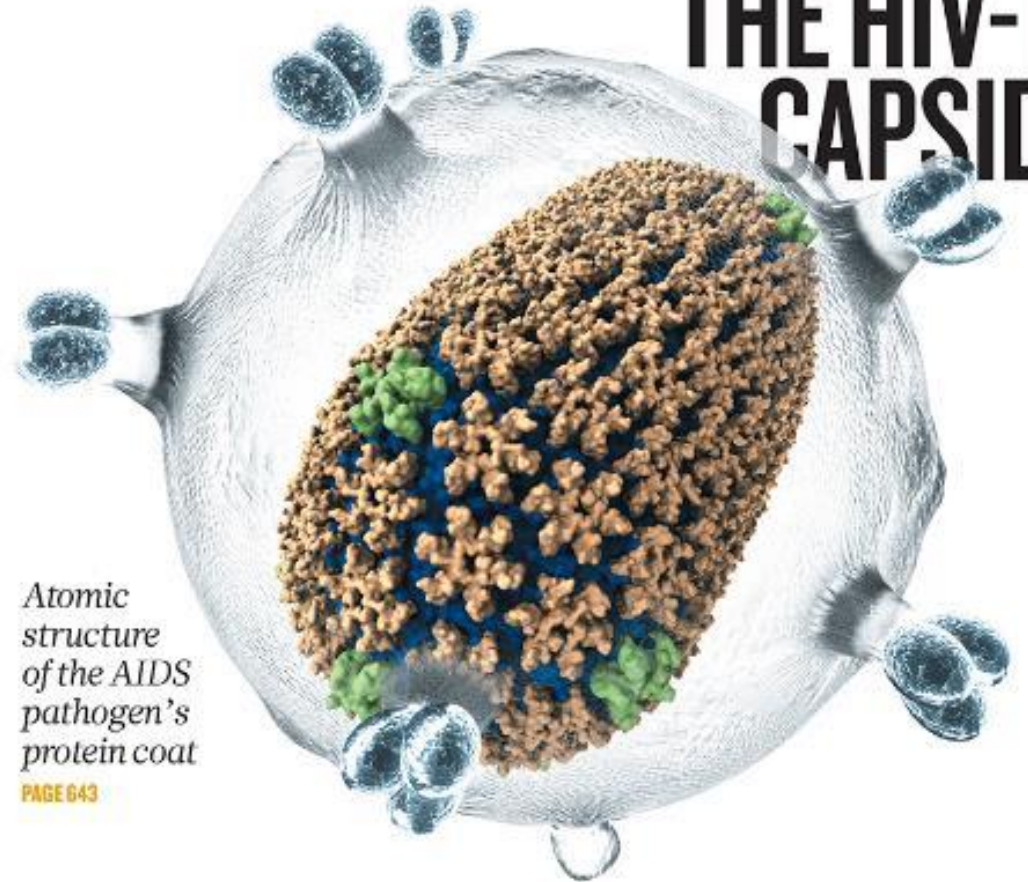
Using a supercomputer powered by the Tesla Platform with over 3,000 Tesla accelerators, University of Illinois scientists performed the first all-atom simulation of the HIV virus and discovered the chemical structure of its capsid – “the perfect target for fighting the infection.”

Without gpu, the supercomputer would need to be 5x larger for similar performance.

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

THE HIV-1 CAPSID



*Atomic
structure
of the AIDS
pathogen's
protein coat*

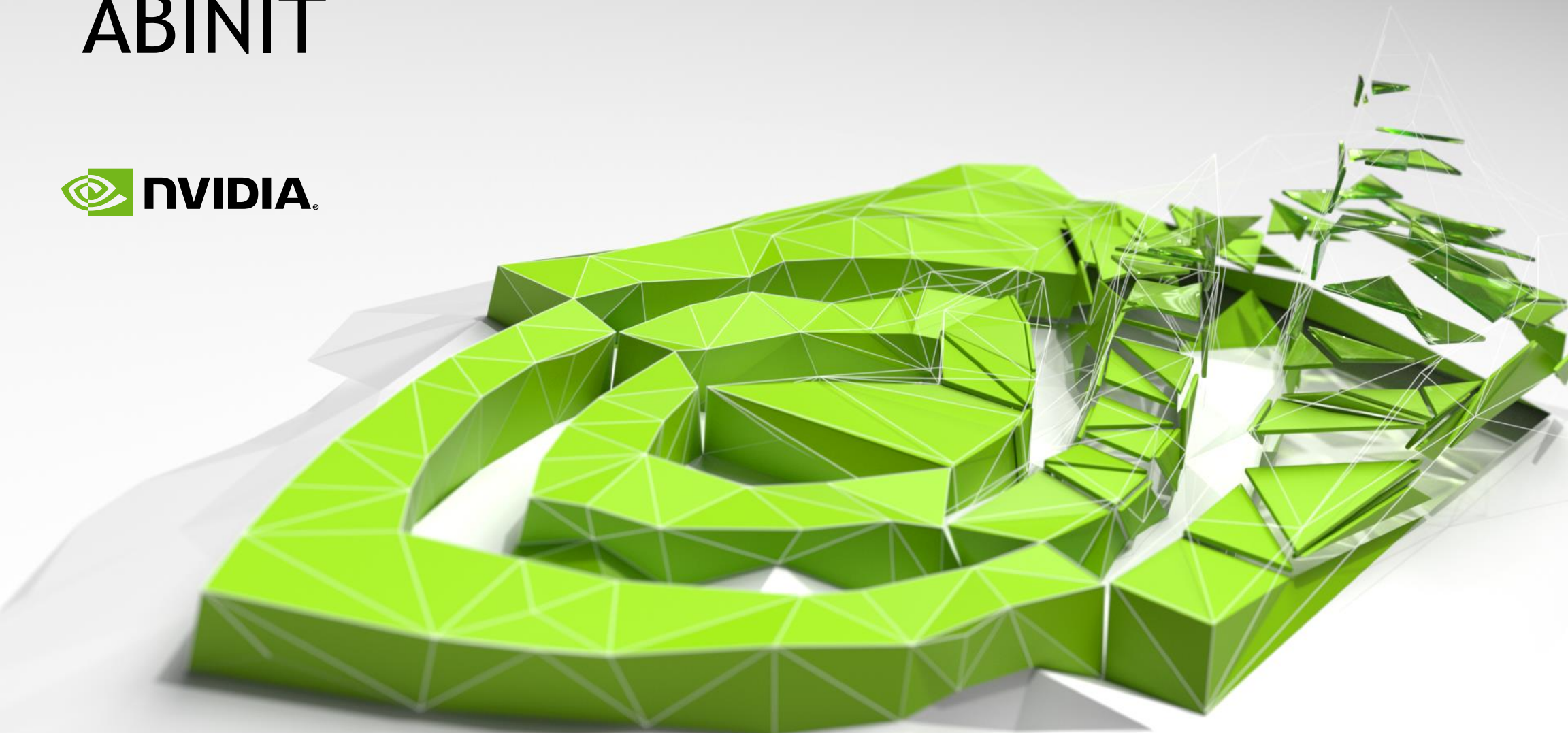
PAGE 643

GPU-Accelerated Quantum Chemistry Apps

Green Lettering Indicates Performance Slides Included

- ▶ Abinit
- ▶ ACES III
- ▶ ADF
- ▶ BigDFT
- ▶ CP2K
- ▶ GAMESS-US
- ▶ Gaussian
- ▶ GPAW
- ▶ LATTE
- ▶ LSDalton
- ▶ MOLCAS
- ▶ Mopac2012
- ▶ NWChem
- ▶ Octopus
- ▶ ONETEP
- ▶ Petot
- ▶ Q-Chem
- ▶ QMCPACK
- ▶ Quantum Espresso
- ▶ Quantum SuperCharger Library
- ▶ RMG
- ▶ TeraChem
- ▶ UNM
- ▶ VASP
- ▶ WL-LSMS

ABINIT

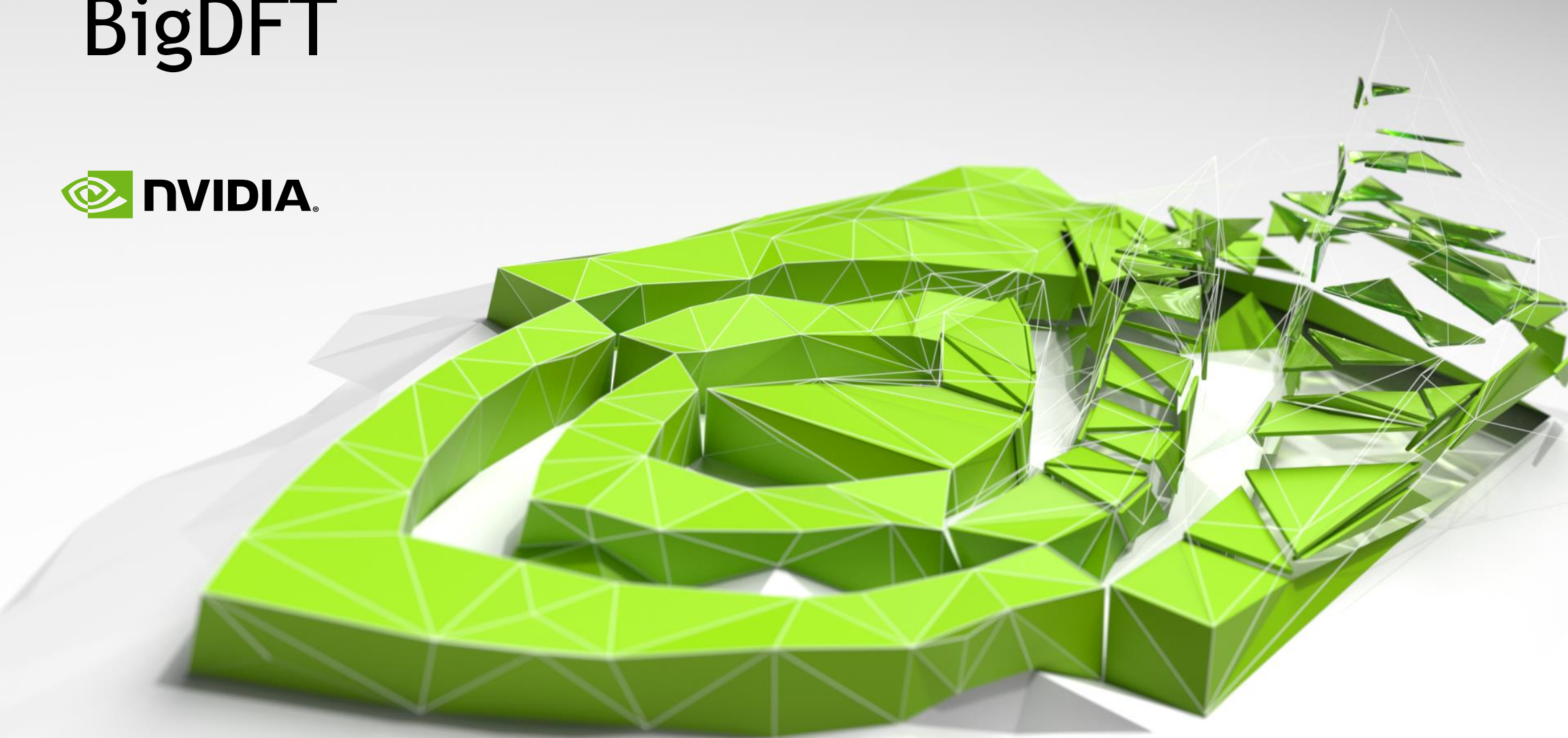


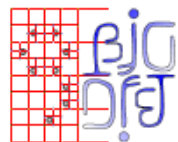
ABINIT on GPUS



- Speed in the parallel version:
 - For ground-state calculations, GPUs can be used. This is based on CUDA+MAGMA
 - For ground-state calculations, the wavelet part of ABINIT (which is BigDFT) is also very well parallelized : MPI band parallelism, combined with GPUs

BigDFT





Multiscale Modelling Methods for Applications in Materials Science CECAM JÜLICH, GERMANY

Introduction to Electronic Structure Calculations with BigDFT

Thierry Deutsch, Damien Caliste, Luigi Genovese

L_Sim - CEA Grenoble

17 September 2013

Courtesy of
BigDFT
team @ CEA

BigDFT
<http://bigdft.org>

Introduction

BigDFT run

Atom positions

Basis set

Pseudopotential

XC

SCF Loop

Performances

Poisson Solver

Relaxation

HPC

Perspectives

Order N

Resonant states

Conclusion



BigDFT version 1.7: capabilities

<http://bigdft.org>

- Free, surface and periodic boundary conditions
- Geometry optimizations (with constraints)
- Born-Oppenheimer Molecular Dynamics
- Saddle point searches (Nudged-Elastic Band Method)
- Vibrations
- External electric fields
- Unoccupied KS orbitals
- Collinear and Non-collinear magnetism
- All XC functionals of the ABINIT package
- Hybrid functionals
- Empirical van der Waals interactions (many flavors)
- **Also available within the ABINIT package**

Courtesy of
BigDFT
team @ CEA



BigDFT

<http://bigdft.org>

Introduction

BigDFT run

Atom positions

Basis set

Pseudopotential

XC

SCF Loop

Performances

Poisson Solver

Relaxation

HPC

Perspectives

Order N

Resonant states

Conclusion

BigDFT version 1.7: capabilities

<http://bigdft.org>

- Free, surface and periodic boundary conditions
- Geometry optimizations (with constraints)
- Born-Oppenheimer Molecular Dynamics
- Saddle point searches (Nudged-Elastic Band Method)
- Vibrations
- External electric fields
- Unoccupied KS orbitals
- Collinear and Non-collinear magnetism
- All XC functionals of the ABINIT package
- Hybrid functionals
- Empirical van der Waals interactions (many flavors)
- **Also available within the ABINIT package**

Courtesy of
BigDFT
team @ CEA



BigDFT

<http://bigdft.org>

Introduction

BigDFT run

Atom positions

Basis set

Pseudopotential

XC

SCF Loop

Performances

Poisson Solver

Relaxation

HPC

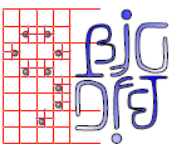
Perspectives

Order N

Resonant states

Conclusion

GPU-ported operations in BigDFT (double precision)



BigDFT
<http://bigdft.org>

Introduction

BigDFT run

- Atom positions
- Basis set
- Pseudopotential
- XC
- SCF Loop

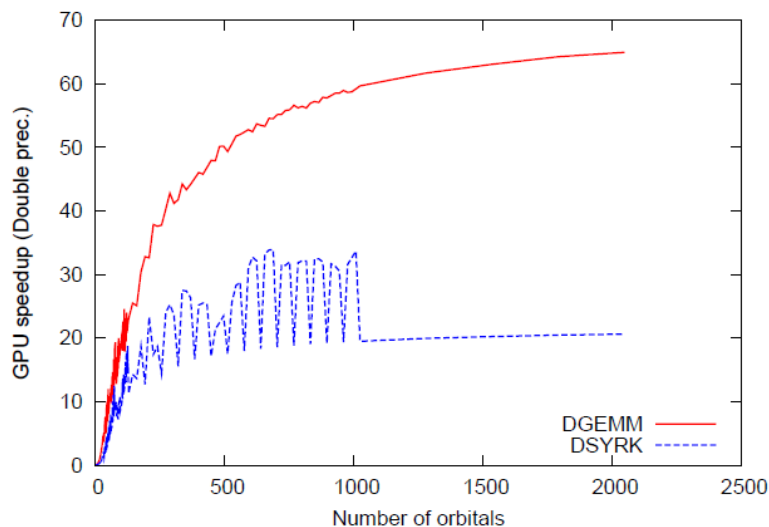
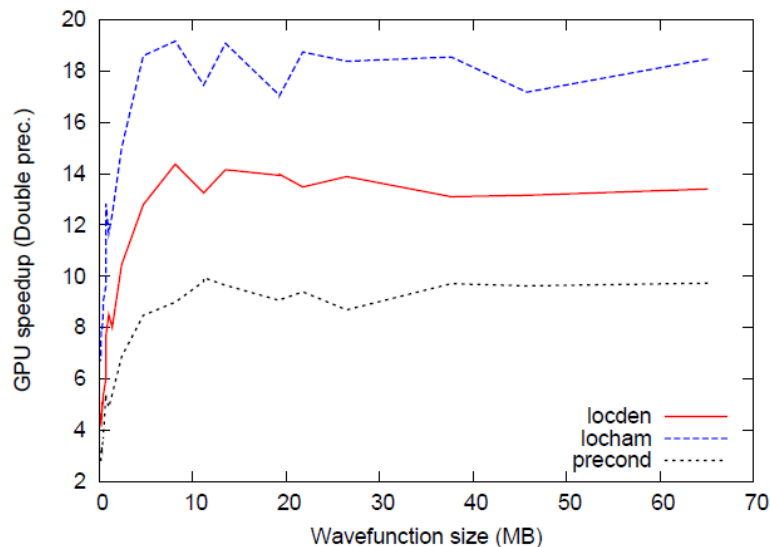
Performances

- Poisson Solver
- Relaxation
- HPC

Perspectives

- Order N
- Resonant states
- Conclusion

Convolutions
(OpenCL rewritten)
GPU speedups between
10 and 20 can be
obtained for different
sections



Linear algebra
(CUBLAS library)
The interfacing with
CUBLAS is immediate,
with considerable
speedups

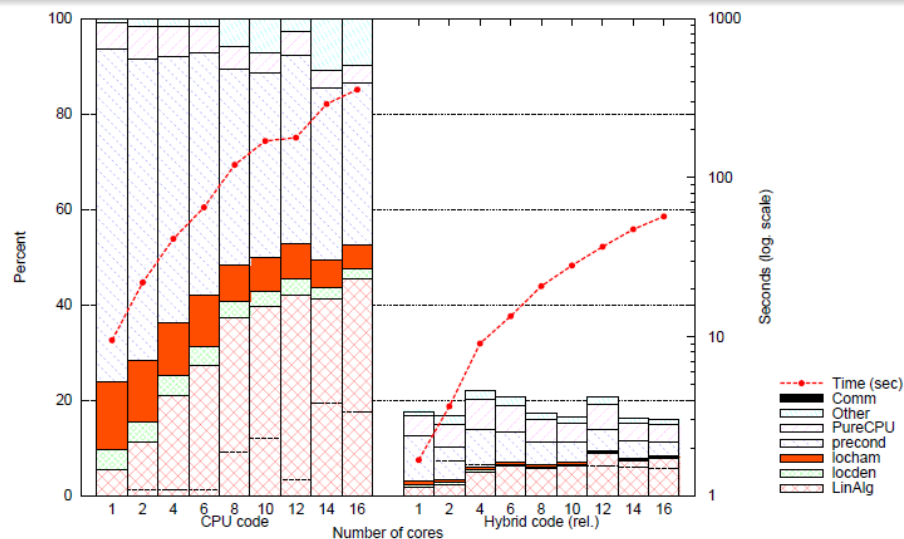
Courtesy of
BigDFT
team @ CEA

BigDFT code on Hybrid architectures

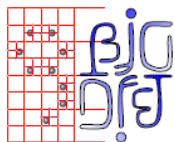
BigDFT code can run on hybrid CPU/GPU supercomputers
In multi-GPU environments, **double precision** calculations

No Hot-spot operations

Different code sections can be ported on GPU
up to 20x speedup for some operations,
7x for the full parallel code



Courtesy of
BigDFT
team @ CEA



BigDFT

<http://bigdft.org>

Introduction

BigDFT run

Atom positions

Basis set

Pseudopotential

XC

SCF Loop

Performances

Poisson Solver

Relaxation

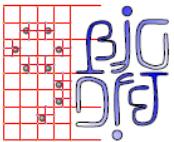
HPC

Perspectives

Order N

Resonant states

Conclusion



See

<http://bigdft.org/Wiki/index.php?title=Category:Tutorials>

- First runs with BigDFT
- Basis-set convergence
- Acceleration example on different platforms:
Kohn-Sham DFT Operation with GPU acceleration

Courtesy of
BigDFT
team @ CEA

BigDFT

<http://bigdft.org>

Introduction

BigDFT run

Atom positions

Basis set

Pseudopotential

XC

SCF Loop

Performances

Poisson Solver

Relaxation

HPC

Perspectives

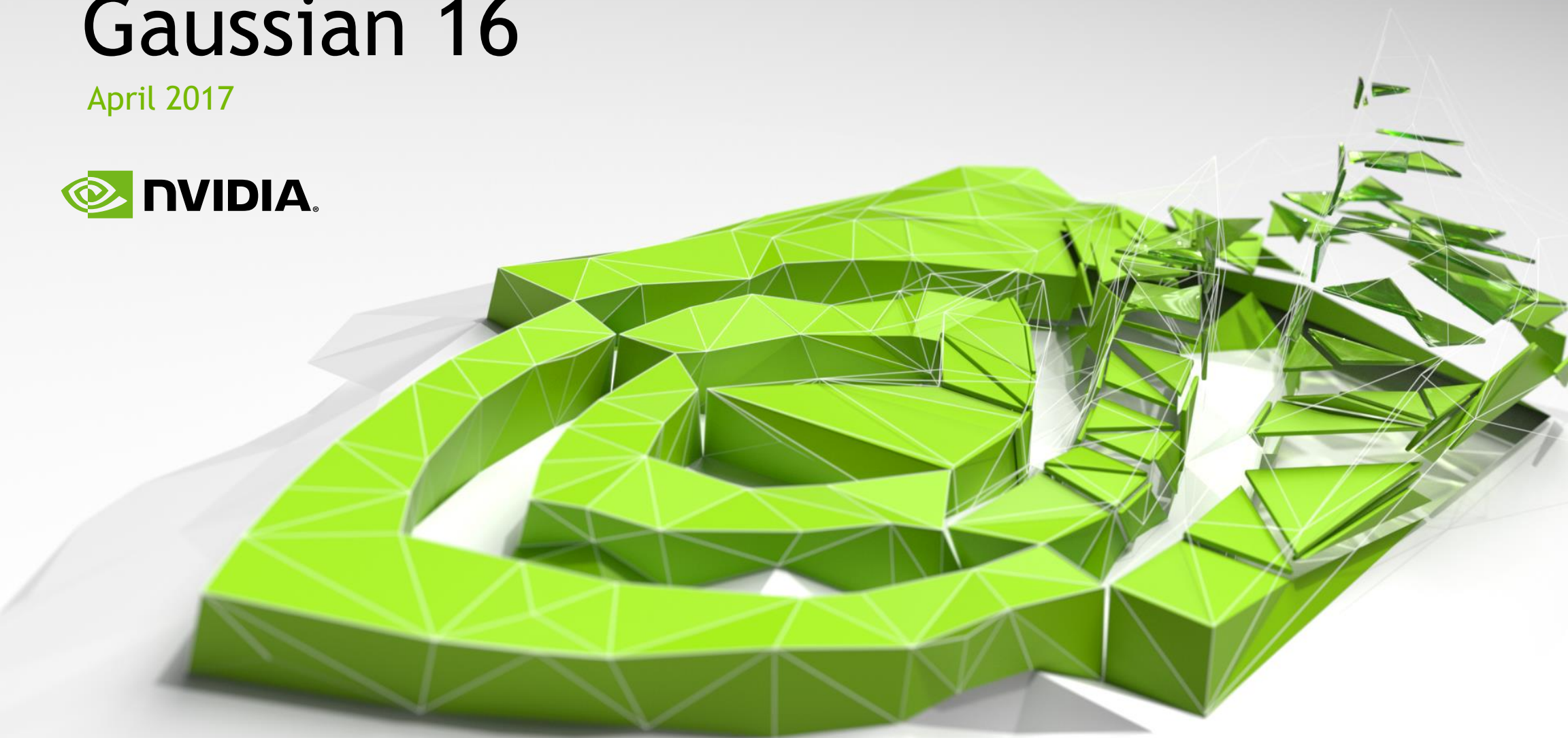
Order N

Resonant states

Conclusion

Gaussian 16

April 2017

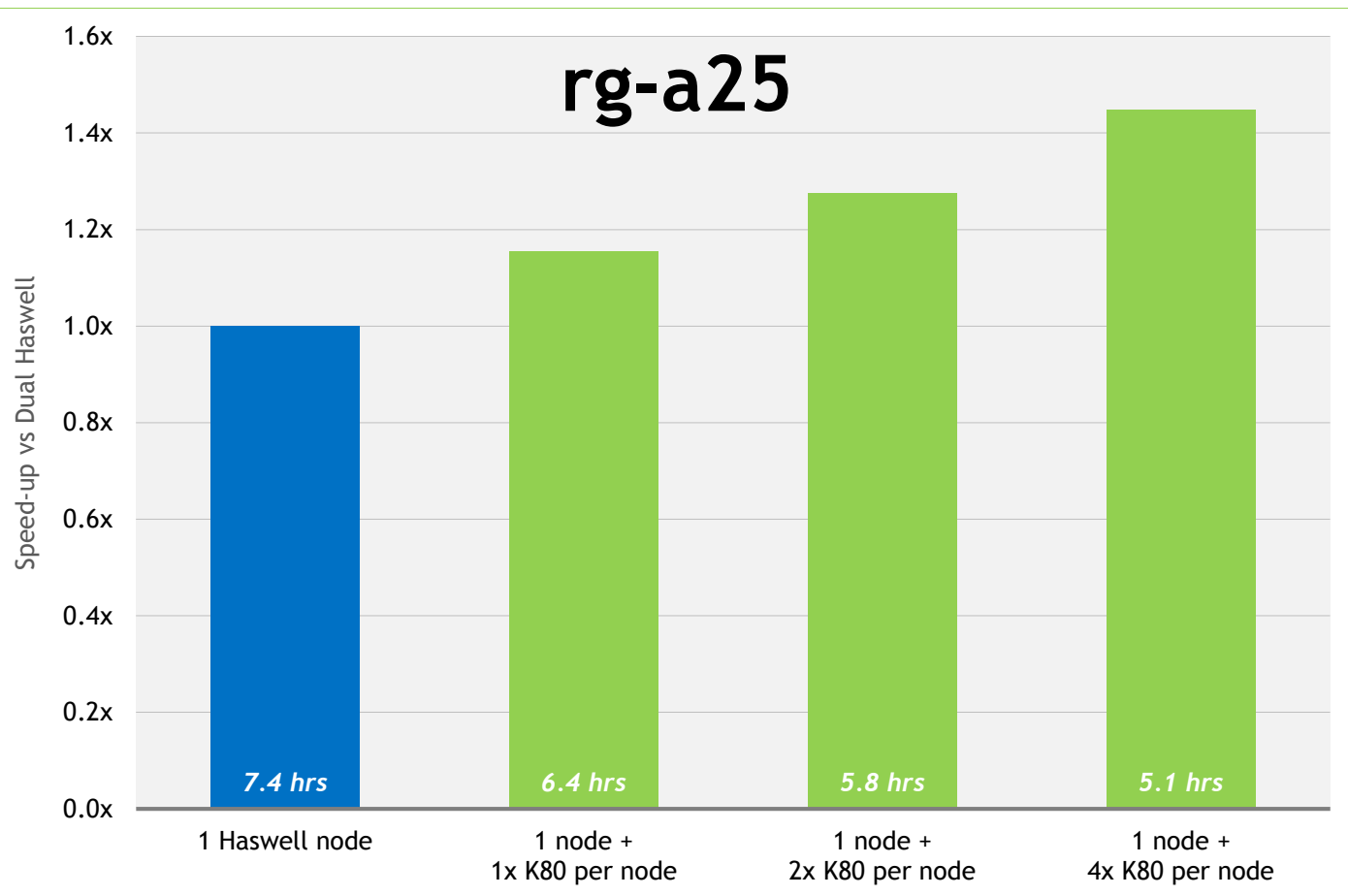


GPU-ACCELERATED GAUSSIAN 16 AVAILABLE

100% PGI OpenACC Port (no CUDA)

- Gaussian is a Top Ten HPC (Quantum Chemistry) Application.
- **80-85% of use cases are GPU-accelerated** (Hartree-Fock and DFT: energies, 1st derivatives (gradients) and 2nd derivatives). More functionality to come.
- K40, K80 support; P100 support coming as a minor release, performance “good”, **faster wall clock times**. Early P100 results promising.
- **No pricing difference** between Gaussian CPU and GPU versions.
- Existing Gaussian 09 customers under maintenance contract **get (free) upgrade**.
- Existing non-maintenance customers required to pay upgrade fee.
- To get the bits or to ask about the upgrade fee, please contact Gaussian, Inc.’s Jim Hess, Operations Manager; jhess@gaussian.com.

rg-a25 on K80s

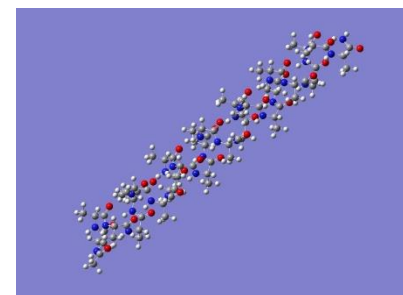


Running **Gaussian** version 16

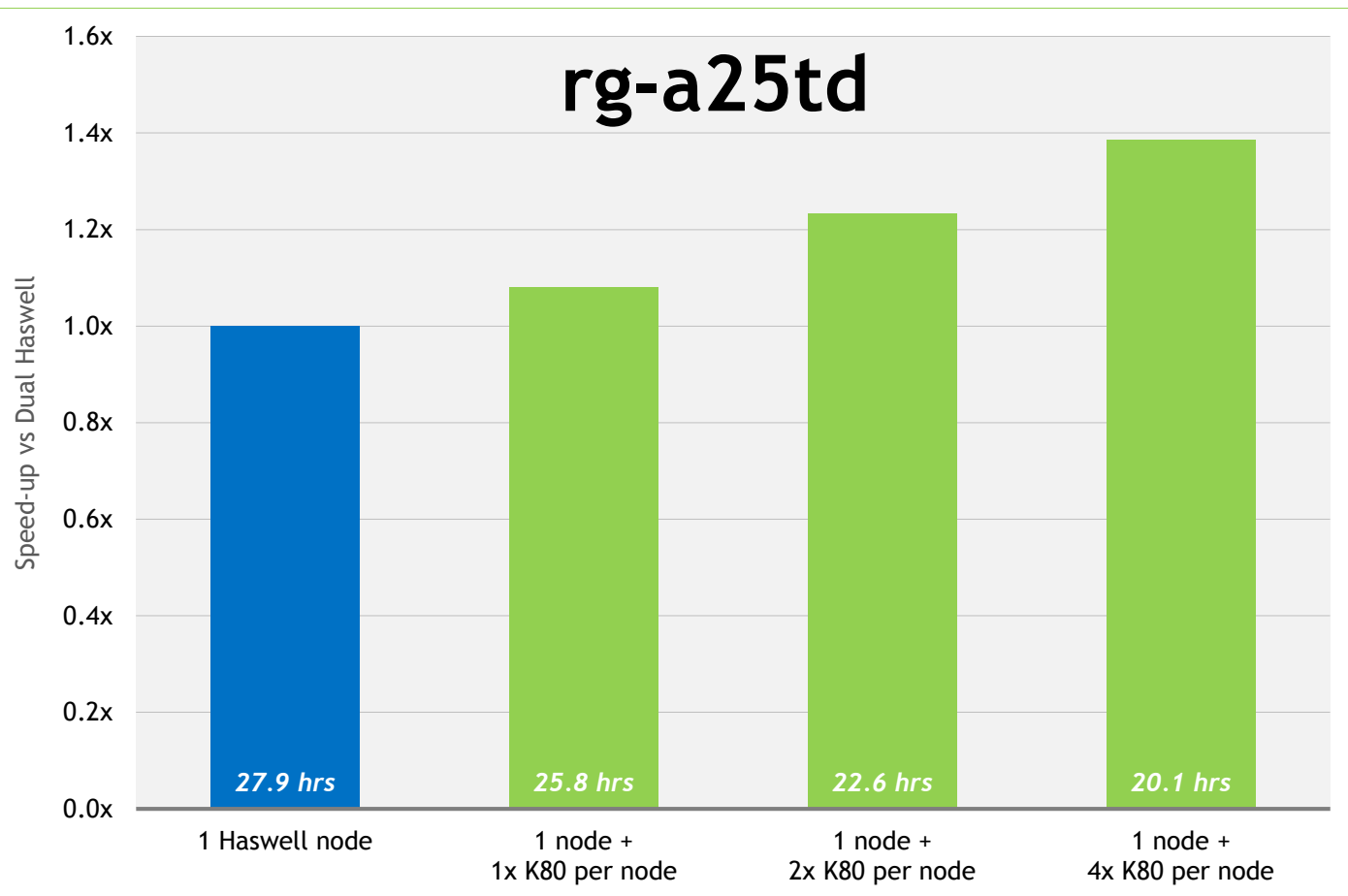
The **blue node** contains Dual Intel Xeon E5-2698 v3@2.3GHz (Haswell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2698 v3@2.3GHz (Haswell) CPUs + Tesla K80 (autoboost) GPUs

