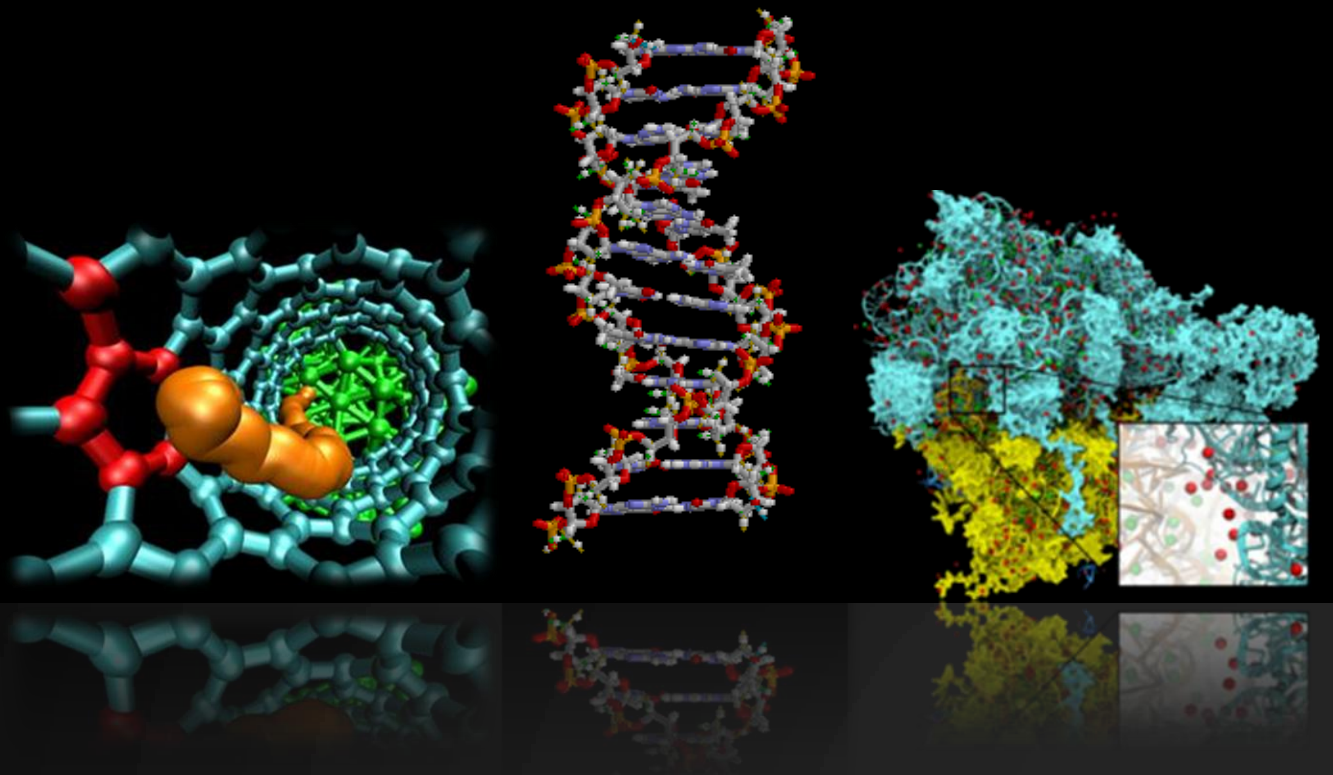
NVIDIA Computational Chemistry & Biology



Mark Berger
Life and Material Science Alliances

Updated: February 4, 2013

Molecular Dynamics (MD) Applications

Application	Features Supported	GPU Perf	Release Status	Notes/Benchmarks
AMBER	PMEMD Explicit Solvent & GB Implicit Solvent	> 100 ns/day JAC NVE on 2X K20s	Released Multi-GPU, multi-node	AMBER 12, GPU Revision Support 12.2 http://ambermd.org/gpus/benchmarks.htm#Benchmarks
CHARMM	Implicit (5x), Explicit (2x) Solvent via OpenMM	2x C2070 equals 32-35x X5667 CPUs	Released Single & multi-GPU in single node	Release C37b1; http://www.charmm.org/news/c37b1.html#postjump
DL_POLY	Two-body Forces, Link-cell Pairs, Ewald SPME forces, Shake VV	4x	Release V 4.03 Multi-GPU, multi-node	Source only, Results Published http://www.stfc.ac.uk/CSE/randd/ccg/software/DL_POLY/25526.aspx
GROMACS	Implicit (5x), Explicit (2x)	165 ns/Day DHFR on 4X C2075s	Released Multi-GPU, multi-node	Release 4.6; 1st Multi-GPU support
LAMMPS	Lennard-Jones, Gay-Berne, Tersoff & many more potentials	3.5-18x on Titan	Released. Multi-GPU, multi-node	http://lammps.sandia.gov/bench.html#desktop and http://lammps.sandia.gov/bench.html#titan
NAMD	Full electrostatics with PME and most simulation features	4.0 ns/days F1-ATPase on 1x K20X	Released 100M atom capable Multi-GPU, multi-node	NAMD 2.9

GPU Perf compared against Multi-core x86 CPU socket.
GPU Perf benchmarked on GPU supported features
and may be a kernel to kernel perf comparison

New/Additional MD Applications Ramping

Application	Features Supported	GPU Perf	Release Status	Notes
Abalone	Simulations (on 1060 GPU)	4-29X (on 1060 GPU)	Released, Version 1.8.51 Single GPU	Agile Molecule, Inc.
Ascalaph	Computation of non-valent interactions	4-29X (on 1060 GPU)	Released, Version 1.1.4 Single GPU	Agile Molecule, Inc.
ACEMD	Written for use only on GPUs	150 ns/day DHFR on 1x K20	Released Single and multi-GPUs	Production bio-molecular dynamics (MD) software specially optimized to run on GPUs
Folding@Home	Powerful distributed computing molecular dynamics system; implicit solvent and folding	Depends upon number of GPUs	Released; GPUs and CPUs	http://folding.stanford.edu GPUs get 4X the points of CPUs
GPUGrid.net	High-performance all-atom biomolecular simulations; explicit solvent and binding	Depends upon number of GPUs	Released; NVIDIA GPUs only	http://www.gpugrid.net/
HALMD	Simple fluids and binary mixtures (pair potentials, high-precision NVE and NVT, dynamic correlations)	Up to 66x on 2090 vs. 1 CPU core	Released, Version 0.2.0 Single GPU	http://halmd.org/benchmarks.html#supercooled-binary-mixture-kob-andersen
HOOMD-Blue	Written for use only on GPUs	Kepler 2X faster than Fermi	Released, Version 0.11.2 Single and multi-GPU on 1 node	http://codeblue.umich.edu/hoomd-blue/ Multi-GPU w/ MPI in March 2013
OpenMM	Implicit and explicit solvent, custom forces	Implicit: 127-213 ns/day Explicit: 18-55 ns/day DHFR	Released Version 4.1.1 Multi-GPU	Library and application for molecular dynamics on high-performance

GPU Perf compared against Multi-core x86 CPU socket.
GPU Perf benchmarked on GPU supported features and may be a kernel to kernel perf comparison

Quantum Chemistry Applications

Application	Features Supported	GPU Perf	Release Status	Notes
Abinit	Local Hamiltonian, non-local Hamiltonian, LOBPCG algorithm, diagonalization / orthogonalization	1.3-2.7X	Released; Version 7.0.5 Multi-GPU support	www.abinit.org
ACES III	Integrating scheduling GPU into SIAL programming language and SIP runtime environment	10X on kernels	Under development Multi-GPU support	http://www.olcf.ornl.gov/wp-content/training/electronic-structure-2012/deumens_ESaccel_2012.pdf
ADF	Fock Matrix, Hessians	TBD	Pilot project completed, Under development Multi-GPU support	www.scm.com
BigDFT	DFT; Daubechies wavelets, part of Abinit	5-25X (1 CPU core to GPU kernel)	Released June 2009, current release 1.6.0 Multi-GPU support	http://inac.cea.fr/L_Sim/BigDFT/news.html, http://www.olcf.ornl.gov/wp-content/training/electronic-structure-2012/BigDFT-Formalism.pdf and http://www.olcf.ornl.gov/wp-content/training/electronic-structure-2012/BigDFT-HPC-tues.pdf
Casino	TBD	TBD	Under development, Spring 2013 release Multi-GPU support	http://www.tcm.phy.cam.ac.uk/~mdt26/casino.html
CP2K	DBCSR (spare matrix multiply library)	2-7X	Under development Multi-GPU support	http://www.olcf.ornl.gov/wp-content/training/ascc_2012/friday/ACSS_2012_VandeVondele_s.pdf
GAMESS-US	Libqc with Rys Quadrature Algorithm, Hartree-Fock, MP2 and CCSD in Q4 2012	1.3-1.6X, 2.3-2.9x HF	Released Multi-GPU support	Next release Q4 2012. http://www.msg.ameslab.gov/gamess/index.html

GPU Perf compared against Multi-core x86 CPU socket.
GPU Perf benchmarked on GPU supported features and may be a kernel to kernel perf comparison

Quantum Chemistry Applications

Application	Features Supported	GPU Perf	Release Status	Notes
GAMESS-UK	(ss|ss) type integrals within calculations using Hartree Fock *ab initio* methods and density functional theory. Supports organics & inorganics.	8x	Release in 2012 Multi-GPU support	http://www.ncbi.nlm.nih.gov/pubmed/21541963
Gaussian	Joint PGI, NVIDIA & Gaussian Collaboration	TBD	Under development Multi-GPU support	Announced Aug. 29, 2011 http://www.gaussian.com/g_press/nvidia_press.htm
GPAW	Electrostatic poisson equation, orthonormalizing of vectors, residual minimization method (rmm-diis)	8x	Released Multi-GPU support	https://wiki.fysik.dtu.dk/gpaw/devel/projects/gpu.html, Samuli Hakala (CSC Finland) & Chris O'Grady (SLAC)
Jaguar	Investigating GPU acceleration	TBD	Under development Multi-GPU support	Schrodinger, Inc. http://www.schrodinger.com/kb/278
MOLCAS	CU_BLAS support	1.1x	Released, Version 7.8 Single GPU. Additional GPU support coming in Version 8	www.molcas.org
MOLPRO	Density-fitted MP2 (DF-MP2), density fitted local correlation methods (DF-RHF, DF-KS), DFT	1.7-2.3X projected	Under development Multiple GPU	www.molpro.net Hans-Joachim Werner

GPU Perf compared against Multi-core x86 CPU socket.
GPU Perf benchmarked on GPU supported features and may be a kernel to kernel perf comparison

Quantum Chemistry Applications

Application	Features Supported	GPU Perf	Release Status	Notes
MOPAC2009	pseudodiagonalization, full diagonalization, and density matrix assembling	3.8-14X	Under Development Single GPU	Academic port. http://openmopac.net
NWChem	Triples part of Reg-CCSD(T), CCSD & EOMCCSD task schedulers	3-10X projected	Release targeting March 2013 Multiple GPUs	Development GPGPU benchmarks: www.nwchem-sw.org And http://www.olcf.ornl.gov/wp-content/training/electronic-structure-2012/Krishnamoorthy-ESCMA12.pdf
Octopus	DFT and TDDFT	TBD	Released	http://www.tddft.org/programs/octopus/
PEtot	Density functional theory (DFT) plane wave pseudopotential calculations	6-10X	Released Multi-GPU	First principles materials code that computes the behavior of the electron structures of materials
Q-CHEM	RI-MP2	8x-14x	Released, Version 4.0	http://www.q-chem.com/doc_for_web/qchem_manual_4.0.pdf

GPU Perf compared against Multi-core x86 CPU socket.
GPU Perf benchmarked on GPU supported features
and may be a kernel to kernel perf comparison

Quantum Chemistry Applications

Application	Features Supported	GPU Perf	Release Status	Notes
QMCPACK	Main features	3-4x	Released Multiple GPUs	NCSA University of Illinois at Urbana-Champaign http://cms.mcc.uiuc.edu/qmcpack/index.php/GPU_version_of_QMCPACK
Quantum Espresso/PWscf	PWscf package: linear algebra (matrix multiply), explicit computational kernels, 3D FFTs	2.5-3.5x	Released Version 5.0 Multiple GPUs	Created by Irish Centre for High-End Computing http://www.quantum-espresso.org/index.php and http://www.quantum-espresso.org/
TeraChem	"Full GPU-based solution"	44-650X vs. GAMESS CPU version	Released Version 1.5 Multi-GPU/single node	Completely redesigned to exploit GPU parallelism. YouTube: http://youtu.be/EJODzk6RFxE?hd=1 and http://www.olcf.ornl.gov/wp-content/training/electronic-structure-2012/Luehr-ESCMA.pdf
VASP	Hybrid Hartree-Fock DFT functionals including exact exchange	2x 2 GPUs comparable to 128 CPU cores	Available on request Multiple GPUs	By Carnegie Mellon University http://arxiv.org/pdf/1111.0716.pdf
WL-LSMS	Generalized Wang-Landau method	3x with 32 GPUs vs. 32 (16-core) CPUs	Under development Multi-GPU support	NICS Electronic Structure Determination Workshop 2012: http://www.olcf.ornl.gov/wp-content/training/electronic-structure-2012/Eisenbach_OakRidge_February.pdf

GPU Perf compared against Multi-core x86 CPU socket.
GPU Perf benchmarked on GPU supported features and may be a kernel to kernel perf comparison

Viz, "Docking" and Related Applications Growing

Related Applications	Features Supported	GPU Perf	Release Status	Notes
Amira 5®	3D visualization of volumetric data and surfaces	70x	Released, Version 5.3.3 Single GPU	Visualization from Visage Imaging. Next release, 5.4, will use GPU for general purpose processing in some functions http://www.visageimaging.com/overview.html
BINDSURF	Allows fast processing of large ligand databases	100X	Available upon request to authors; single GPU	High-Throughput parallel blind Virtual Screening, http://www.biomedcentral.com/1471-2105/13/S14/S13
BUDE	Empirical Free Energy Forcefield	6.5-13.4X	Released Single GPU	University of Bristol http://www.bris.ac.uk/biochemistry/cpfg/bude/bude.htm
Core Hopping	GPU accelerated application	3.75-5000X	Released, Suite 2011 Single and multi-GPUs.	Schrodinger, Inc. http://www.schrodinger.com/products/14/32/
FastROCS	Real-time shape similarity searching/comparison	800-3000X	Released Single and multi-GPUs.	Open Eyes Scientific Software http://www.eyesopen.com/fastrocs
PyMol	Lines: 460% increase Cartoons: 1246% increase Surface: 1746% increase Spheres: 753% increase Ribbon: 426% increase	1700x	Released, Version 1.5 Single GPUs	http://pymol.org/
VMD	High quality rendering, large structures (100 million atoms), analysis and visualization tasks, multiple GPU support for display of molecular orbitals	100-125X or greater on kernels	Released, Version 1.9	Visualization from University of Illinois at Urbana-Champaign http://www.ks.uiuc.edu/Research/vmd/

GPU Perf compared against Multi-core x86 CPU socket.
GPU Perf benchmarked on GPU supported features
and may be a kernel to kernel perf comparison

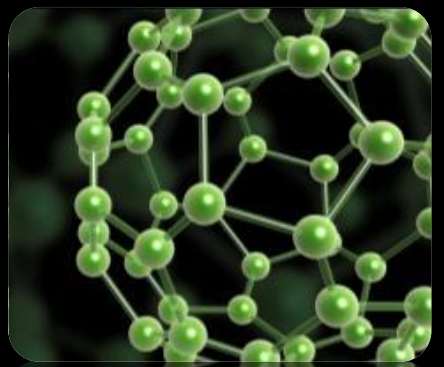
Bioinformatics Applications

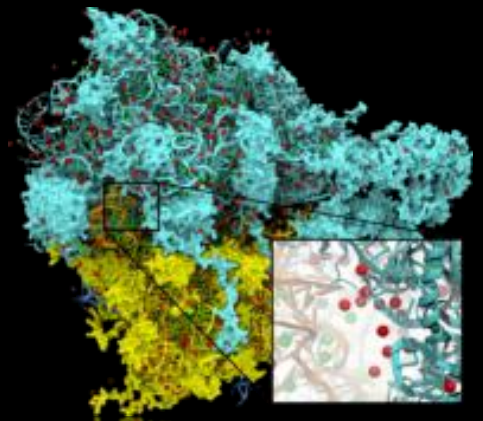
Application	Features Supported	GPU Speedup	Release Status	Website
BarraCUDA	Alignment of short sequencing reads	6-10x	Version 0.6.2 – 3/2012 Multi-GPU, multi-node	http://seqbarracuda.sourceforge.net/
CUDASW++	Parallel search of Smith-Waterman database	10-50x	Version 2.0.8 – Q1/2012 Multi-GPU, multi-node	http://sourceforge.net/projects/cudasw/
CUSHAW	Parallel, accurate long read aligner for large genomes	10x	Version 1.0.40 – 6/2012 Multiple-GPU	http://cushaw.sourceforge.net/
GPU-BLAST	Protein alignment according to BLASTP	3-4x	Version 2.2.26 – 3/2012 Single GPU	http://eudoxus.cheme.cmu.edu/gpublast/gpublast.html
GPU-HMMER	Parallel local and global search of Hidden Markov Models	60-100x	Version 2.3.2 – Q1/2012 Multi-GPU, multi-node	http://www.mpihmmer.org/installguideGPUHMMER.htm
mCUDA-MEME	Scalable motif discovery algorithm based on MEME	4-10x	Version 3.0.12 Multi-GPU, multi-node	https://sites.google.com/site/yongchaosoftware/mcuda-meme
SeqNFind	Hardware and software for reference assembly, blast, SW, HMM, de novo assembly	400x	Released. Multi-GPU, multi-node	http://www.seqnfind.com/
UGENE	Fast short read alignment	6-8x	Version 1.11 – 5/2012 Multi-GPU, multi-node	http://ugene.unipro.ru/
WideLM	Parallel linear regression on multiple similarly-shaped models	150x	Version 0.1-1 – 3/2012 Multi-GPU, multi-node	http://insilicos.com/products/widelm

GPU Perf compared against same or similar code running on single CPU machine
Performance measured internally or independently

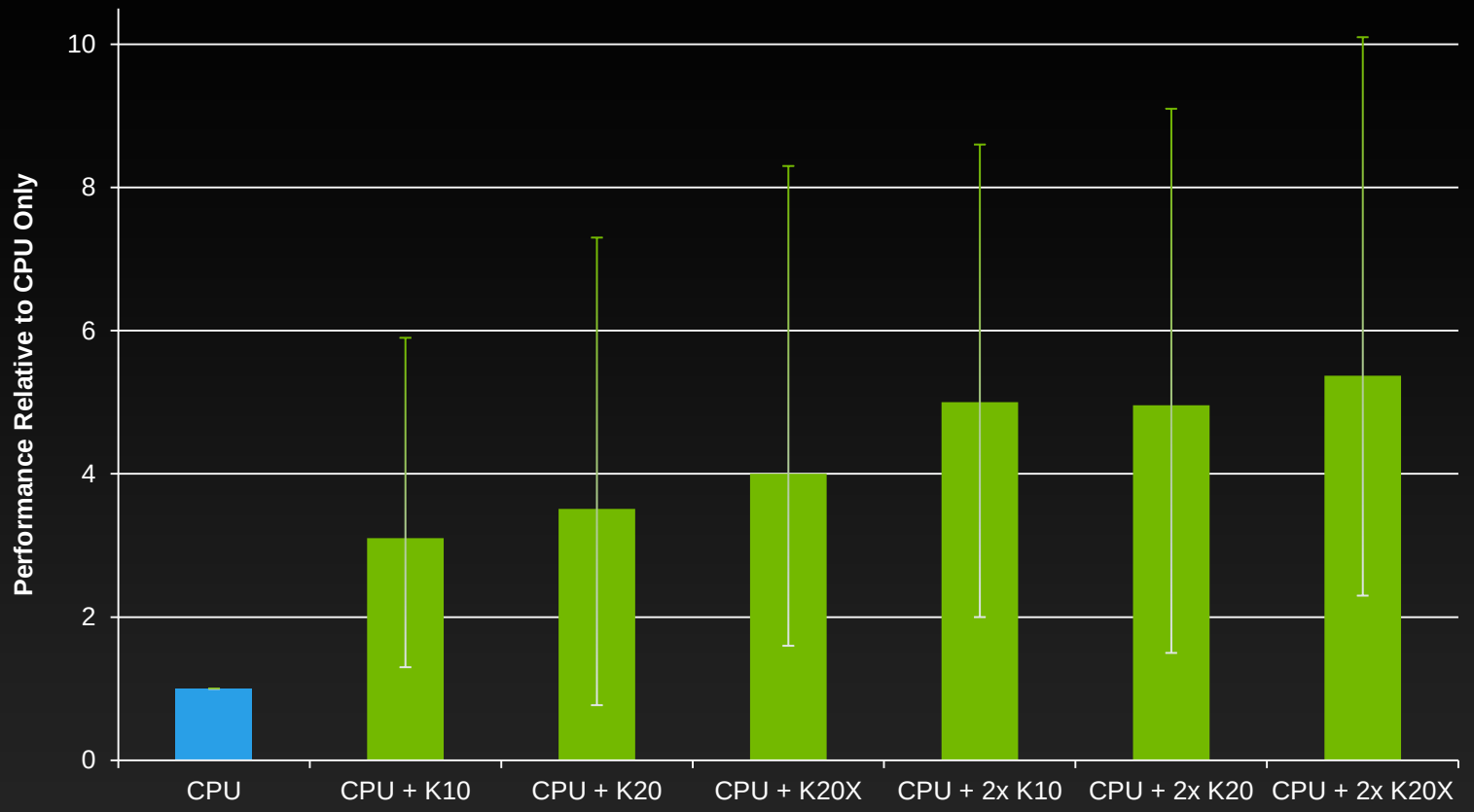
TESLA

Molecular Dynamics Module





MD Average Speedups



The blue node contains Dual E5-2687W CPUs (8 Cores per CPU).

The green nodes contain Dual E5-2687W CPUs (8 Cores per CPU) and 1 or 2 NVIDIA K10, K20, or K20X GPUs.

Average speedup calculated from 4 AMBER, 3 NAMD, 3 LAMMPS, and 1 GROMACS test cases.

Error bars show the maximum and minimum speedup for each hardware configuration.