Alanine 25. Two steps: Force and Frequency. APFD 6-31G*
nAtoms = 259, nBasis = 2195



rg-a25td on K80s

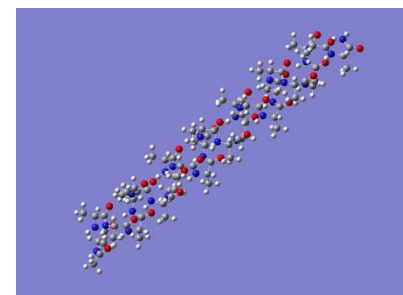


Running **Gaussian** version 16

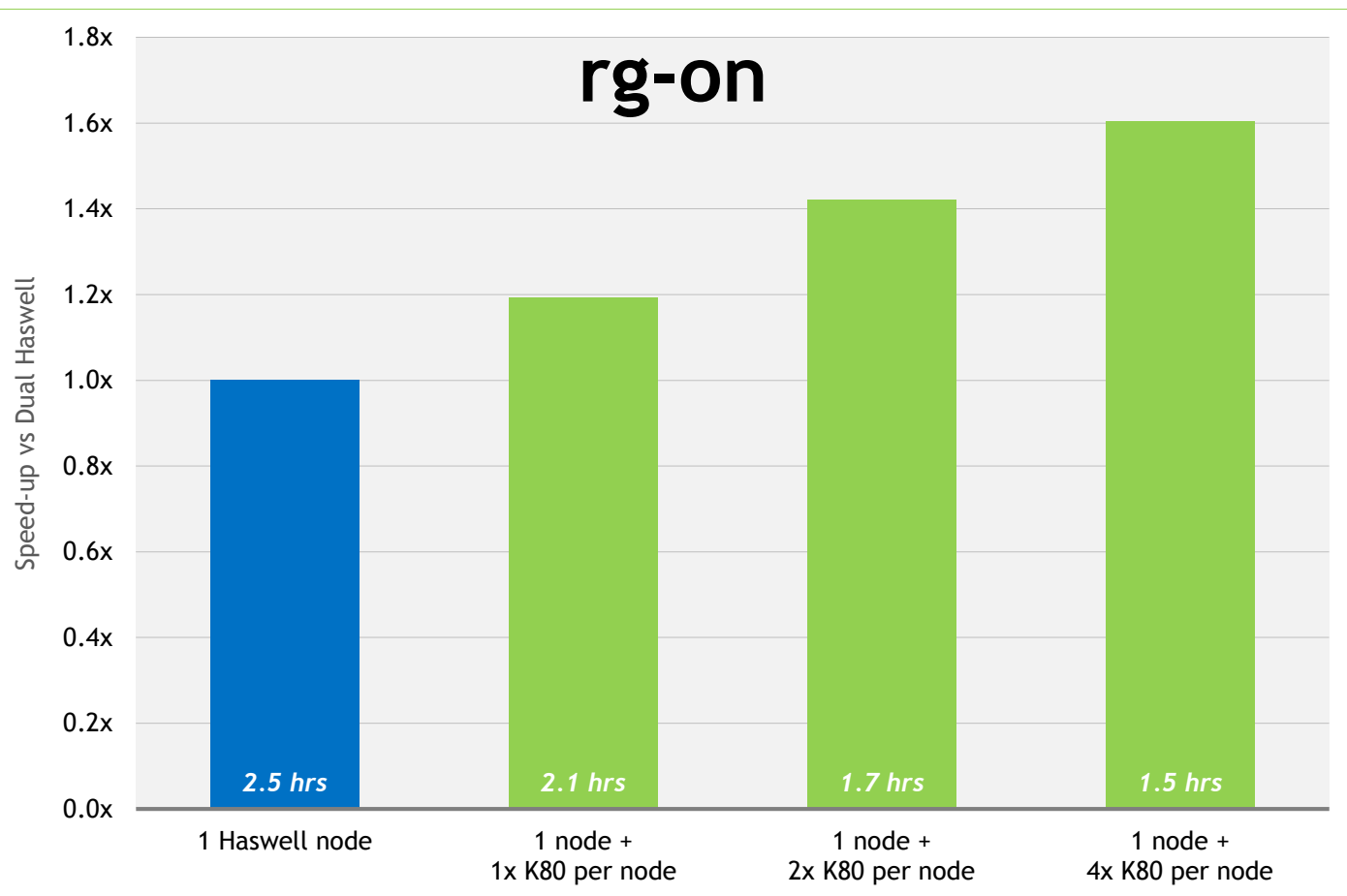
The **blue node** contains Dual Intel Xeon E5-2698 v3@2.3GHz (Haswell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2698 v3@2.3GHz (Haswell) CPUs + Tesla K80 (autoboost) GPUs

Alanine 25. Two Time-Dependent (TD) steps: Force and Frequency. APFD 6-31G*
nAtoms = 259, nBasis = 2195



rg-on on K80s

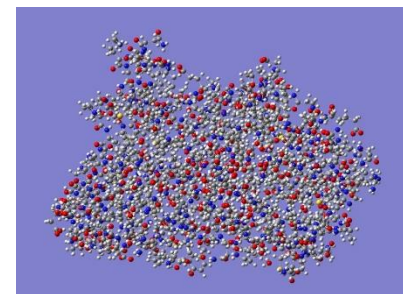


Running **Gaussian** version 16

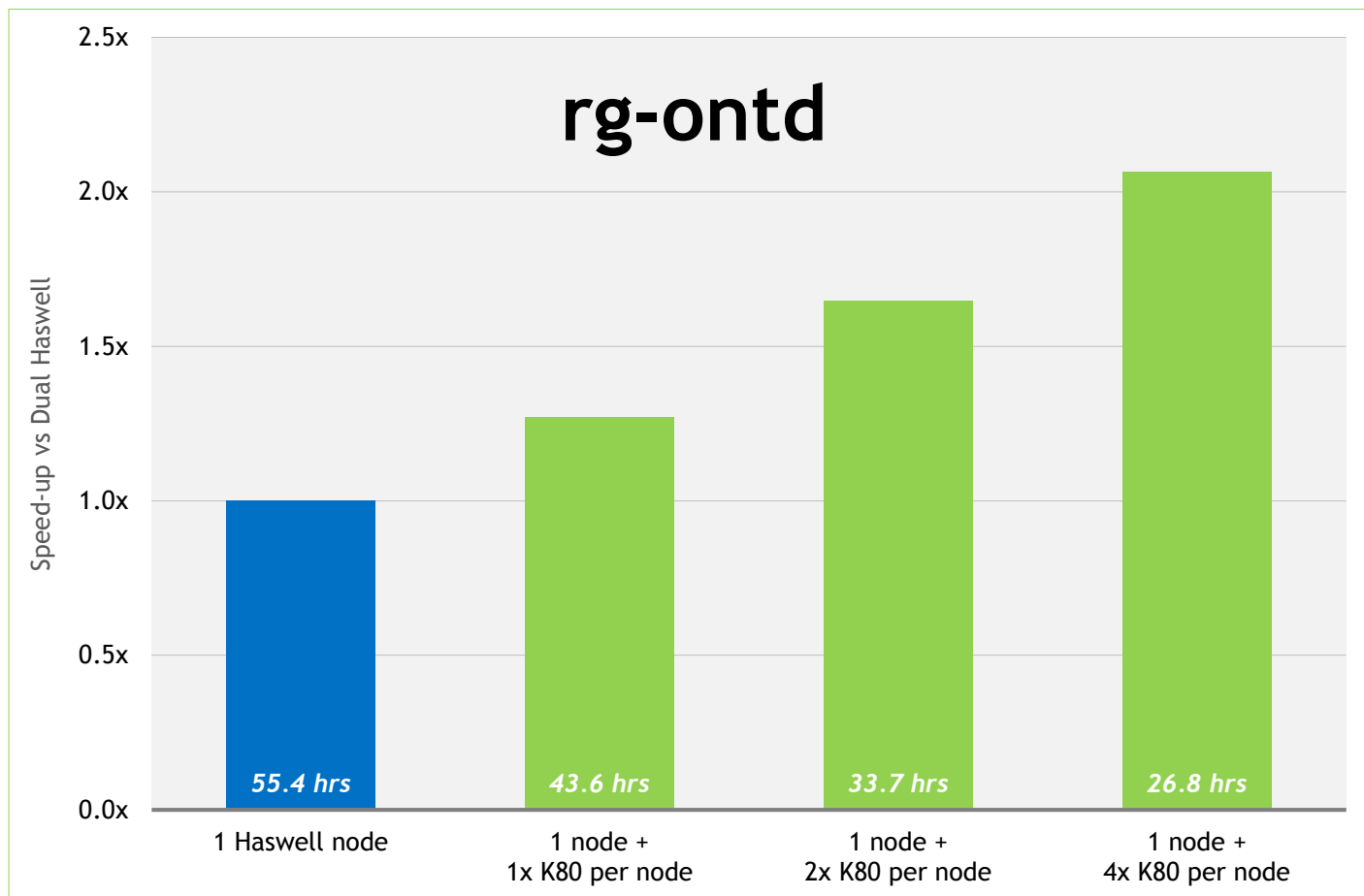
The **blue node** contains Dual Intel Xeon E5-2698 v3@2.3GHz (Haswell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2698 v3@2.3GHz (Haswell) CPUs + Tesla K80 (autoboost) GPUs

GFP ONIOM. Two steps: Force and Frequency. APFD/6-311+G(2d,p):amber=softfirst)=embed
nAtoms = 3715 (48/3667), nBasis = 813



rg-ontd on K80s

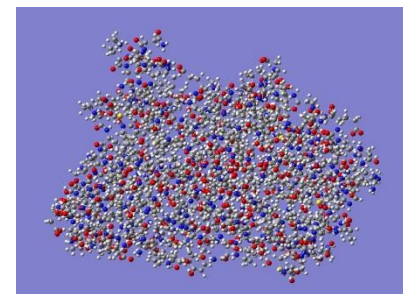


Running **Gaussian** version 16

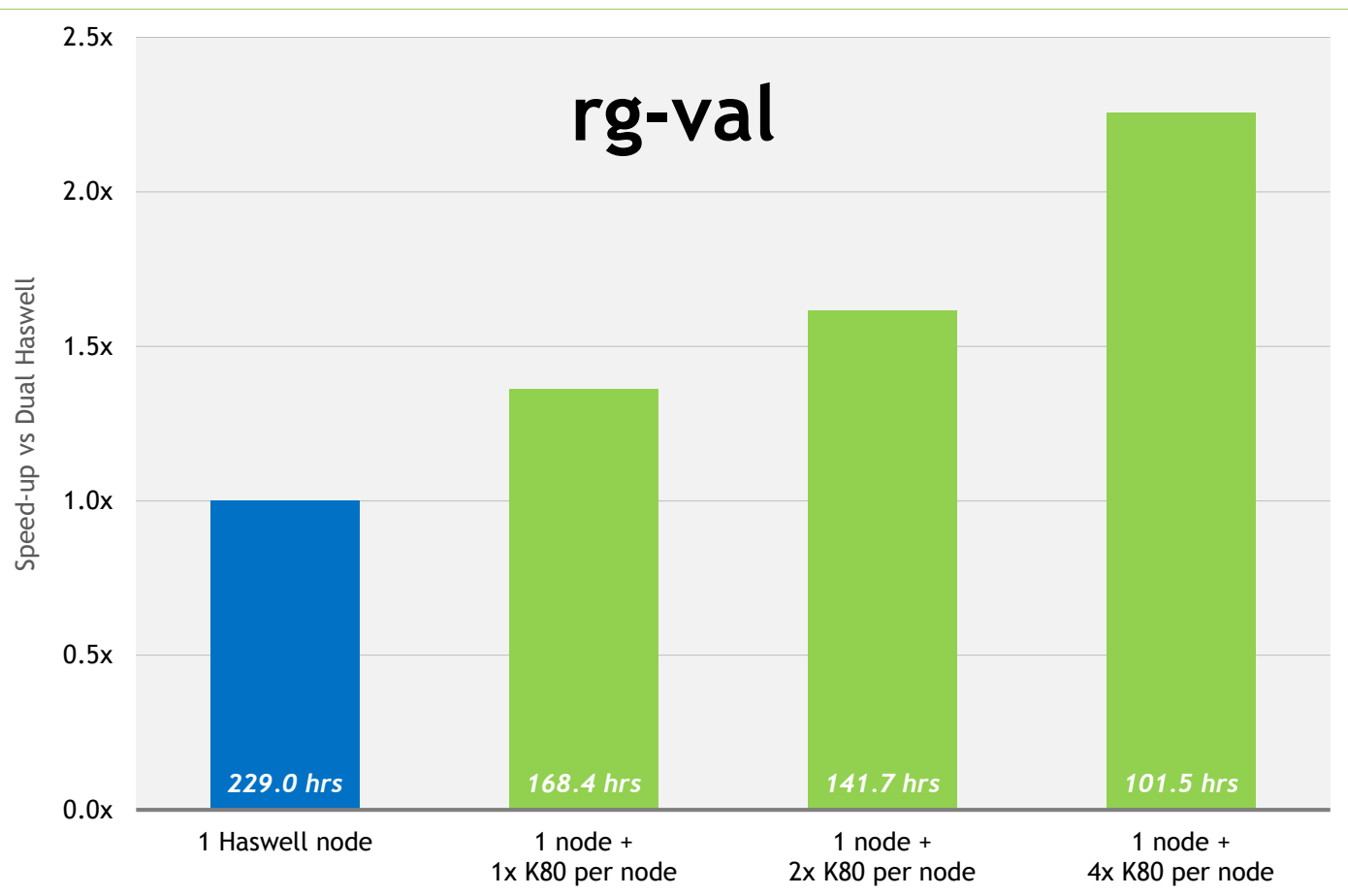
The **blue node** contains Dual Intel Xeon E5-2698 v3@2.3GHz (Haswell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2698 v3@2.3GHz (Haswell) CPUs + Tesla K80 (autoboost) GPUs

GFP ONIOM. Two Time-Dependent (TD) steps: Force and Frequency. APFD/6-311+G(2d,p):amber=softfirst)=embed
nAtoms = 3715 (48/3667), nBasis = 813



rg-val on K80s

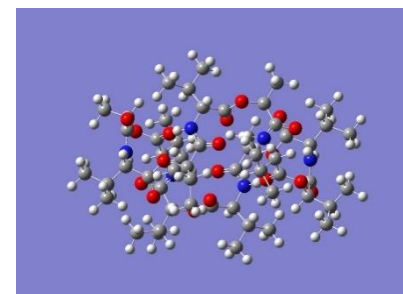


Running **Gaussian** version 16

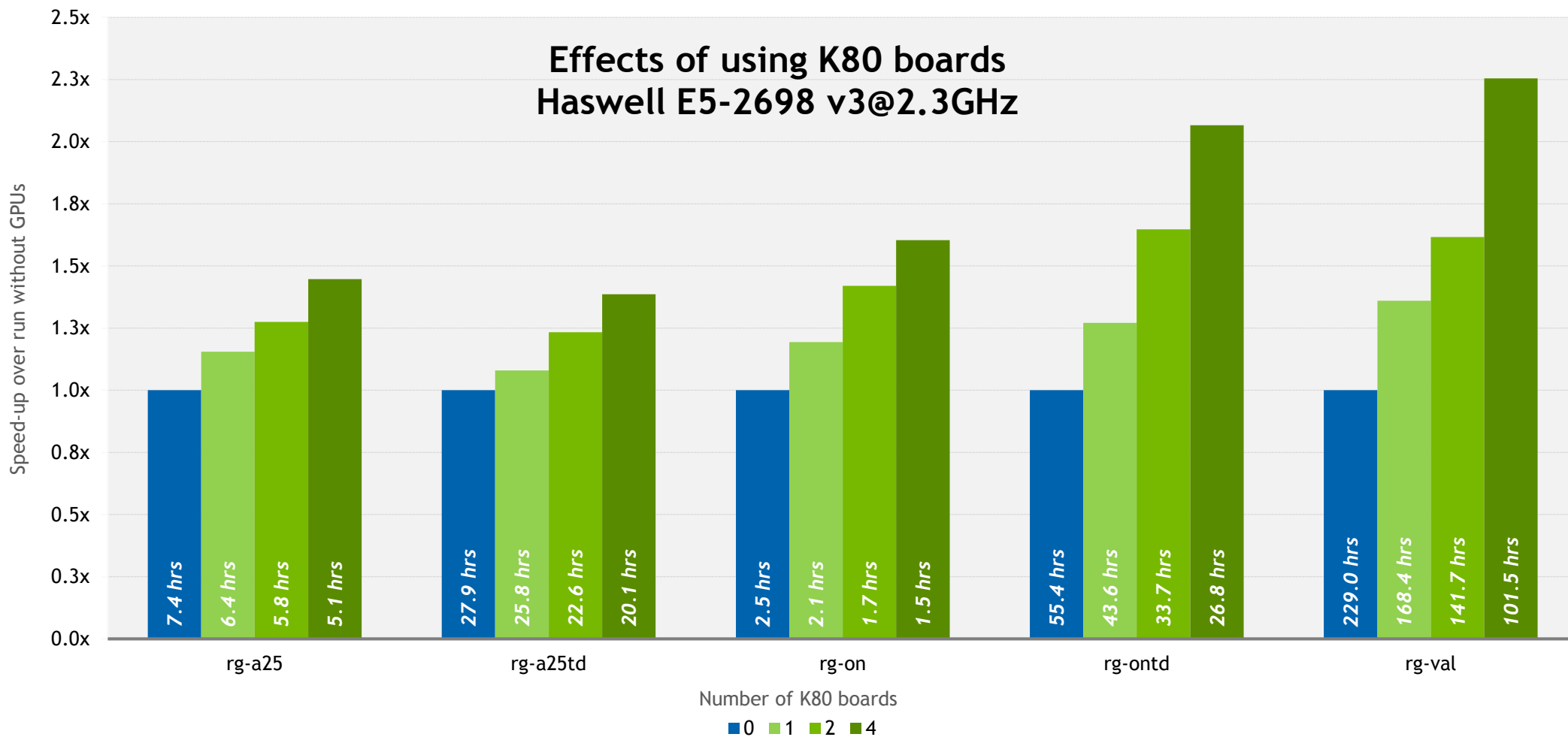
The **blue node** contains Dual Intel Xeon E5-2698 v3@2.3GHz (Haswell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2698 v3@2.3GHz (Haswell) CPUs + Tesla K80 (autoboost) GPUs

Valinomycin. Two steps: Force and Frequency. APFD 6-311+G(2d,p)
nAtoms = 168, nBasis = 2646



Effects of using K80s



Gaussian 16 Supported Platforms

- 4-way collaboration; Gaussian, Inc., PGI, NVIDIA and HPE
- HPE, NVIDIA and PGI is the development platform
- All released/certified x86_64 versions of Gaussian 16 use the PGI compilers
- Certified versions of Gaussian 16 use Intel only for Itanium, XLF for some IBM platforms, Fujitsu compilers for some SPARC-based machines and PGI for the rest (including some Apple products)
- GINC is collaborating with IBM, PGI (and NVIDIA) to release an OpenPower version of Gaussian that also uses the PGI compiler
- See Gaussian Supported Platforms for more details:
http://gaussian.com/g16/g16_plat.pdf

CLOSING REMARKS

Significant Progress has been made in enabling Gaussian on GPUs with OpenACC

OpenACC is increasingly becoming more versatile

Significant work lies ahead to improve performance

Expand feature set:

PBC, Solvation, MP2, ONIOM, triples-Corrections

ACKNOWLEDGEMENTS

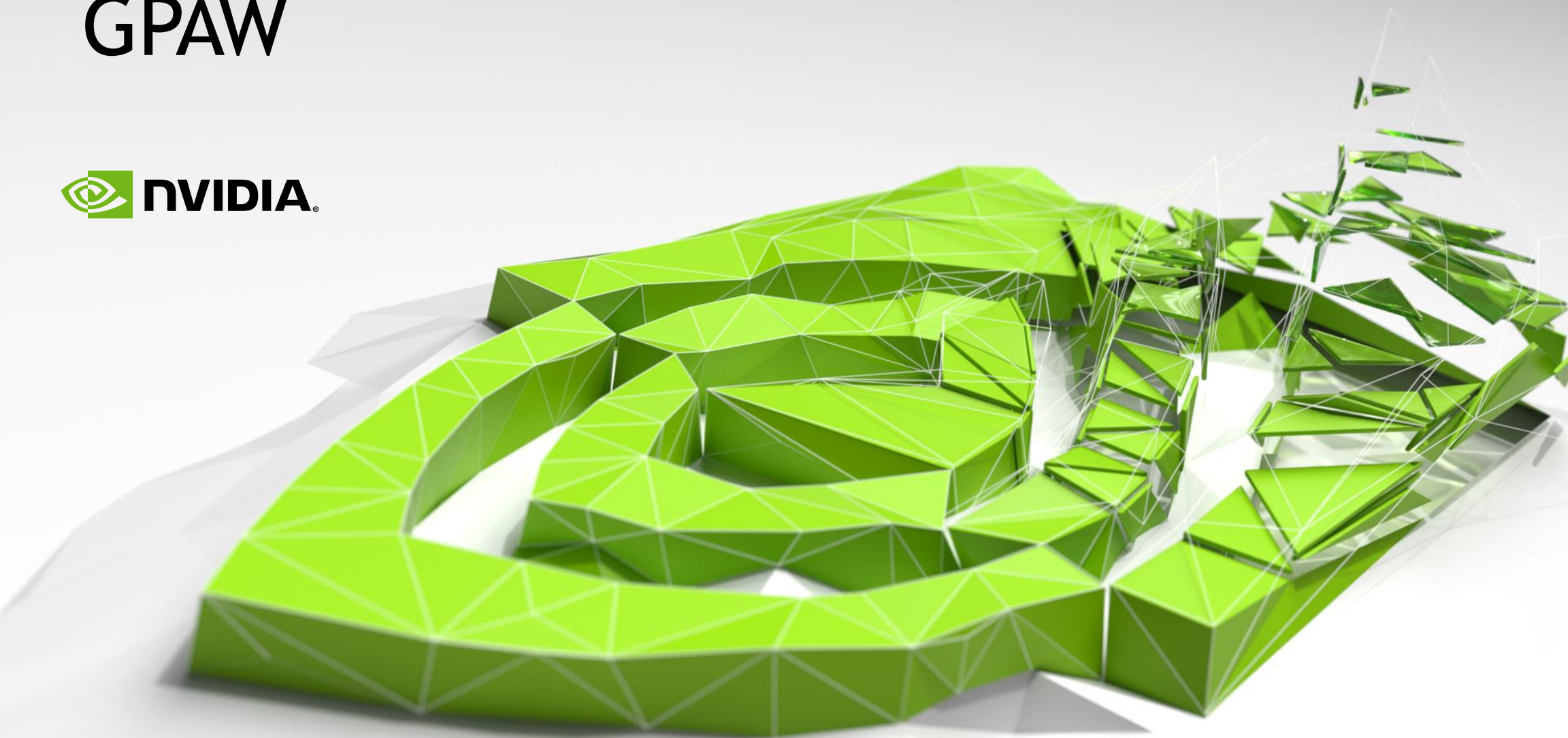
Development is taking place with:

Hewlett-Packard (HP) Series SL2500 Servers (Intel® Xeon® E5-2680 v2 (2.8GHz/10-core/25MB/8.0GT-s QPI/115W, DDR3-1866)

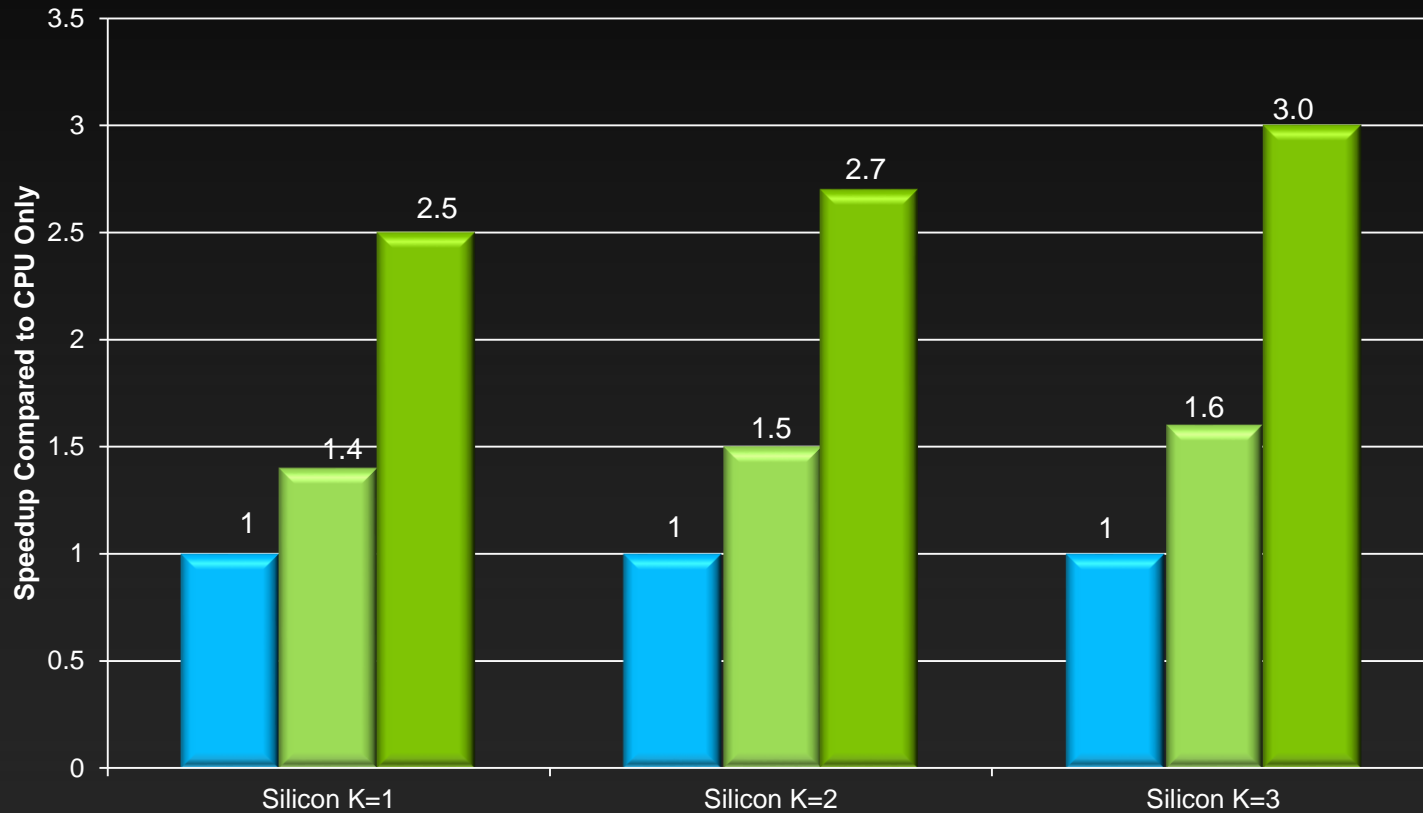
NVIDIA® Tesla® GPUs (K40 and later)

PGI Accelerator Compilers (16.x) with OpenACC (2.5 standard)

GPAW



Increase Performance with Kepler

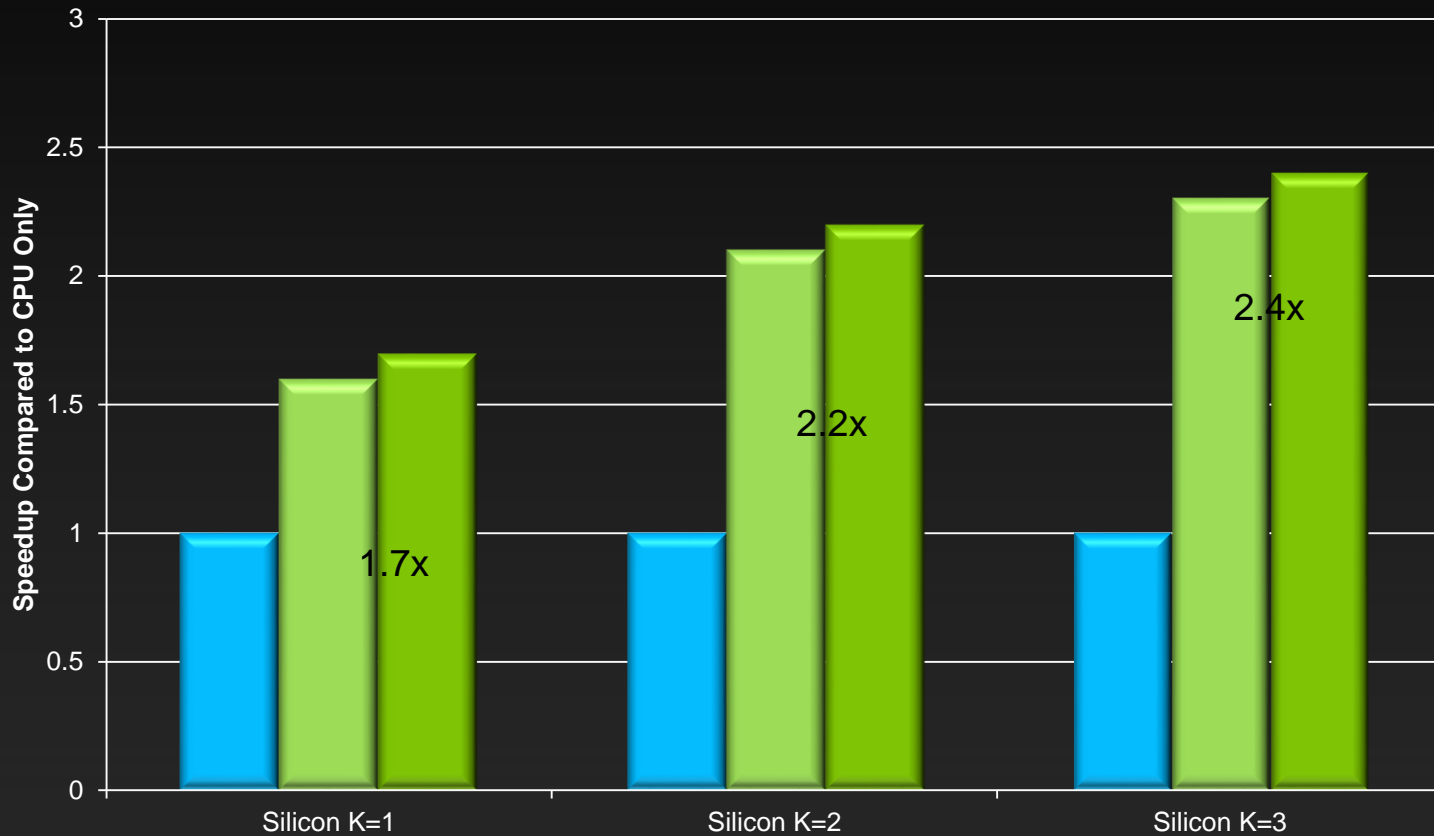


Running **GPAW** 10258

The **blue nodes** contain 1x E5-2687W CPU (8 Cores per CPU).