Molecular Dynamics (MD) Applications

Application	Features Supported	GPU Perf	Release Status	Notes/Benchmarks
AMBER	PMEMD Explicit Solvent & GB Implicit Solvent	> 100 ns/day JAC NVE on 2X K20s	Released Multi-GPU, multi-node	AMBER 12, GPU Revision Support 12.2 http://ambermd.org/gpus/benchmarks.htm#Benchmarks
CHARMM	Implicit (5x), Explicit (2x) Solvent via OpenMM	2x C2070 equals 32-35x X5667 CPUs	Released Single & multi-GPU in single node	Release C37b1; http://www.charmm.org/news/c37b1.html#postjump
DL_POLY	Two-body Forces, Link-cell Pairs, Ewald SPME forces, Shake VV	4x	Release V 4.03 Multi-GPU, multi-node	Source only, Results Published http://www.stfc.ac.uk/CSE/randd/ccg/software/DL_POLY/25526.aspx
GROMACS	Implicit (5x), Explicit (2x)	165 ns/Day DHFR on 4X C2075s	Released Multi-GPU, multi-node	Release 4.6; 1st Multi-GPU support
LAMMPS	Lennard-Jones, Gay-Berne, Tersoff & many more potentials	3.5-18x on Titan	Released. Multi-GPU, multi-node	http://lammps.sandia.gov/bench.html#desktop and http://lammps.sandia.gov/bench.html#titan
NAMD	Full electrostatics with PME and most simulation features	4.0 ns/days F1-ATPase on 1x K20X	Released 100M atom capable Multi-GPU, multi-node	NAMD 2.9

GPU Perf compared against Multi-core x86 CPU socket.
GPU Perf benchmarked on GPU supported features
and may be a kernel to kernel perf comparison

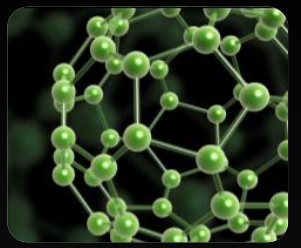
New/Additional MD Applications Ramping

Application	Features Supported	GPU Perf	Release Status	Notes
Abalone	Simulations (on 1060 GPU)	4-29X (on 1060 GPU)	Released, Version 1.8.51 Single GPU	Agile Molecule, Inc.
Ascalaph	Computation of non-valent interactions	4-29X (on 1060 GPU)	Released, Version 1.1.4 Single GPU	Agile Molecule, Inc.
ACEMD	Written for use only on GPUs	150 ns/day DHFR on 1x K20	Released Single and multi-GPUs	Production bio-molecular dynamics (MD) software specially optimized to run on GPUs
Folding@Home	Powerful distributed computing molecular dynamics system; implicit solvent and folding	Depends upon number of GPUs	Released; GPUs and CPUs	http://folding.stanford.edu GPUs get 4X the points of CPUs
GPUGrid.net	High-performance all-atom biomolecular simulations; explicit solvent and binding	Depends upon number of GPUs	Released; NVIDIA GPUs only	http://www.gpugrid.net/
HALMD	Simple fluids and binary mixtures (pair potentials, high-precision NVE and NVT, dynamic correlations)	Up to 66x on 2090 vs. 1 CPU core	Released, Version 0.2.0 Single GPU	http://halmd.org/benchmarks.html#supercooled-binary-mixture-kob-andersen
HOOMD-Blue	Written for use only on GPUs	Kepler 2X faster than Fermi	Released, Version 0.11.2 Single and multi-GPU on 1 node	http://codeblue.umich.edu/hoomd-blue/ Multi-GPU w/ MPI in March 2013
OpenMM	Implicit and explicit solvent, custom forces	Implicit: 127-213 ns/day Explicit: 18-55 ns/day DHFR	Released Version 4.1.1 Multi-GPU	Library and application for molecular dynamics on high-performance

GPU Perf compared against Multi-core x86 CPU socket.
GPU Perf benchmarked on GPU supported features and may be a kernel to kernel perf comparison

Built from Ground Up for GPUs

Computational Chemistry



What

- Study disease & discover drugs
- Predict drug and protein interactions

Why

- Speed of simulations is critical
- Enables study of:
  - Longer timeframes
  - Larger systems
  - More simulations

How

- GPUs increase throughput & accelerate simulations

AMBER 11 Application
4.6x performance increase with 2 GPUs with only a 54% added cost*

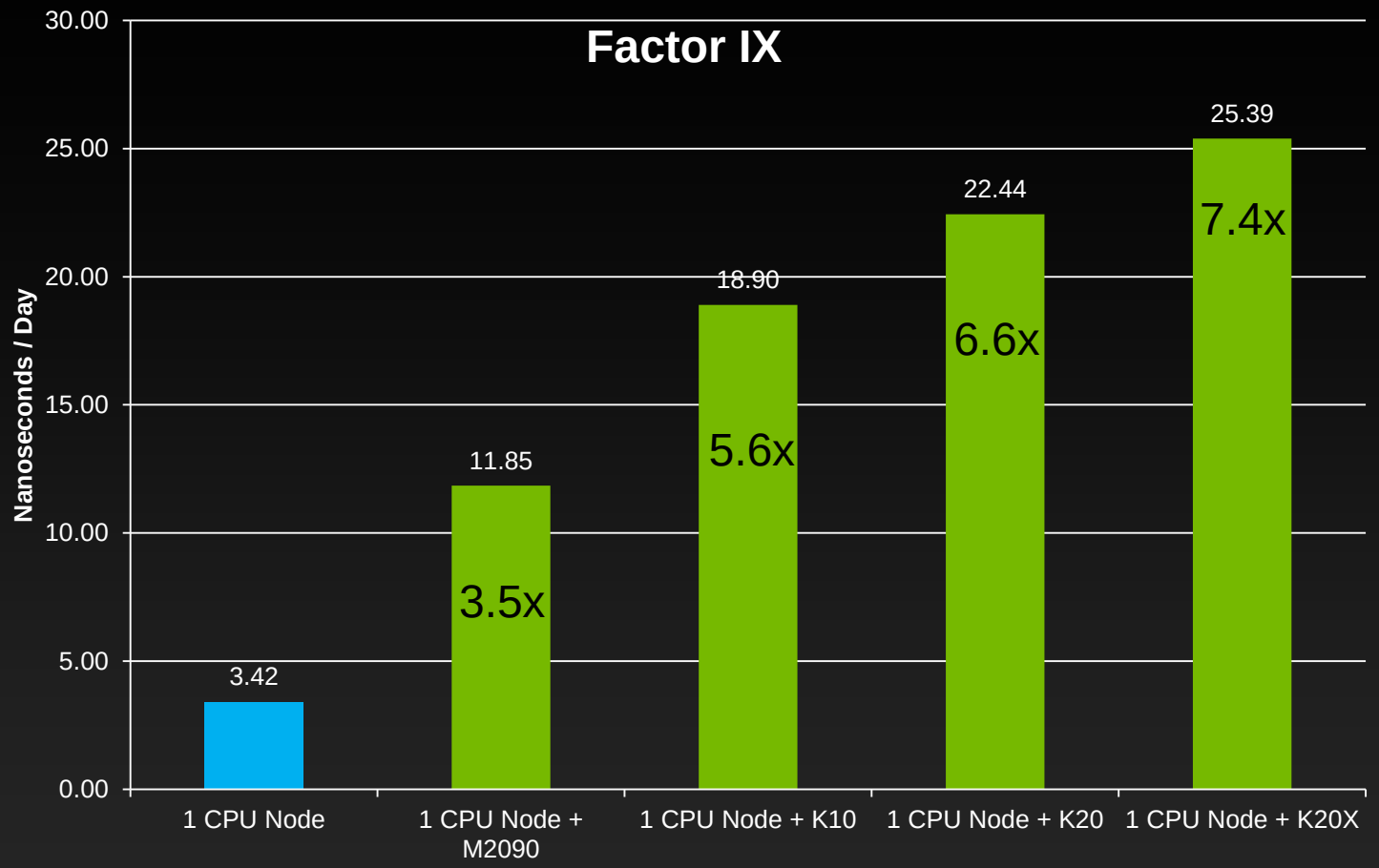
- AMBER 11 Cellulose NPT on 2x E5670 CPUS + 2x Tesla C2090s (per node) vs. 2xcE5670 CPUs (per node)
- Cost of CPU node assumed to be $9333. Cost of adding two (2) 2090s to single node is assumed to be $5333

GPU READY APPLICATIONS

Abalone
ACEMD
AMBER
DL_PLOY
GAMESS
GROMACS
LAMMPS
NAMD
NWChem
Q-CHEM
Quantum Espresso
TeraChem

AMBER 12
GPU Support Revision 12.2
1/22/2013

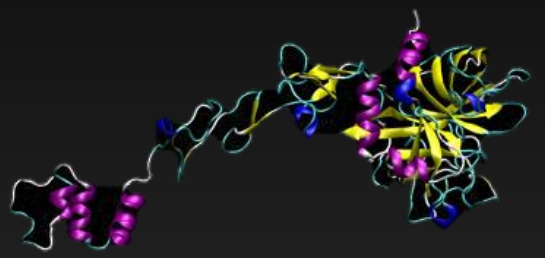
Kepler - Our Fastest Family of GPUs Yet



Running AMBER 12 GPU Support Revision 12.1

The blue node contains Dual E5-2687W CPUs (8 Cores per CPU).

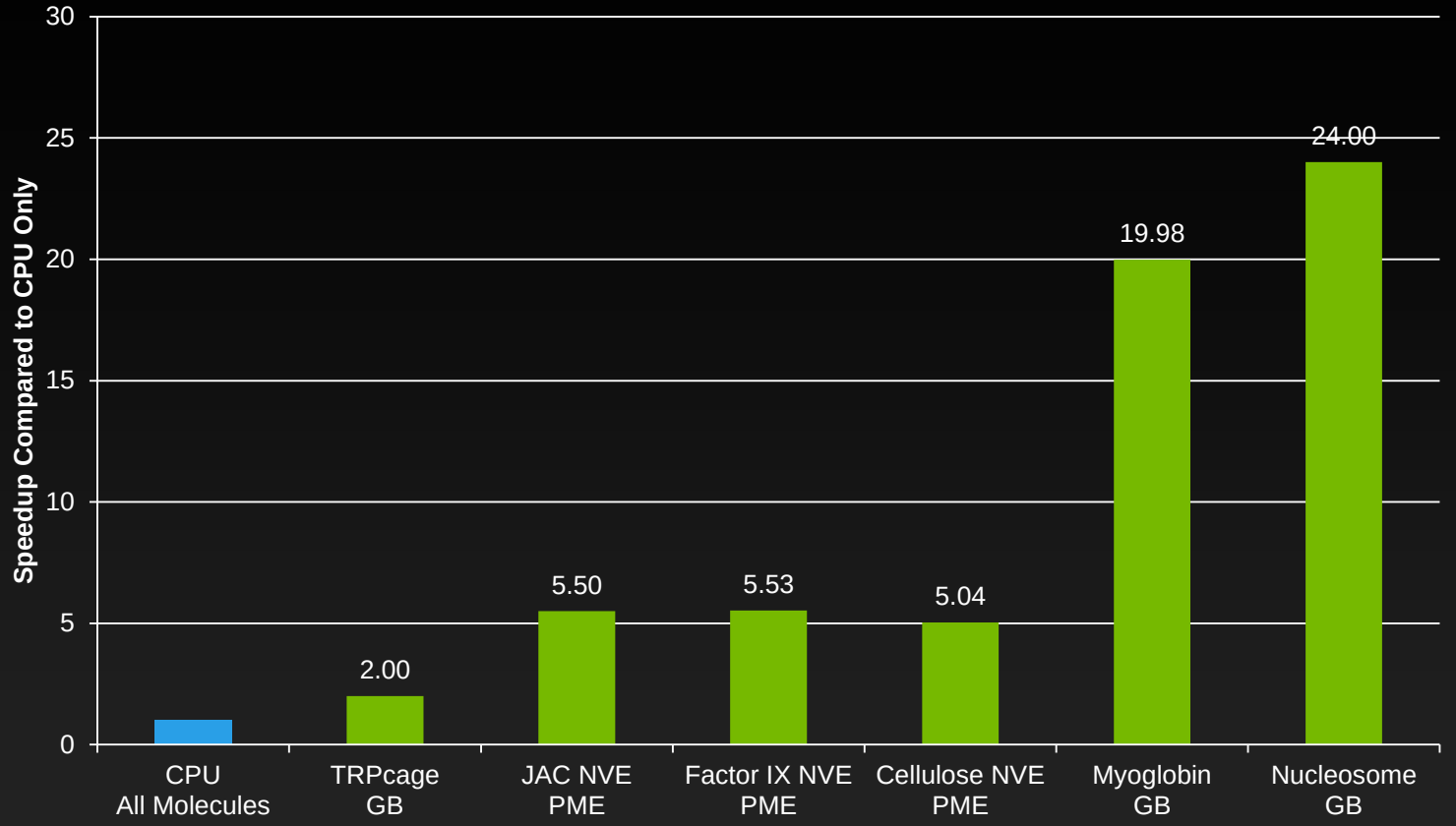
The green nodes contain Dual E5-2687W CPUs (8 Cores per CPU) and either 1x NVIDIA M2090, 1x K10 or 1x K20 for the GPU



Factor IX

GPU speedup/throughput increased from 3.5x (with M2090) to 7.4x (with K20X) when compared to a CPU only node

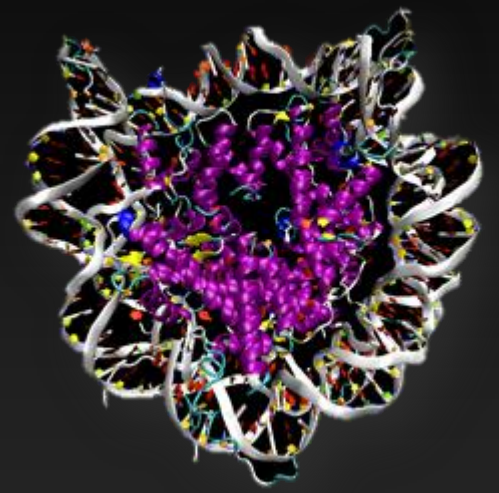
K10 Accelerates Simulations of All Sizes



Running AMBER 12 GPU Support Revision 12.1

The blue node contains Dual E5-2687W CPUs (8 Cores per CPU).

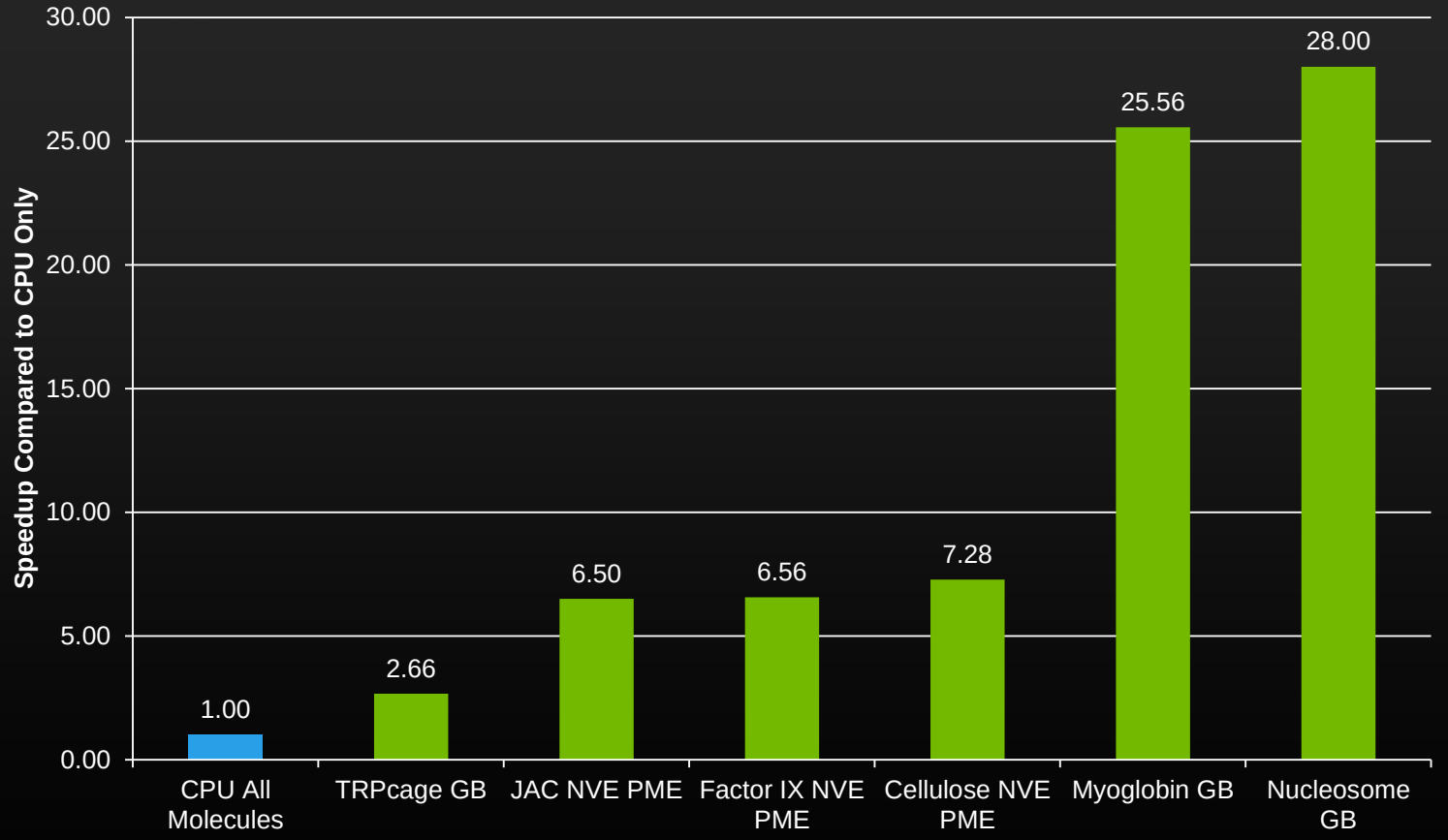
The green nodes contain Dual E5-2687W CPUs (8 Cores per CPU) and 1x NVIDIA K10 GPU



Nucleosome

Gain 24x performance by adding just 1 GPU when compared to dual CPU performance

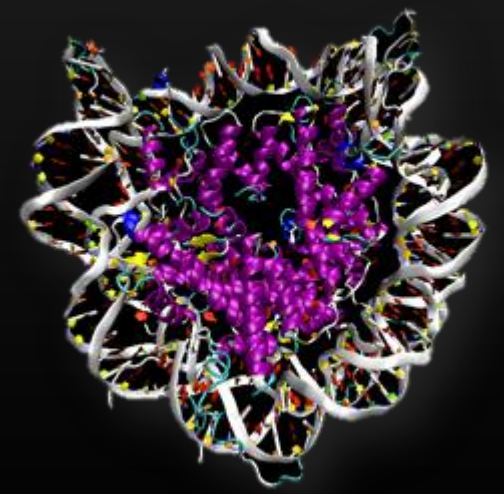
K20 Accelerates Simulations of All Sizes



Running AMBER 12 GPU Support Revision 12.1 SPFP with CUDA 4.2.9 ECC Off

The blue node contains 2x Intel E5-2687W CPUs (8 Cores per CPU)

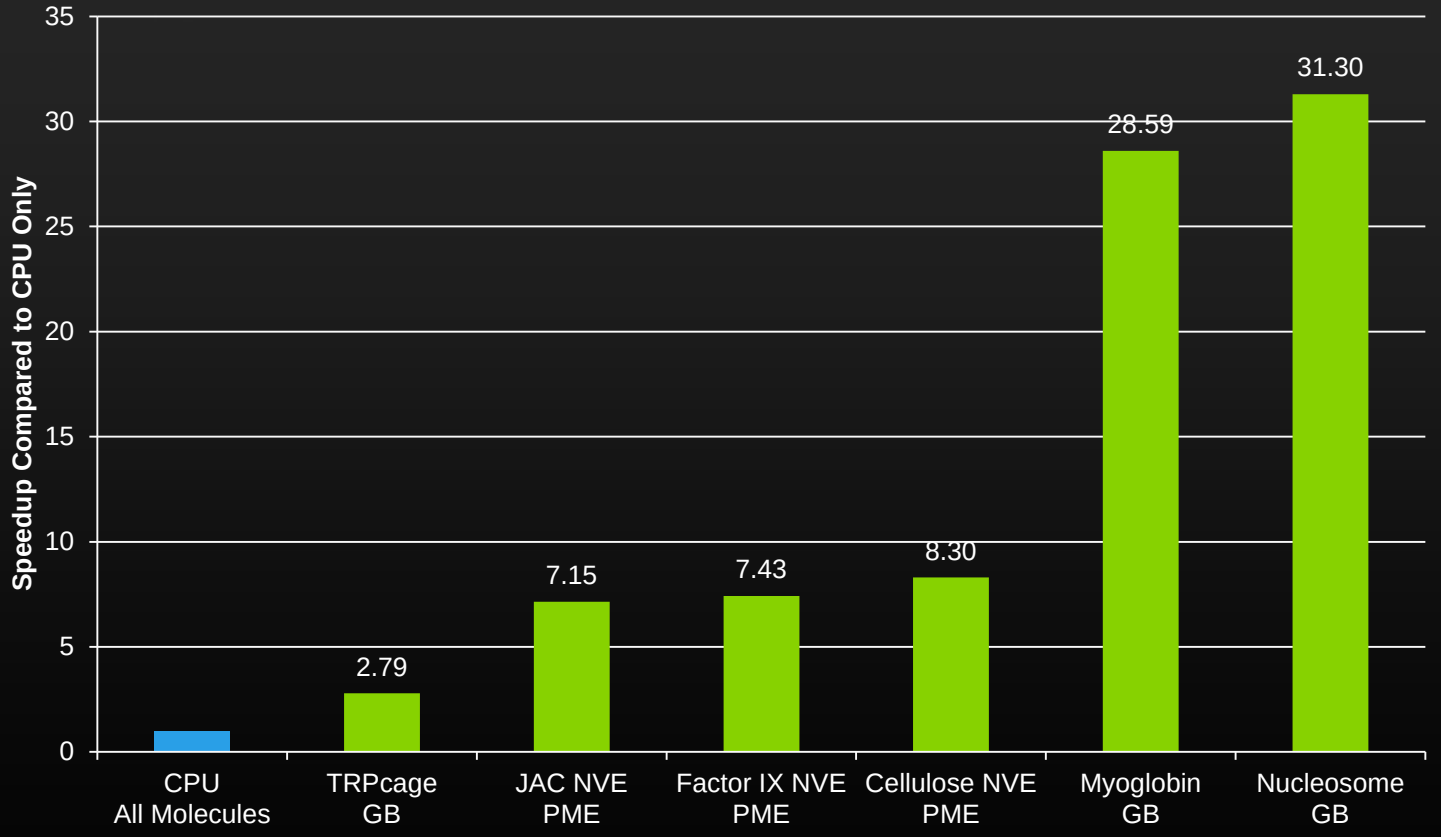
Each green nodes contains 2x Intel E5-2687W CPUs (8 Cores per CPU) plus 1x NVIDIA K20 GPUs



Nucleosome

Gain 28x throughput/performance by adding just one K20 GPU when compared to dual CPU performance

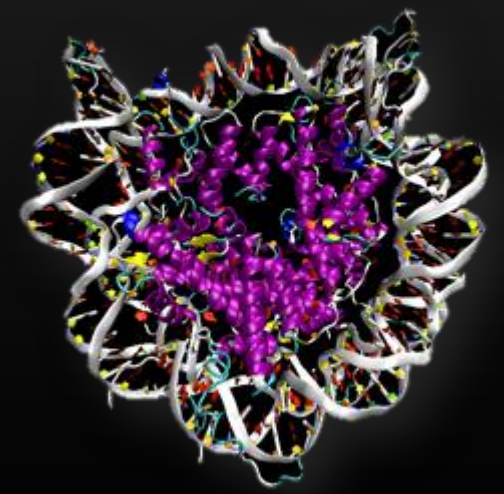
K20X Accelerates Simulations of All Sizes



Running AMBER 12 GPU Support Revision 12.1

The blue node contains Dual E5-2687W CPUs (8 Cores per CPU).