The **green nodes** contain 1x E5-2687W CPU (8 Cores per CPU) and 1x or 2x NVIDIA K20X for the GPU.

Increase Performance with Kepler

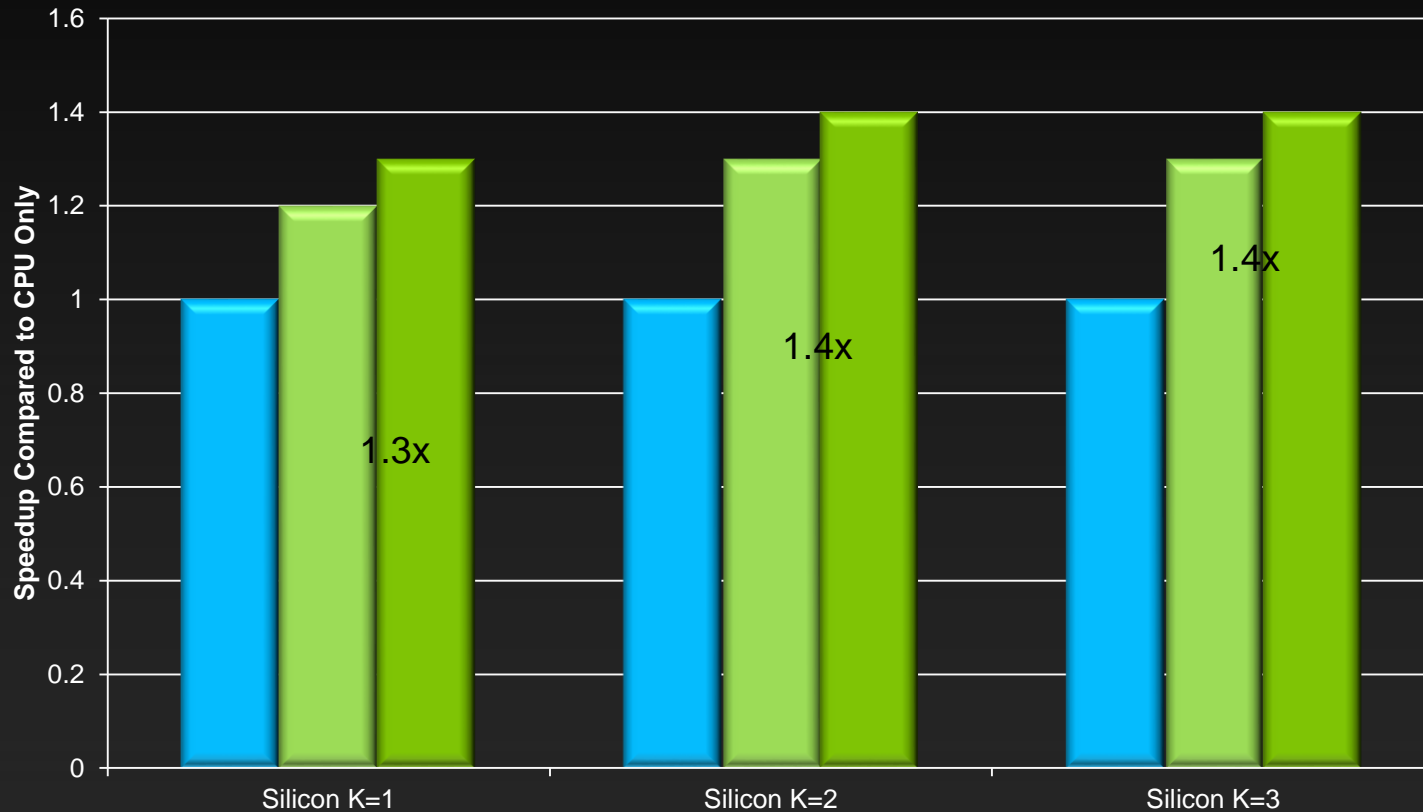


Running **GPAW** 10258

The **blue nodes** contain 1x E5-2687W CPU (8 Cores per CPU).

The **green nodes** contain 1x E5-2687W CPUs (8 Cores per CPU) and 2x NVIDIA K20 or K20X for the GPU.

Increase Performance with Kepler



Running **GPAW** 10258

The **blue nodes** contain 2x E5-2687W CPUs (8 Cores per CPU).

The **green nodes** contain 2x E5-2687W CPUs (8 Cores per CPU) and 2x NVIDIA K20 or K20X for the GPU.

Multi-GPU Accelerated Large Scale Electronic Structure Calculations

Used with
permission from
Samuli Hakala

Samuli Hakala

COMP Centre of Excellence

Department of Applied Physics

Aalto University School of Science

Email: samuli.hakala@aalto.fi

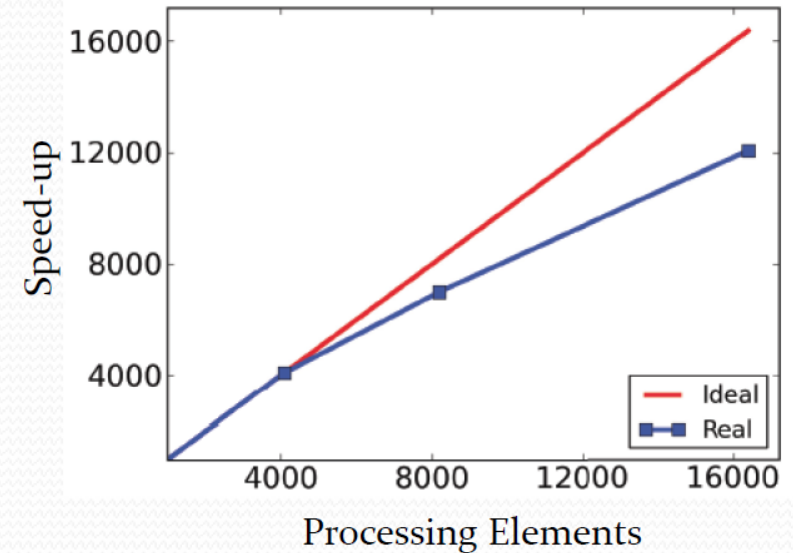
GPU Technology Conference, March 2013



Aalto University
School of Science

GPAW

- Density Functional Theory (DFT) program package for electronic structure calculations
- Time-Dependent Density Functional Theory (TDDFT) is implemented in the linear response and time propagation schemes
- Can use real-space grids, atom centered basis functions or plane waves
- Random Phase Approximation (RPA) also available
- Scales to thousands of cores and suitable for large scale calculations
- Open Source software licensed under GPL



Ground state DFT calculation of 561 Au atom cluster on Blue Gene/P.

LibXC on GPUs

- A reusable library of >250 exchange-correlation functionals
- Used by 15 different codes (Abinit, GPAW, BigDFT, etc.)
- Can be a performance bottleneck for small systems
- Can “clone” existing functionals for GPU use with fairly minimal changes to existing LibXC code and parallelizes well over grid points
- More information:
 - <https://confluence.slac.stanford.edu/display/SUNCAT/libxc+on+GPUs>
- Work by Lin Li, Jun Yan, Christopher O’Grady (Stanford/SLAC)

Functional	Type	Speedup ((GPU+CPU)/CPU)
PW, PW Mode, OB PW, PW RPA	LDA Correlation	23,23,23,37
PBE, PBE sol, xPBE, PBE JRGX, RGE ₂ , APBE	GGA Correlation	56, 58, 58, 58, 58, 58
RPBE	GGA Exchange	95
TPSS	MGGA Exchange	51

Ground State Performance

Bulk Silicon

- 95 atoms with periodic boundary conditions, 380 bands and 1 k-point. Grid size: 56x56x80.
- Time is in seconds per one SCF iteration.
- Intel Xeon X5650, NVIDIA Tesla M2070

Si ₉₅	CPU	GPU	%	S-Up
Poisson Solver	1.8	0.13	1%	14
Orthonormalization	23	3.0	23%	7.7
Precondition	9.4	0.77	6%	12
RMM-DIIS other	32	3.2	25%	10
Subspace Diag	23	2.1	16%	11
Other	2.7	2.7	21%	1.0
Total (SCF-Iter)	93	13		9.7/7.7

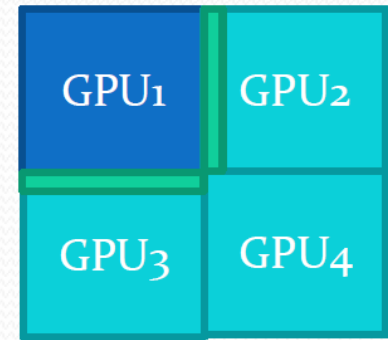
Fullerene

- C₆₀ molecule with 240 valence electrons. Grid size: 84x84x84
- Intel Xeon X5650, NVIDIA Tesla M2070

C ₆₀	CPU	GPU	%	S-Up
	13	0.64	7%	20
	11	1.2	13%	9.2
	16	0.99	11%	16
	8.1	0.6	7%	13
	22	2.1	23%	10
	3.5	3.2	35%	1.1
	76	9.1		13/8.3

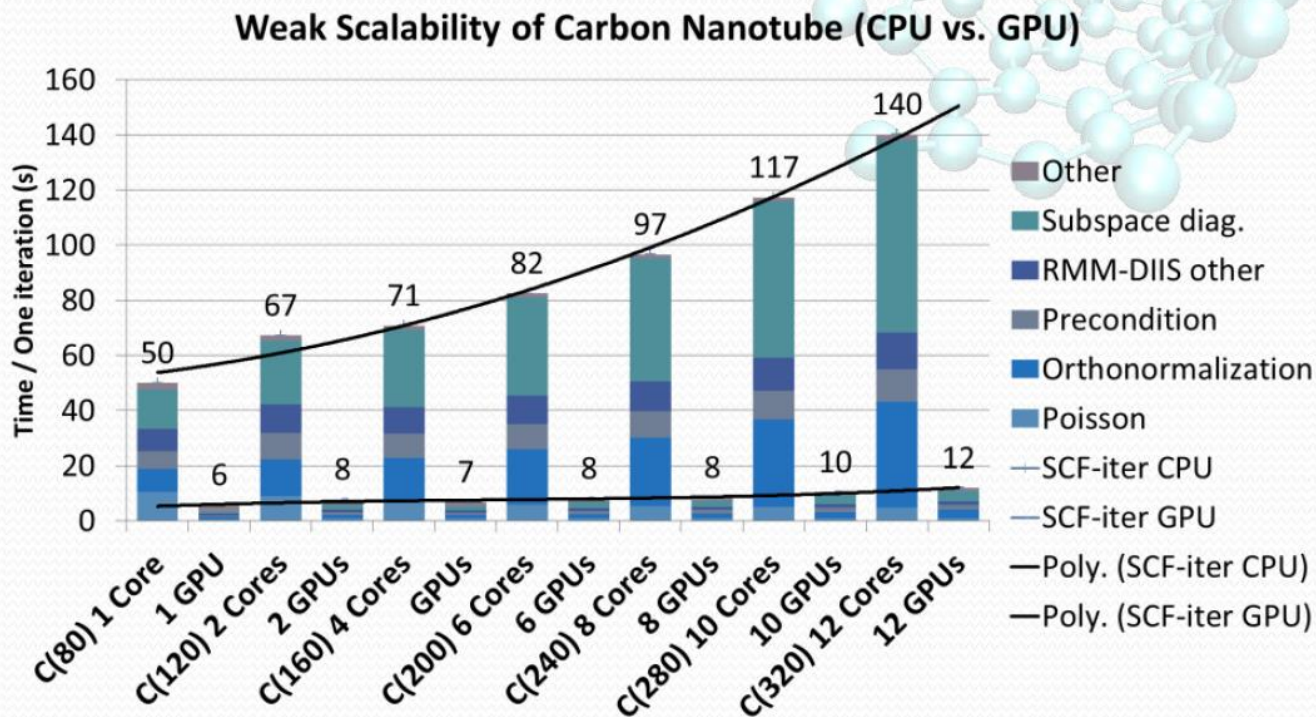
Multi-GPU Parallelization

- Parallelization is done with MPI
- Multiple GPUs can be used by domain decomposition or parallelization over k-points or spins
- Domain decomposition for the stencil operations involves exchanging boundary regions between neighboring nodes
- Communications between nodes require data movement: device memory → host memory → destinations node host memory → destinations node device memory.
- Overlaps receives, sends and computations in the middle part of the grid, BUT this causes issues with small grids
 - Small grids: Synchronous transfers
 - Medium grids: Asynchronous transfers
 - Large grids: Overlap calculations and asynchronous transfers
 - Combine of several wave functions and boundary regions into few large transfers



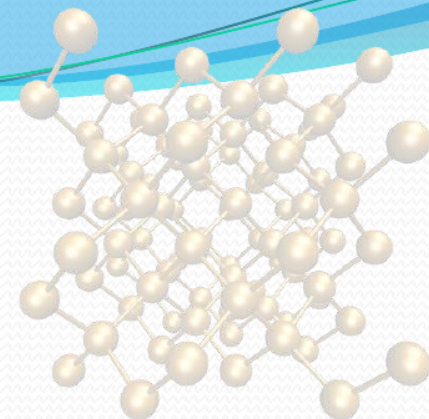
Weak Scalability (Carbon)

- The size of a carbon nanotube and the number of MPI tasks are varied from 80 atoms (240 states) to 320 atoms (1280 states) and 1 task to 12 tasks.
- Comparison between equal number of GPUs and CPU cores.
- CPU: Intel Xeon X5650 GPU: NVIDIA Tesla M2070
- Calculations performed on Vuori cluster at CSC

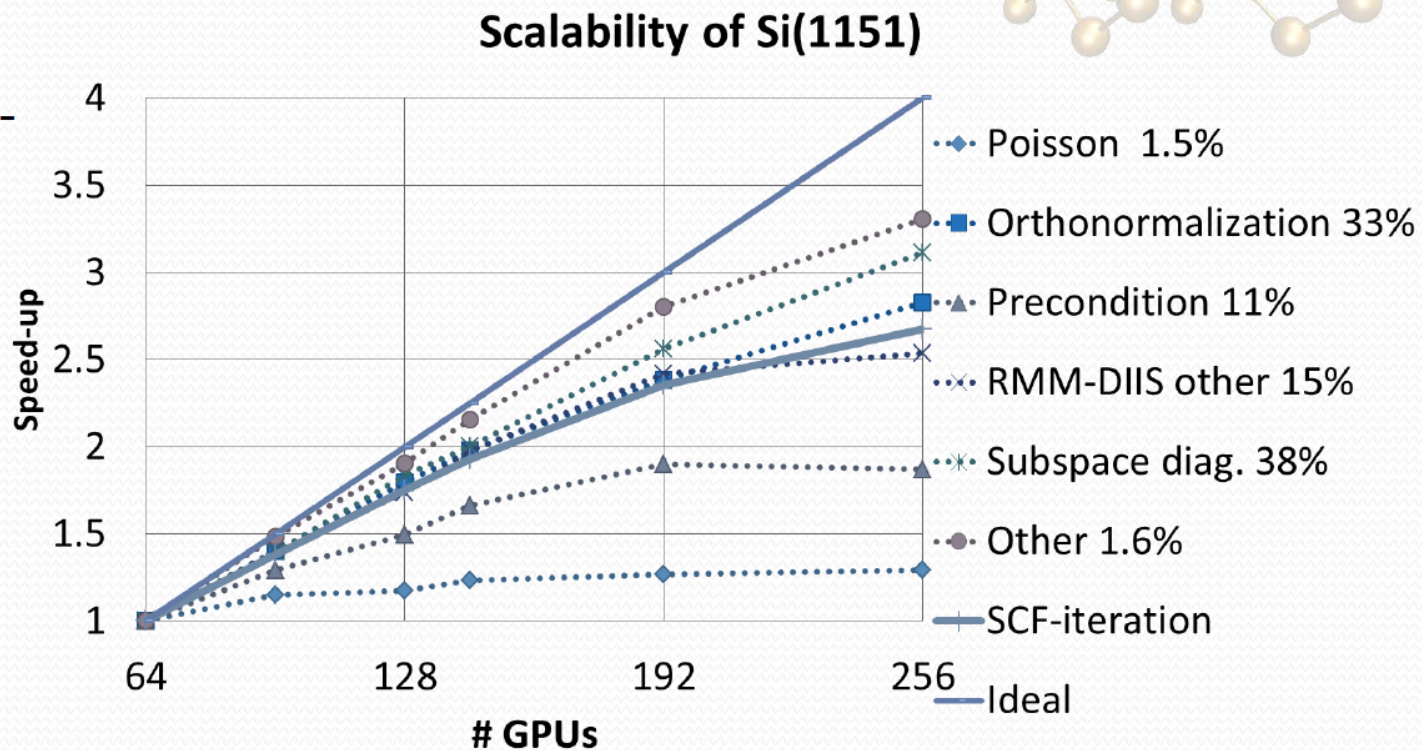


# MPI tasks	1	2	4	6	8	10	12
Speed-Up	8.8	8.7	10.5	10.2	11.5	11.3	11.9

Strong Scalability

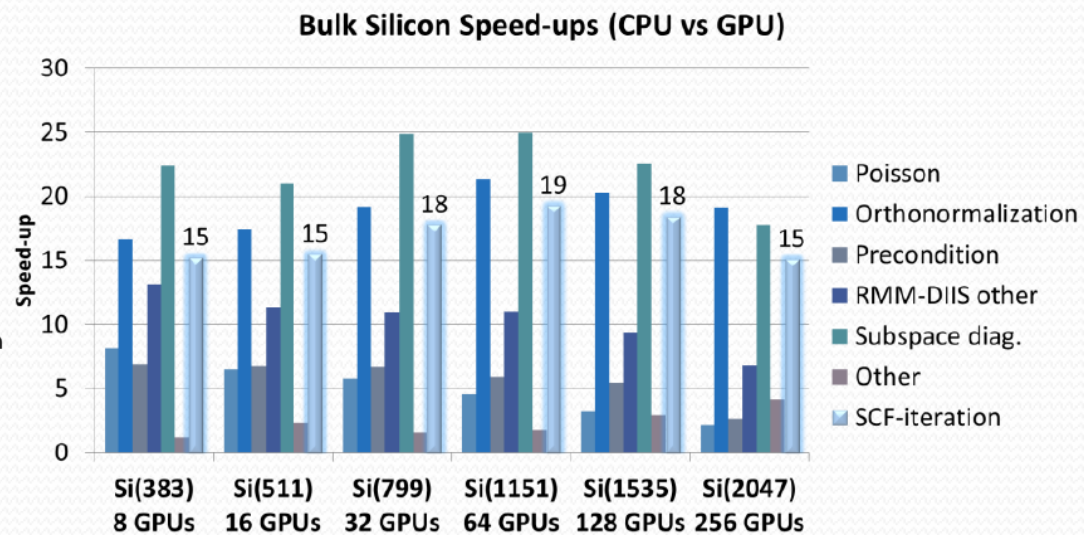
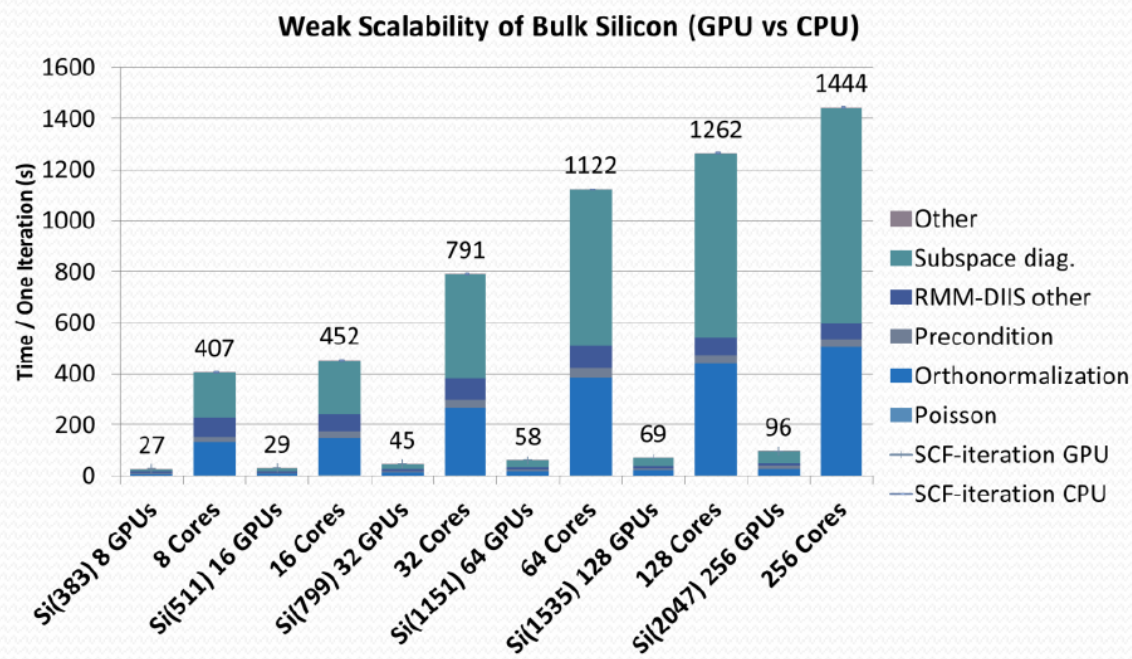


- Bulk silicon with 1151 atoms with periodic boundary conditions, 4604 bands and 1 k-point in the Brillouin zone.
- The number of GPUs is increased from 64 to 256.
- Grid size: $164 \times 164 \times 108$
- Speed-up comparison to 64 GPUs.
- NVIDIA Tesla M2090
- Calculations performed on CURIE cluster in France at GENCI/CEA



Weak Scalability (Silicon)

- The size of bulk silicon system and the number of MPI tasks are varied from 383 atoms (1532 bands) to 2046 atoms (8188 bands) and 8 task to 256 tasks with periodic boundary conditions.
- The largest system requires about 1.3TB of memory for calculations.
- CPU: Intel Xeon E5640 GPU: NVIDIA Tesla M2090



Random Phase Approximation

GPAW Random Phase Approximation (RPA) code:

- 6000 lines of python, 1000 lines of C/CUDA (and re-uses many GPAW functions)
- Better than DFT for correlated materials, but more computationally expensive
- Useful for oxides, Van der Waals systems, etc.

GPU Techniques:

- Use BLAS₃ “zherk” instead of BLAS₂ “zher”
- Batch FFTs
- GPU kernels parallelized over atoms/bands/projector-functions
- No thinking: all calculations on GPU

Preliminary ((GPU+CPU)/CPU) speedup for 202-electron N₂-on-Ru: 30x

Work by Jun Yan, Lin Li, Christopher O’Grady (Stanford/SLAC)

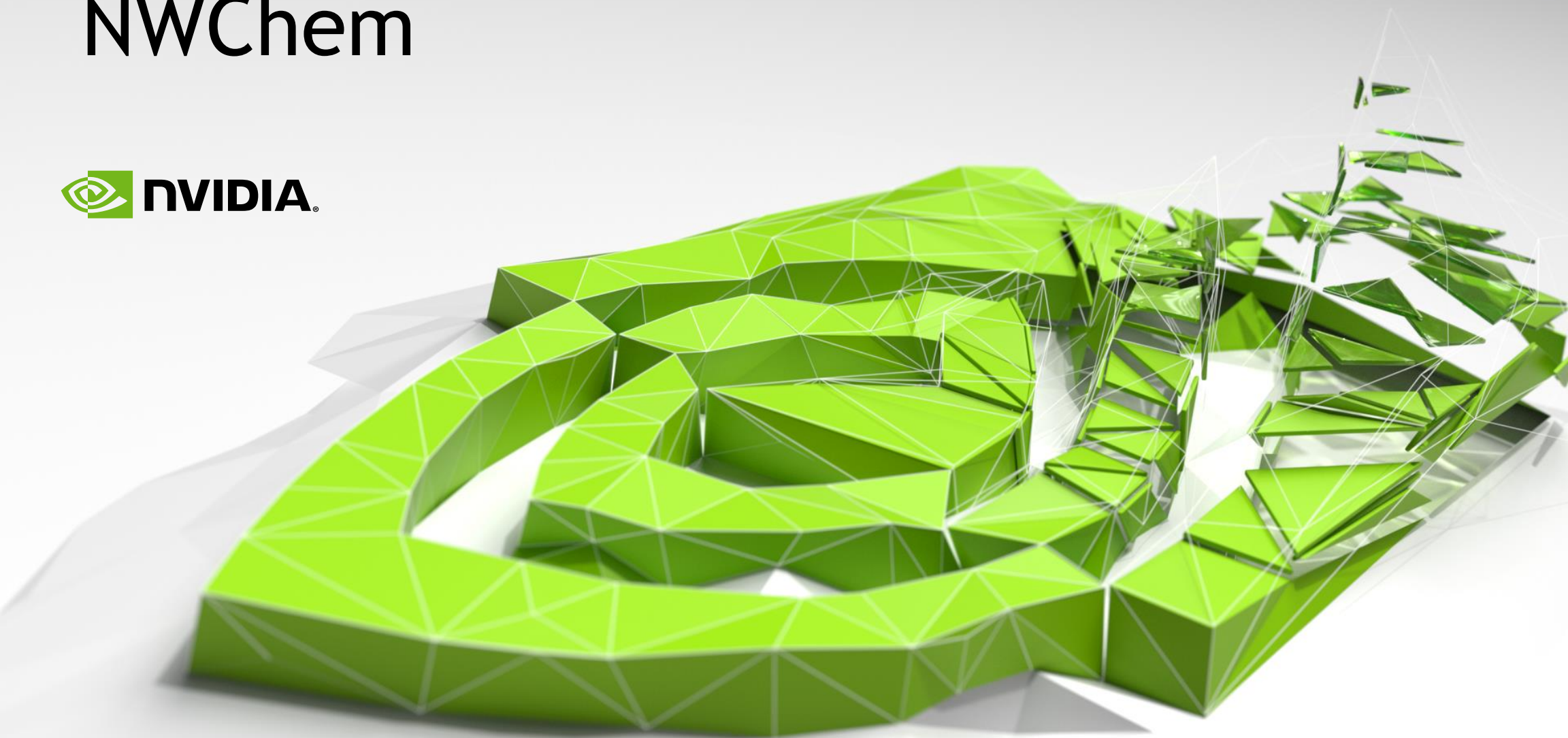


Summary

- We have accelerated the most numerically intensive parts of ground state DFT calculations
- Overall speed-ups in our tests varied from 8.8 to 19 depending on system size.
- Our multi-GPU implementation scales well even on large hybrid clusters.
- Code is available at GPAW Subversion repository.
- Acknowledgements to CSC and PRACE for computing resources

Hakala S., Havu V., Enkovaara J., Nieminen R. M. "Parallel Electronic Structure Calculations Using Multiple Graphics Processing Units (GPUs)" In: Manninen, P., Öster, P. (eds.) PARA 2012. LNCS, vol. 7782, pp. 63--76. Springer, Heidelberg (2013)

NWChem

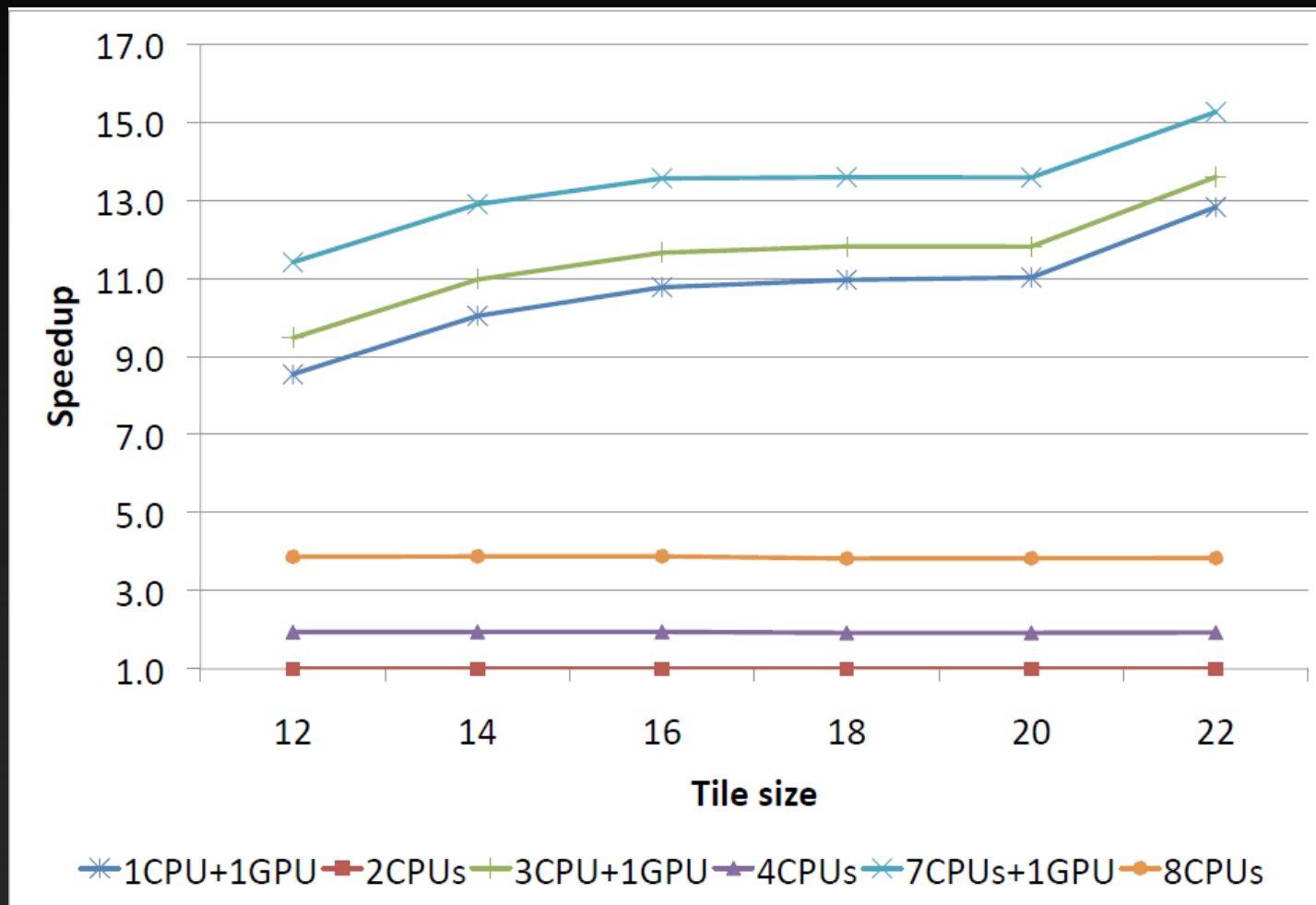


NWChem 6.3 Release with GPU Acceleration



- Addresses large complex and challenging molecular-scale scientific problems in the areas of catalysis, materials, geochemistry and biochemistry on highly scalable, parallel computing platforms to obtain the fastest time-to-solution
- Researchers can for the first time be able to perform large scale coupled cluster with perturbative triples calculations utilizing the NVIDIA GPU technology. A highly scalable multi-reference coupled cluster capability will also be available in NWChem 6.3.
- The software, released under the Educational Community License 2.0, can be downloaded from the NWChem website at www.nwchem-sw.org

NWChem - Speedup of the non-iterative calculation for various configurations/tile sizes



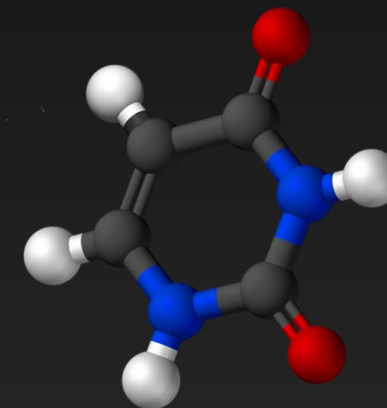
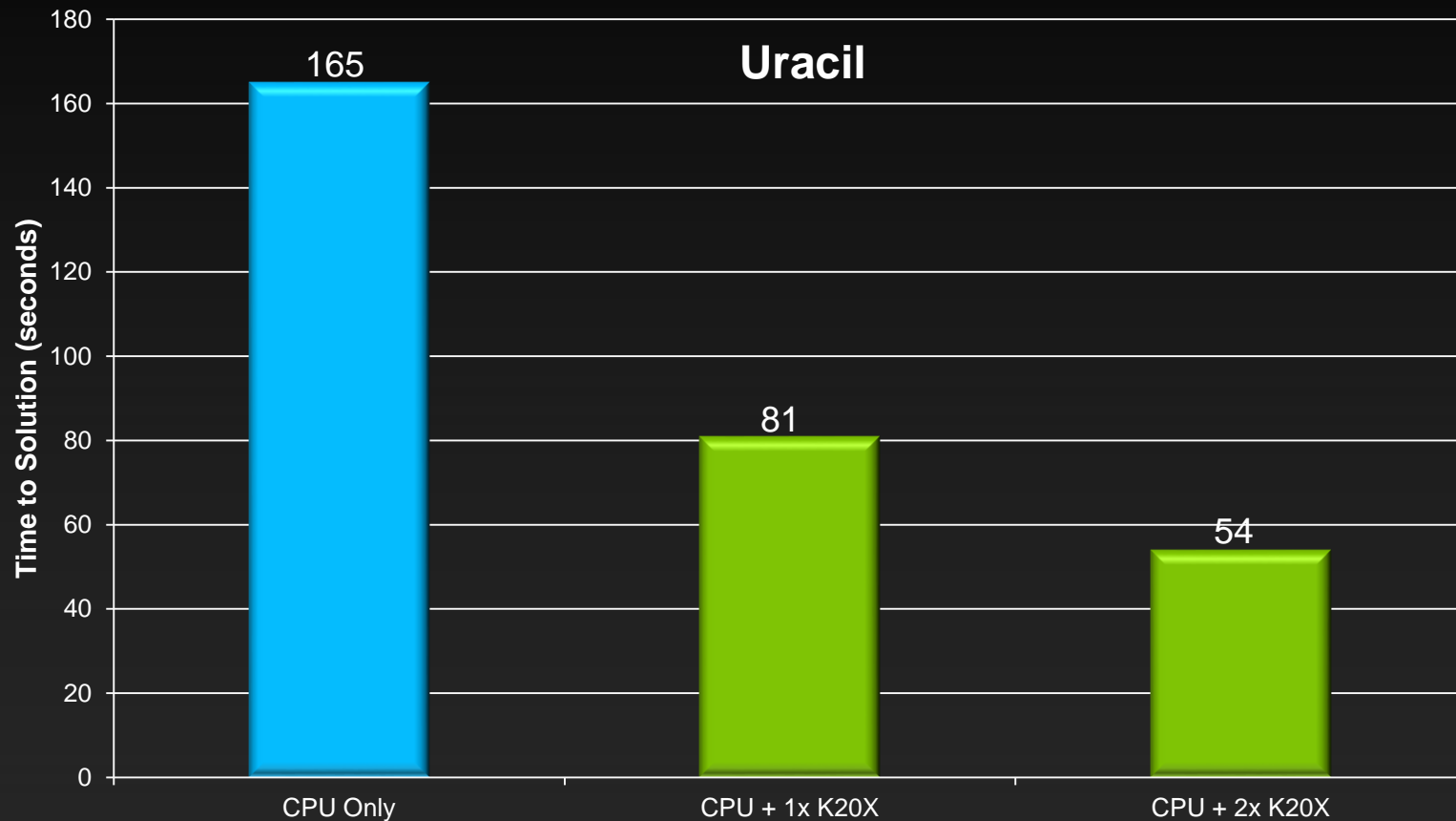
System: cluster consisting of dual-socket nodes constructed from:

- 8-core AMD Interlagos processors
- 64 GB of memory
- Tesla M2090 (Fermi) GPUs

The nodes are connected using a high-performance QDR Infiniband interconnect

Courtesy of Kowolski, K., Bhaskaran-Nair, et al @ PNNL, JCTC (submitted)

Kepler, Faster Performance (NWChem)

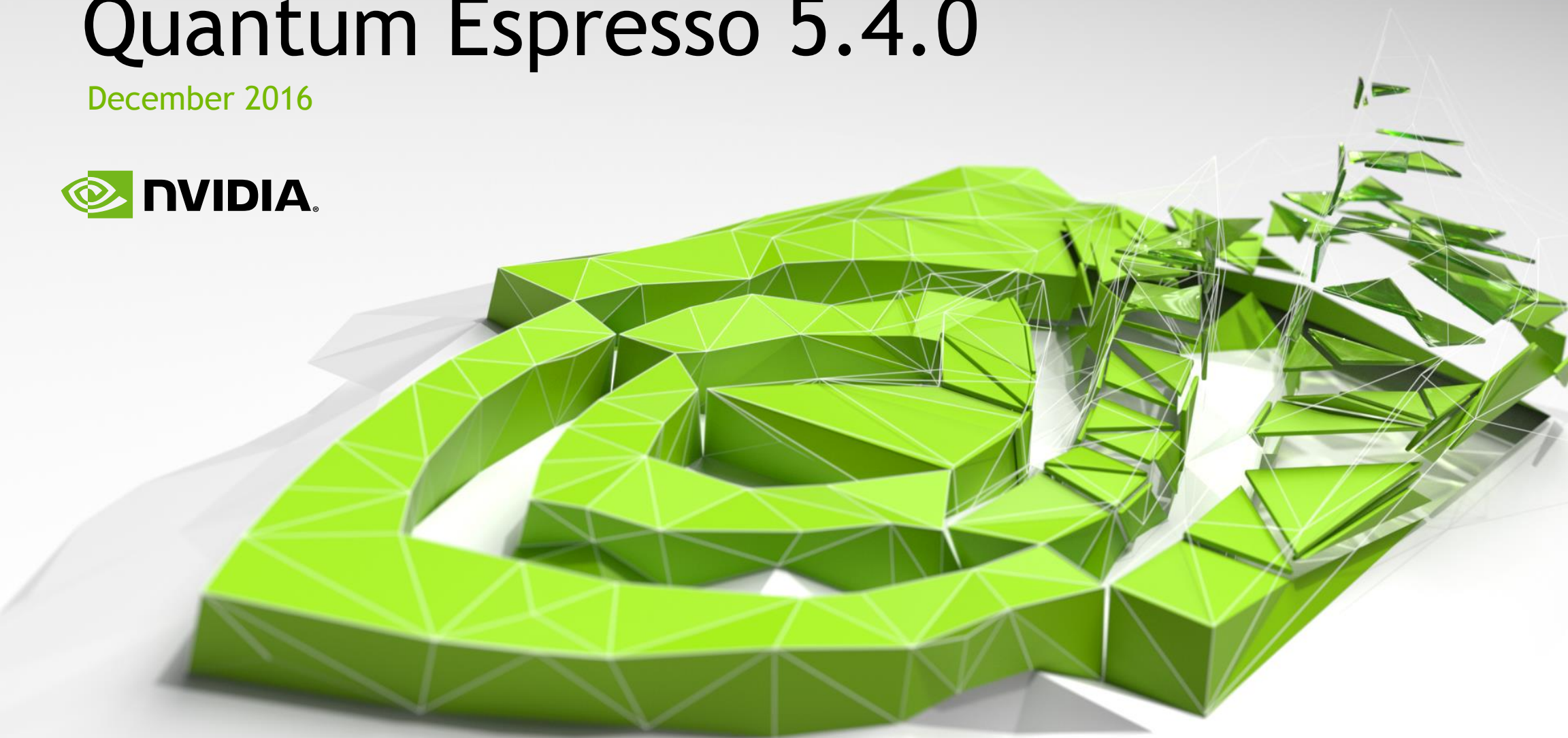


Uracil Molecule

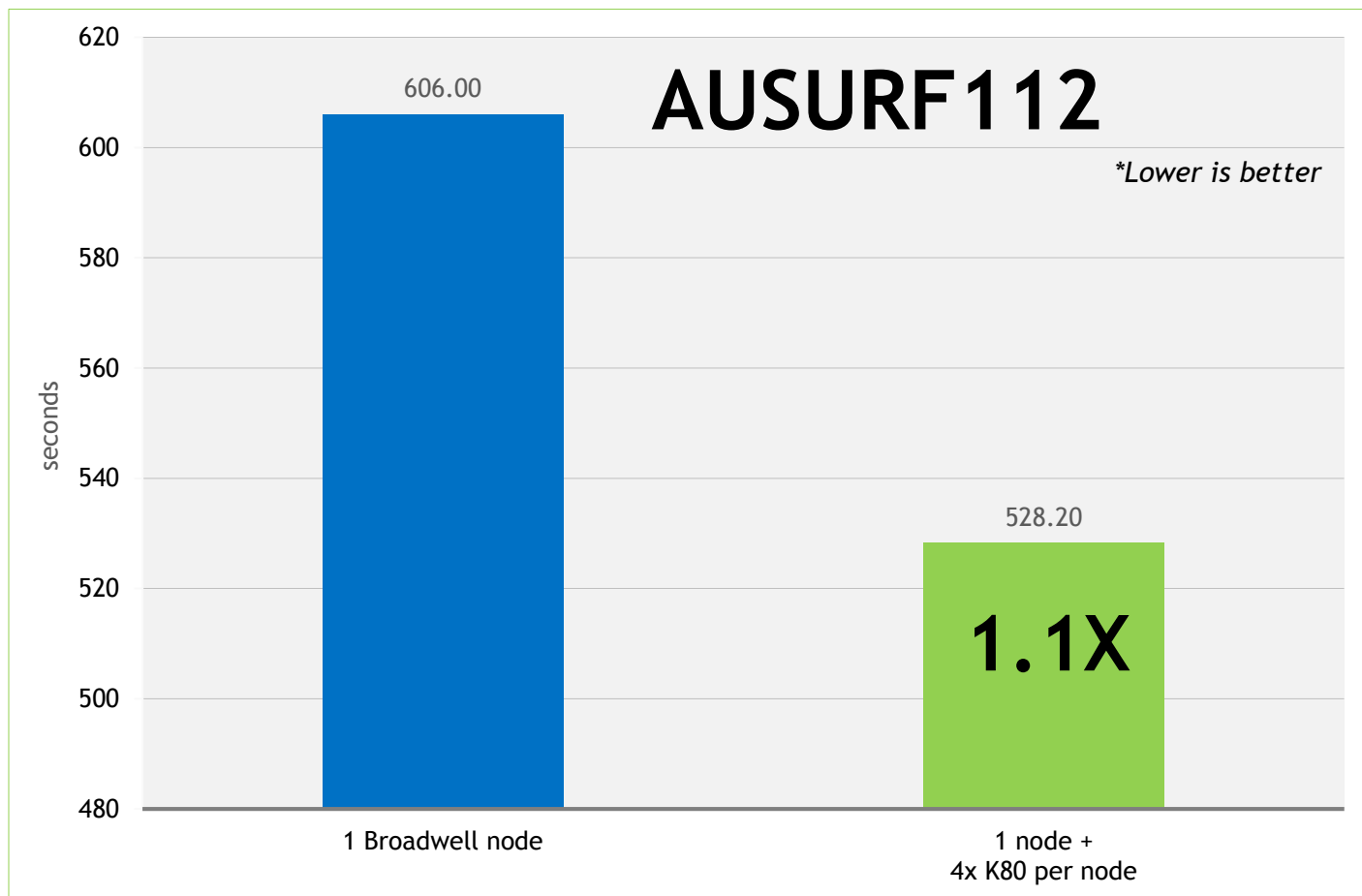
Performance improves by **2x** with one GPU and by **3.1x** with 2 GPUs

Quantum Espresso 5.4.0

December 2016



AUSURF112 on K80s

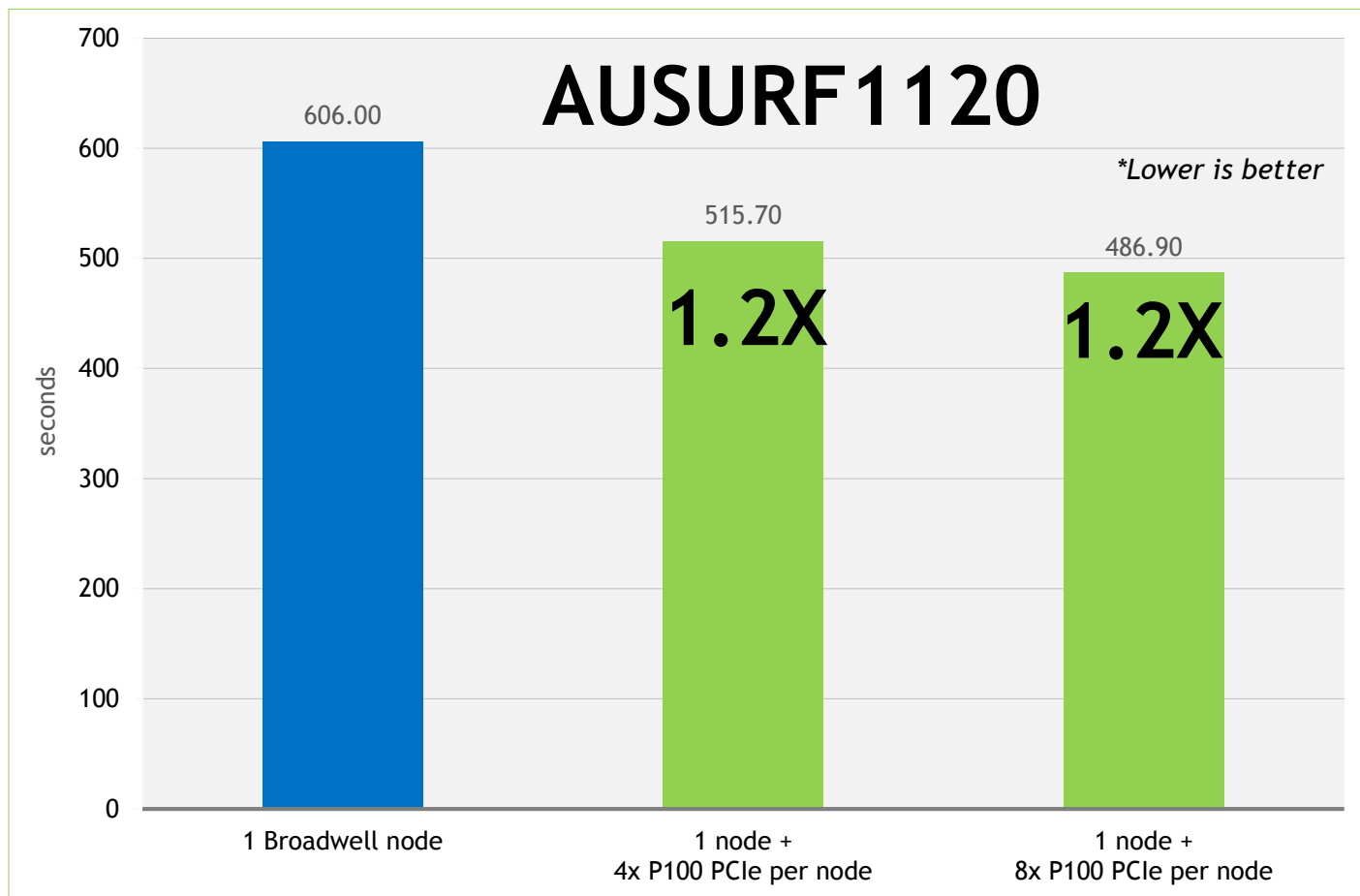


Running **Quantum Espresso** version 5.4.0

The **blue node** contains Dual Intel Xeon E5-2690 v4@2.6GHz [3.5GHz Turbo] (Broadwell) CPUs

The **green node** contains Dual Intel Xeon E5-2690 v4@2.6GHz [3.5GHz Turbo] (Broadwell) CPUs + Tesla K80 (autoboost) GPUs

AUSURF112 on P100s PCIe



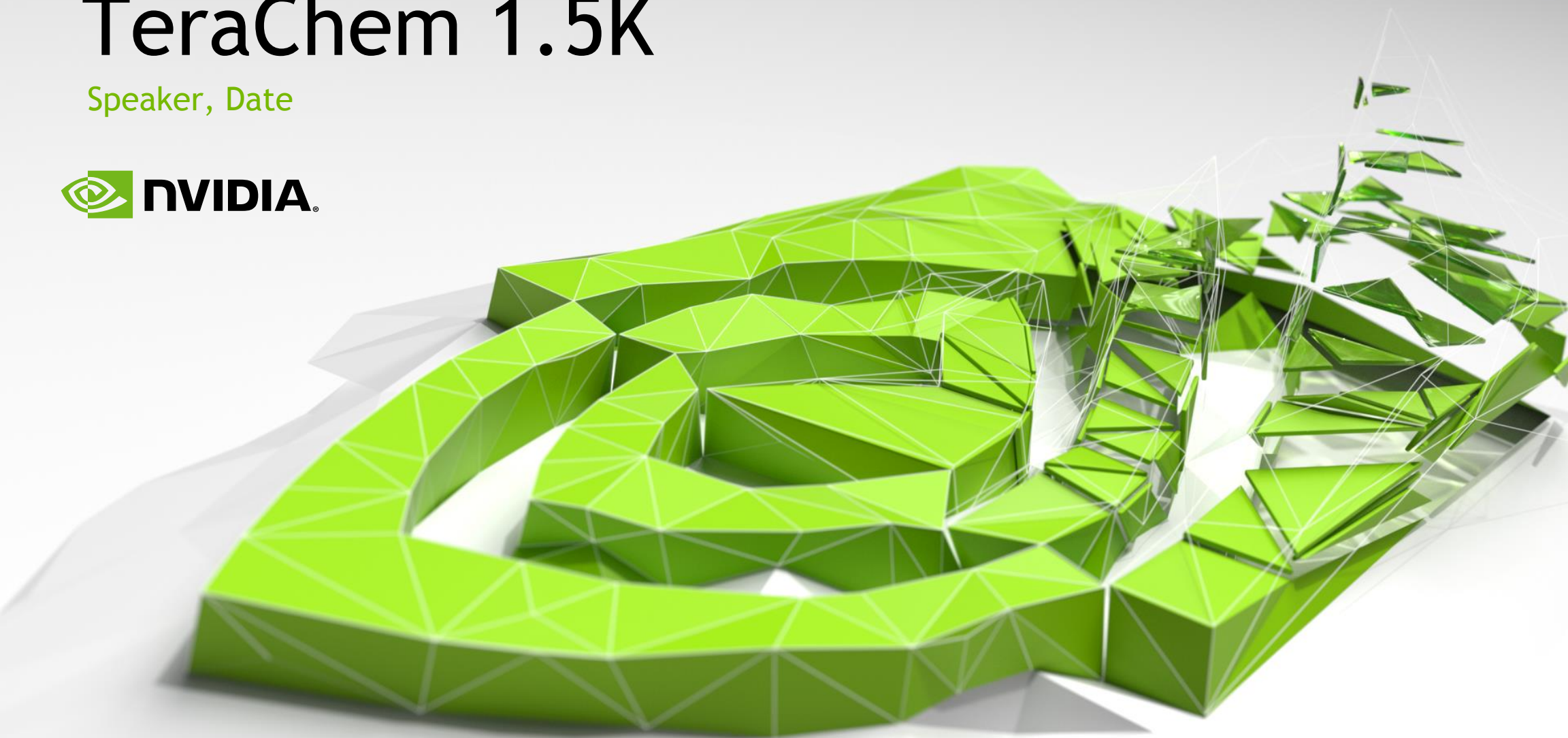
Running **Quantum Espresso** version 5.4.0

The **blue node** contains Dual Intel Xeon E5-2690 v4@2.6GHz [3.5GHz Turbo] (Broadwell) CPUs

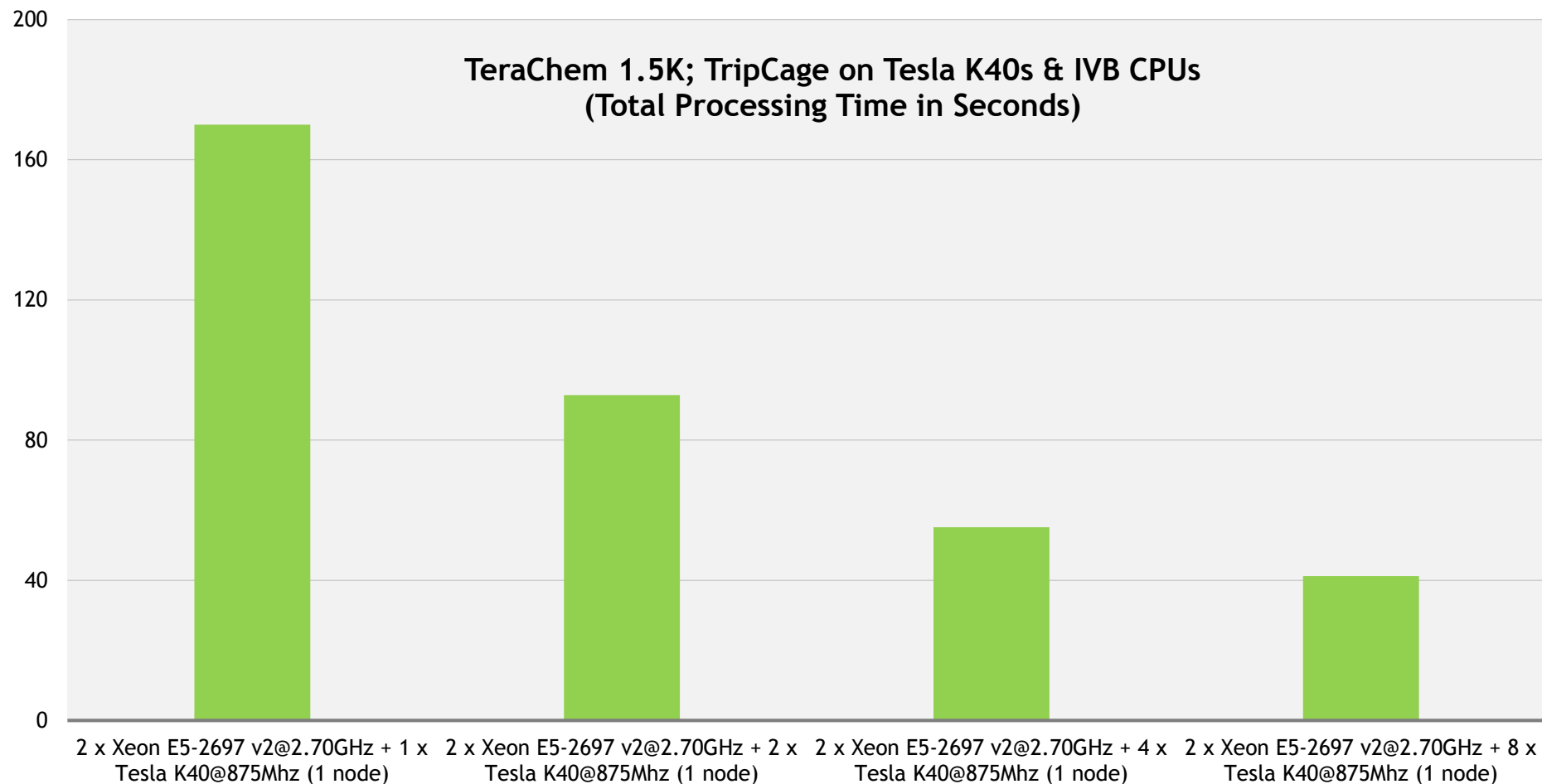
The **green nodes** contain Dual Intel Xeon E5-2690 v4@2.6GHz [3.5GHz Turbo] (Broadwell) CPUs + Tesla P100 PCIe GPUs