The green nodes contain Dual E5-2687W CPUs (8 Cores per CPU) and 1x NVIDIA K20X GPU

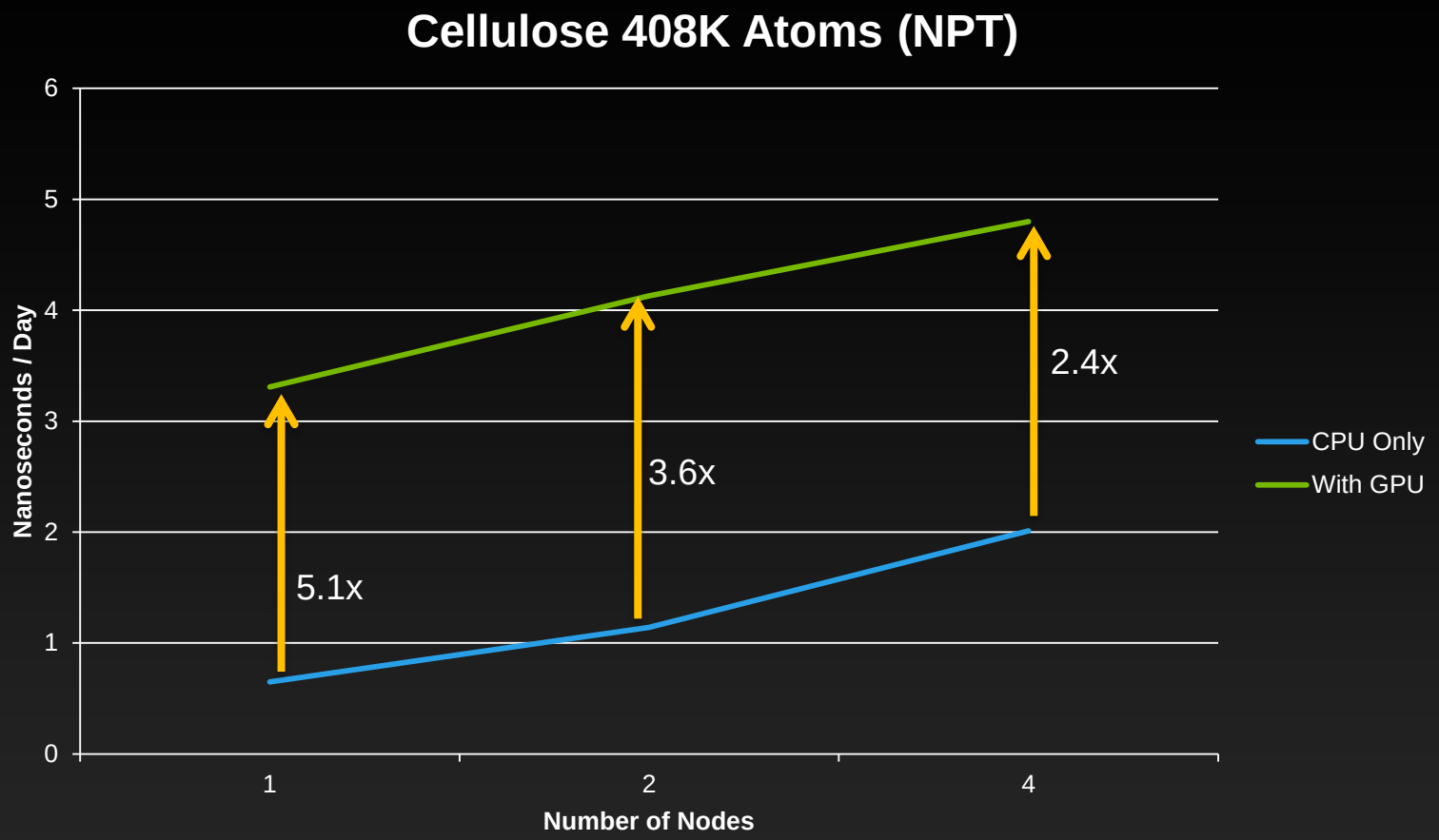


Nucleosome

Gain 31x performance by adding just one K20X GPU when compared to dual CPU performance

AMBER Benchmark Report, Revision 2.0, dated Nov. 5, 2012

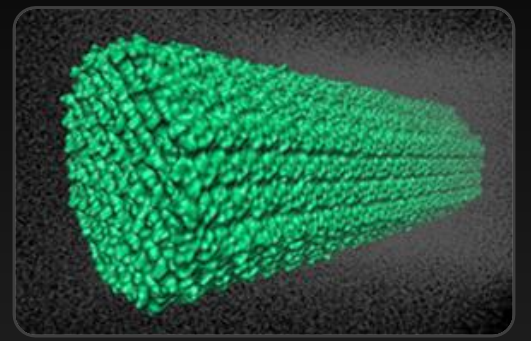
K10 Strong Scaling over Nodes



Running AMBER 12 with CUDA 4.2 ECC Off

The blue nodes contains 2x Intel X5670 CPUs (6 Cores per CPU)

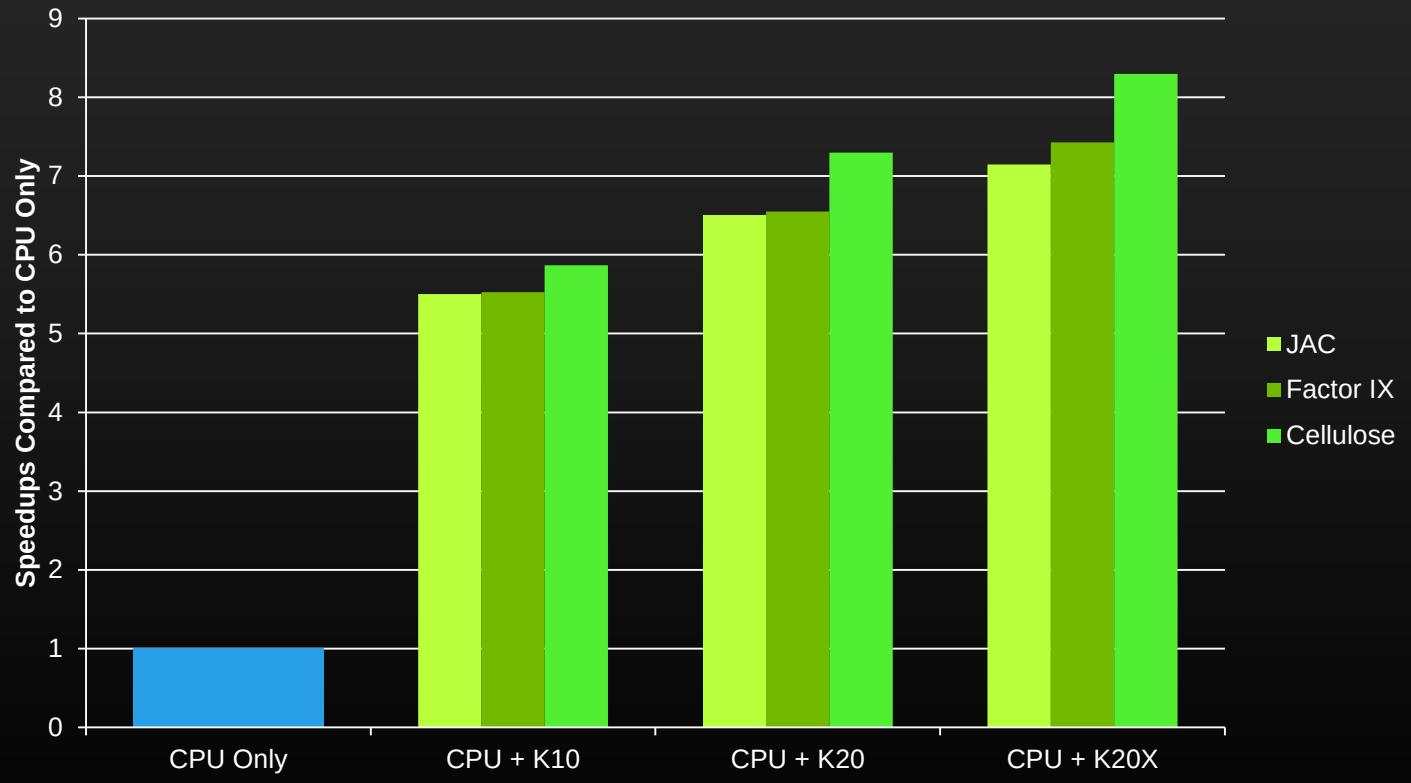
The green nodes contains 2x Intel X5670 CPUs (6 Cores per CPU) plus 2x NVIDIA K10 GPUs



Cellulose

GPUs significantly outperform CPUs while scaling over multiple nodes

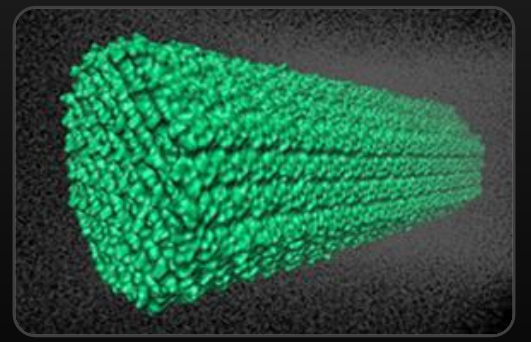
Kepler – Universally Faster



Running AMBER 12 GPU Support Revision 12.1

The CPU Only node contains Dual E5-2687W CPUs (8 Cores per CPU).

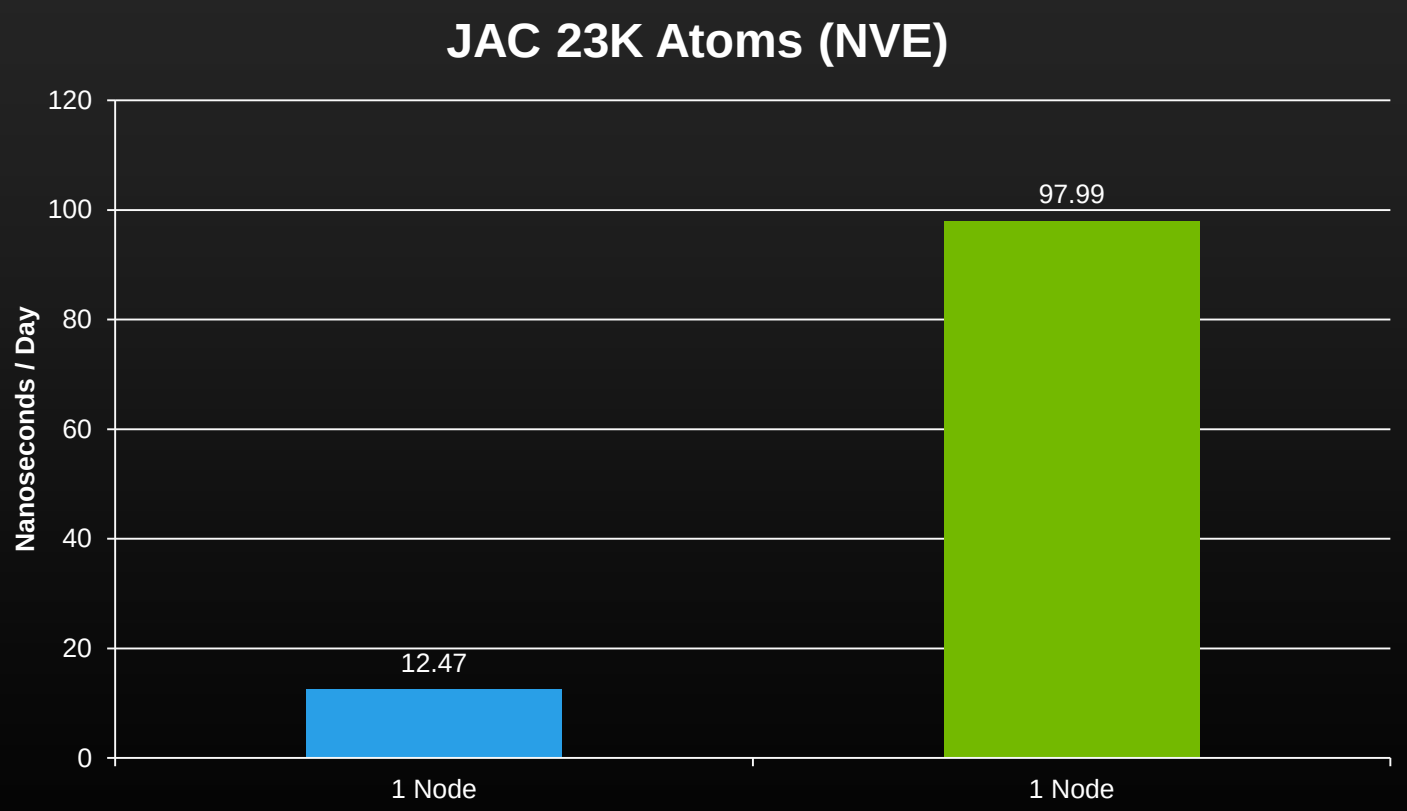
The Kepler nodes contain Dual E5-2687W CPUs (8 Cores per CPU) and 1x NVIDIA K10, K20, or K20X GPUs



Cellulose

The Kepler GPUs accelerated all simulations, up to 8x

K10 Extreme Performance

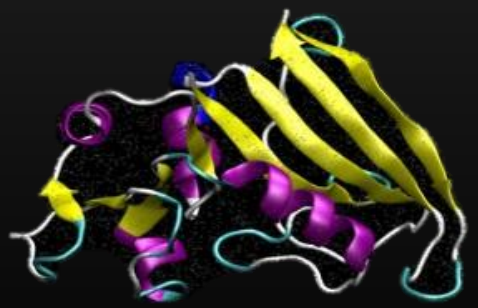


Gain 7.8X performance by adding just 2 GPUs when compared to dual CPU performance

Running AMBER 12 GPU Support Revision 12.1

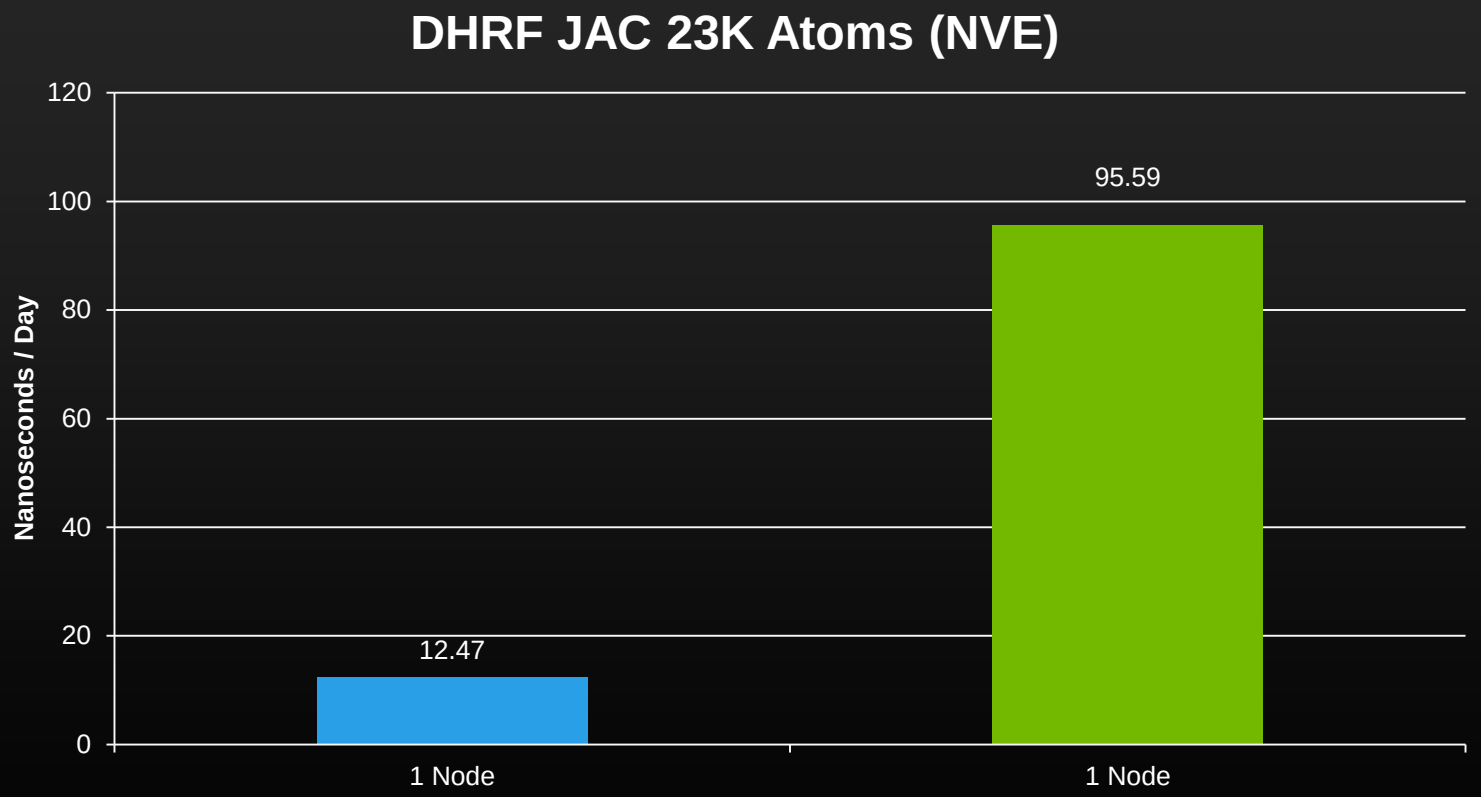
The blue node contains Dual E5-2687W CPUs (8 Cores per CPU).

The green node contain Dual E5-2687W CPUs (8 Cores per CPU) and 2x NVIDIA K10 GPUs



DHFR

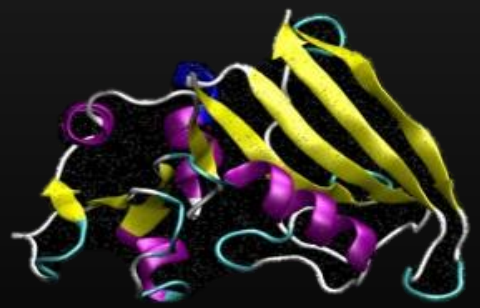
K20 Extreme Performance



Running AMBER 12 GPU Support Revision 12.1 SPFP with CUDA 4.2.9 ECC Off

The blue node contains 2x Intel E5-2687W CPUs (8 Cores per CPU)

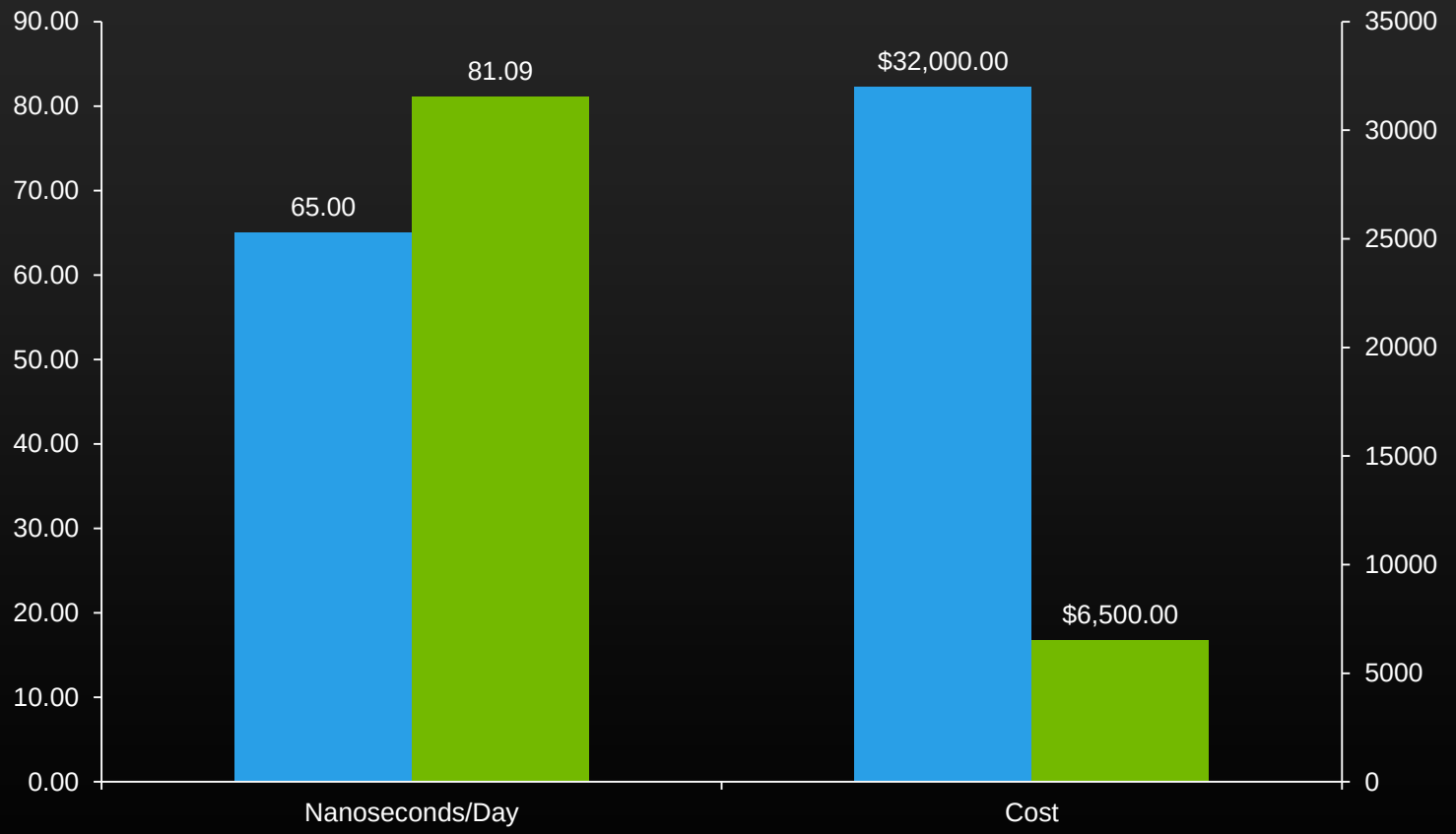
Each green node contains 2x Intel E5-2687W CPUs (8 Cores per CPU) plus 2x NVIDIA K20 GPU



DHFR

Gain > 7.5X throughput/performance by adding just 2 K20 GPUs when compared to dual CPU performance

Replace 8 Nodes with 1 K20 GPU

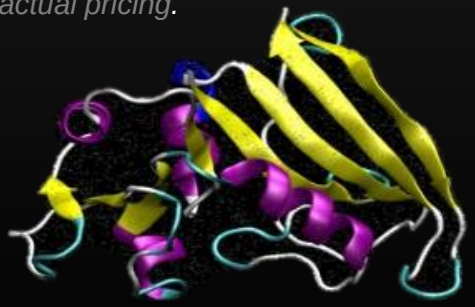


Running AMBER 12 GPU Support Revision 12.1 SPFP with CUDA 4.2.9 ECC Off

The eight (8) blue nodes each contain 2x Intel E5-2687W CPUs (8 Cores per CPU)

Each green node contains 2x Intel E5-2687W CPUs (8 Cores per CPU) plus 1x NVIDIA K20 GPU

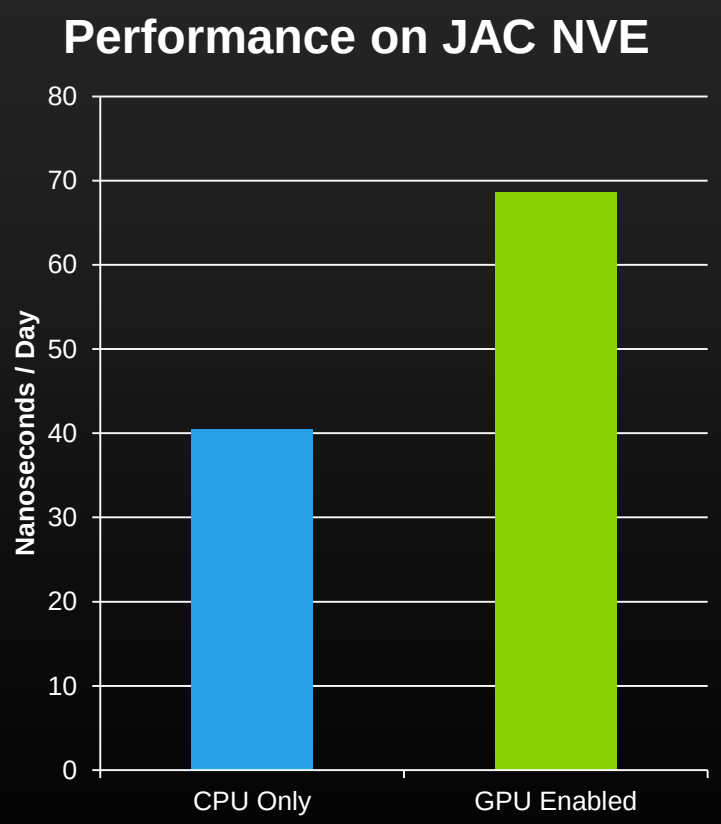
Note: Typical CPU and GPU node pricing used. Pricing may vary depending on node configuration. Contact your preferred HW vendor for actual pricing.