TeraChem 1.5K

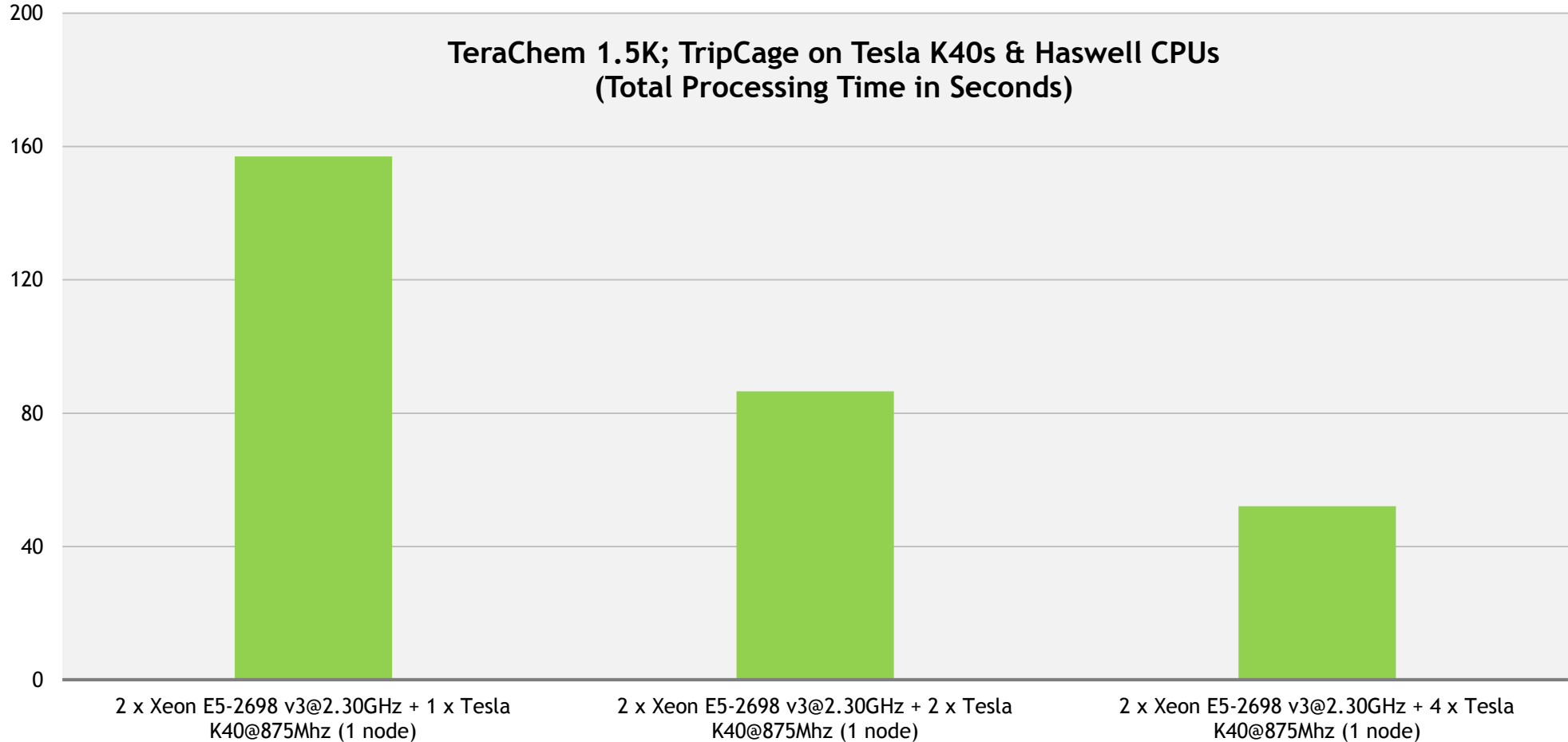
Speaker, Date



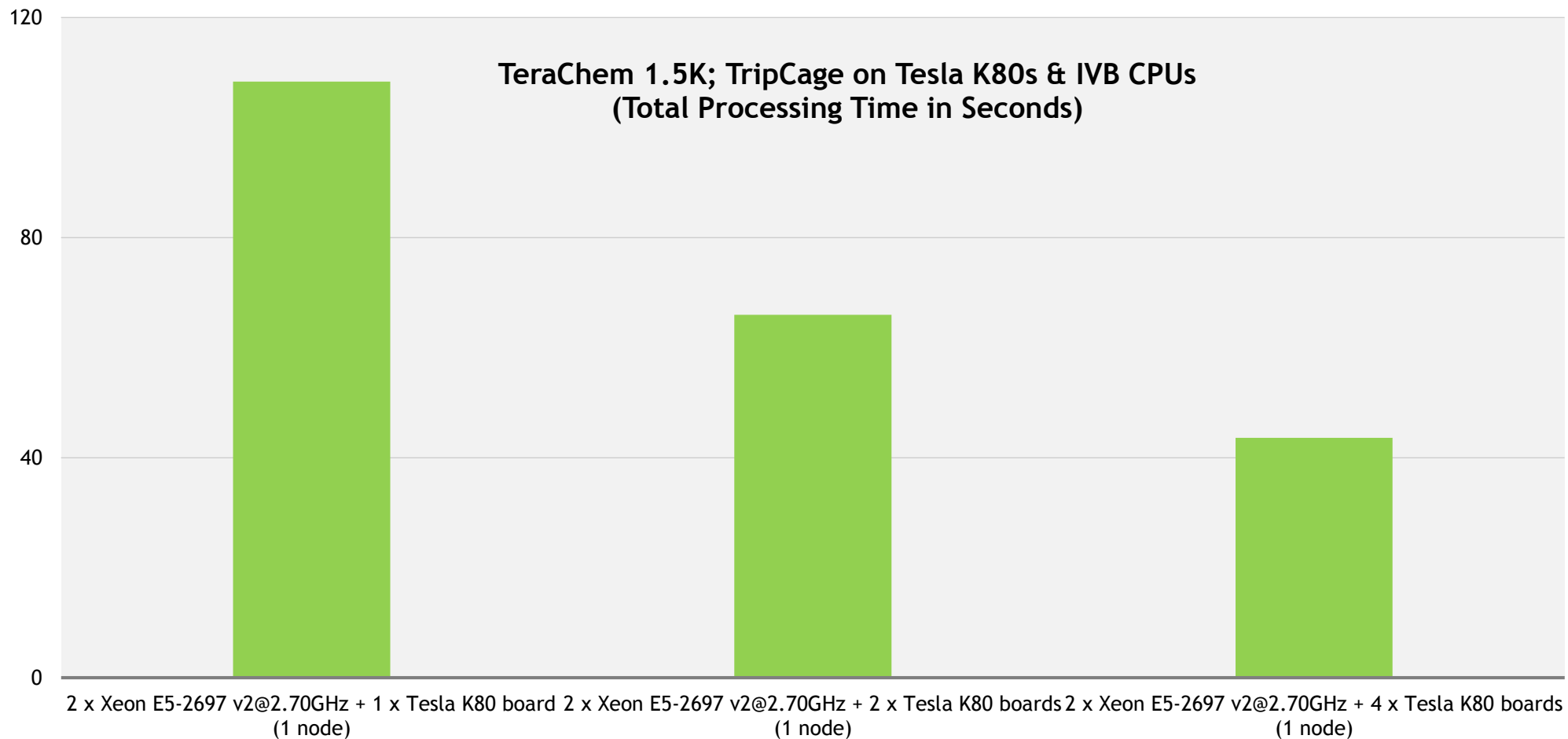
TERACHEM 1.5K; TRIPCAGE ON TESLA K40S



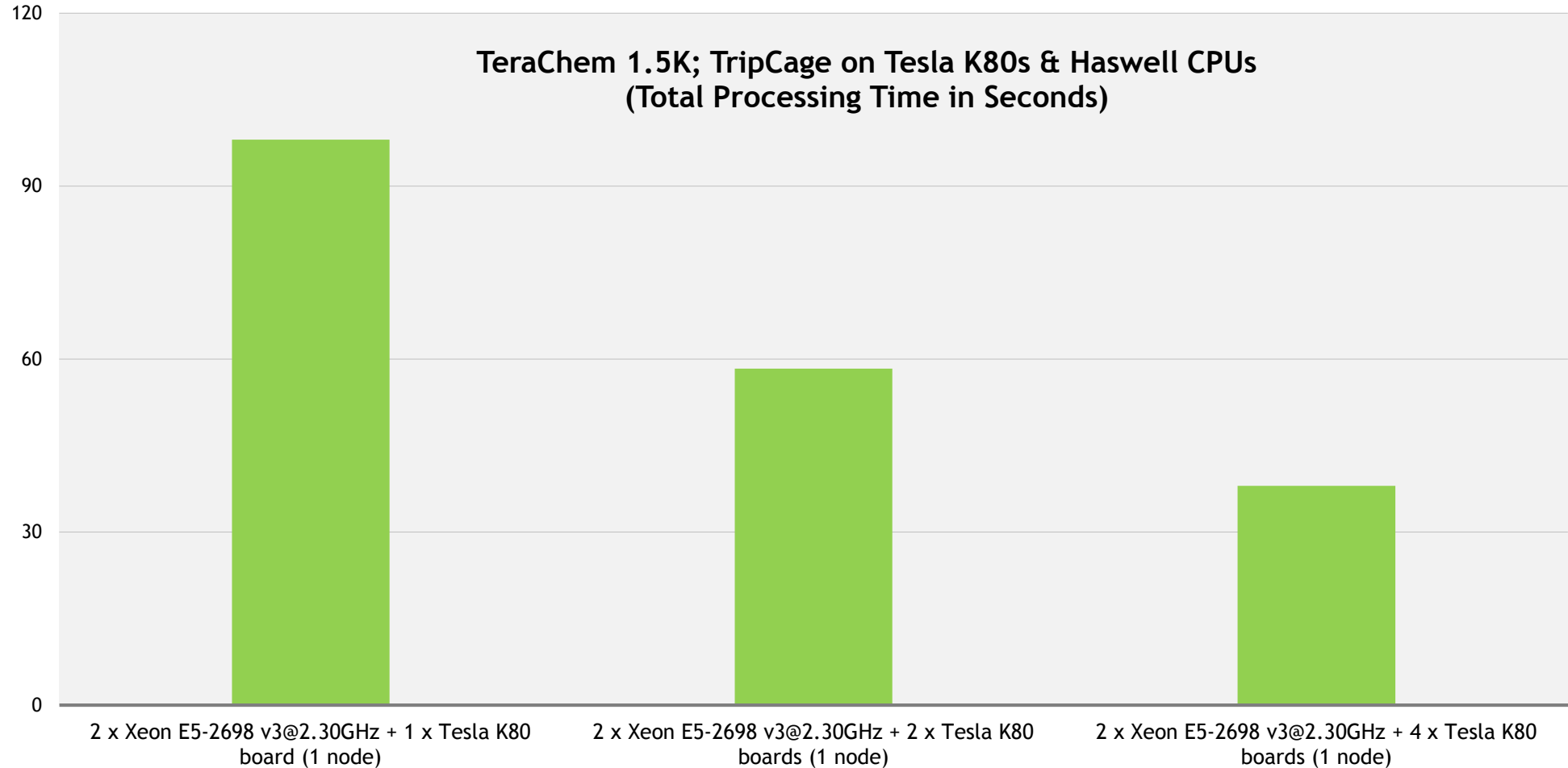
TERACHEM 1.5K; TRIPCAGE ON TESLA K40S & HASWELL CPUS



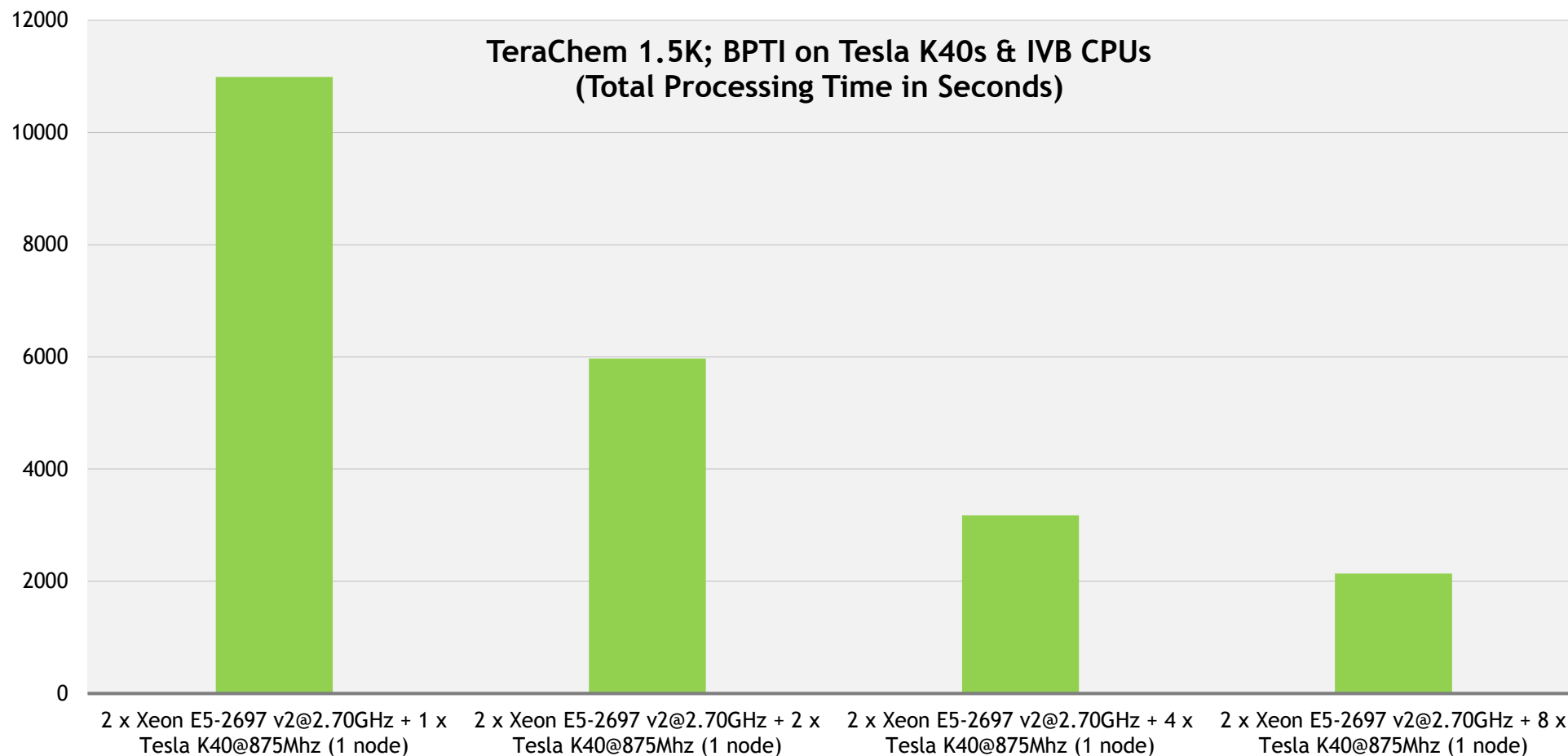
TERACHEM 1.5K; TRIPCAGE ON TESLA K80S & IVB CPUS



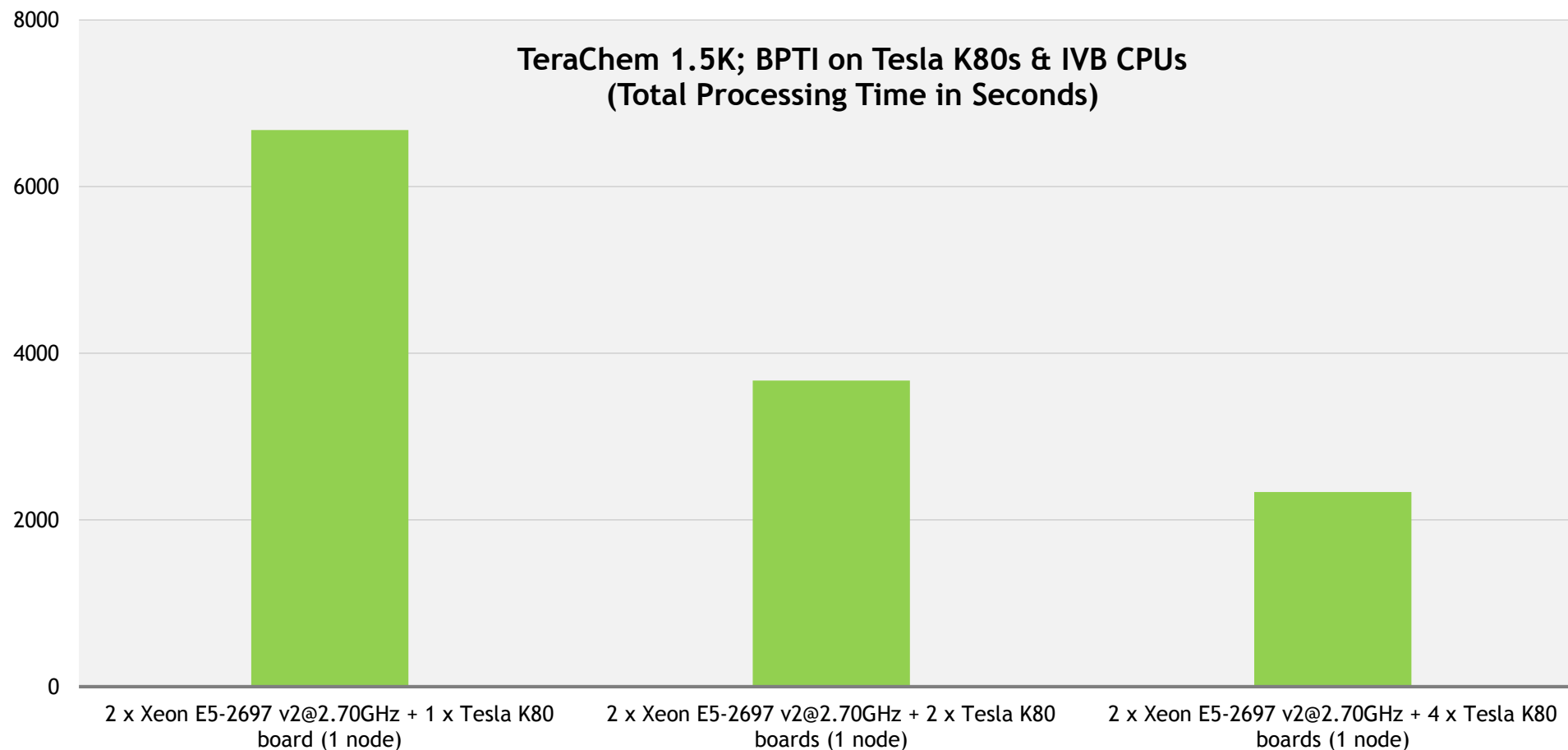
TERACHEM 1.5K; TRIPCAGE ON TESLA K80S & HASWELL CPUS



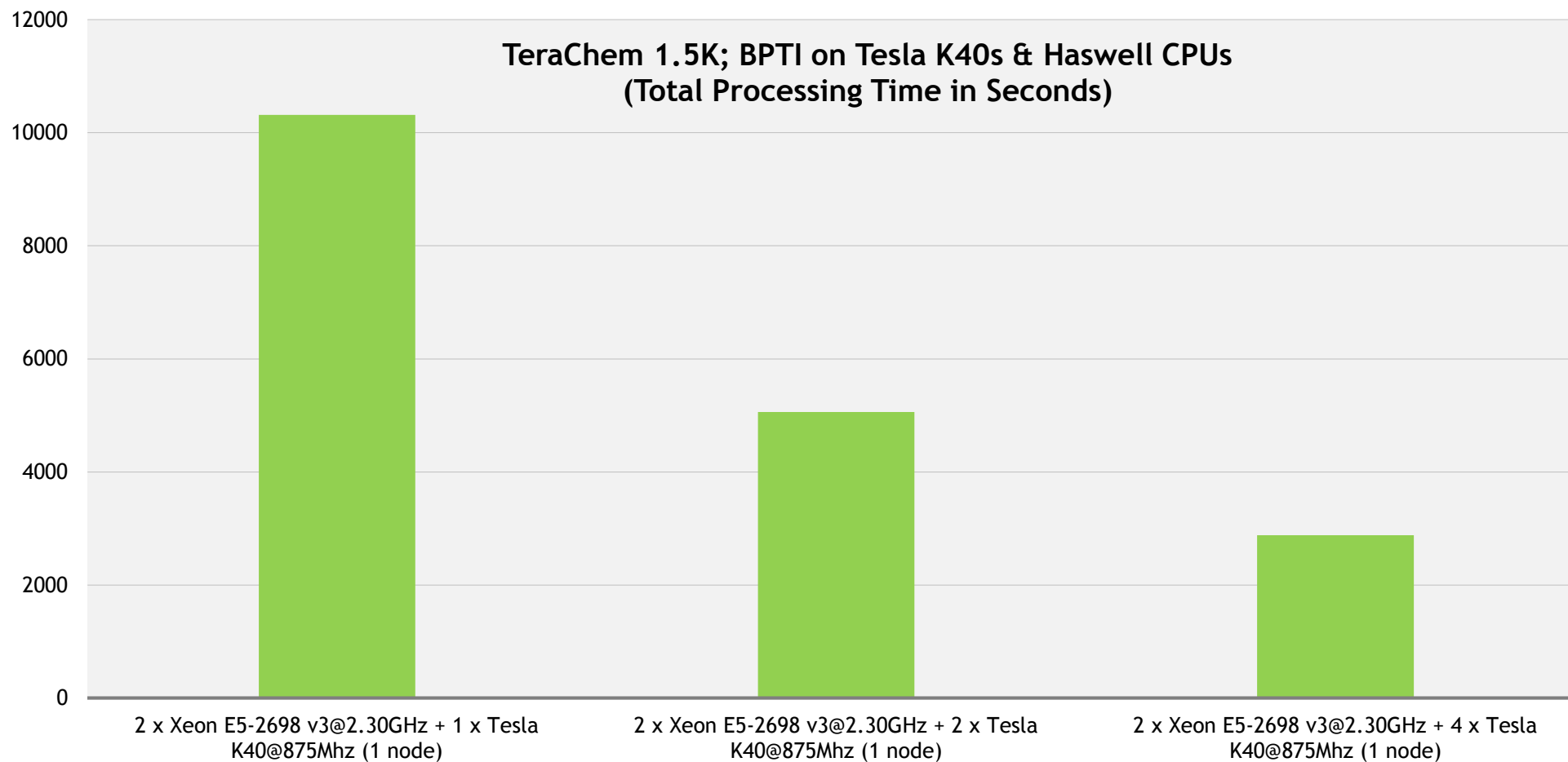
TERACHEM 1.5K; BPTI ON TESLA K40S & IVB CPUS



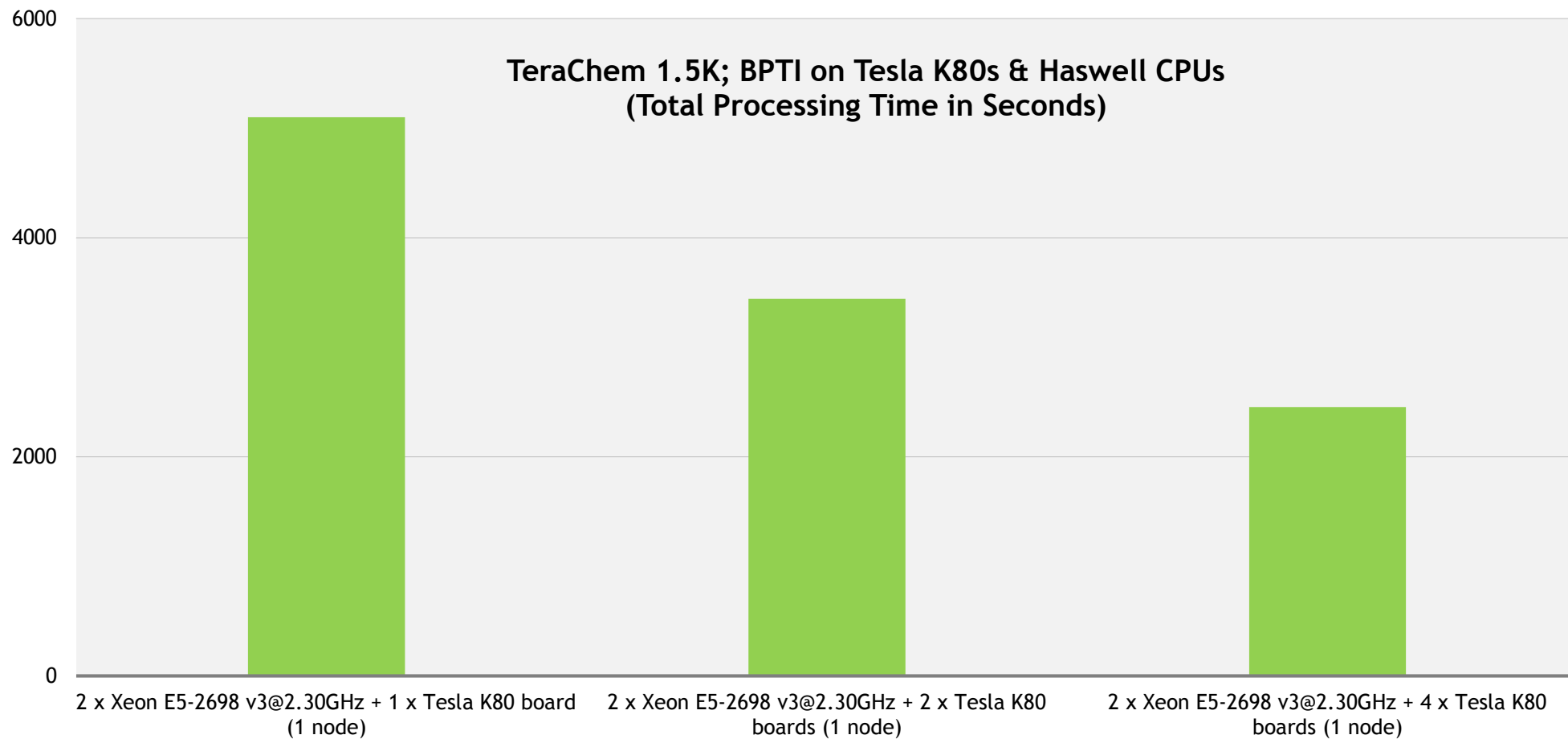
TERACHEM 1.5K; BPTI ON TESLA K80S & IVB CPUS



TERACHEM 1.5K; BPTI ON TESLA K40S & IVB CPUS



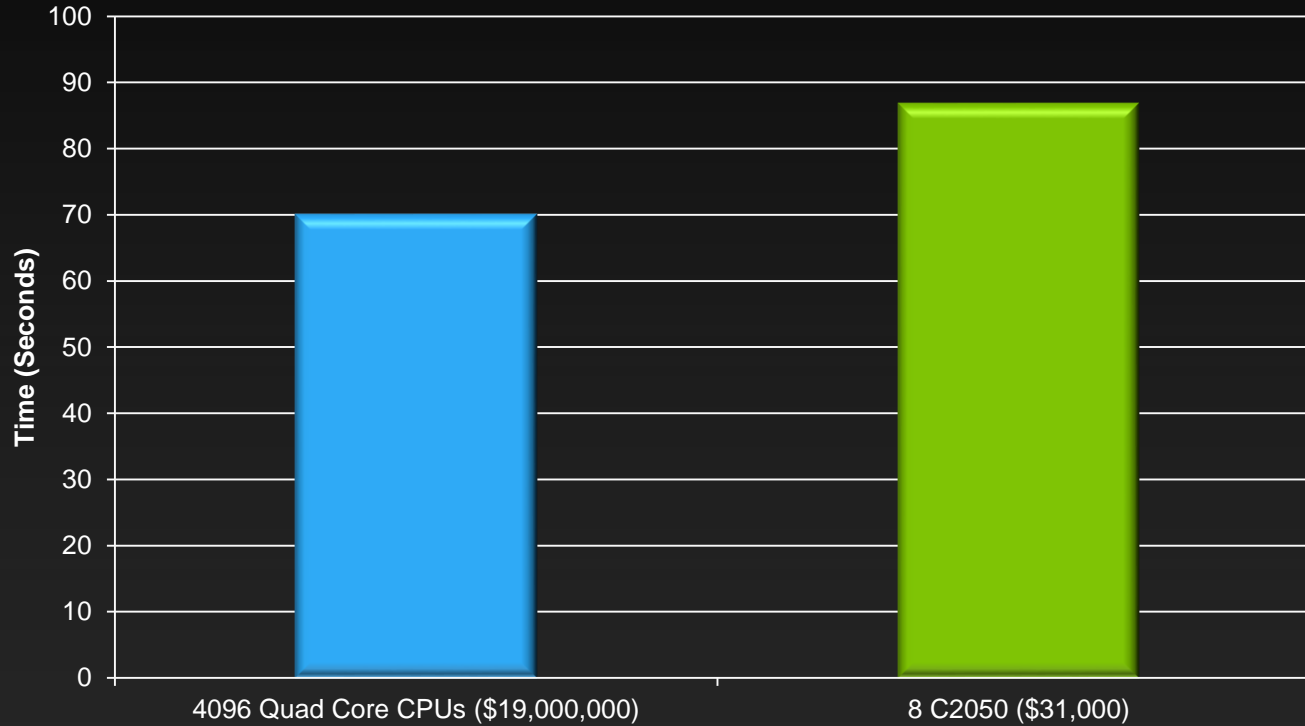
TERACHEM 1.5K; BPTI ON TESLA K80S & HASWELL CPUS



TeraChem Supercomputer Speeds on GPUs



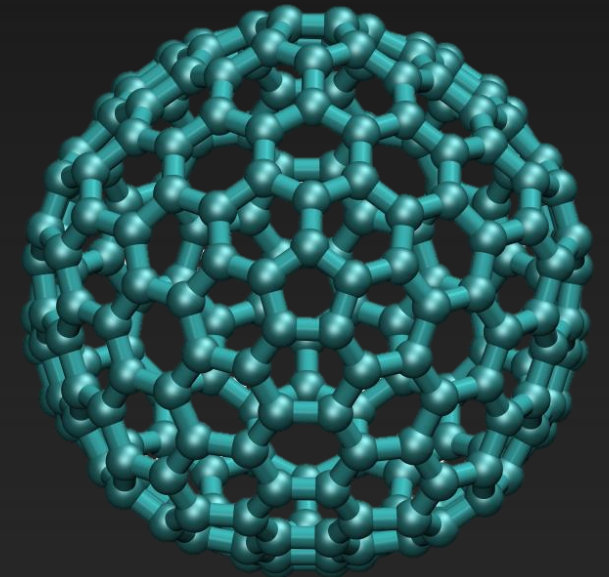
Time for SCF Step



TeraChem running on 8 C2050s on 1 node

NWChem running on 4096 Quad Core CPUs
In the Chinook Supercomputer

Giant Fullerene C₂₄₀ Molecule



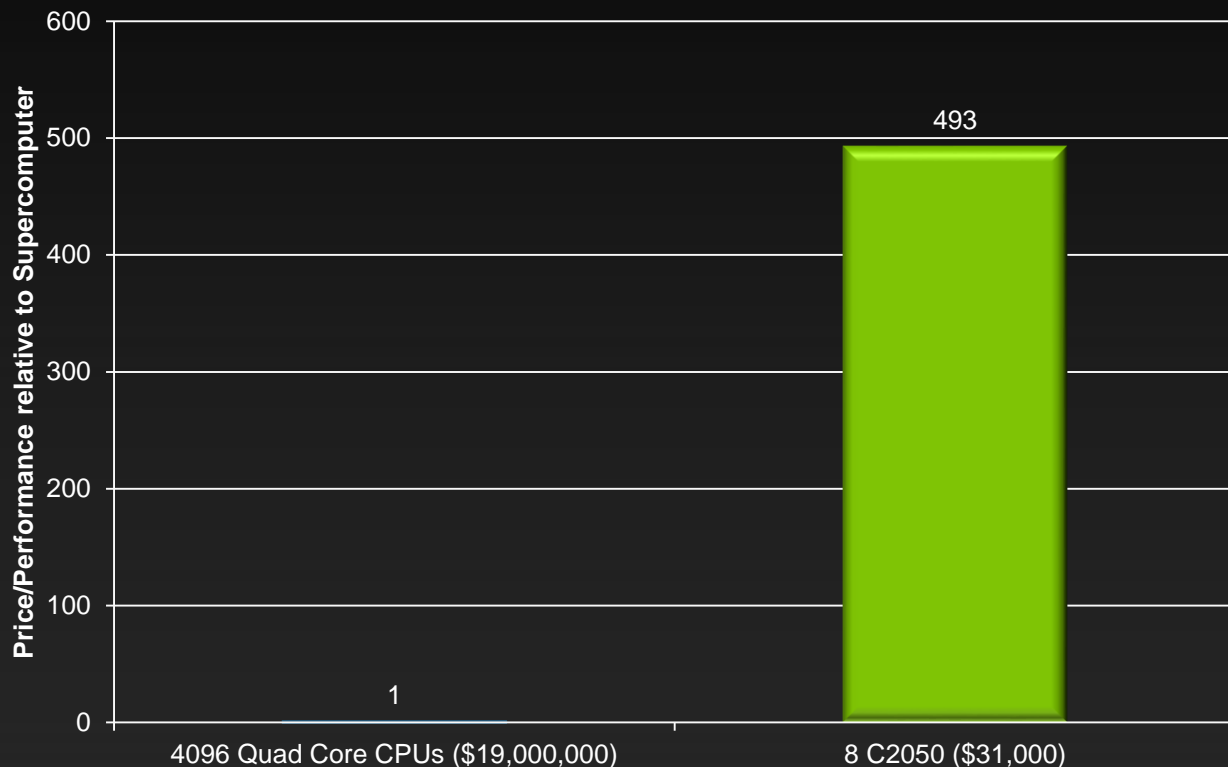
Similar performance from just a handful of GPUs

TeraChem

Bang for the Buck



Performance/Price

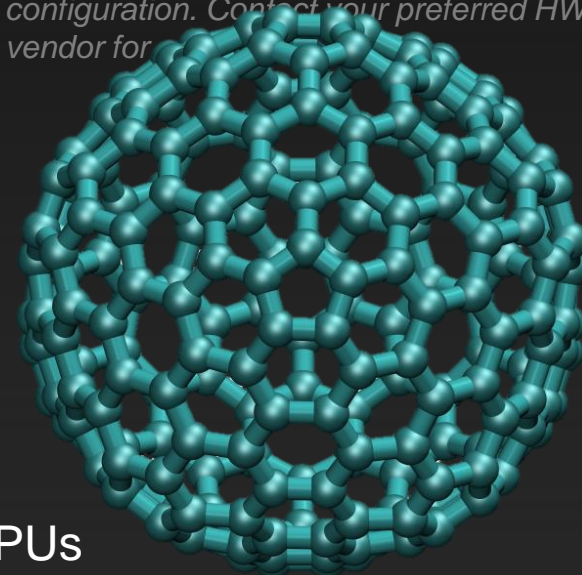


TeraChem running on 8 C2050s on 1 node

NWChem running on 4096 Quad Core CPUs
In the Chinook Supercomputer

Giant Fullerene C₂₄₀ Molecule

Note: Typical CPU and GPU node pricing used. Pricing may vary depending on node configuration. Contact your preferred HW vendor for