DHFR

Cut down simulation costs to ¼ and gain higher performance

Replace 7 Nodes with 1 K10 GPU

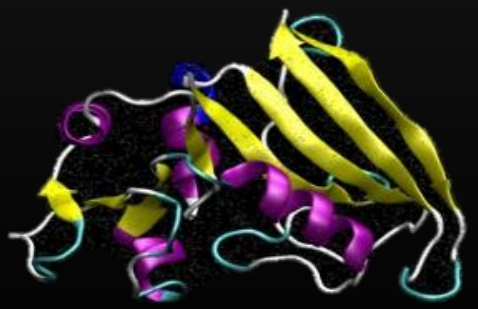


Running AMBER 12 GPU Support Revision 12.1 SPFP with CUDA 4.2.9 ECC Off

The eight (8) blue nodes each contain 2x Intel E5-2687W CPUs (8 Cores per CPU)

The green node contains 2x Intel E5-2687W CPUs (8 Cores per CPU) plus 1x NVIDIA K10 GPU

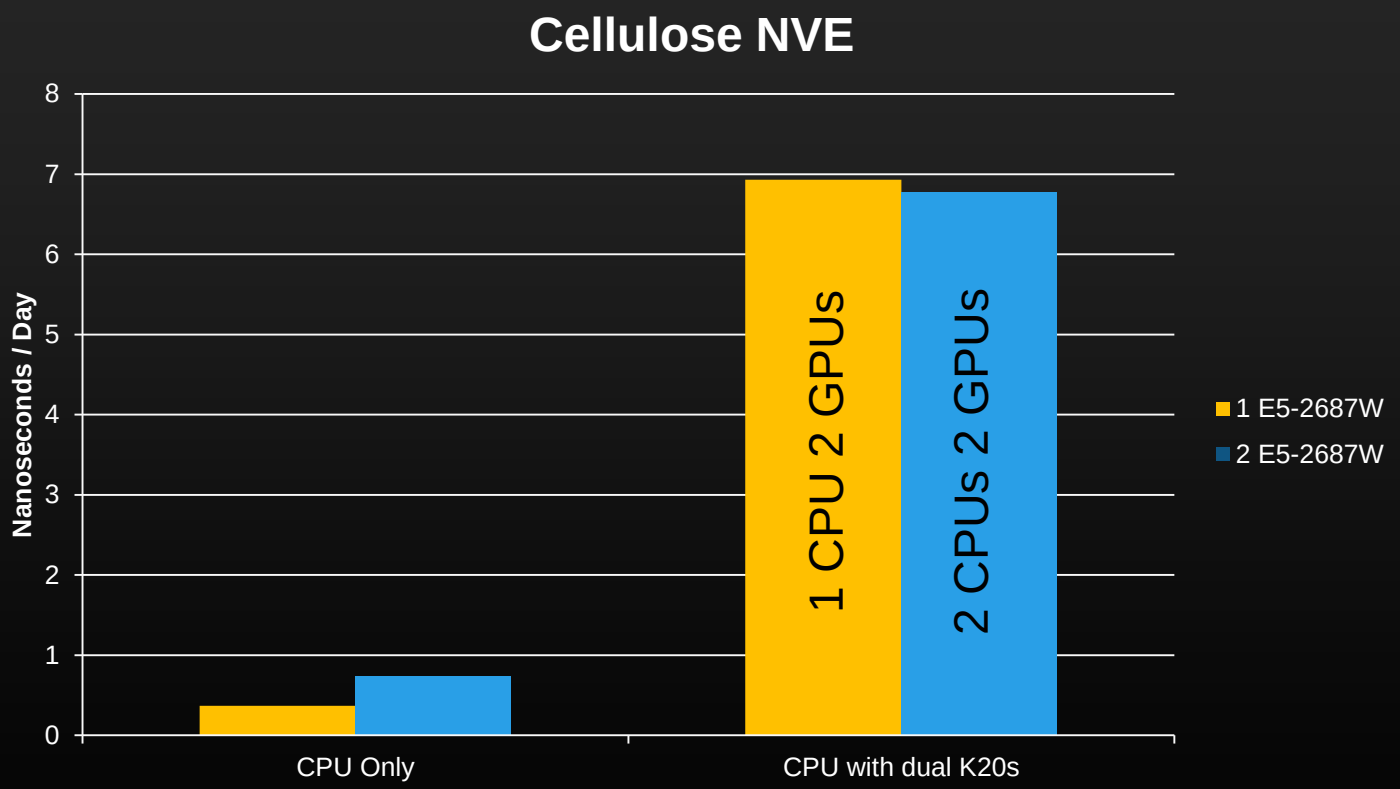
Note: Typical CPU and GPU node pricing used. Pricing may vary depending on node configuration. Contact your preferred HW vendor for actual pricing.



DHFR

Cut down simulation costs to ¼ and increase performance by 70%

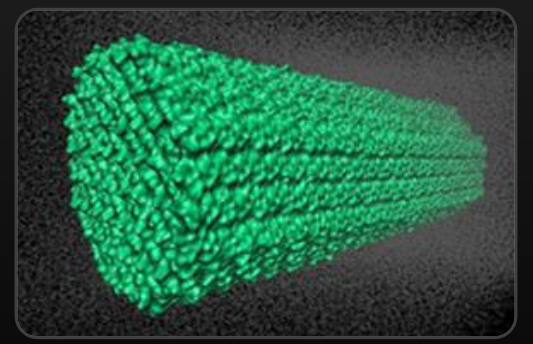
Extra CPUs decrease Performance



Running AMBER 12 GPU Support Revision 12.1

The orange bars contains one E5-2687W CPUs (8 Cores per CPU).

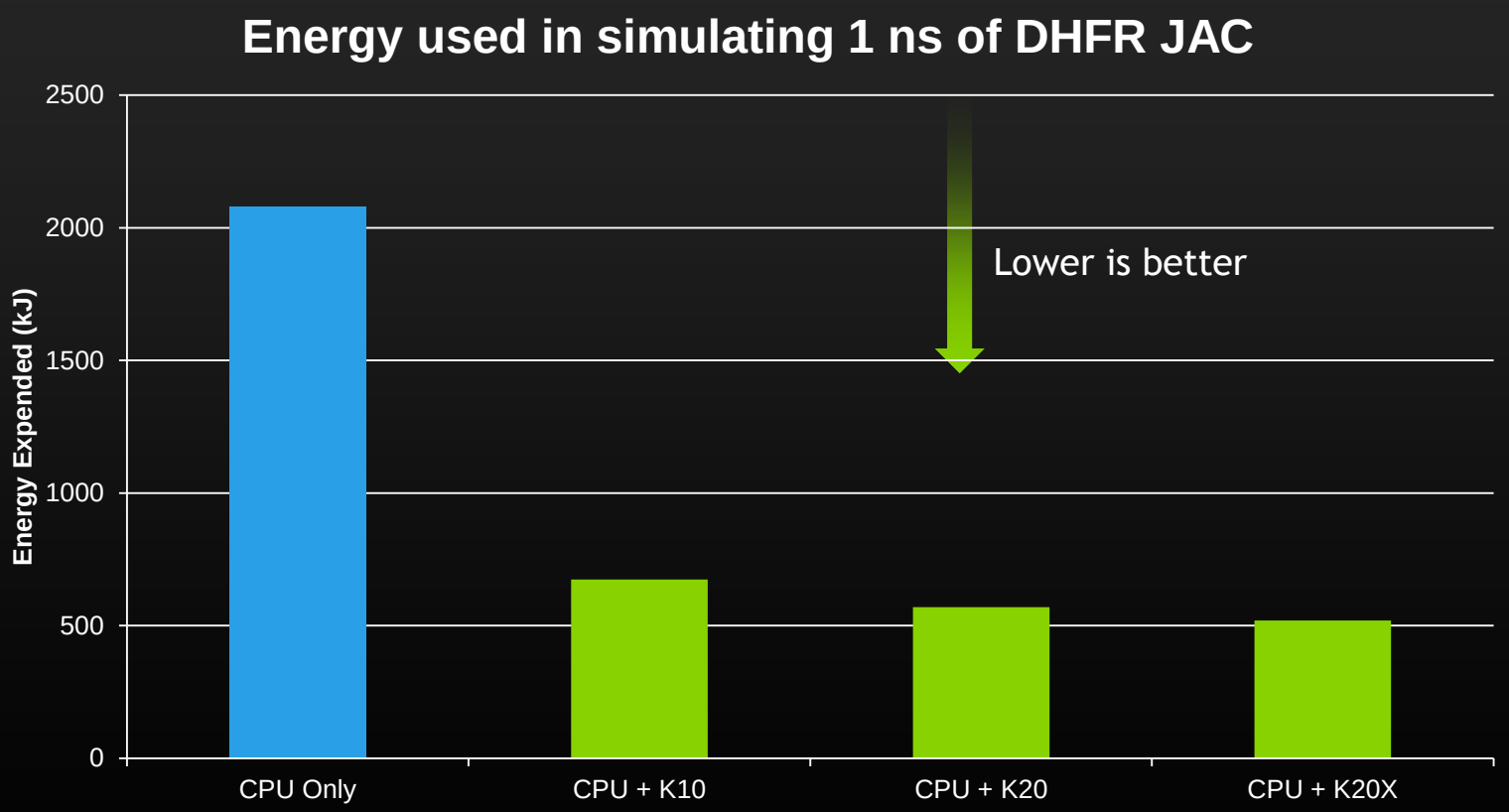
The blue bars contain Dual E5-2687W CPUs (8 Cores per CPU)



Cellulose

When used with GPUs, dual CPU sockets perform worse than single CPU sockets.

Kepler - Greener Science



Running AMBER 12 GPU Support Revision 12.1

The blue node contains Dual E5-2687W CPUs (150W each, 8 Cores per CPU).

The green nodes contain Dual E5-2687W CPUs (8 Cores per CPU) and 1x NVIDIA K10, K20, or K20X GPUs (235W each).

Energy Expended = Power x Time

The GPU Accelerated systems use 65-75% less energy

Recommended GPU Node Configuration for AMBER Computational Chemistry

Workstation or Single Node Configuration	
# of CPU sockets	2
Cores per CPU socket	4+ (1 CPU core drives 1 GPU)
CPU speed (Ghz)	2.66+
System memory per node (GB)	16
GPUs	Kepler K10, K20, K20X Fermi M2090, M2075, C2075
# of GPUs per CPU socket	1-2 (4 GPUs on 1 socket is good to do 4 fast serial GPU runs)
GPU memory preference (GB)	6
GPU to CPU connection	PCIe 2.0 16x or higher
Server storage	2 TB
Network configuration	Infiniband QDR or better

Scale to multiple nodes with same single node configuration

AMBER Benchmark Report, Revision 2.0, dated Nov. 5, 2012

Benefits of GPU AMBER Accelerated Computing

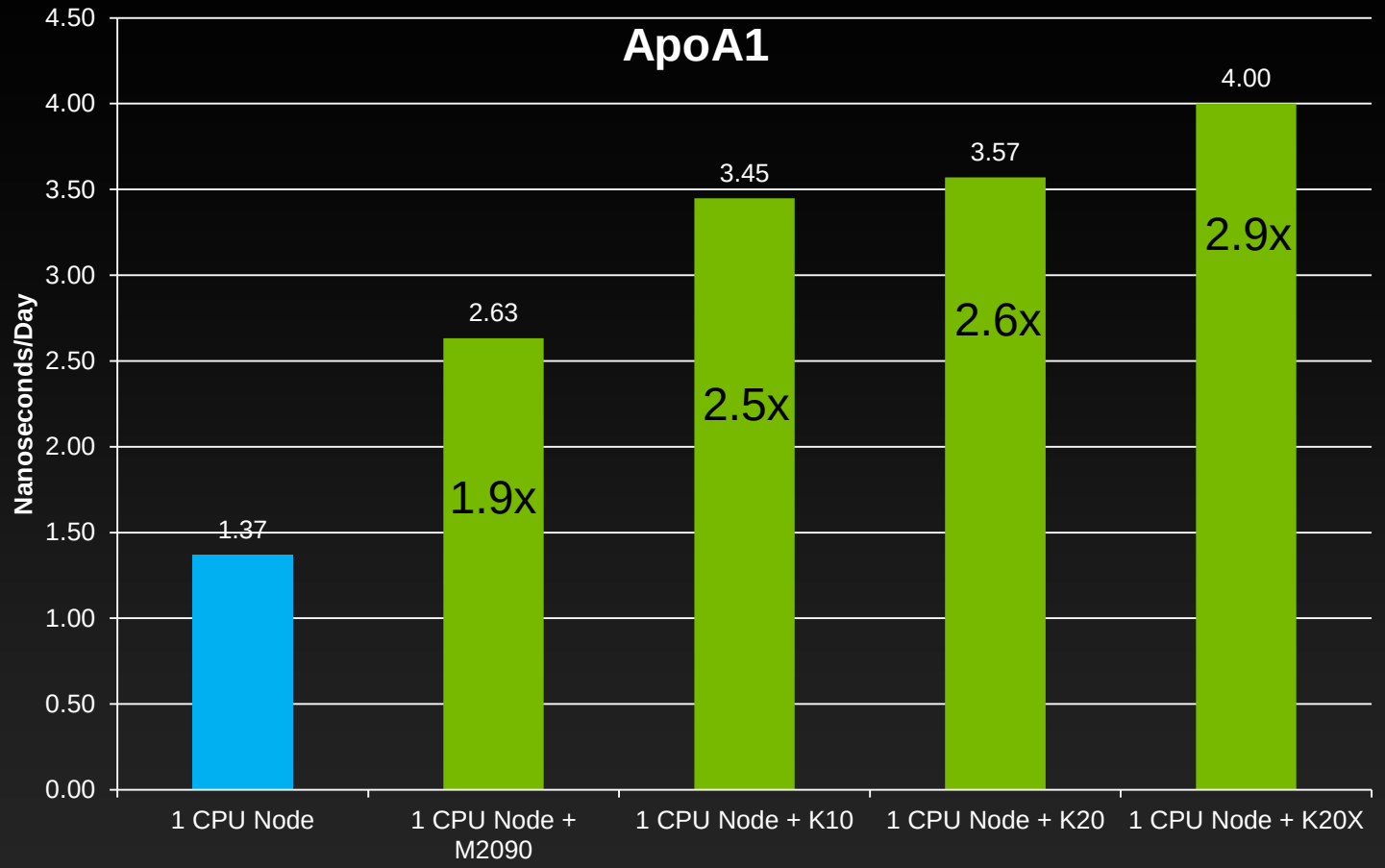
- Faster than CPU only systems in all tests
- Most major compute intensive aspects of classical MD ported
- Large performance boost with marginal price increase
- Energy usage cut by more than half
- GPUs scale well within a node and over multiple nodes
- K20 GPU is our fastest and lowest power high performance GPU yet

Try GPU accelerated AMBER for free – www.nvidia.com/GPUTestDrive

AMBER Benchmark Report, Revision 2.0, dated Nov. 5, 2012

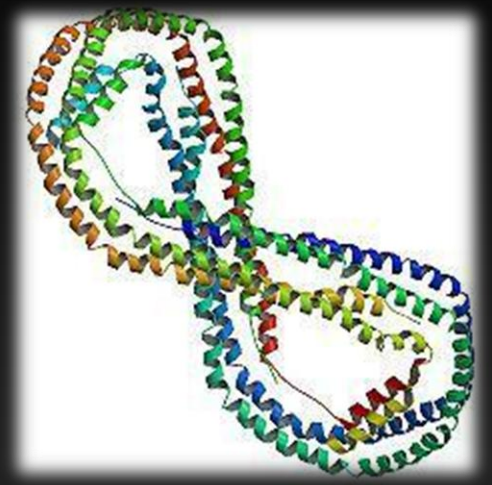
NAMD 2.9

Kepler - Our Fastest Family of GPUs Yet



Running NAMD version 2.9
The blue node contains Dual E5-2687W CPUs (8 Cores per CPU).

The green nodes contain Dual E5-2687W CPUs (8 Cores per CPU) and either 1x NVIDIA M2090, 1x K10 or 1x K20 for the GPU

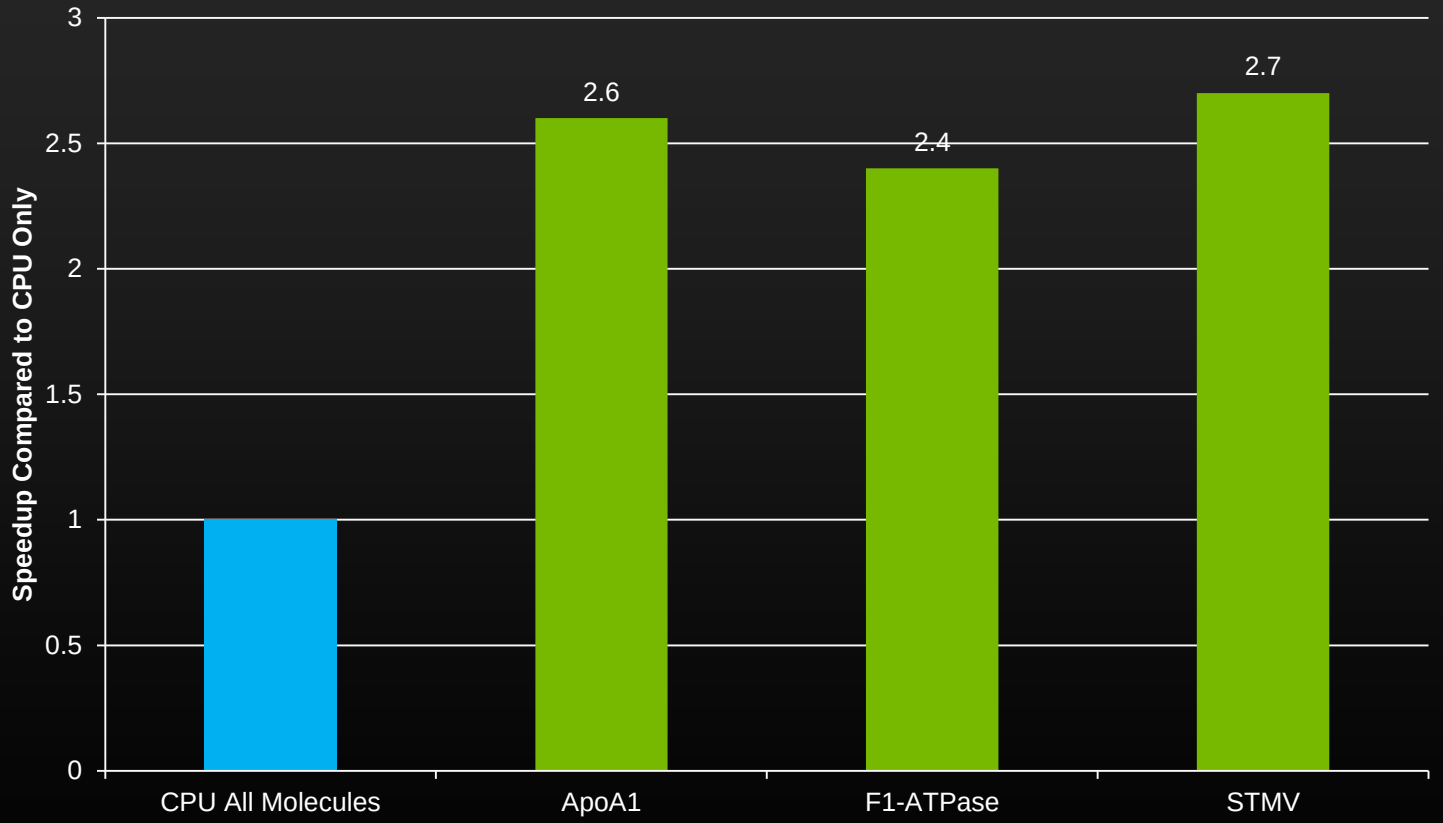


Apolipoprotein A1

GPU speedup/throughput increased from 1.9x (with M2090) to 2.9x (with K20X) when compared to a CPU only node

NAMD Benchmark Report, Revision 2.0, dated Nov. 5, 2012

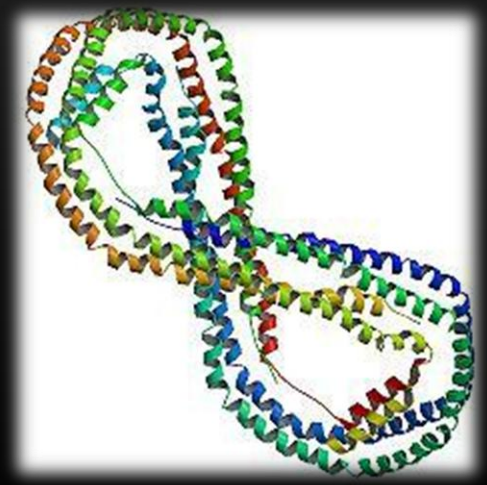
Accelerates Simulations of All Sizes



Running NAMD 2.9 with CUDA 4.0 ECC Off

The blue node contains 2x Intel E5-2687W CPUs (8 Cores per CPU)

Each green node contains 2x Intel E5-2687W CPUs (8 Cores per CPU) plus 1x NVIDIA K20 GPUs

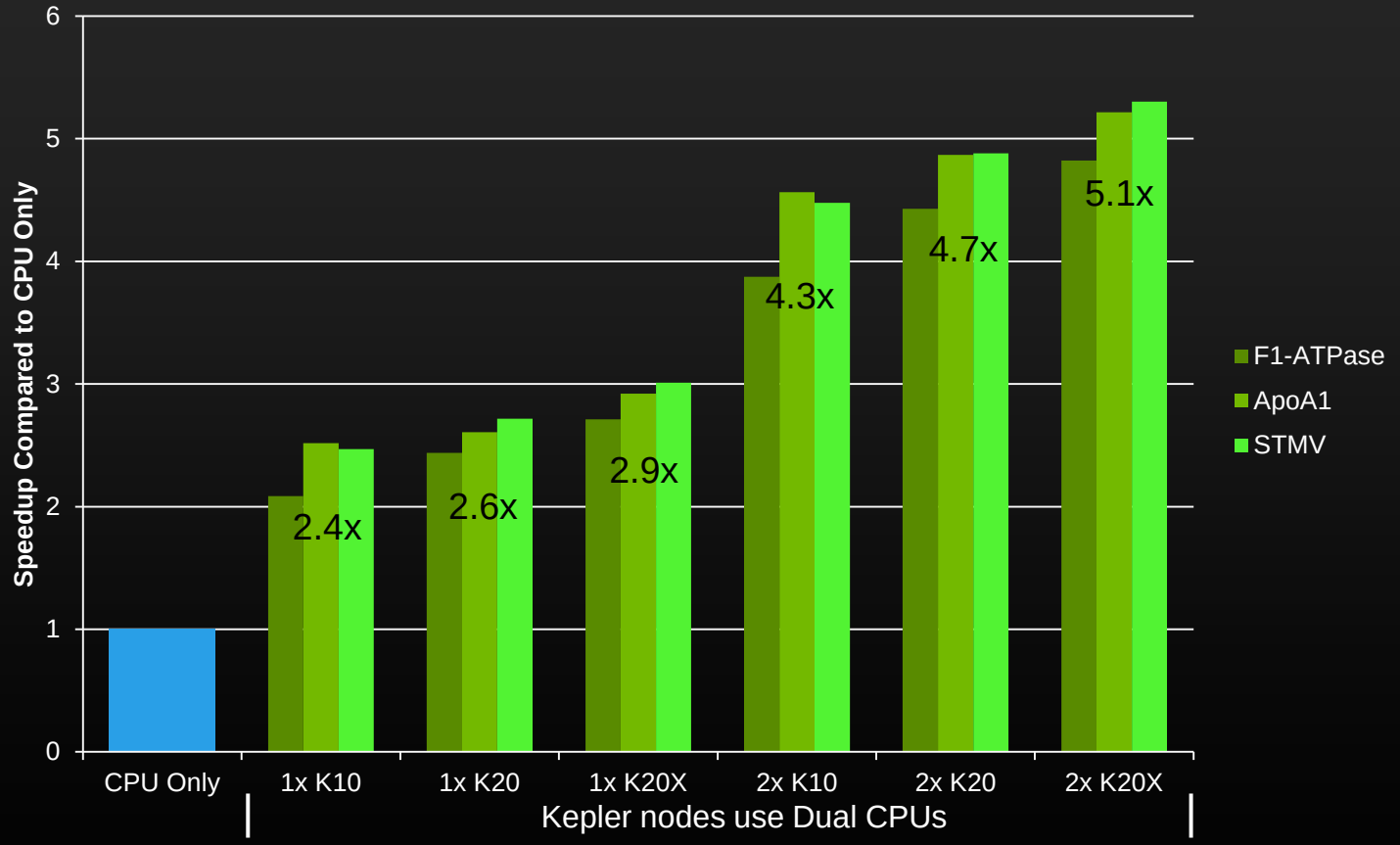


Apolipoprotein A1

Gain 2.5x throughput/performance by adding just 1 GPU when compared to dual CPU performance

NAMD Benchmark Report, Revision 2.0, dated Nov. 5, 2012

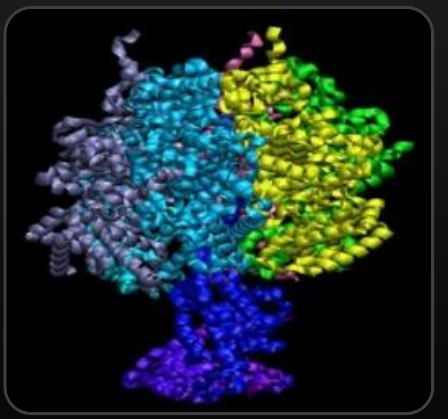
Kepler – Universally Faster



Running NAMD version 2.9

The CPU Only node contains Dual E5-2687W CPUs (8 Cores per CPU).

The Kepler nodes contain Dual E5-2687W CPUs (8 Cores per CPU) and 1 or two NVIDIA K10, K20, or K20X GPUs.

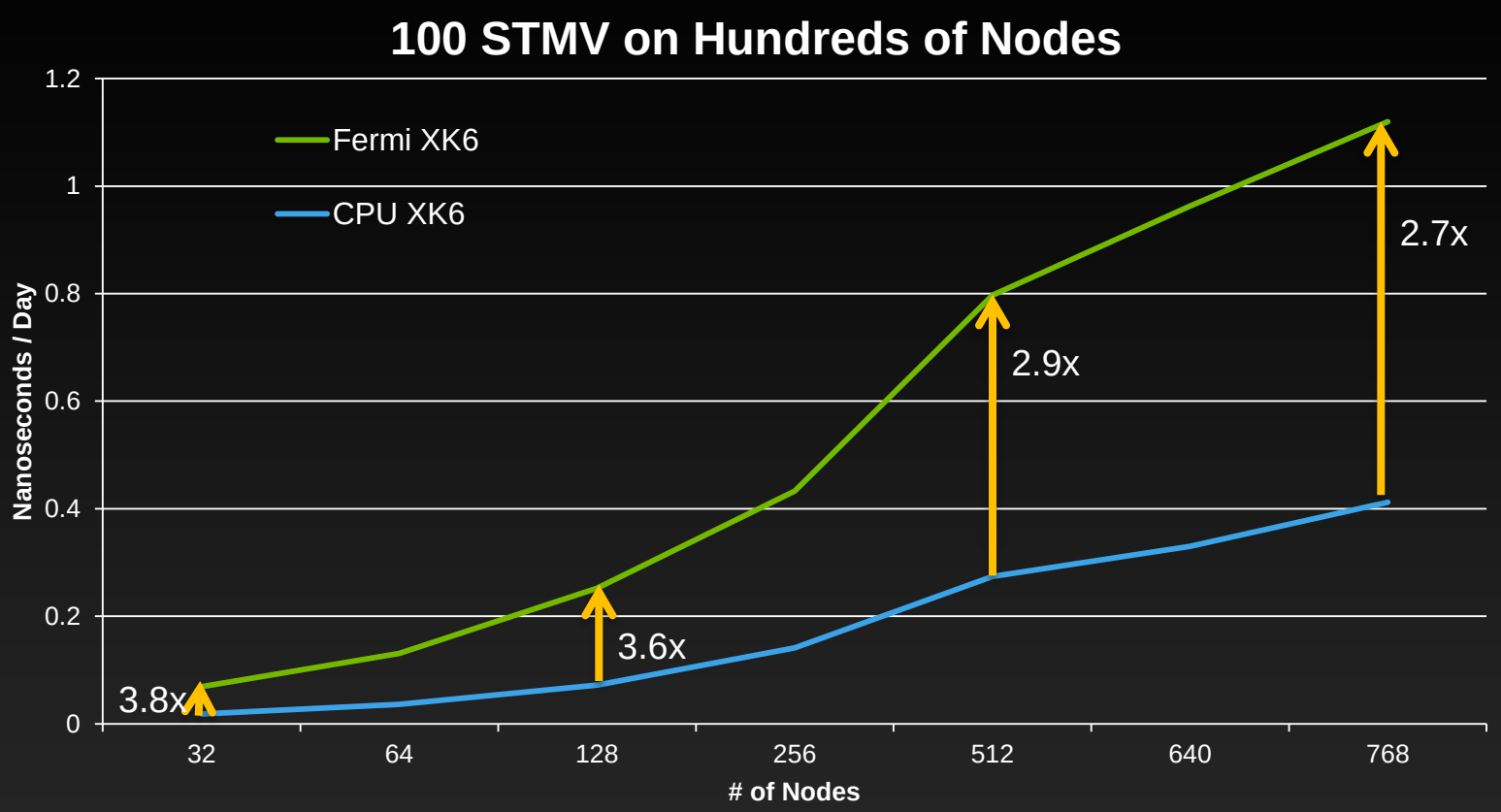


F1-ATPase

The Kepler GPUs accelerate all simulations, up to 5x

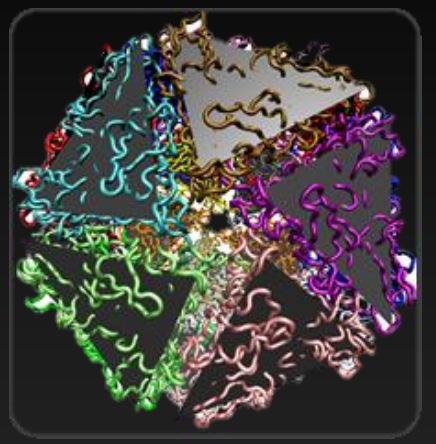
Average acceleration printed in bars

Outstanding Strong Scaling with Multi-STMV



Running NAMD version 2.9
Each blue XE6 CPU node contains 1x AMD 1600 Opteron (16 Cores per CPU).

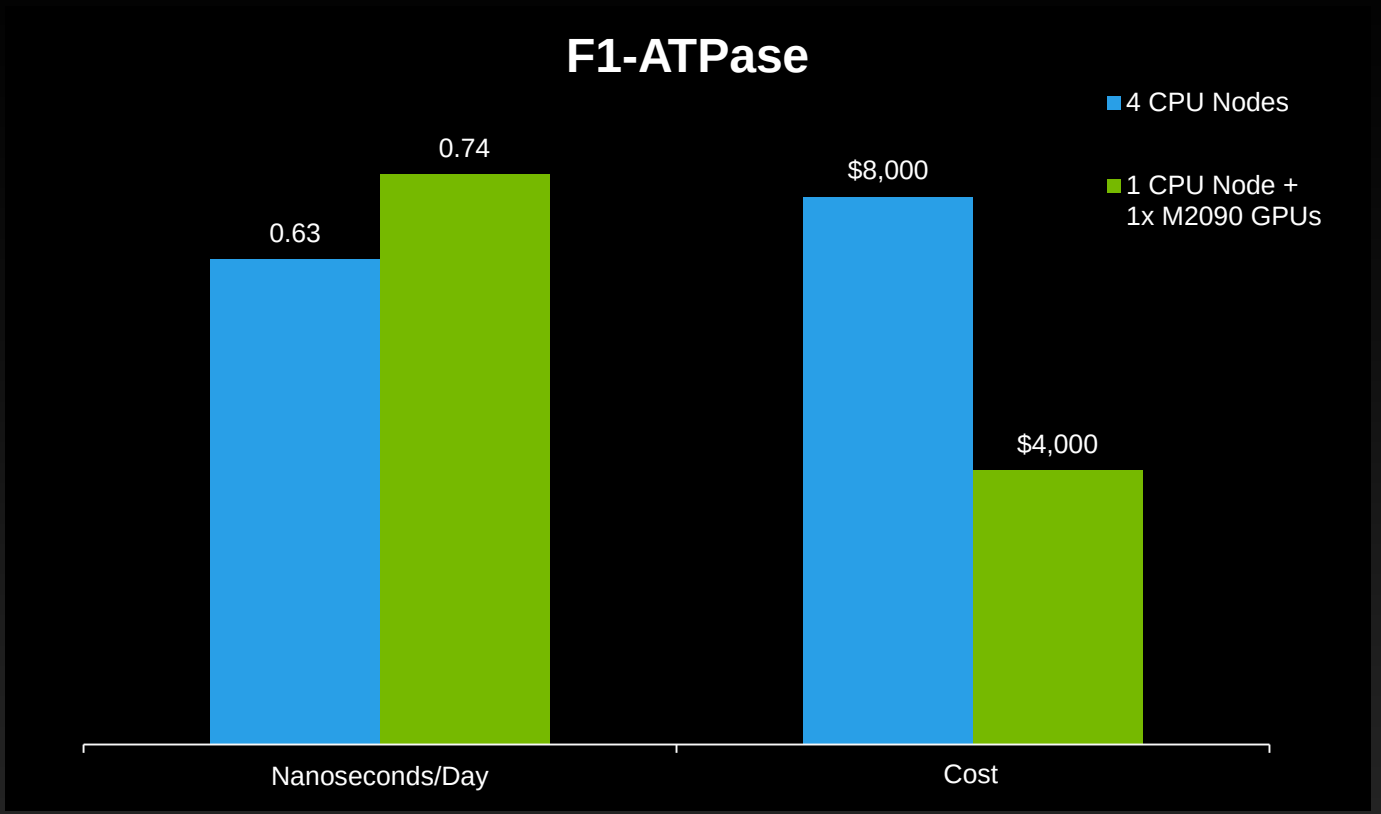
Each green XK6 CPU+GPU node contains 1x AMD 1600 Opteron (16 Cores per CPU) and an additional 1x NVIDIA X2090 GPU.



Concatenation of 100
Satellite Tobacco Mosaic Virus

Accelerate your science by 2.7-3.8x when compared to CPU-based supercomputers

Replace 3 Nodes with 1 2090 GPU



Running NAMD version 2.9
Each blue node contains 2x Intel Xeon X5550 CPUs (4 Cores per CPU).

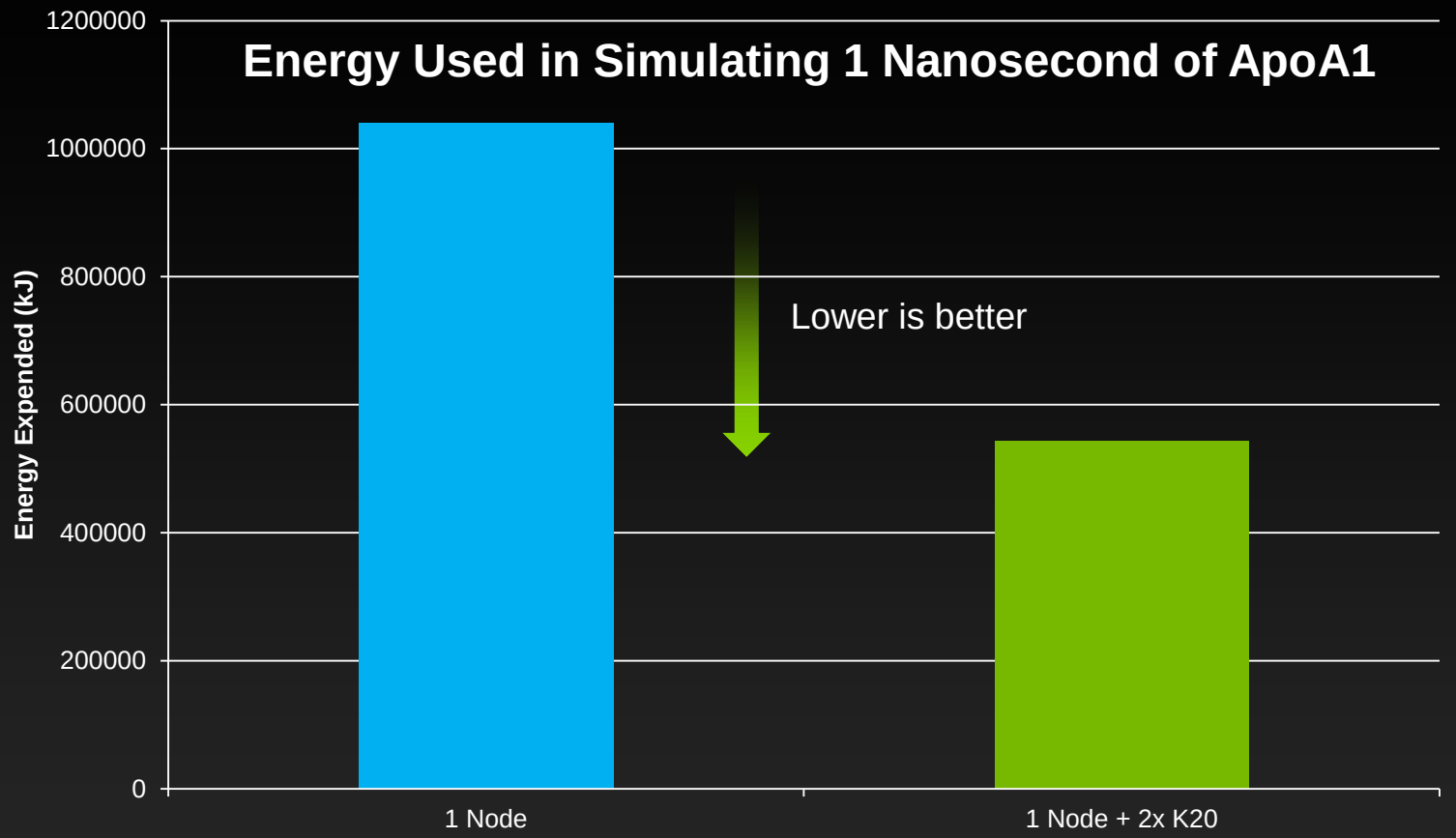
The green node contains 2x Intel Xeon X5550 CPUs (4 Cores per CPU) and 1x NVIDIA M2090 GPU

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

F1-ATPase

Speedup of 1.2x for 50% the cost

K20 - Greener: Twice The Science Per Watt



Cut down energy usage by ½ with GPUs

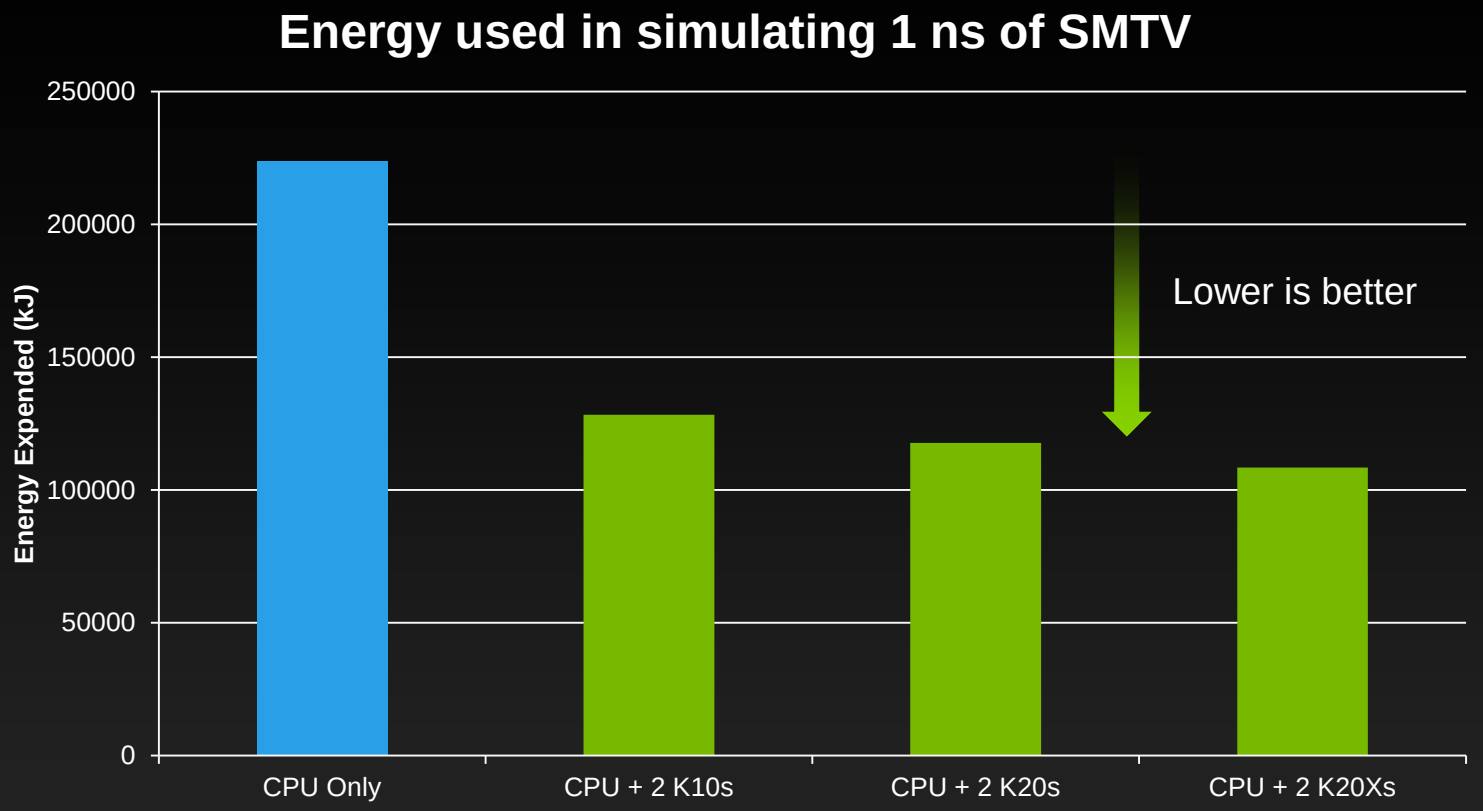
Running NAMD version 2.9
Each blue node contains Dual E5-2687W CPUs (95W, 4 Cores per CPU).