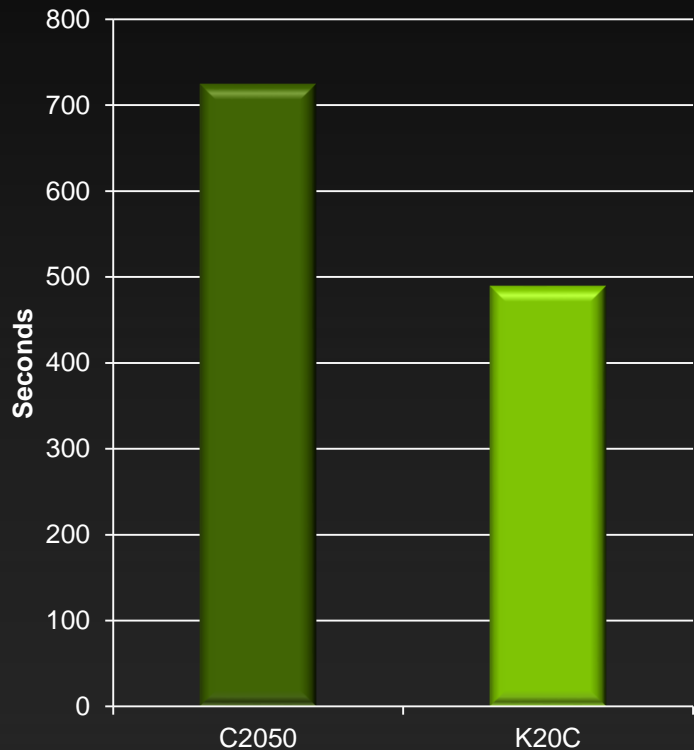


Dollars spent on GPUs do 500x more science than those spent on CPUs

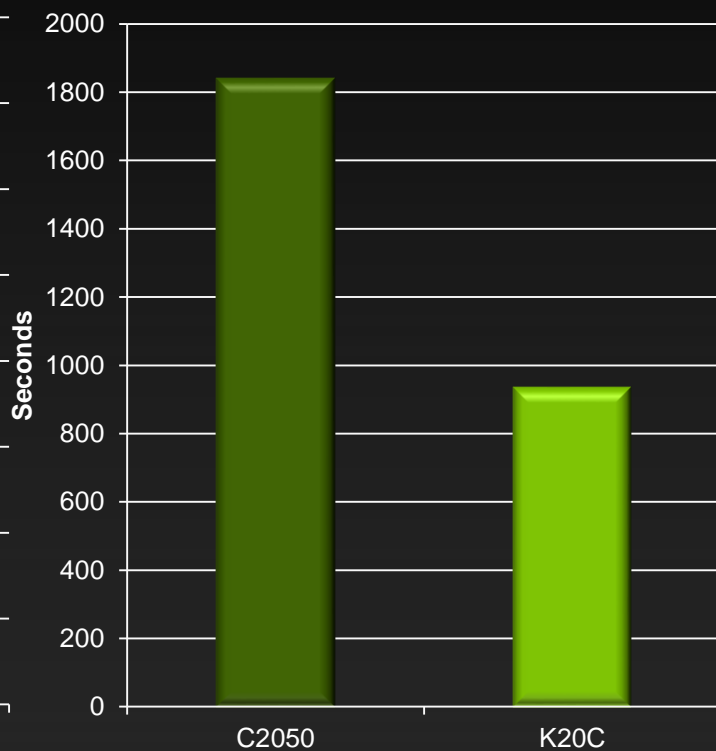
Kepler's Even Better



Olestra BLYP 453 Atoms

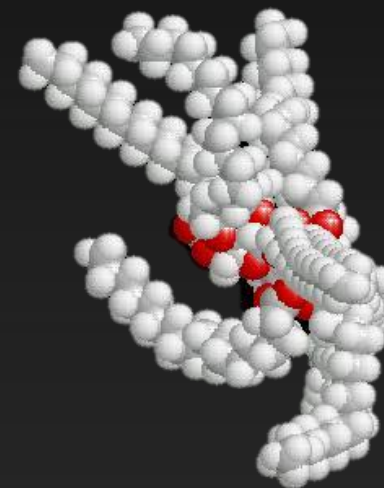


B3LYP/6-31G(d)



TeraChem running on C2050 and K20C

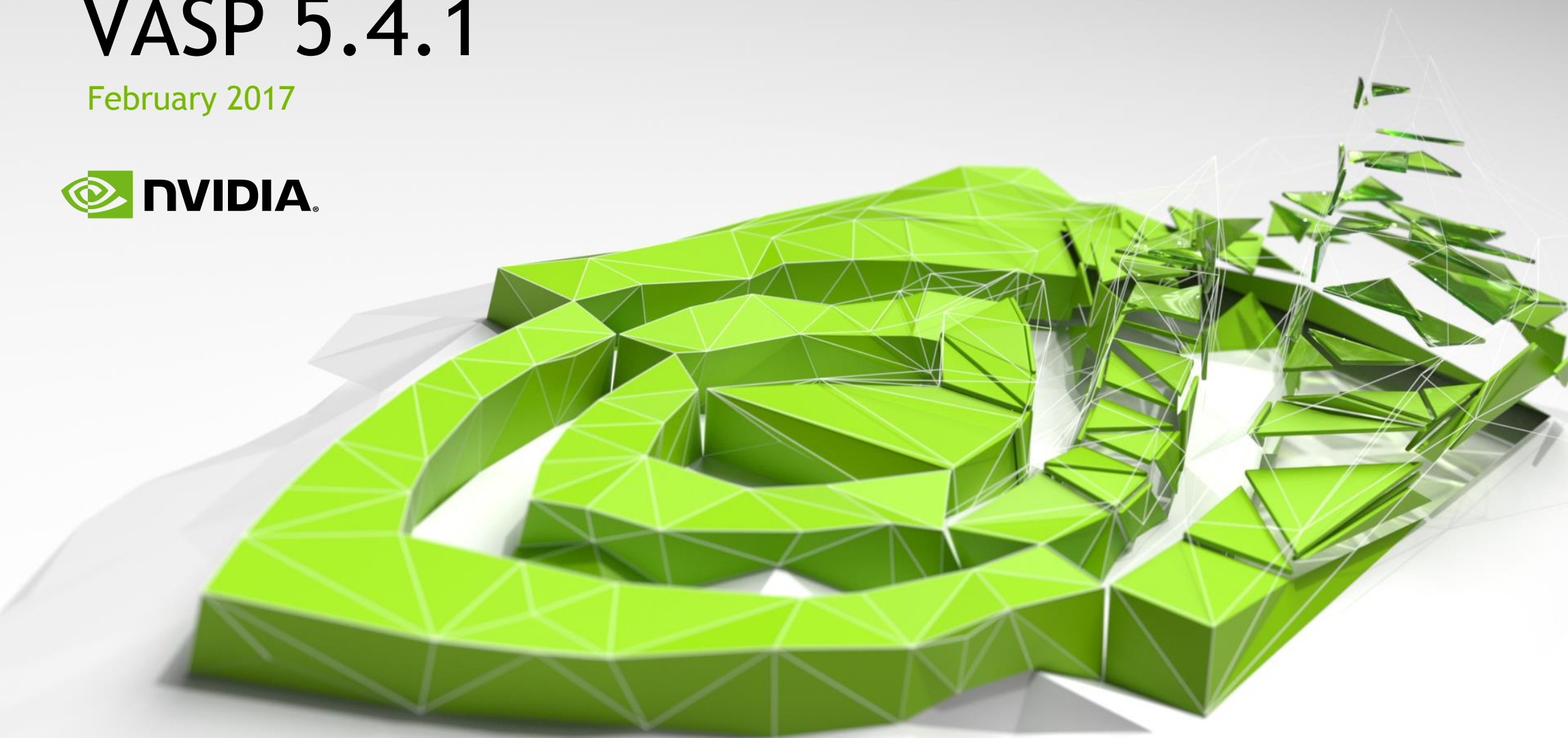
First graph is of BLYP/G-31(d)
Second is B3LYP/6-31G(d)



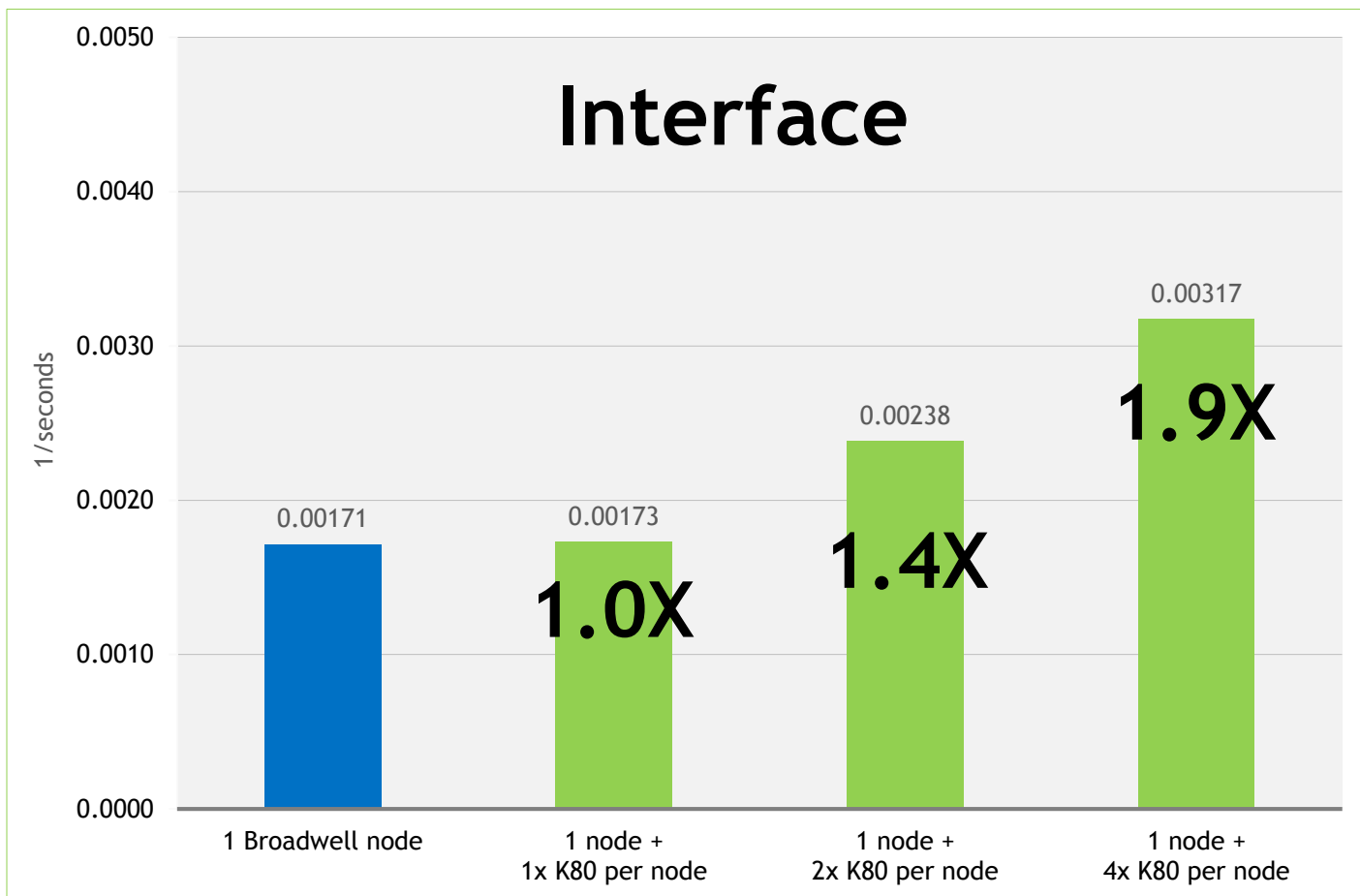
Kepler performs **2x faster** than Tesla

VASP 5.4.1

February 2017



Interface on K80s



Running **VASP** version 5.4.1

The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

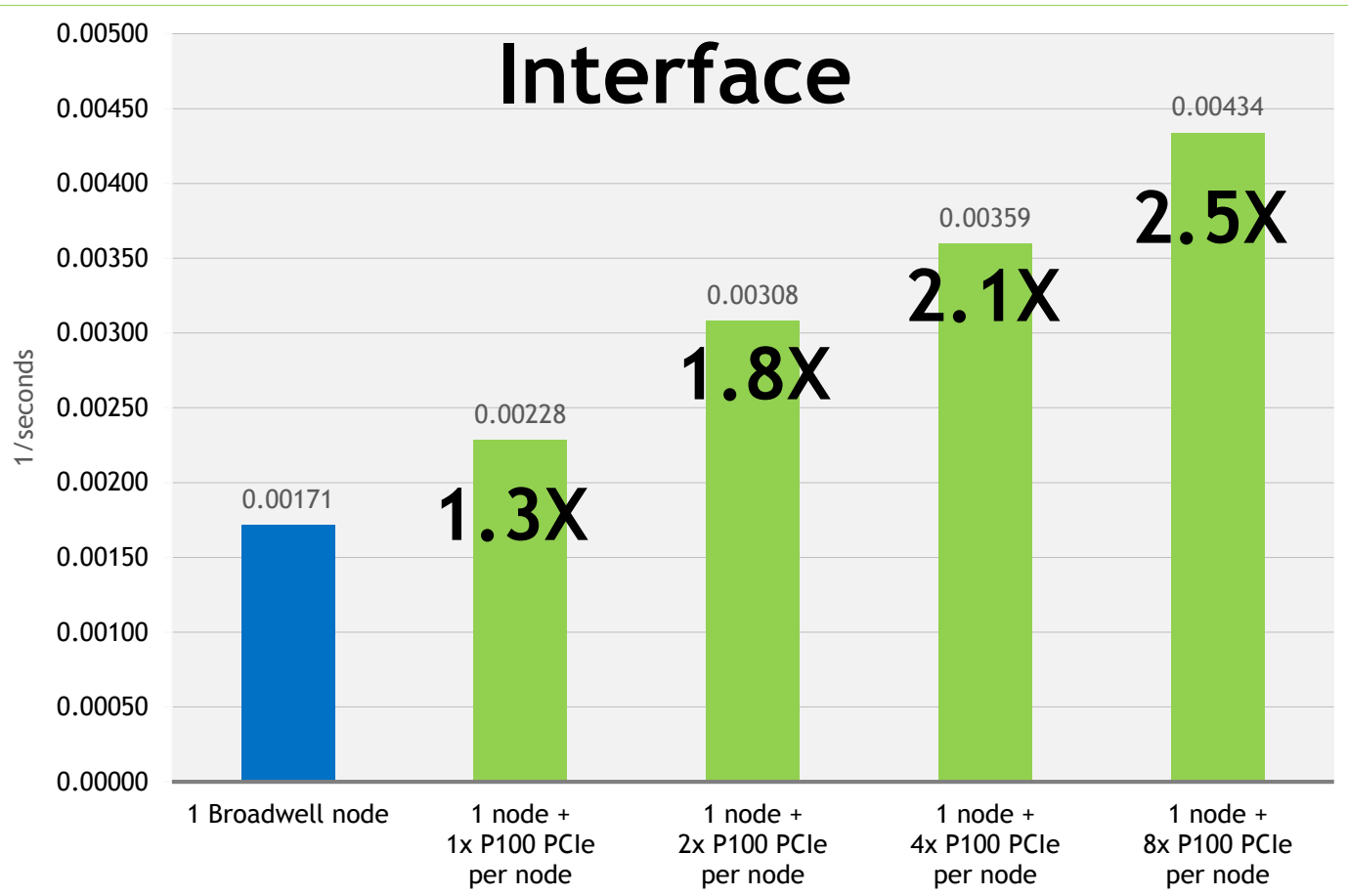
The **green nodes** contain Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla K80 (autoboost) GPUs

➤ 1x K80 is paired with Single Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

Interface between a platinum slab Pt(111) (108 atoms) and liquid water (120 water molecules) (468 ions)

1256 bands
762048 plane waves
ALGO = Fast (Davidson + RMM-DIIS)

Interface on P100s PCIe



Running **VASP** version 5.4.1

The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

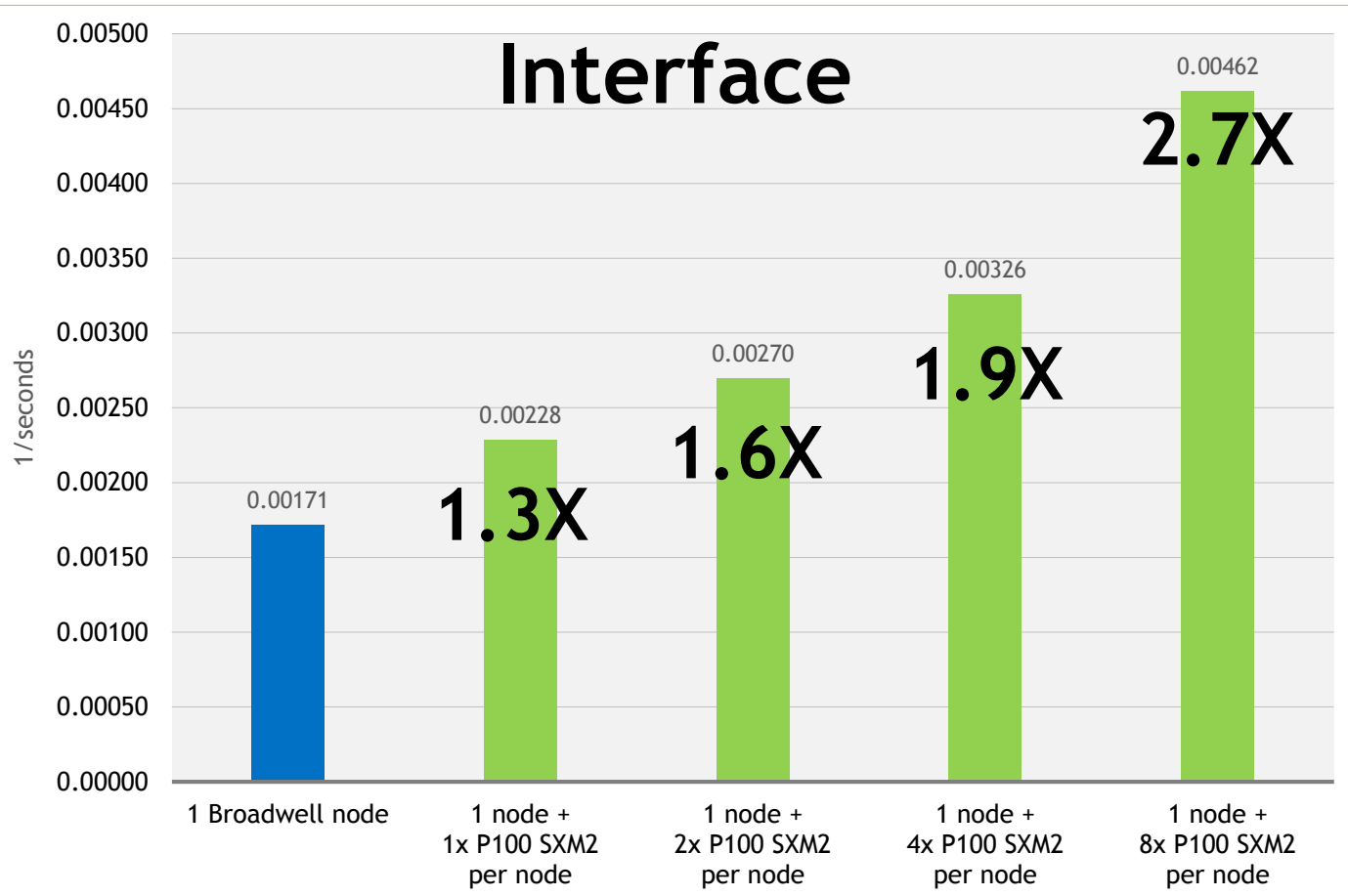
The **green nodes** contain Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla P100 PCIe GPUs

➤ 1x P100 PCIe is paired with Single Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

Interface between a platinum slab Pt(111) (108 atoms) and liquid water (120 water molecules) (468 ions)

1256 bands
762048 plane waves
ALGO = Fast (Davidson + RMM-DIIS)

Interface on P100s SXM2



Running **VASP** version 5.4.1

The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

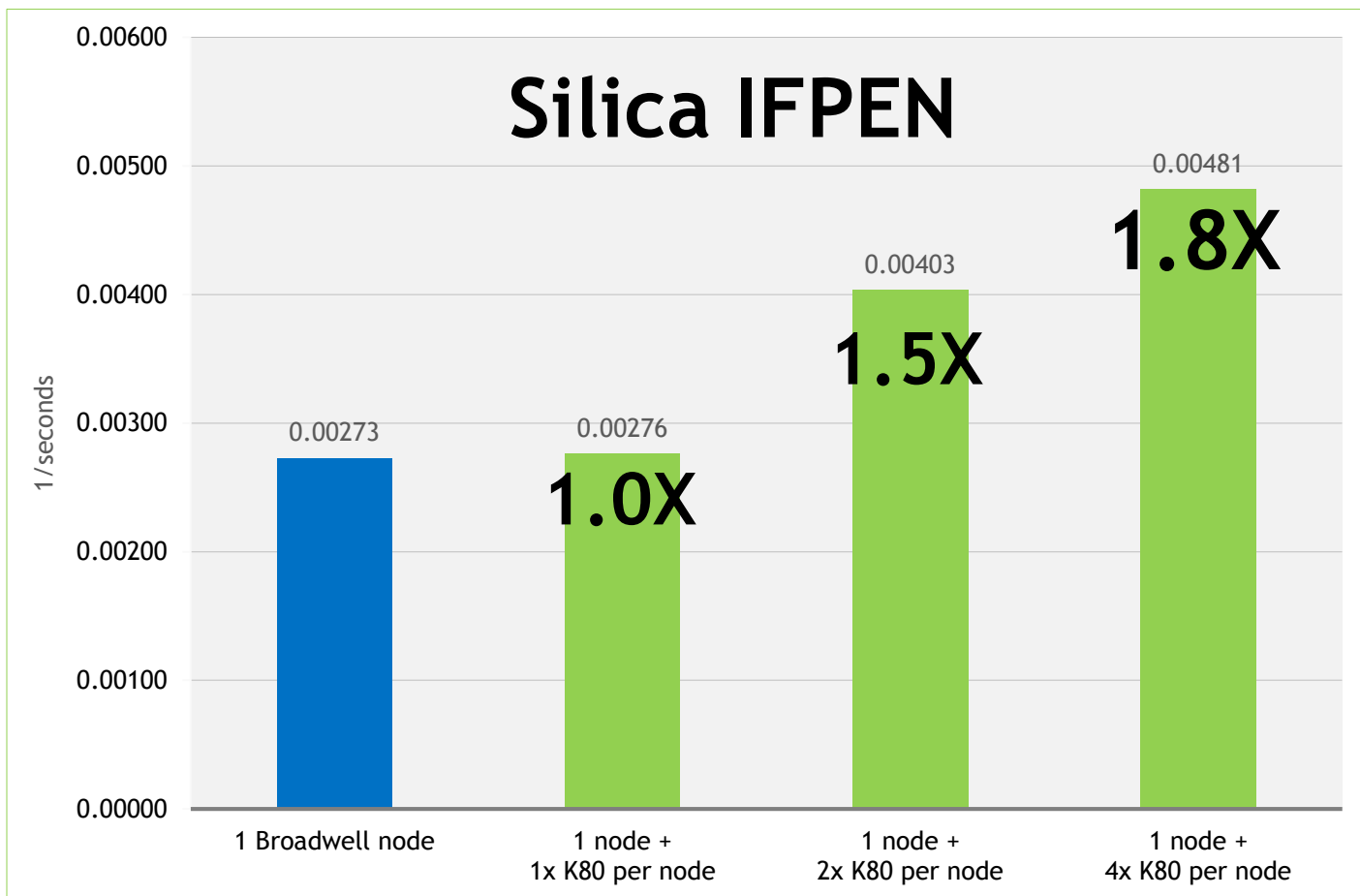
The **green nodes** contain Dual Intel Xeon E5-2698 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla P100 SXM2 GPUs

➤ 1x P100 SXM2 is paired with Single Intel Xeon E5-2698 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

Interface between a platinum slab Pt(111) (108 atoms) and liquid water (120 water molecules) (468 ions)

1256 bands
762048 plane waves
ALGO = Fast (Davidson + RMM-DIIS)

Silica IFPEN on K80s



Running **VASP** version 5.4.1

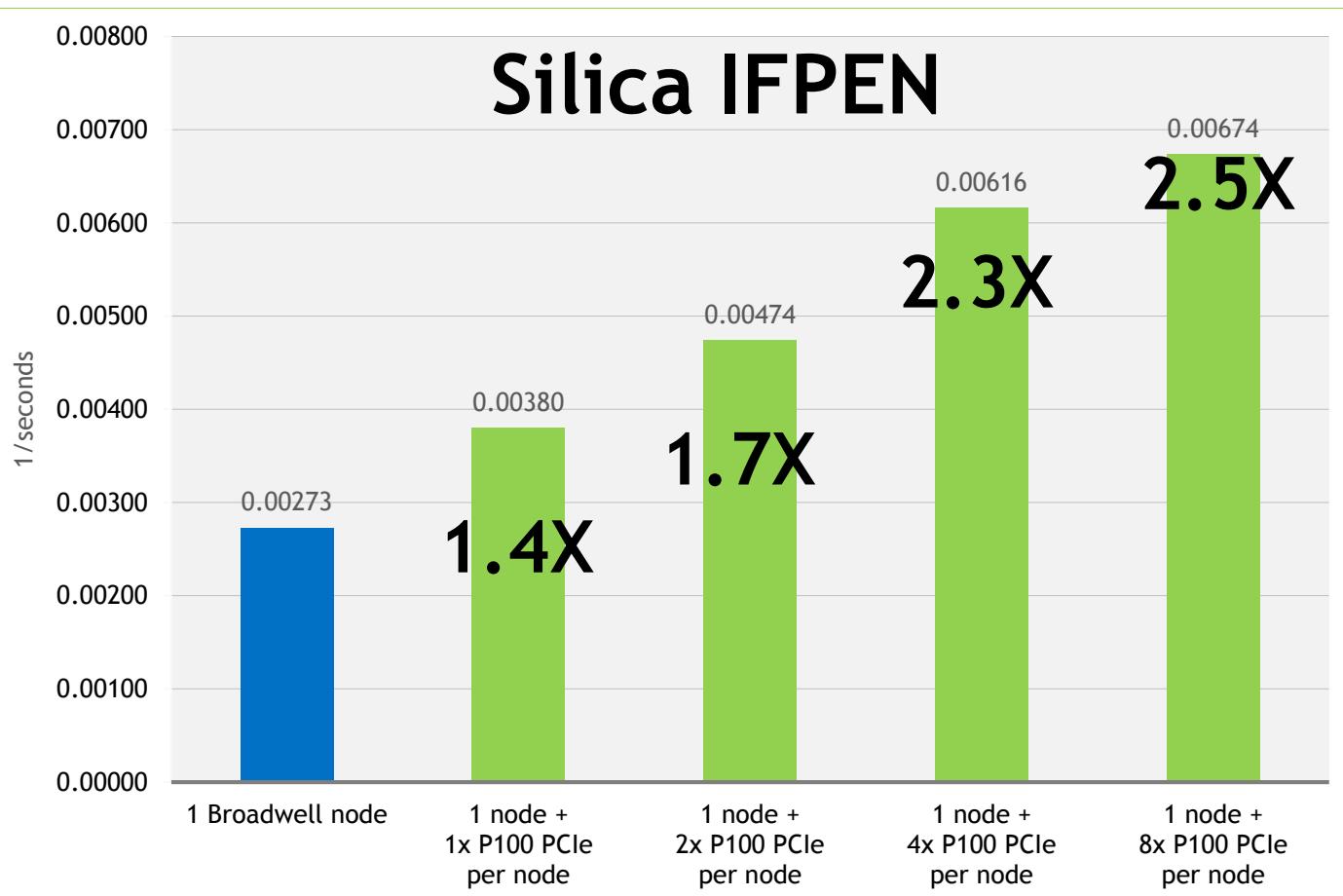
The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla K80 (autoboost) GPUs

➤ 1x K80 is paired with Single Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

240 ions, cristobalite (high) bulk
720 bands
? plane waves
ALGO = Very Fast (RMM-DIIS)

Silica IFPEN on P100s PCIe



Running **VASP** version 5.4.1

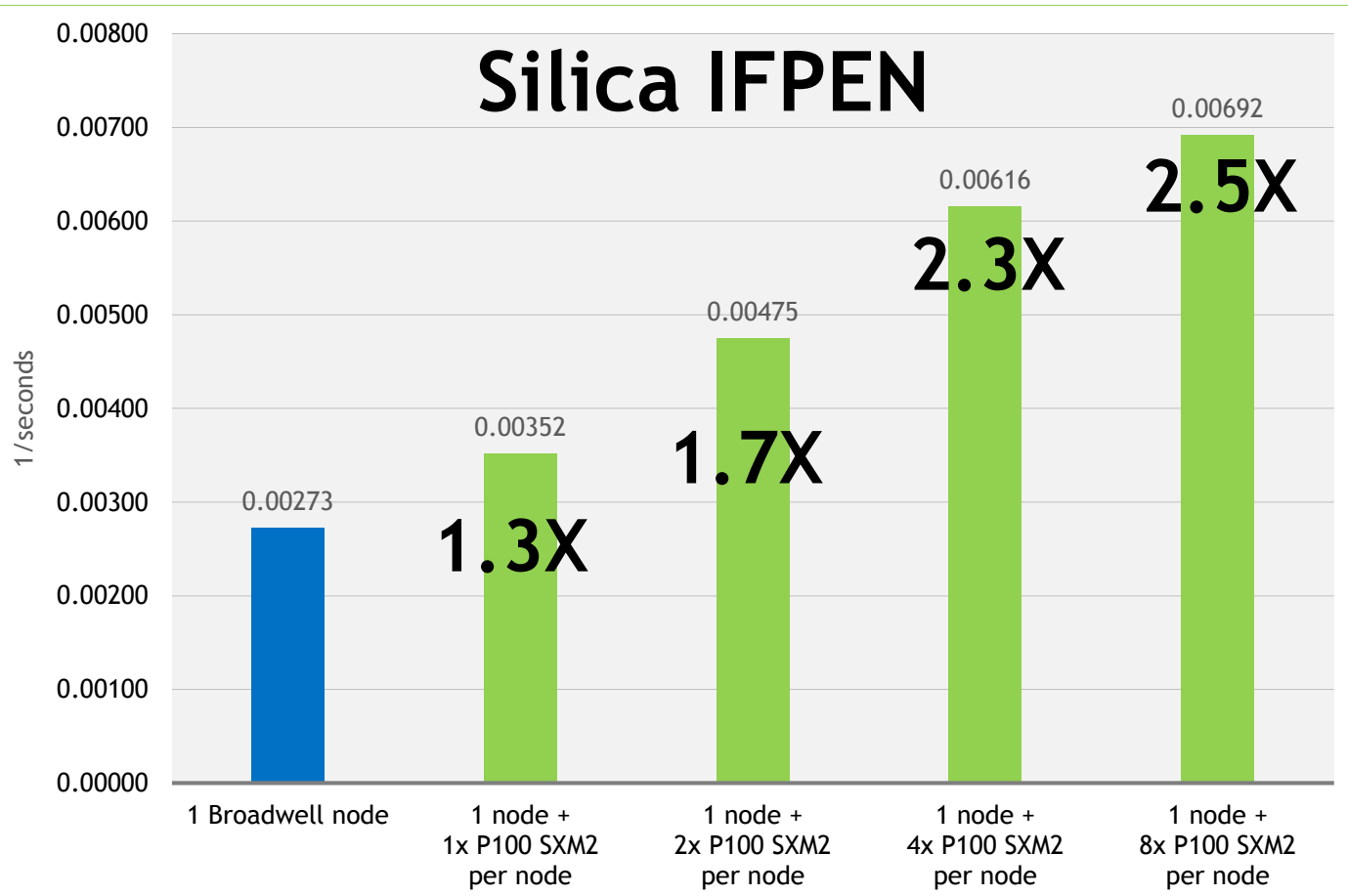
The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla P100 PCIe GPUs

➤ 1x P100 PCIe is paired with Single Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

240 ions, cristobalite (high) bulk
720 bands
? plane waves
ALGO = Very Fast (RMM-DIIS)

Silica IFPEN on P100s SXM2



Running **VASP** version 5.4.1

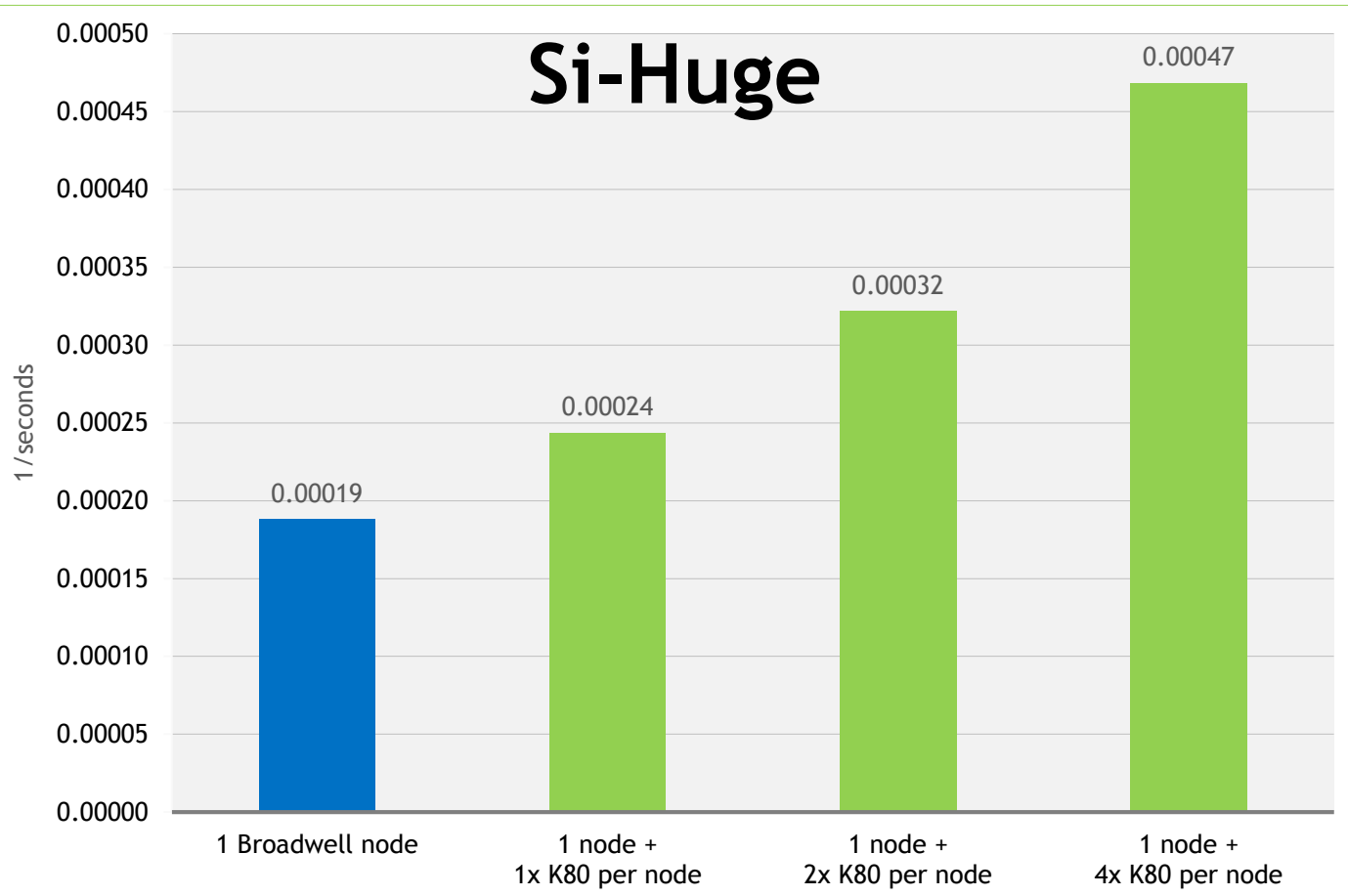
The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2698 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla P100 SXM2 GPUs

- 1x P100 SXM2 is paired with Single Intel Xeon E5-2698 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

240 ions, cristobalite (high) bulk
720 bands
? plane waves
ALGO = Very Fast (RMM-DIIS)

Si-Huge on K80s



Running **VASP** version 5.4.1

The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla K80 (autoboost) GPUs

➤ 1x K80 is paired with Single Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

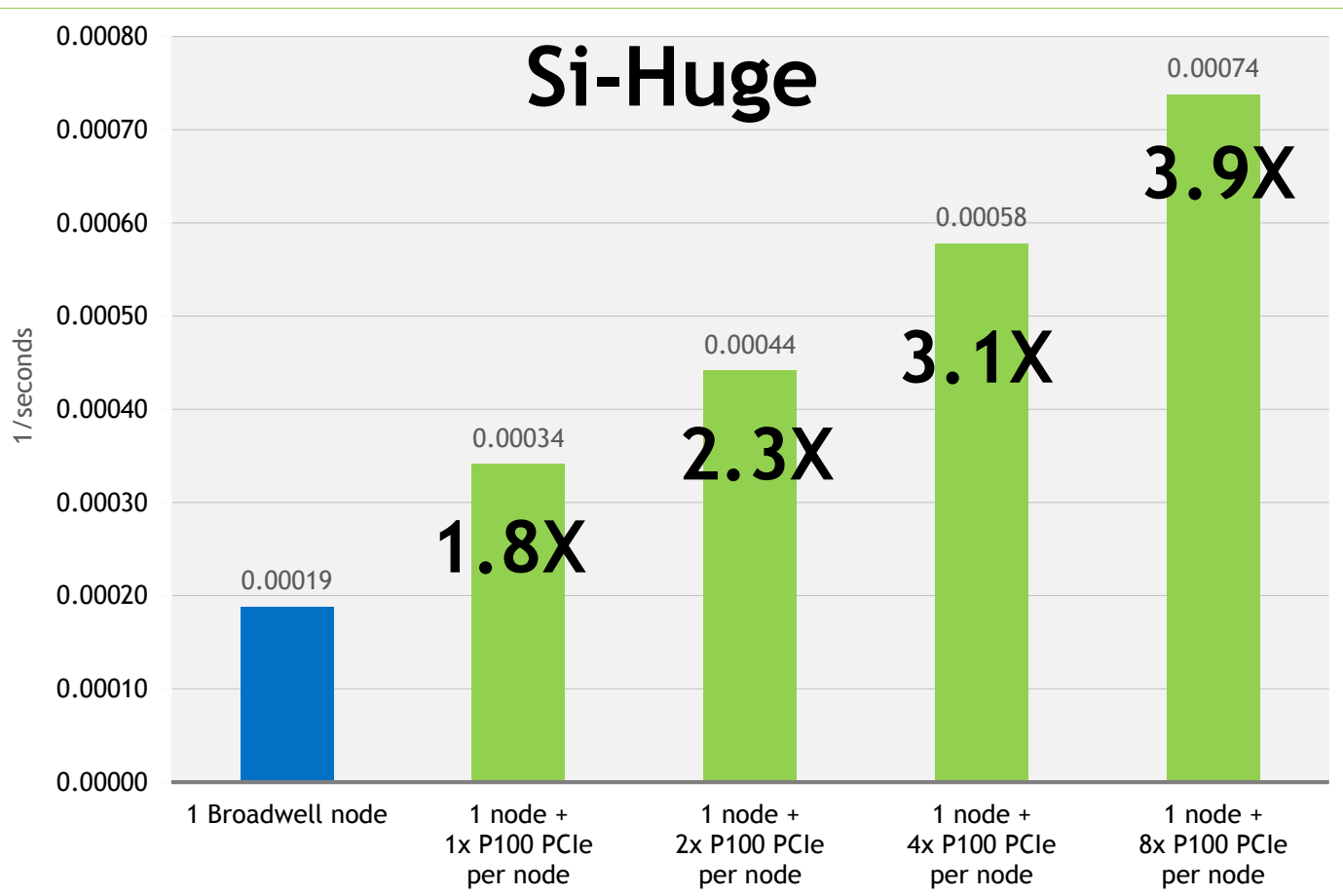
512 Si atoms

1282 bands

864000 Plane Waves

Algo = Normal (blocked Davidson)

Si-Huge on P100s PCIe



Running **VASP** version 5.4.1

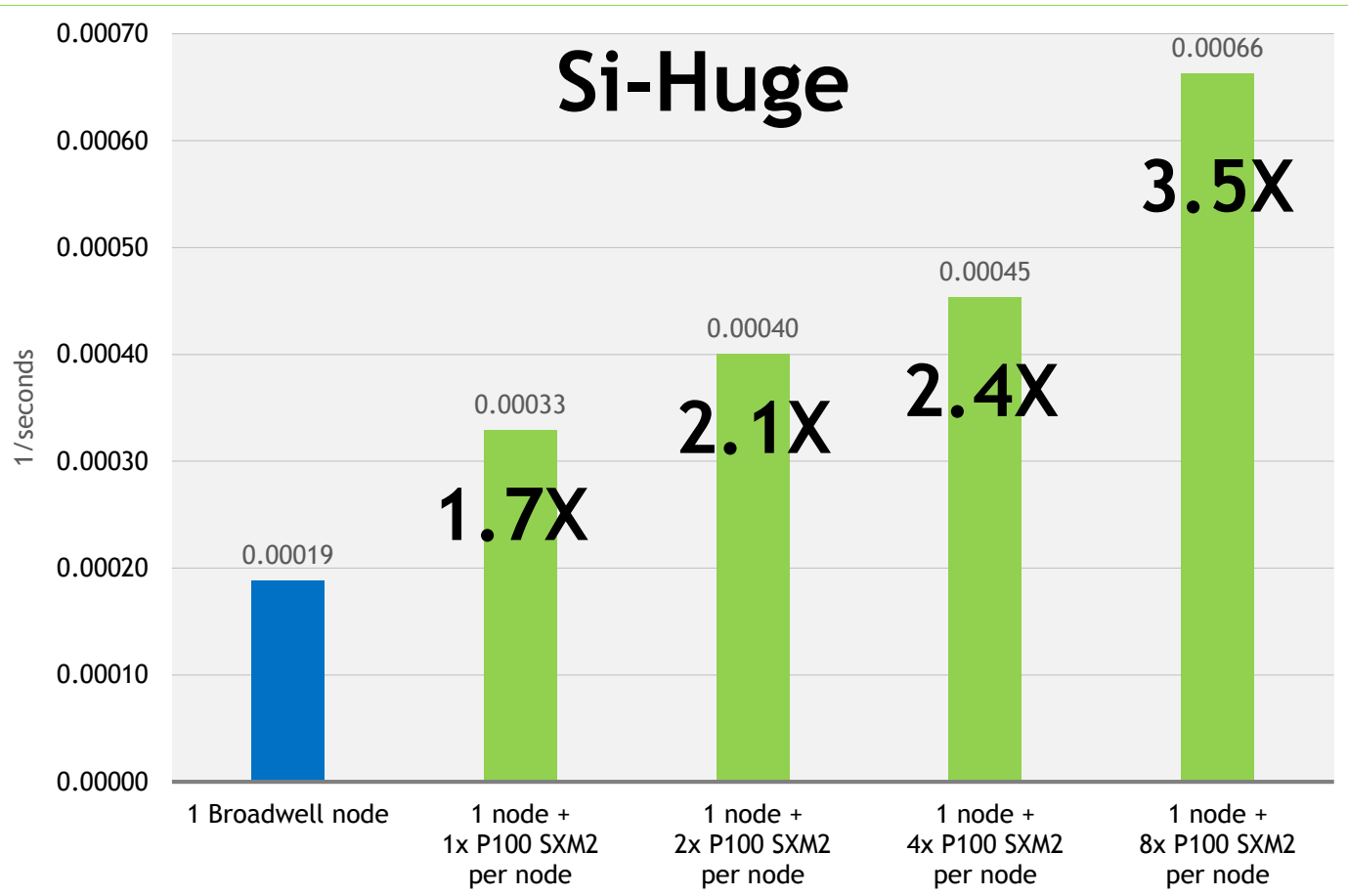
The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla P100 PCIe GPUs

➤ 1x P100 PCIe is paired with Single Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

512 Si atoms
1282 bands
864000 Plane Waves
Algo = Normal (blocked Davidson)

Si-Huge on P100s SXM2



Running **VASP** version 5.4.1

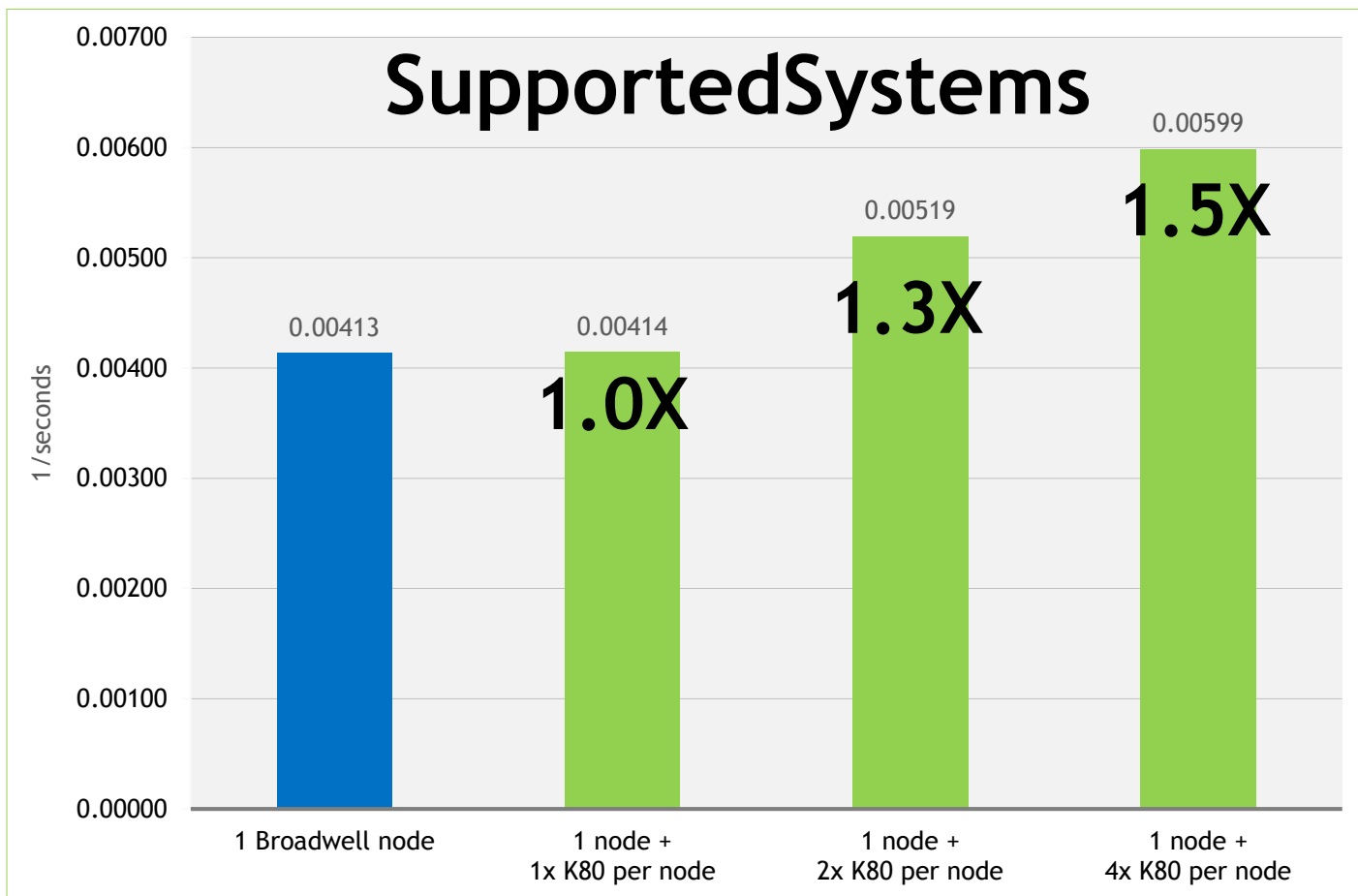
The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2698 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla P100 SXM2 GPUs

- 1x P100 SXM2 is paired with Single Intel Xeon E5-2698 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

512 Si atoms
1282 bands
864000 Plane Waves
Algo = Normal (blocked Davidson)

SupportedSystems on K80s



Running **VASP** version 5.4.1

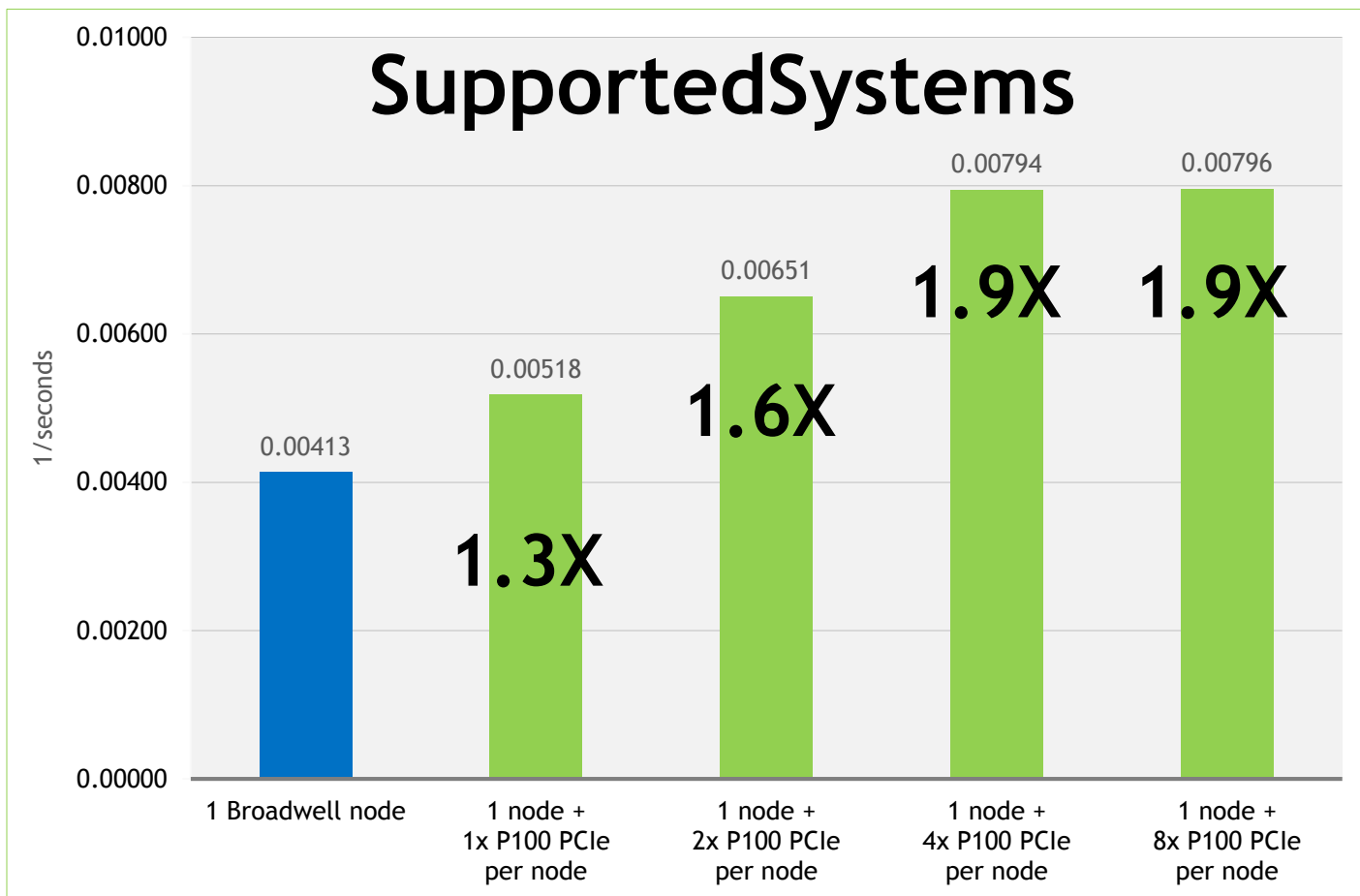
The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla K80 (autoboost) GPUs

➤ 1x K80 is paired with Single Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

*267 ions
788 bands
762048 plane waves
ALGO = Fast (Davidson + RMM-DIIS)*

SupportedSystems on P100s PCIe



Running **VASP** version 5.4.1

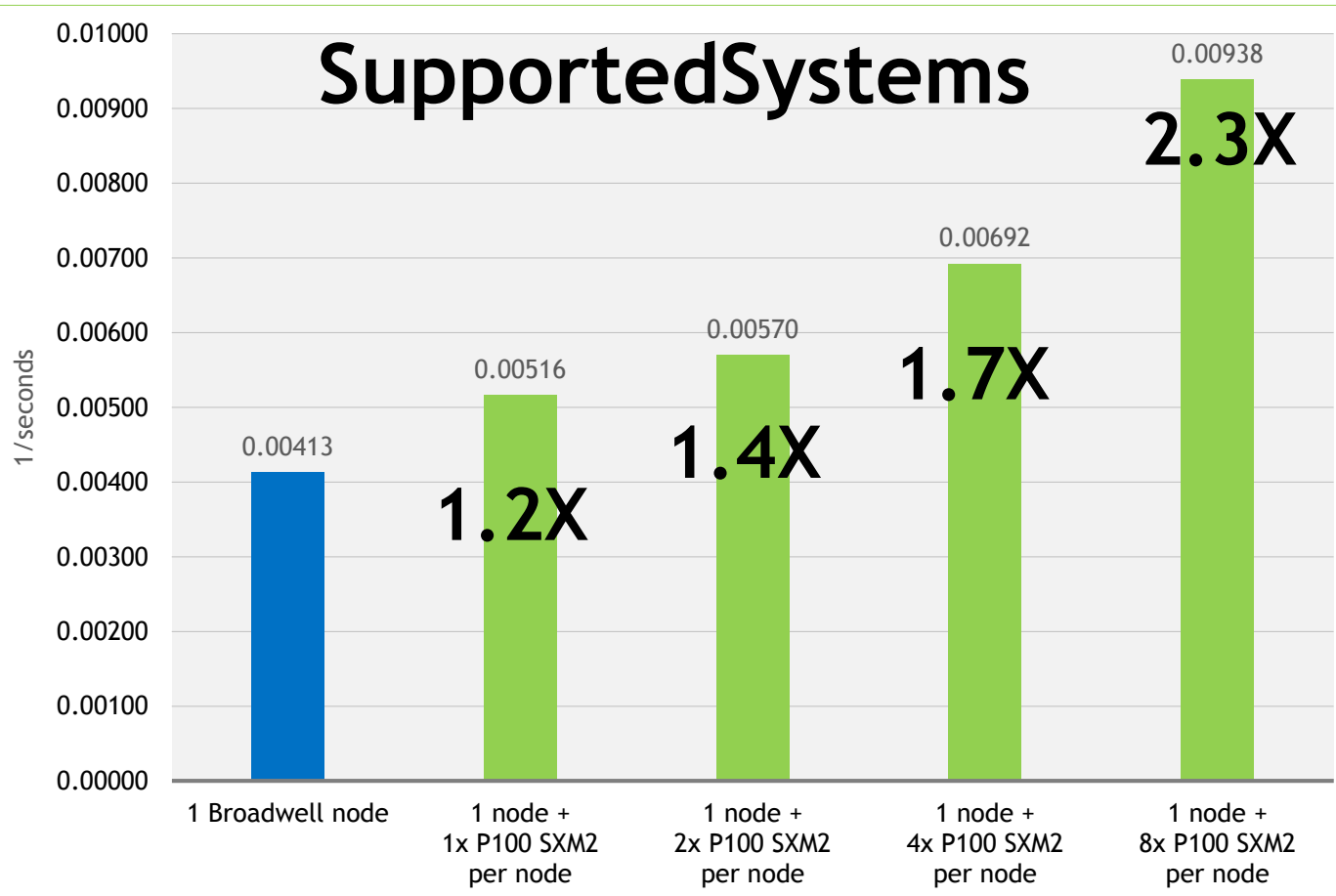
The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla P100 PCIe GPUs

➤ 1x P100 PCIe is paired with Single Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

*267 ions
788 bands
762048 plane waves
ALGO = Fast (Davidson + RMM-DIIS)*

SupportedSystems on P100s SXM2



Running **VASP** version 5.4.1

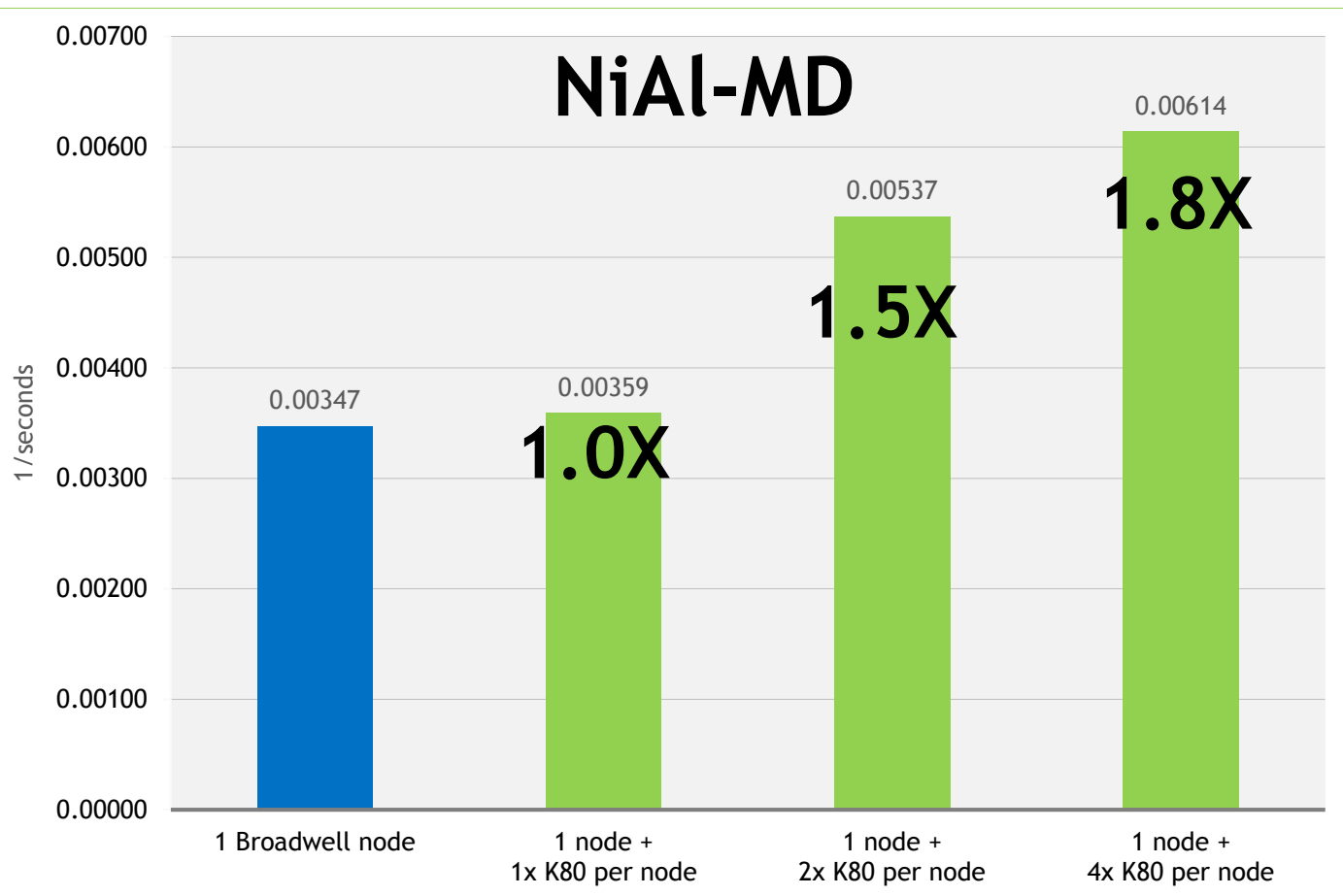
The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2698 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla P100 SXM2 GPUs

- 1x P100 SXM2 is paired with Single Intel Xeon E5-2698 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

*267 ions
788 bands
762048 plane waves
ALGO = Fast (Davidson + RMM-DIIS)*

NiAl-MD on K80s



Running **VASP** version 5.4.1

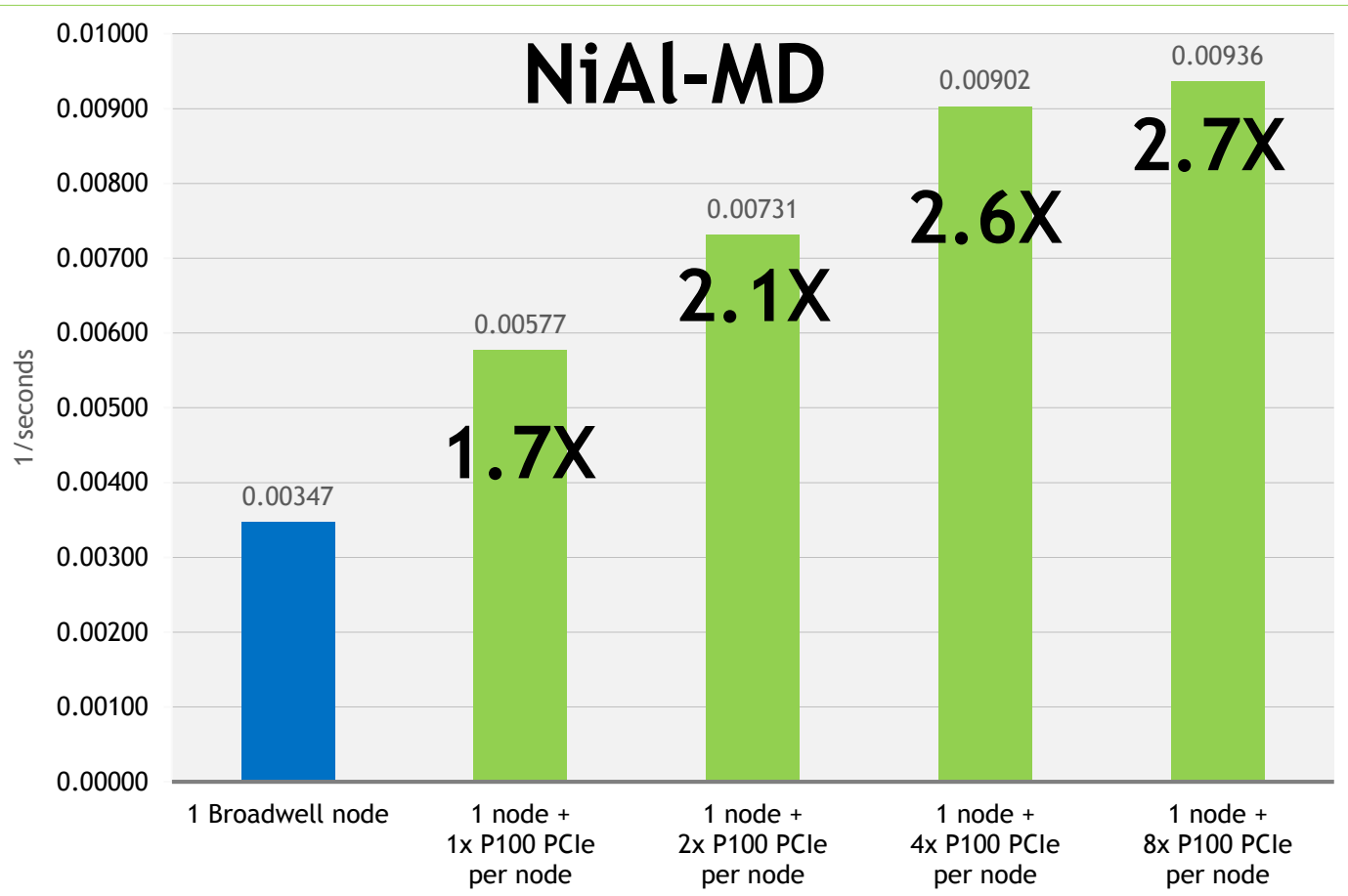
The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla K80 (autoboost) GPUs