Each green node contains 2x Intel Xeon X5550 CPUs (95W, 4 Cores per CPU) and 2x NVIDIA K20 GPUs (225W per GPU)

Energy Expended
= Power x Time

NAMD Benchmark Report, Revision 2.0, dated Nov. 5, 2012

Kepler - Greener: Twice The Science/Joule



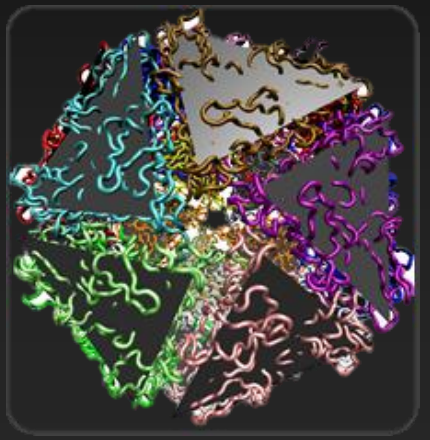
Cut down energy usage by ½ with GPUs

Running NAMD version 2.9

The blue node contains Dual E5-2687W CPUs (150W each, 8 Cores per CPU).

The green nodes contain Dual E5-2687W CPUs (8 Cores per CPU) and 2x NVIDIA K10, K20, or K20X GPUs (235W each).

Energy Expended
= Power x Time



Satellite Tobacco Mosaic Virus

Recommended GPU Node Configuration for NAMD Computational Chemistry

Workstation or Single Node Configuration	
# of CPU sockets	2
Cores per CPU socket	6+
CPU speed (Ghz)	2.66+
System memory per socket (GB)	32
GPUs	Kepler K10, K20, K20X Fermi M2090, M2075, C2075
# of GPUs per CPU socket	1-2
GPU memory preference (GB)	6
GPU to CPU connection	PCIe 2.0 or higher
Server storage	500 GB or higher
Network configuration	Gemini, InfiniBand

Scale to multiple nodes with same single node configuration

NAMD Benchmark Report, Revision 2.0, dated Nov. 5, 2012

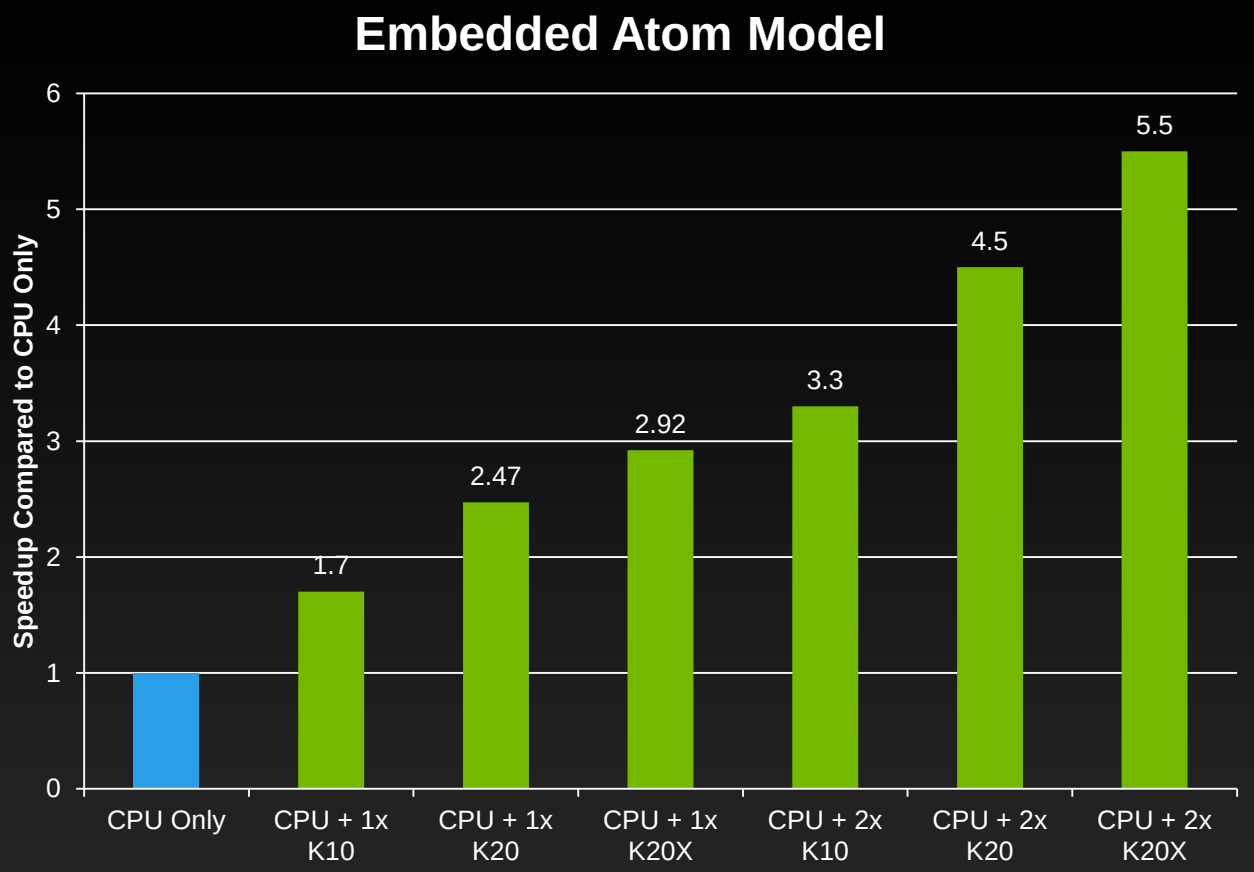
Summary/Conclusions
Benefits of GPU Accelerated Computing

- Faster than CPU only systems in all tests
- Large performance boost with small marginal price increase
- Energy usage cut in half
- GPUs scale very well within a node and over multiple nodes
- Tesla K20 GPU is our fastest and lowest power high performance GPU to date

Try GPU accelerated NAMD for free – www.nvidia.com/GPUTestDrive

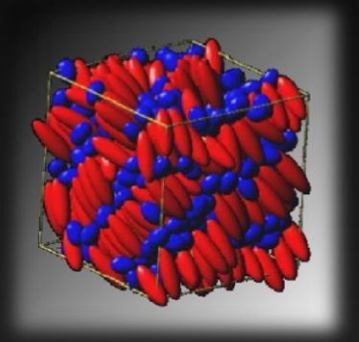
LAMMPS, Jan. 2013 or later

More Science for Your Money



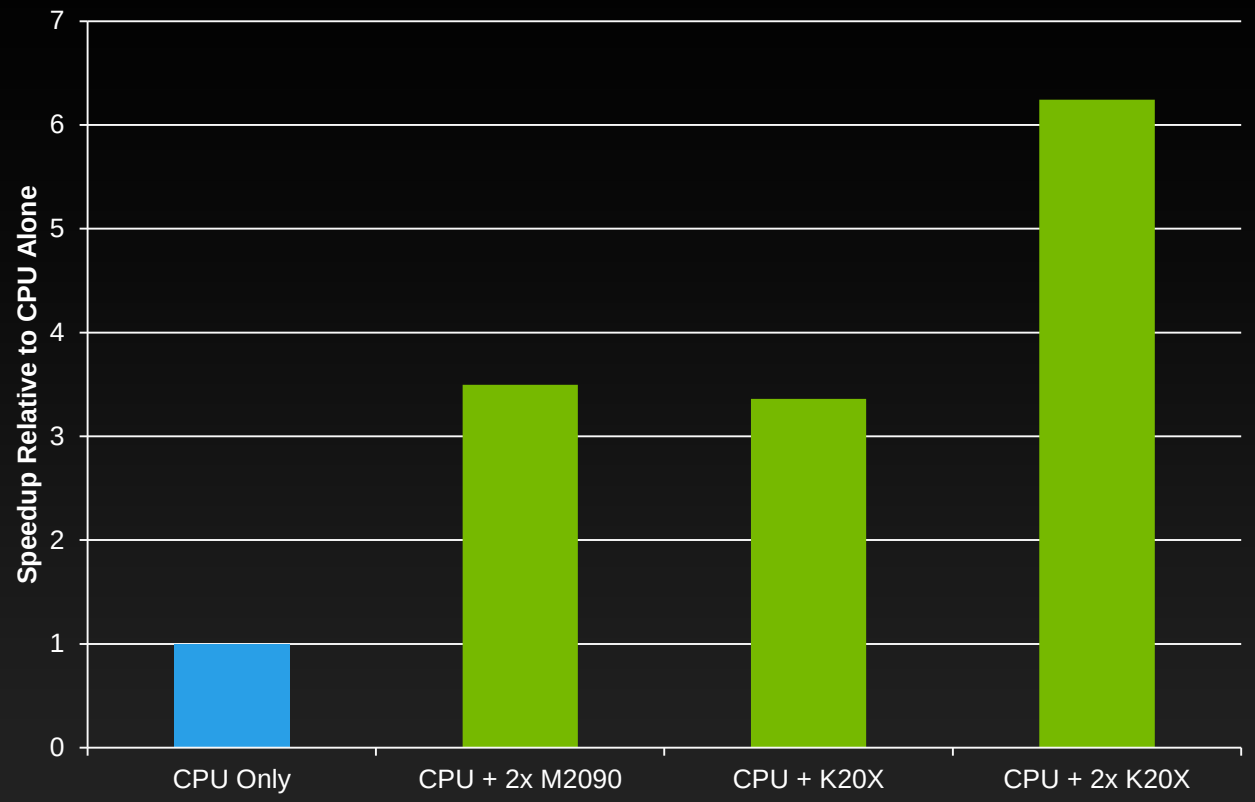
Blue node uses 2x E5-2687W (8 Cores and 150W per CPU).

Green nodes have 2x E5-2687W and 1 or 2 NVIDIA K10, K20, or K20X GPUs (235W).



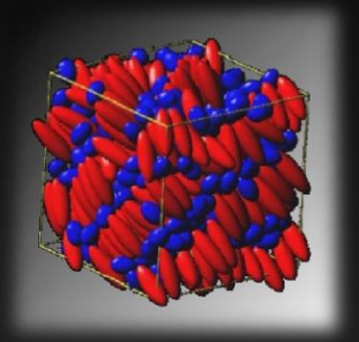
Experience performance increases of up to 5.5x with Kepler GPU nodes.

K20X, the Fastest GPU Yet



Blue node uses 2x E5-2687W (8 Cores and 150W per CPU).

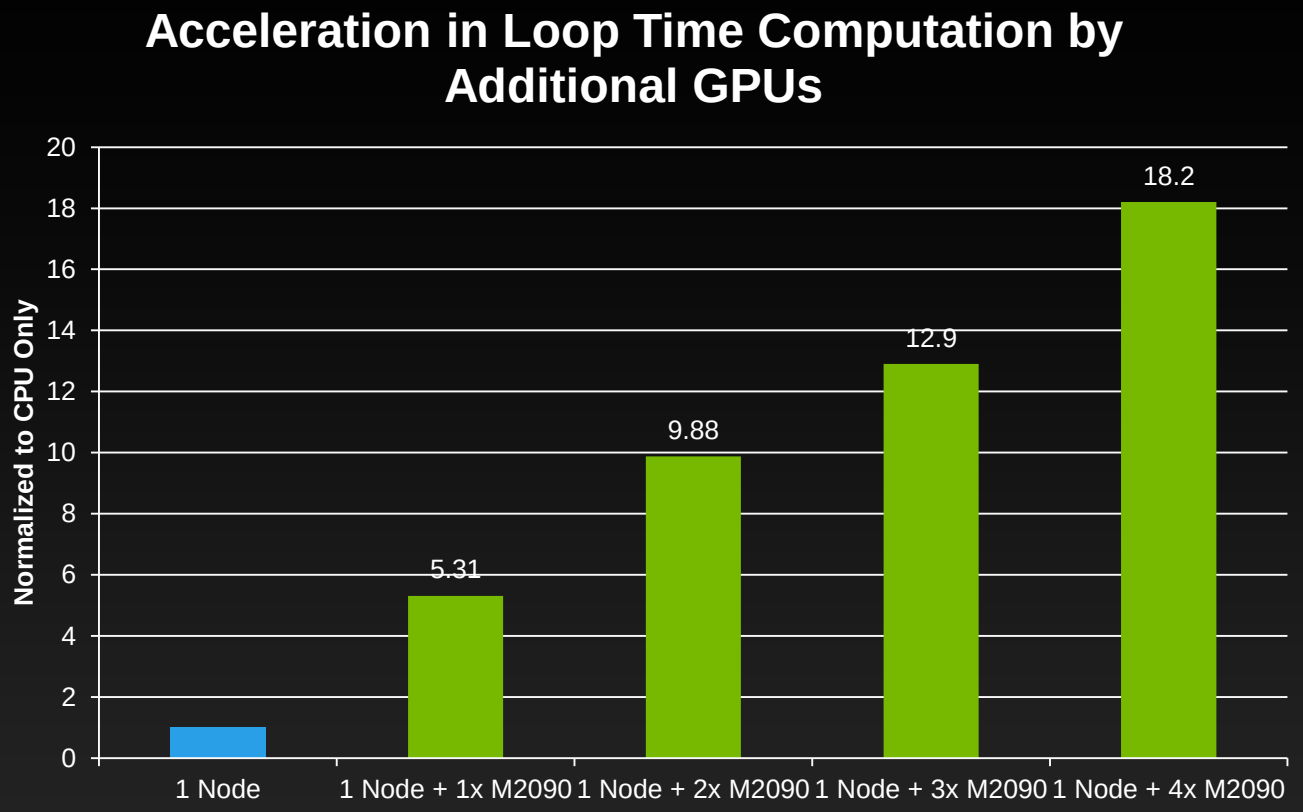
Green nodes have 2x E5-2687W and 2 NVIDIA M2090s or K20X GPUs (235W).



Experience performance increases of up to 6.2x with Kepler GPU nodes.

One K20X performs as well as two M2090s

Get a CPU Rebate to Fund Part of Your GPU Budget



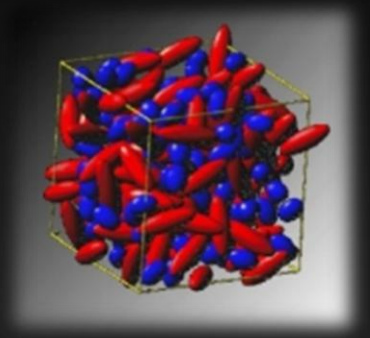
Increase performance 18x when compared to CPU-only nodes

Cheaper CPUs used with GPUs AND still faster overall performance when compared to more expensive CPUs!

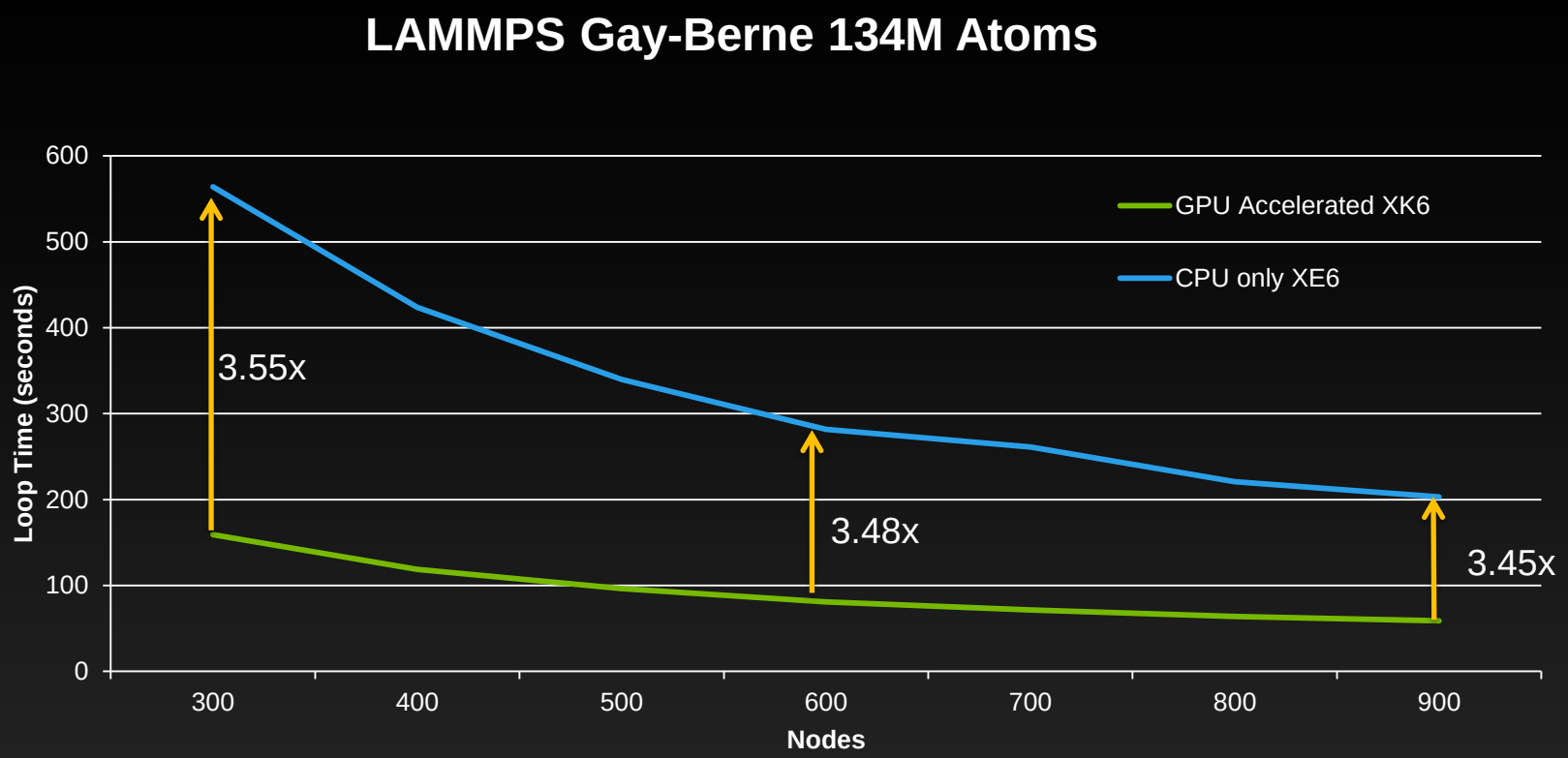
Running NAMD version 2.9

The blue node contains Dual X5670 CPUs (6 Cores per CPU).

The green nodes contain Dual X5570 CPUs (4 Cores per CPU) and 1-4 NVIDIA M2090 GPUs.



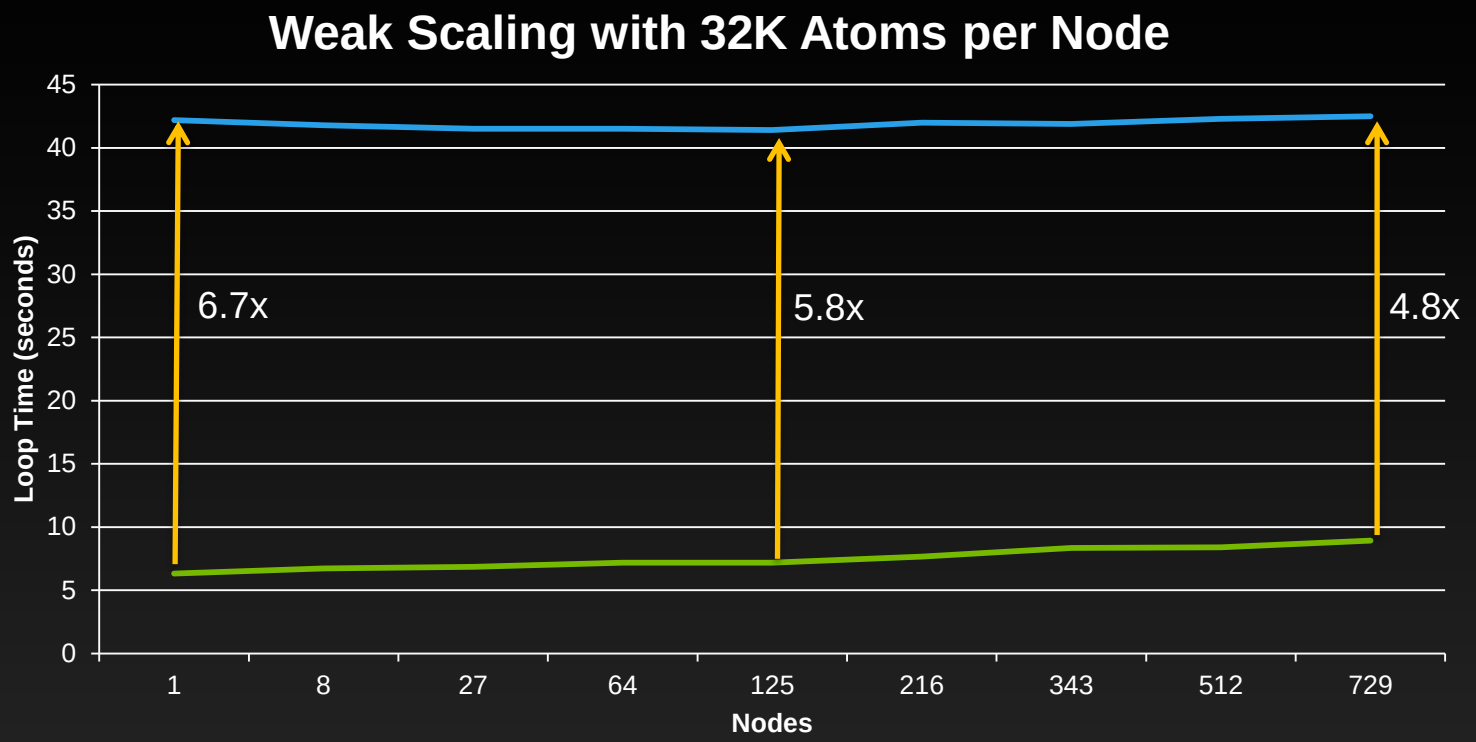
Excellent Strong Scaling on Large Clusters



From 300-900 nodes, the NVIDIA GPU-powered XK6 maintained 3.5x performance compared to XE6 CPU nodes

Each blue Cray XE6 Nodes have 2x AMD Opteron CPUs (16 Cores per CPU)
Each green Cray XK6 Node has 1x AMD Opteron 1600 CPU (16 Cores per CPU) and 1x NVIDIA X2090

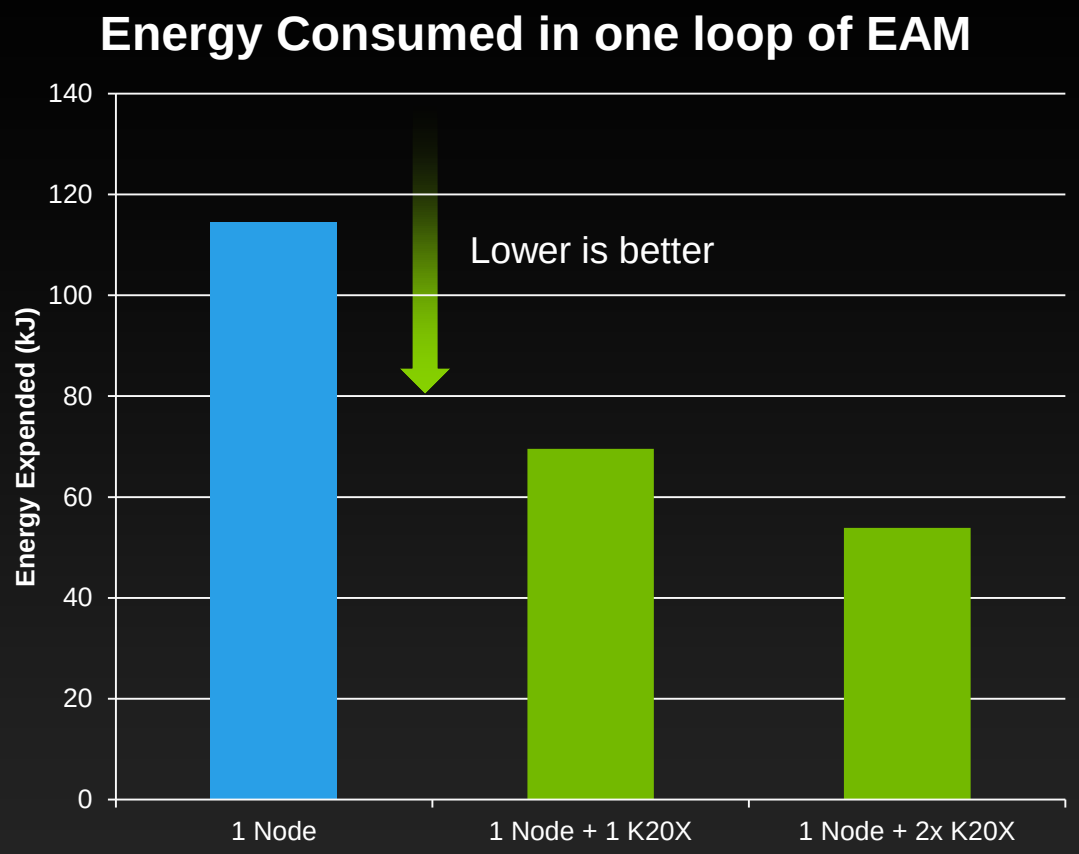
GPUs Sustain 5x Performance for Weak Scaling



Performance of 4.8x-6.7x with GPU-accelerated nodes when compared to CPUs alone

Each blue Cray XE6 Node have 2x AMD Opteron CPUs (16 Cores per CPU)
Each green Cray XK6 Node has 1x AMD Opteron 1600 CPU (16 Core per CPU) and 1x NVIDIA X2090

Faster, Greener – Worth It!



GPU-accelerated computing uses
53% less energy than CPU only

Energy Expended = Power x Time

Power calculated by combining the component's TDPs

Blue node uses 2x E5-2687W (8 Cores and 150W per CPU) and CUDA 4.2.9.
Green nodes have 2x E5-2687W and 1 or 2 NVIDIA K20X GPUs (235W) running CUDA 5.0.36.

Try GPU accelerated LAMMPS for free – www.nvidia.com/GPUTestDrive

Molecular Dynamics with LAMMPS on a ~~Hybrid Cray~~ Supercomputer

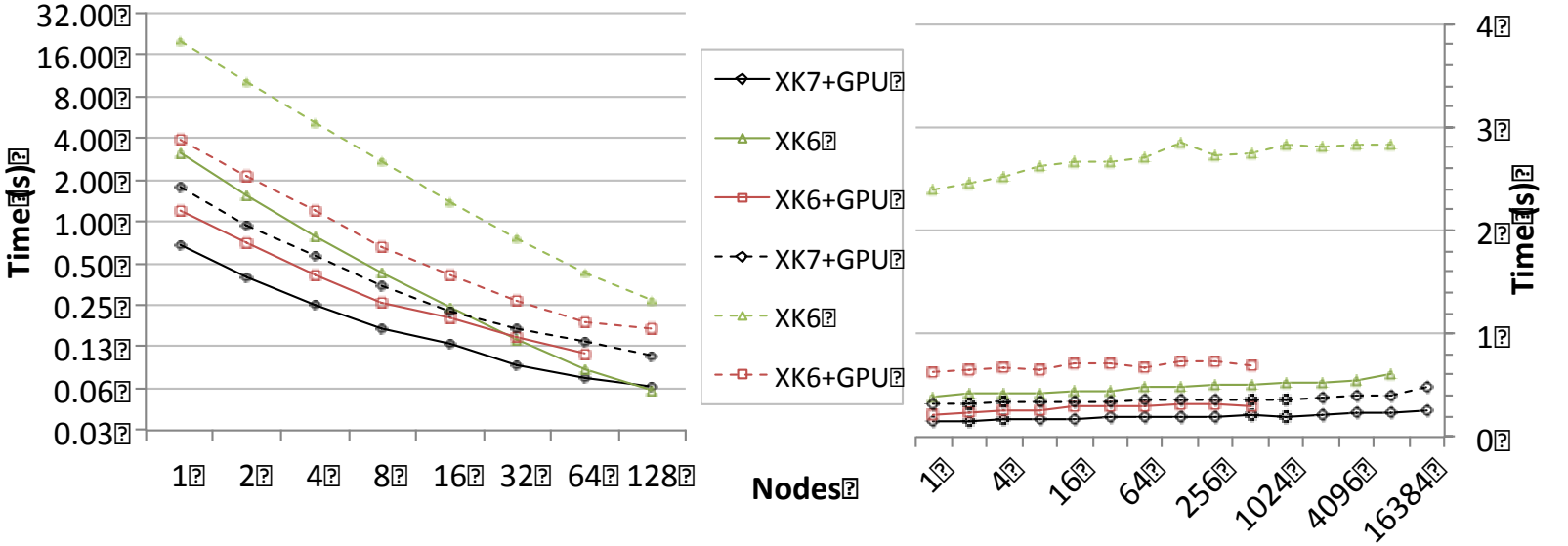
the World's Most Powerful

W. Michael Brown
National Center for Computational Sciences
Oak Ridge National Laboratory

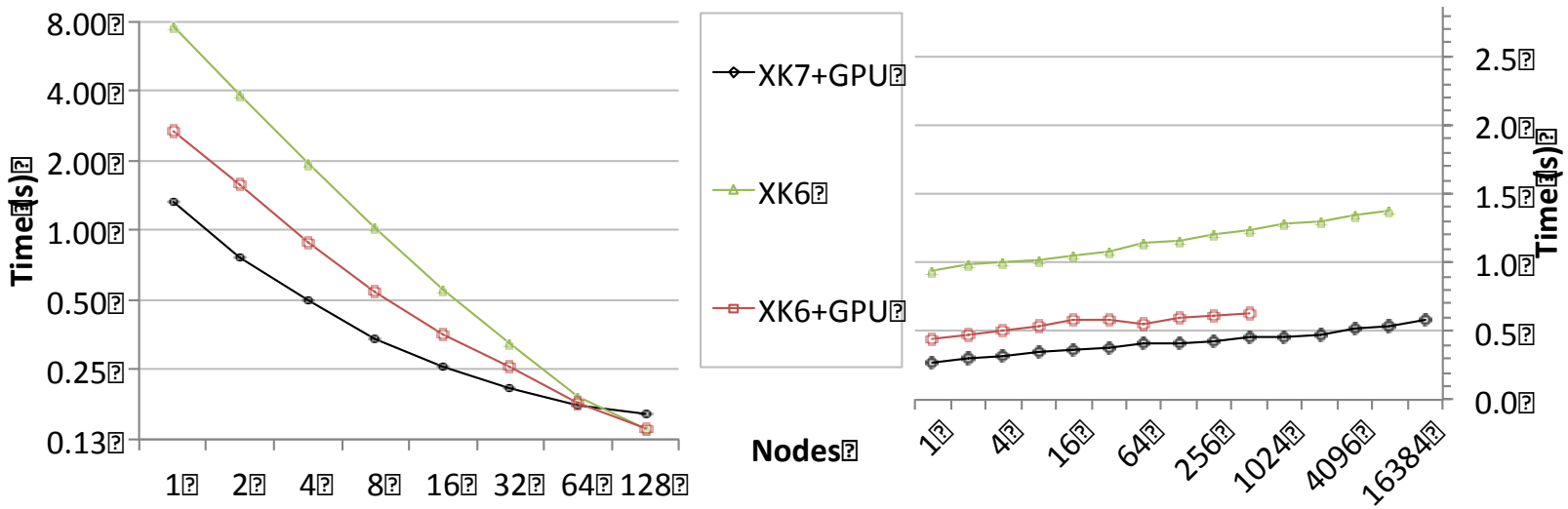
NVIDIA Technology Theater, Supercomputing 2012
November 14, 2012

Early Kepler Benchmarks on Titan

Atomic Fluid

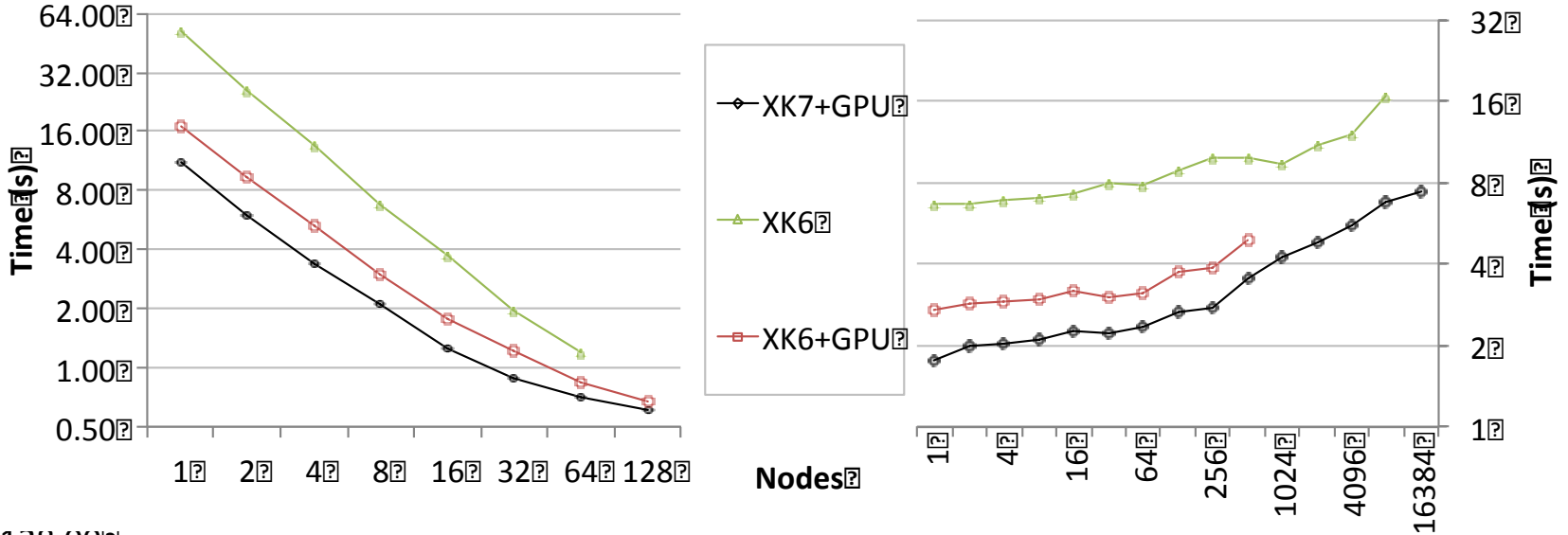


Bulk Copper

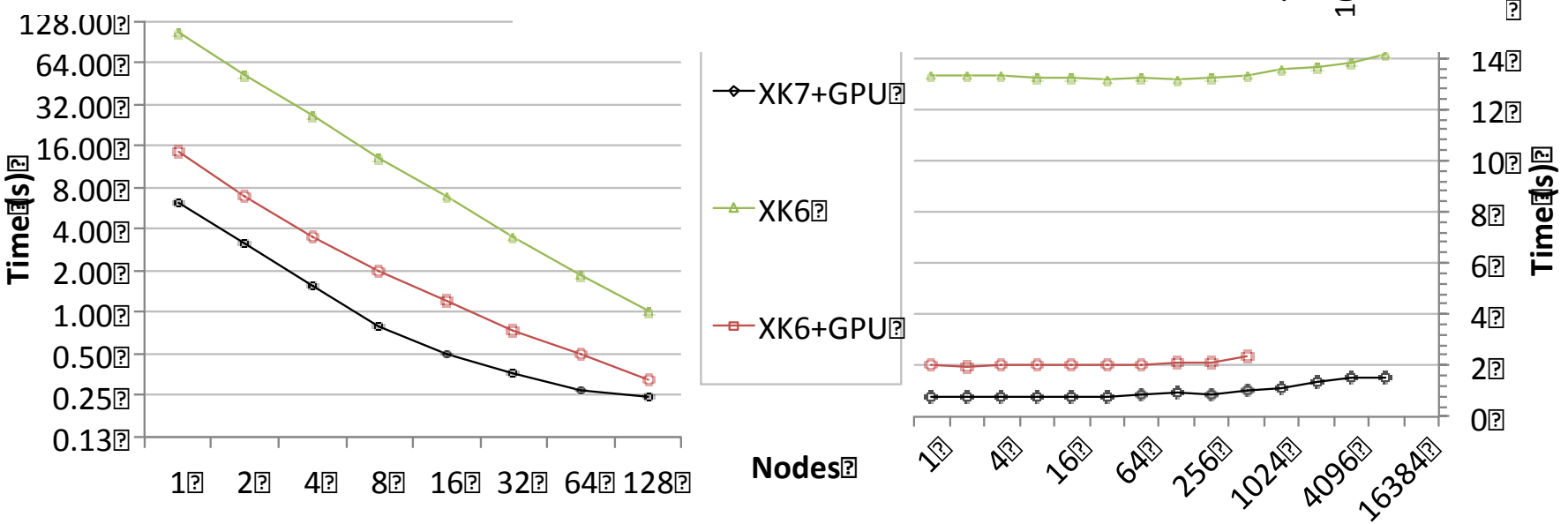


Early Kepler Benchmarks on Titan

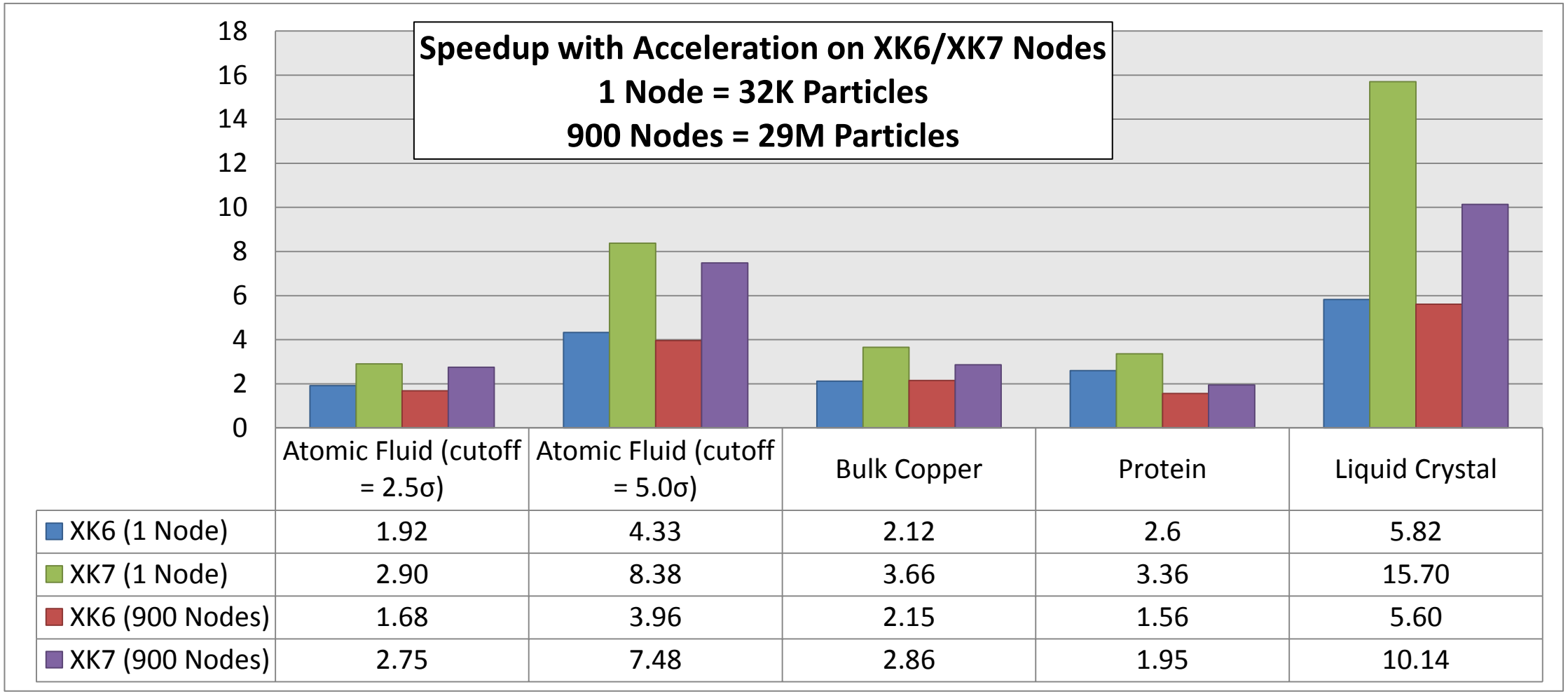
Protein



Liquid Crystal



Early Titan XK6/XK7 Benchmarks



	Atomic Fluid (cutoff = 2.5σ)	Atomic Fluid (cutoff = 5.0σ)	Bulk Copper	Protein	Liquid Crystal
XK6 (1 Node)	1.92	4.33	2.12	2.6	5.82
XK7 (1 Node)	2.90	8.38	3.66	3.36	15.70
XK6 (900 Nodes)	1.68	3.96	2.15	1.56	5.60
XK7 (900 Nodes)	2.75	7.48	2.86	1.95	10.14

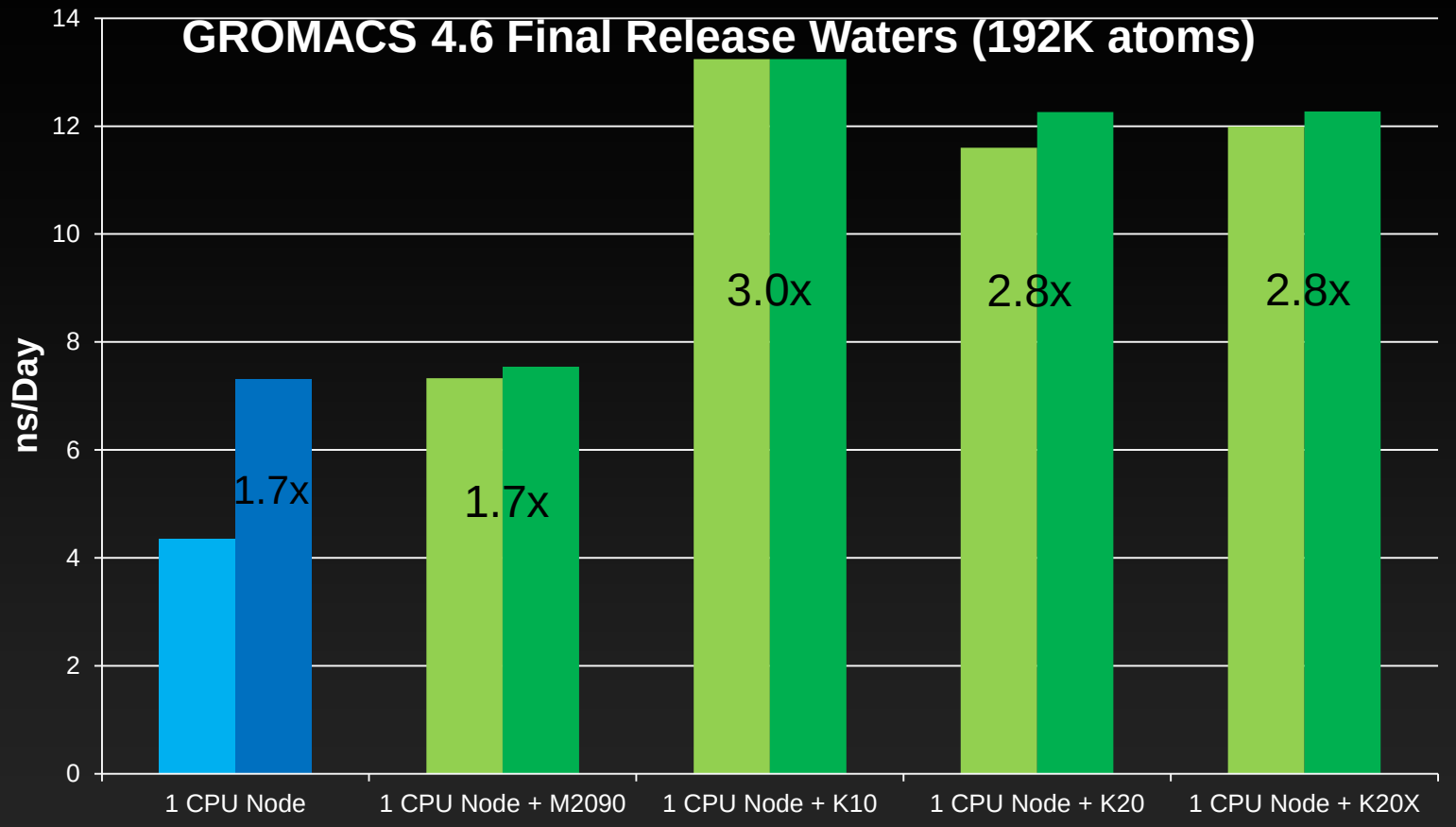
Recommended GPU Node Configuration for LAMMPS Computational Chemistry

Workstation or Single Node Configuration	
# of CPU sockets	2
Cores per CPU socket	6+
CPU speed (Ghz)	2.66+
System memory per socket (GB)	32
GPUs	Kepler K10, K20, K20X Fermi M2090, M2075, C2075
# of GPUs per CPU socket	1-2
GPU memory preference (GB)	6
GPU to CPU connection	PCIe 2.0 or higher
Server storage	500 GB or higher
Network configuration	Gemini, InfiniBand