➤ 1x K80 is paired with Single Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

*500 ions
3200 bands
729000 plane waves
ALGO = Fast (Davidson + RMM-DIIS)*

NiAl-MD on P100s PCIe



Running **VASP** version 5.4.1

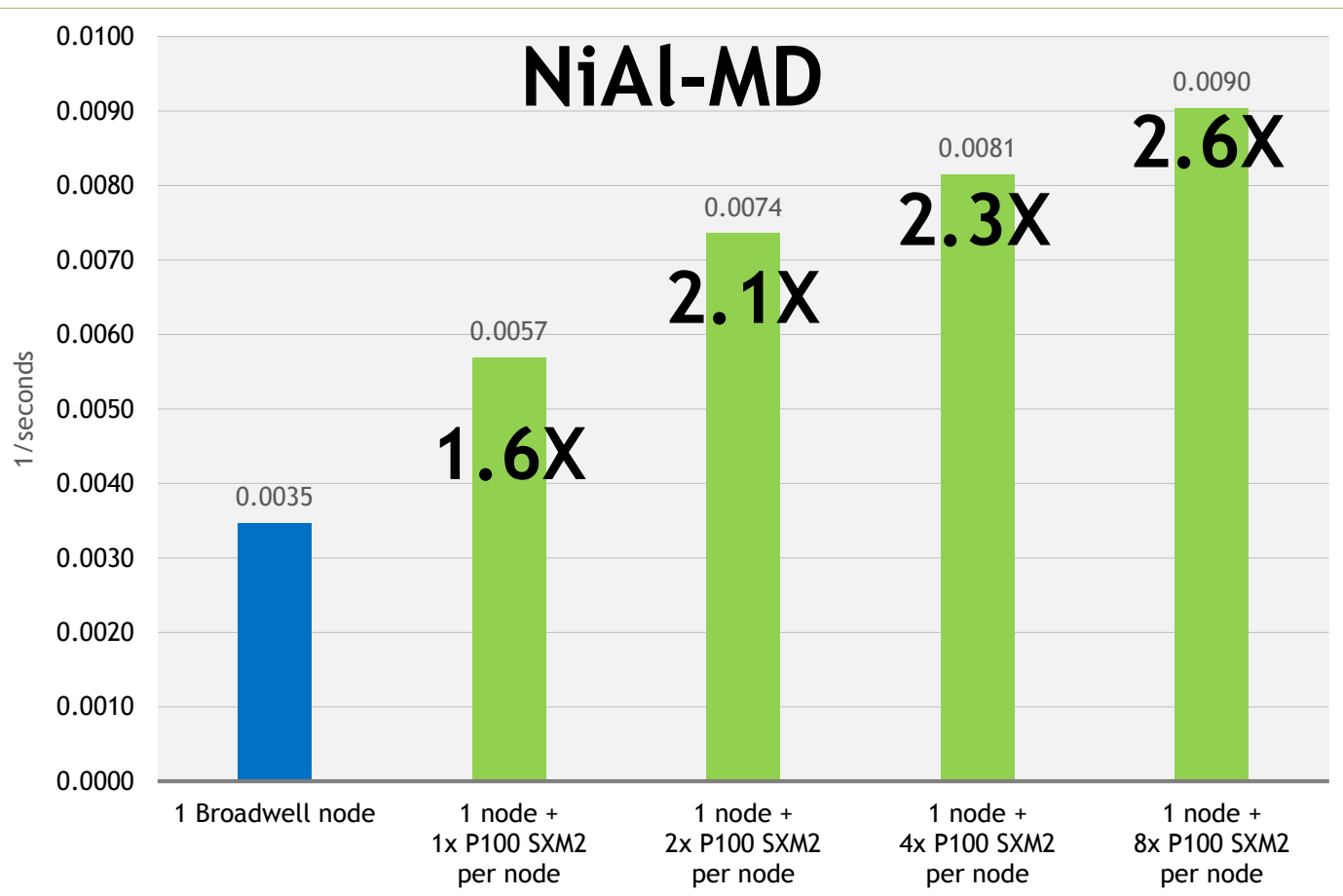
The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla P100 PCIe GPUs

➤ 1x P100 PCIe is paired with Single Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

*500 ions
3200 bands
729000 plane waves
ALGO = Fast (Davidson + RMM-DIIS)*

NiAl-MD on P100s SXM2



Running **VASP** version 5.4.1

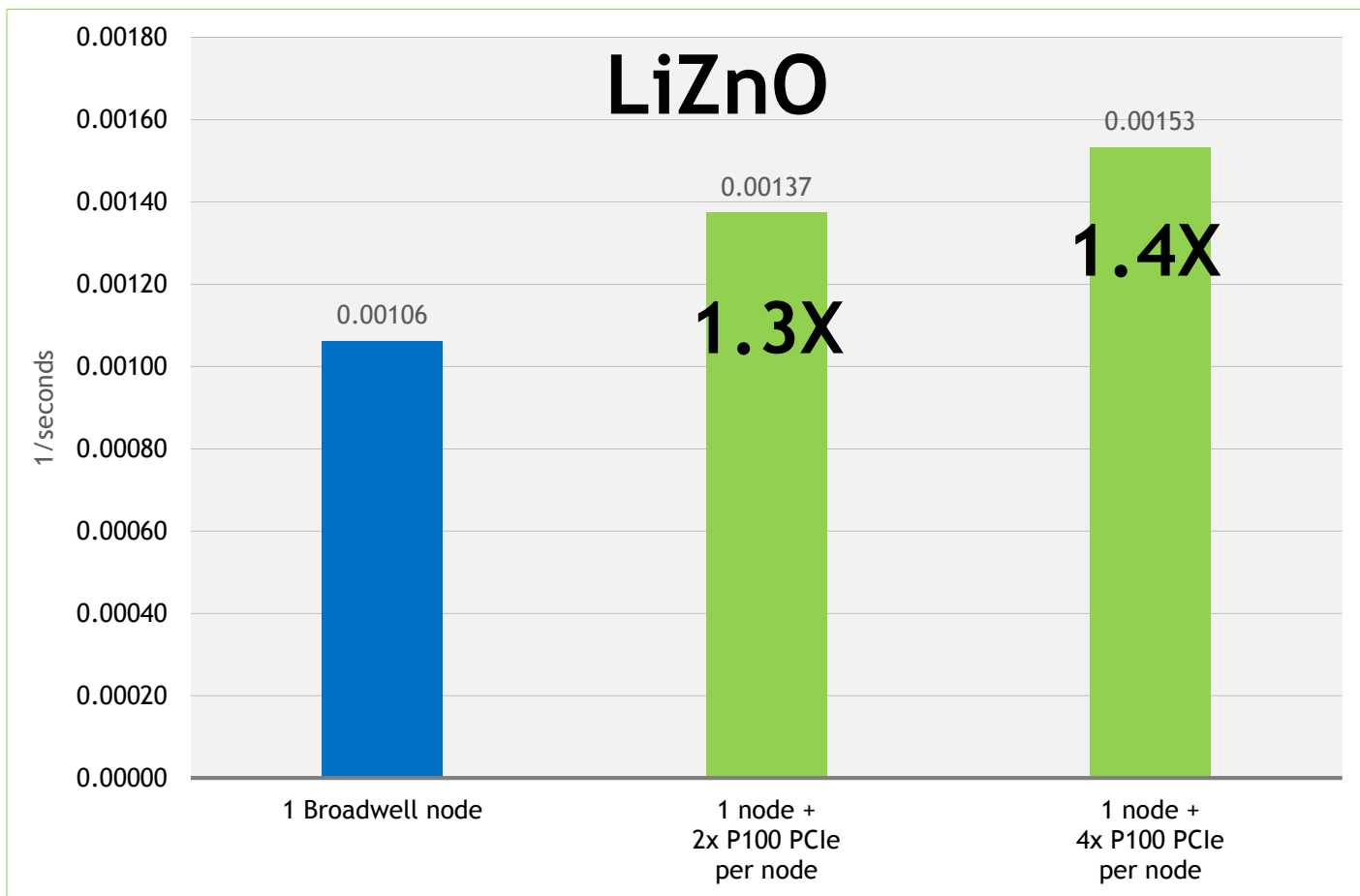
The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2698 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla P100 SXM2 GPUs

- 1x P100 SXM2 is paired with Single Intel Xeon E5-2698 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

*500 ions
3200 bands
729000 plane waves
ALGO = Fast (Davidson + RMM-DIIS)*

LiZnO on P100s PCIe



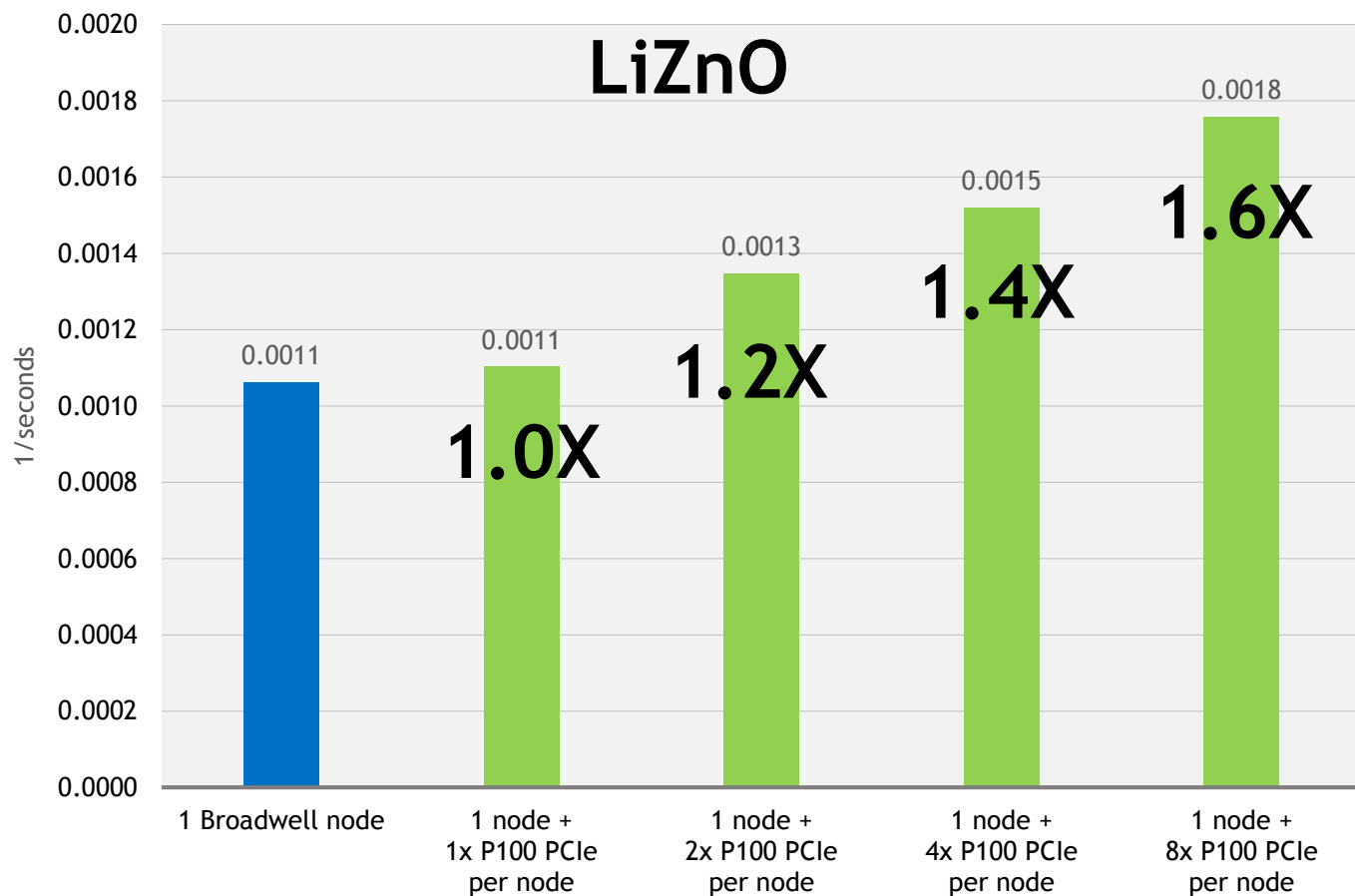
Running **VASP** version 5.4.1

The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla P100 PCIe GPUs

*500 ions
3200 bands
729000 plane waves
ALGO = Fast (Davidson + RMM-DIIS)*

LiZnO on P100s SXM2



Running **VASP** version 5.4.1

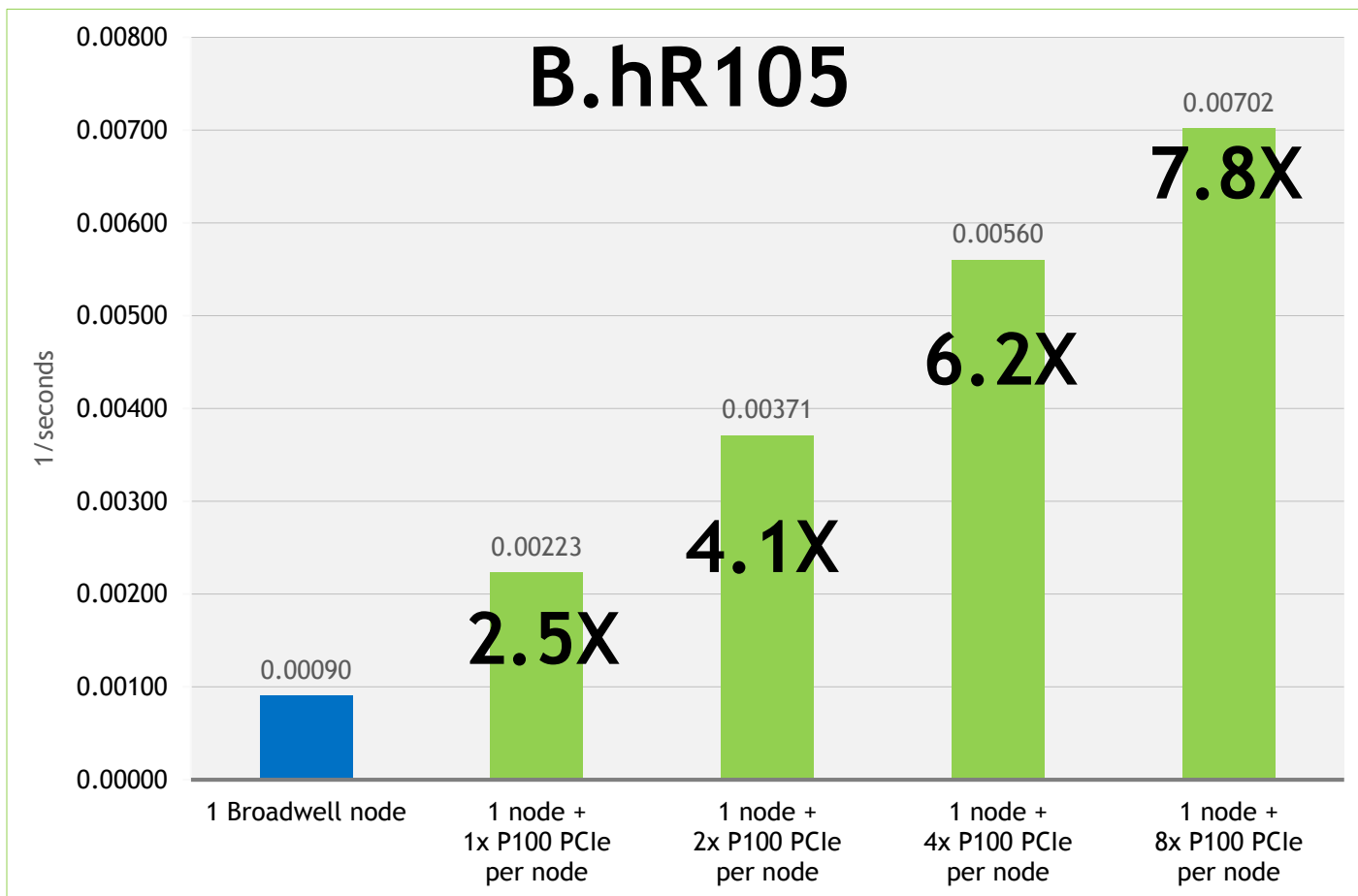
The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2698 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla P100 SXM2 GPUs

- 1x P100 SXM2 is paired with Single Intel Xeon E5-2698 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

*500 ions
3200 bands
729000 plane waves
ALGO = Fast (Davidson + RMM-DIIS)*

B.hR105 on P100s PCIe



Running **VASP** version 5.4.1

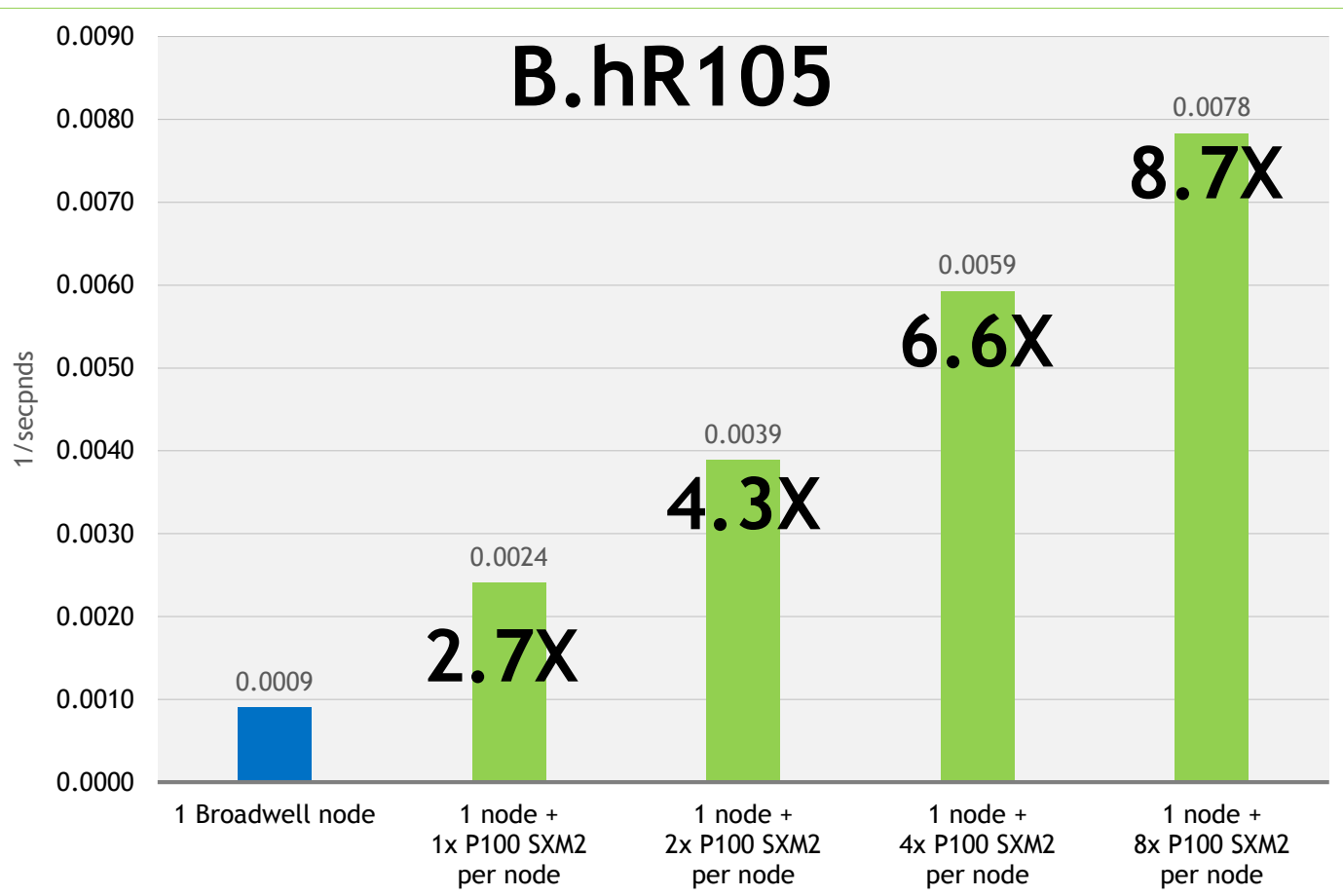
The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla P100 PCIe GPUs

➤ 1x P100 PCIe is paired with Single Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

105 Boron atoms (*B-rhombohedral structure*)
216 bands
110592 plane waves
Hybrid Functional with blocked Davison
(ALGO=Normal)
LHFCALC=.True. (Exact Exchange)

B.hR105 on P100s SXM2



Running **VASP** version 5.4.1

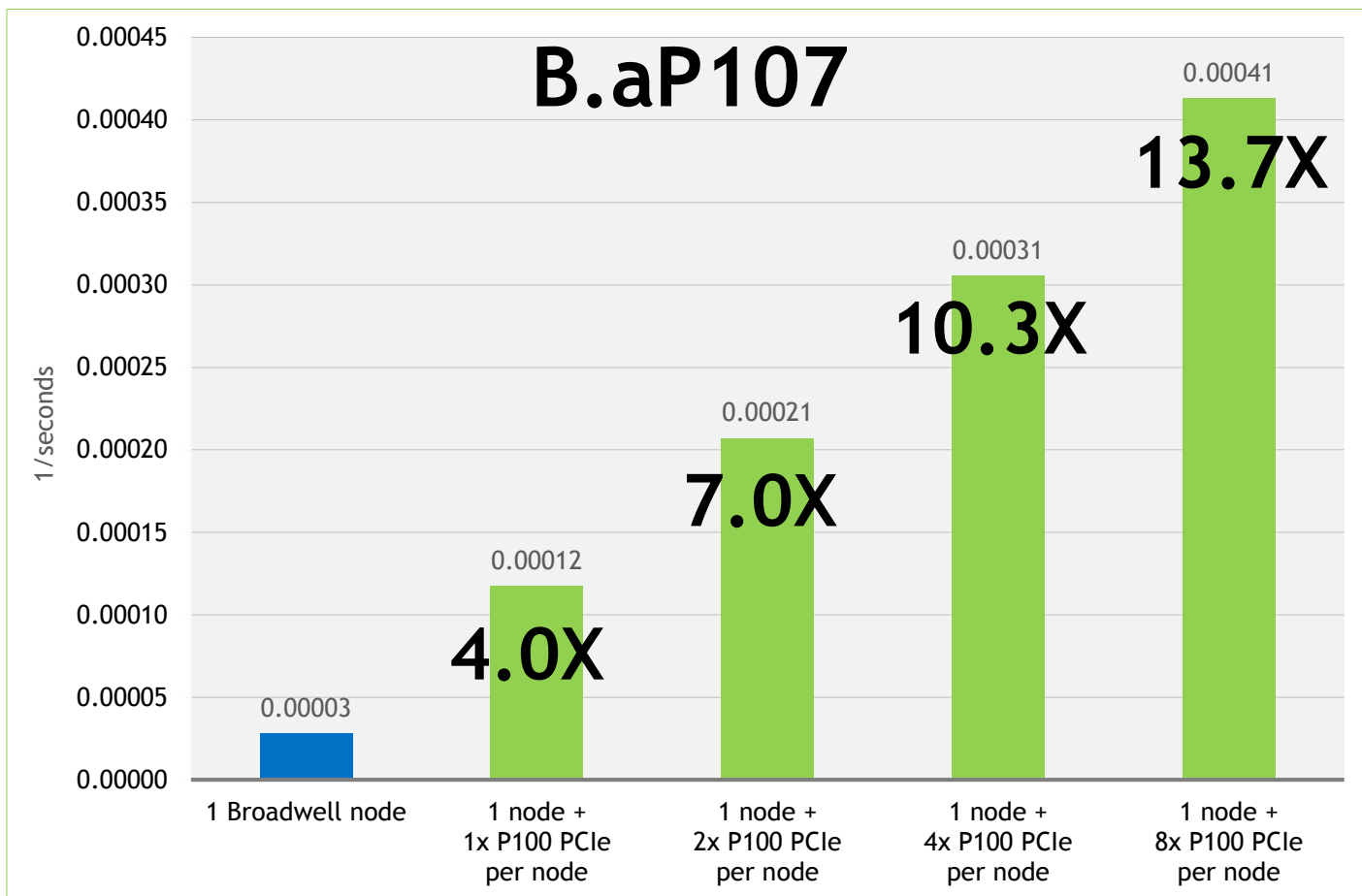
The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2698 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla P100 SXM2 GPUs

- 1x P100 SXM2 is paired with Single Intel Xeon E5-2698 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

105 Boron atoms (*B-rhombohedral structure*)
216 bands
110592 plane waves
Hybrid Functional with blocked Davison
(ALGO=Normal)
LHFCALC=.True. (Exact Exchange)

B.aP107 on P100s PCIe



Running **VASP** version 5.4.1

The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla P100 PCIe GPUs

➤ 1x P100 PCIe is paired with Single Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

107 Boron atoms (symmetry broken 107-atom B' variant)

216 bands

110592 plane waves

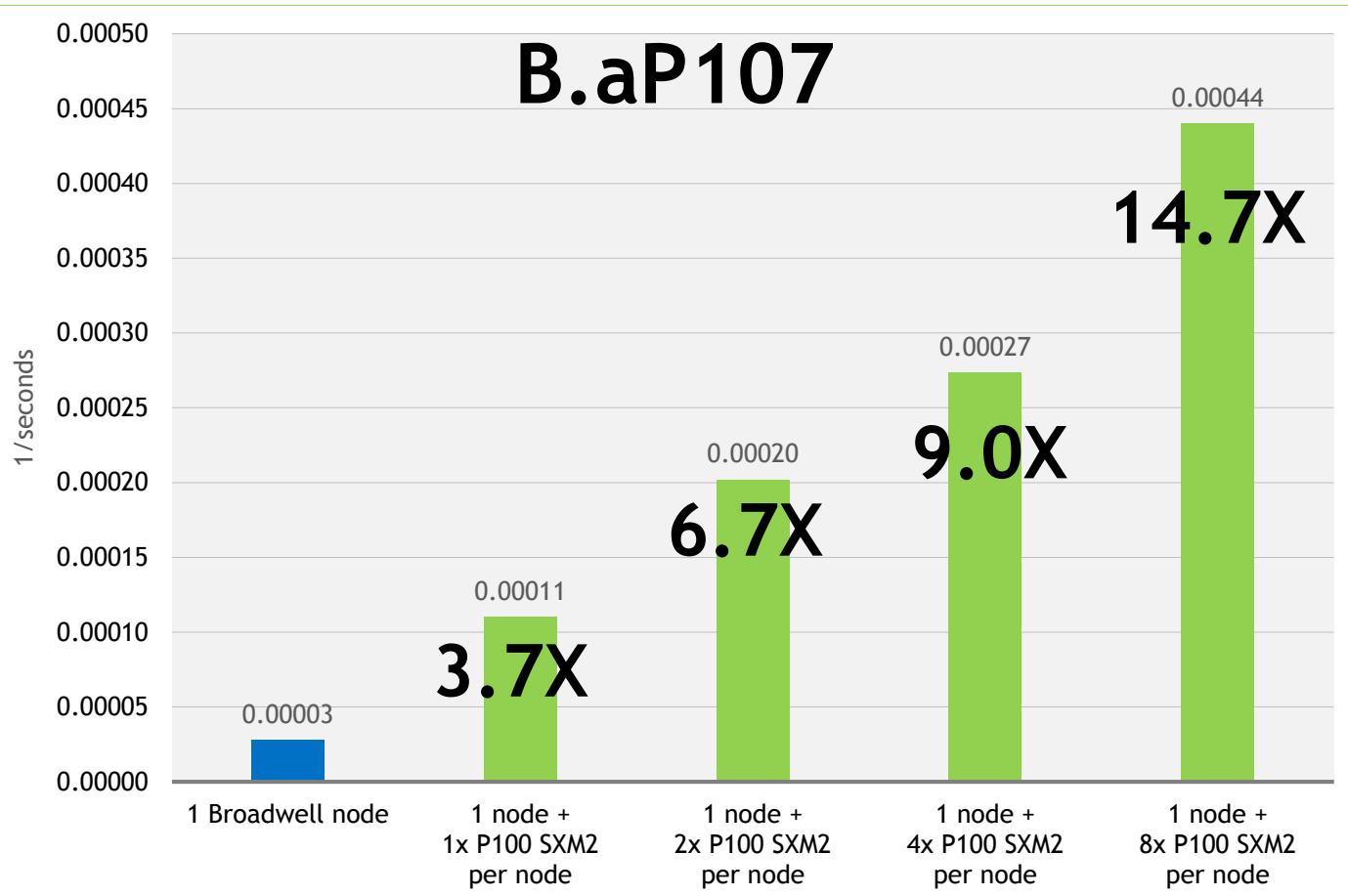
Hybrid functional calculation (exact exchange) with blocked Davidson. No KPoint parallelization.

Hybrid Functional with blocked Davidson

(ALGO=Normal)

LHFCALC=.True. (Exact Exchange)

B.aP107 on P100s SXM2



Running **VASP** version 5.4.1

The **blue node** contains Dual Intel Xeon E5-2699 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs

The **green nodes** contain Dual Intel Xeon E5-2698 v4@2.2GHz [3.6GHz Turbo] (Broadwell) CPUs + Tesla P100 SXM2 GPUs

➤ 1x P100 SXM2 is paired with Single Intel Xeon E5-2698 v4@2.2GHz [3.6GHz Turbo] (Broadwell)

107 Boron atoms (symmetry broken 107-atom B' variant)

216 bands

110592 plane waves

Hybrid functional calculation (exact exchange) with blocked Davidson. No KPoint parallelization.

Hybrid Functional with blocked Davidson (ALGO=Normal)

LHFCALC=.True. (Exact Exchange)

Quantum Chemistry (QC) on GPUs

Dec, 19, 2016