Scale to multiple nodes with same single node configuration

GROMACS 4.6 Final, Pre-Beta and 4.6 Beta

Kepler - Our Fastest Family of GPUs Yet



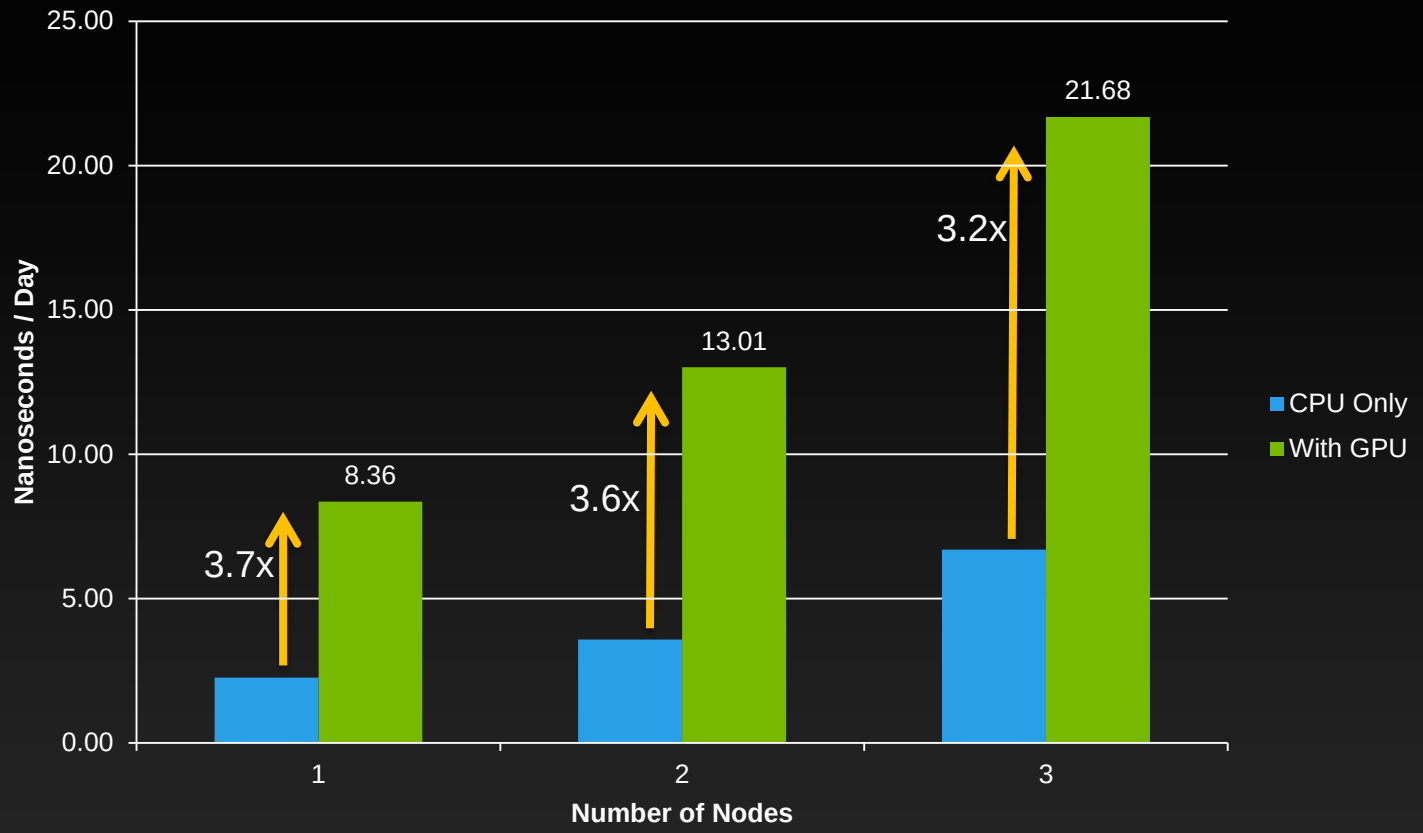
Running GROMACS 4.6 Final Release

The blue nodes contains either single or dual E5-2687W CPUs (8 Cores per CPU).

The green nodes contain either single or dual E5-2687W CPUs (8 Cores per CPU) and either 1x NVIDIA M2090, 1x K10 or 1x K20 for the GPU

Single Sandybridge CPU per node with single K10, K20 or K20X produces best performance

Great Scaling in Small Systems



Running GROMACS 4.6 pre-beta with CUDA 4.1

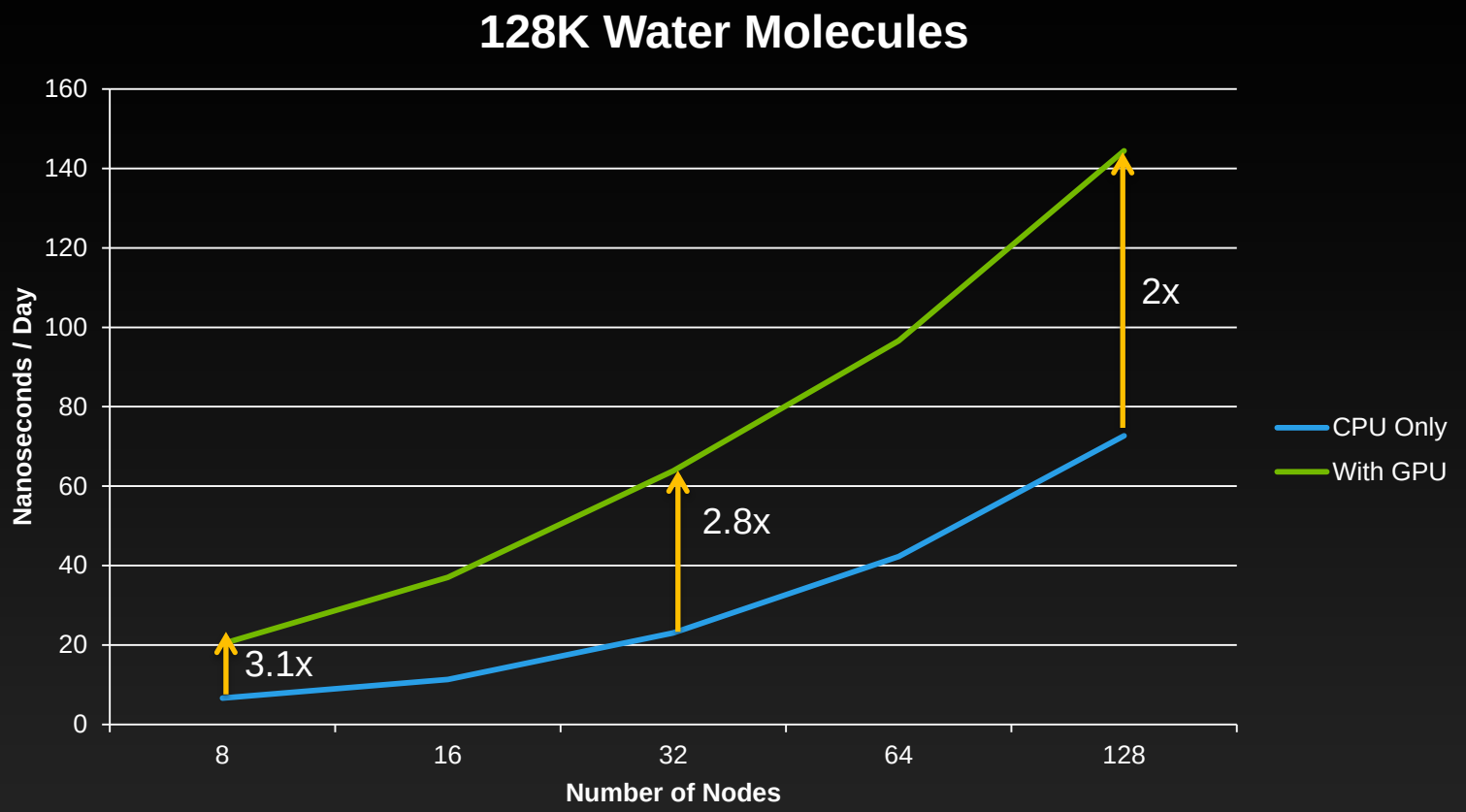
Each blue node contains 1x Intel X5550 CPU (95W TDP, 4 Cores per CPU)

Each green node contains 1x Intel X5550 CPU (95W TDP, 4 Cores per CPU) and 1x NVIDIA M2090 (225W TDP per GPU)

Benchmark systems: RNAse in water with 16,816 atoms in truncated dodecahedron box

Get up to 3.7x performance compared to CPU-only nodes

Additional Strong Scaling on Larger System



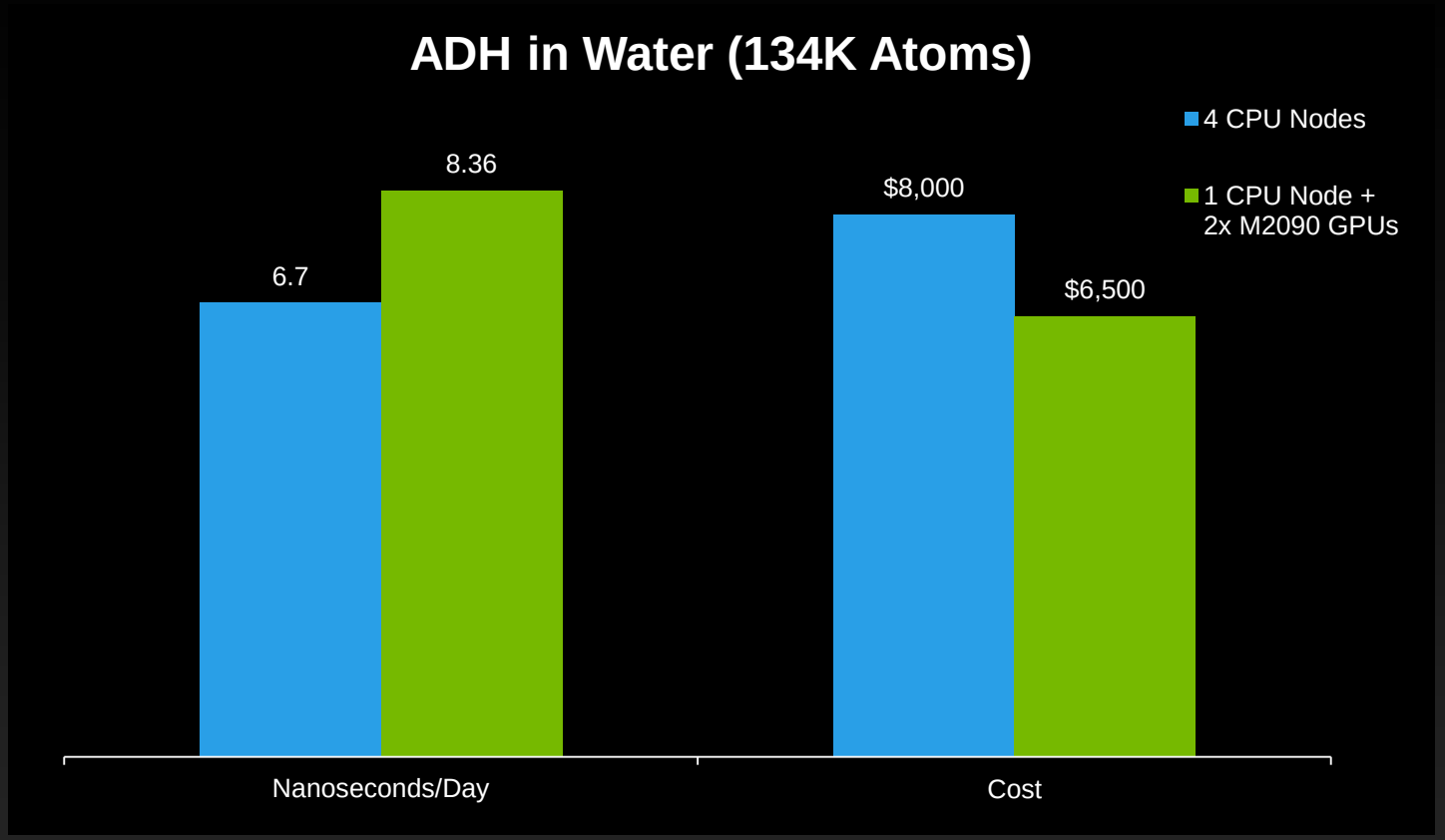
Running GROMACS 4.6 pre-beta with CUDA 4.1

Each blue node contains 1x Intel X5670 (95W TDP, 6 Cores per CPU)

Each green node contains 1x Intel X5670 (95W TDP, 6 Cores per CPU) and 1x NVIDIA M2070 (225W TDP per GPU)

Up to 128 nodes, NVIDIA GPU-accelerated nodes deliver 2-3x performance when compared to CPU-only nodes

Replace 3 Nodes with 2 GPUs



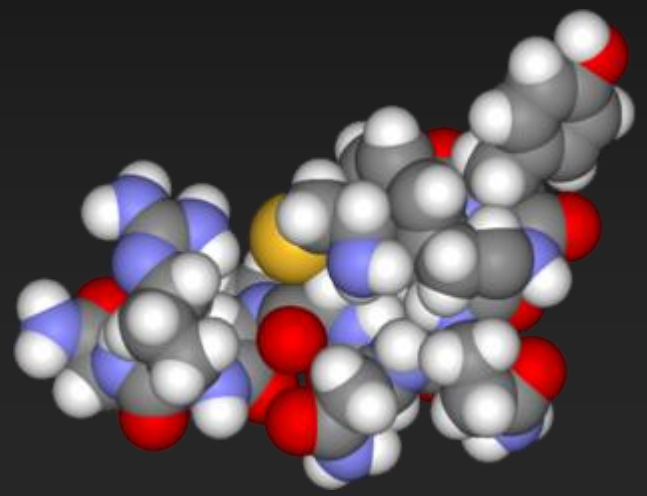
Running GROMACS 4.6 pre-beta with CUDA 4.1

The blue node contains 2x Intel X5550 CPUs (95W TDP, 4 Cores per CPU)

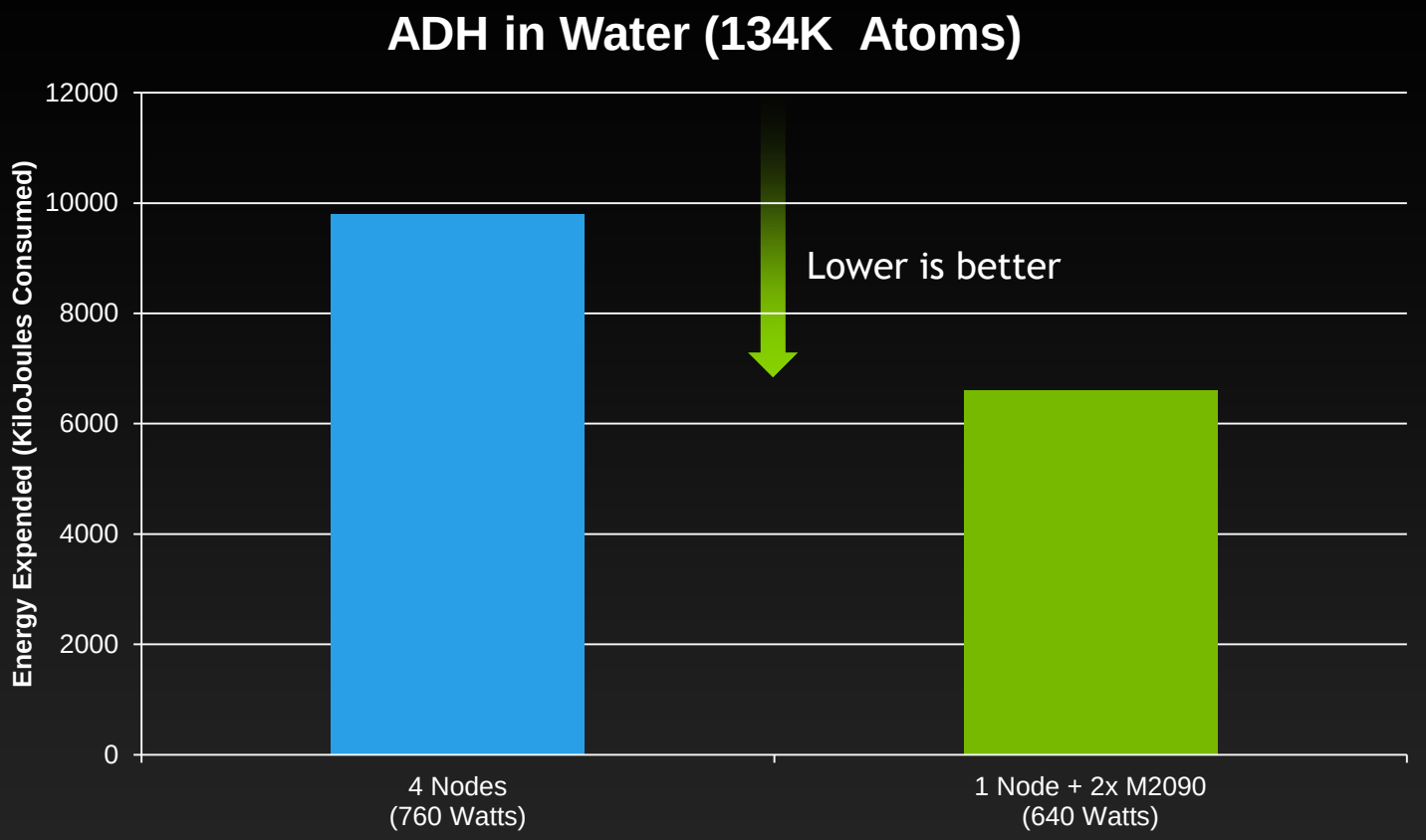
The green node contains 2x Intel X5550 CPUs (95W TDP, 4 Cores per CPU) and 2x NVIDIA M2090s as the GPU (225W TDP per GPU)

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

Save thousands of dollars and perform 25% faster



Greener Science



Running GROMACS 4.6 with CUDA 4.1

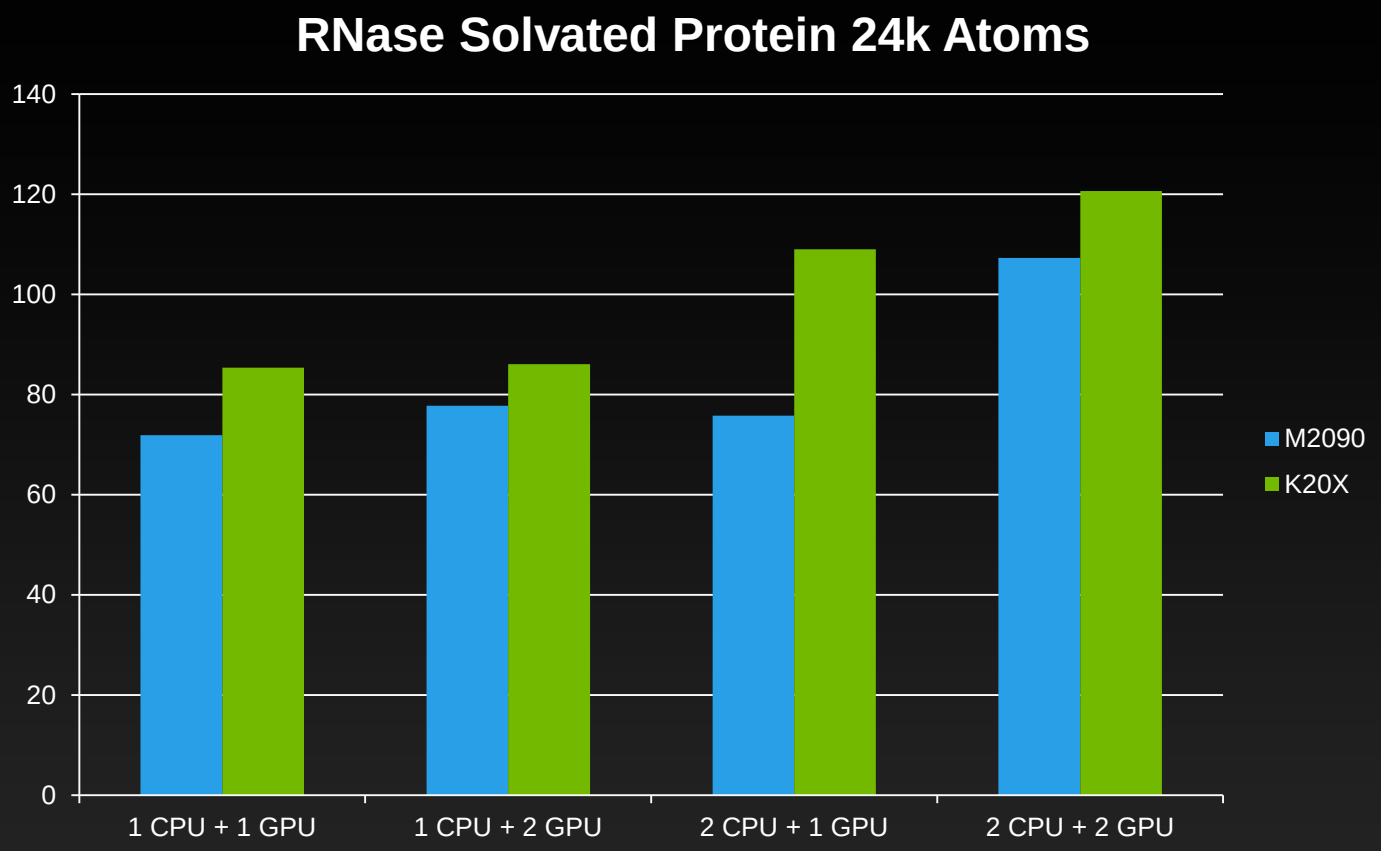
The blue nodes contain 2x Intel X5550 CPUs (95W TDP, 4 Cores per CPU)

The green node contains 2x Intel X5550 CPUs, 4 Cores per CPU) and 2x NVIDIA M2090s GPUs (225W TDP per GPU)

Energy Expended
= Power x Time

In simulating each nanosecond, the GPU-accelerated system uses 33% less energy

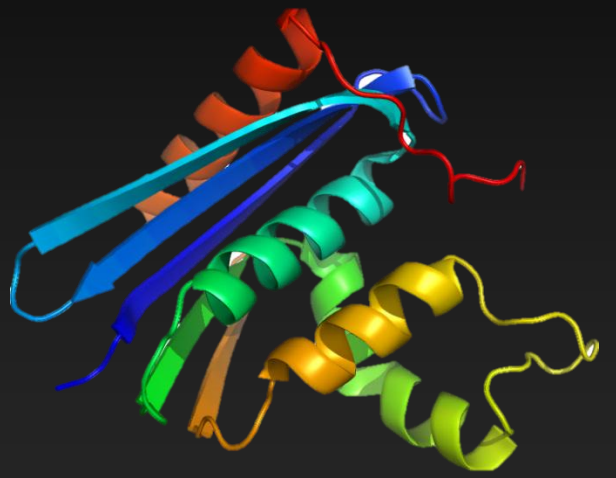
The Power of Kepler



Running GROMACS version 4.6 beta

The grey nodes contain 1 or 2 E5-2687W CPUs (150W each, 8 Cores per CPU) and 1 or 2 NVIDIA M2090s.

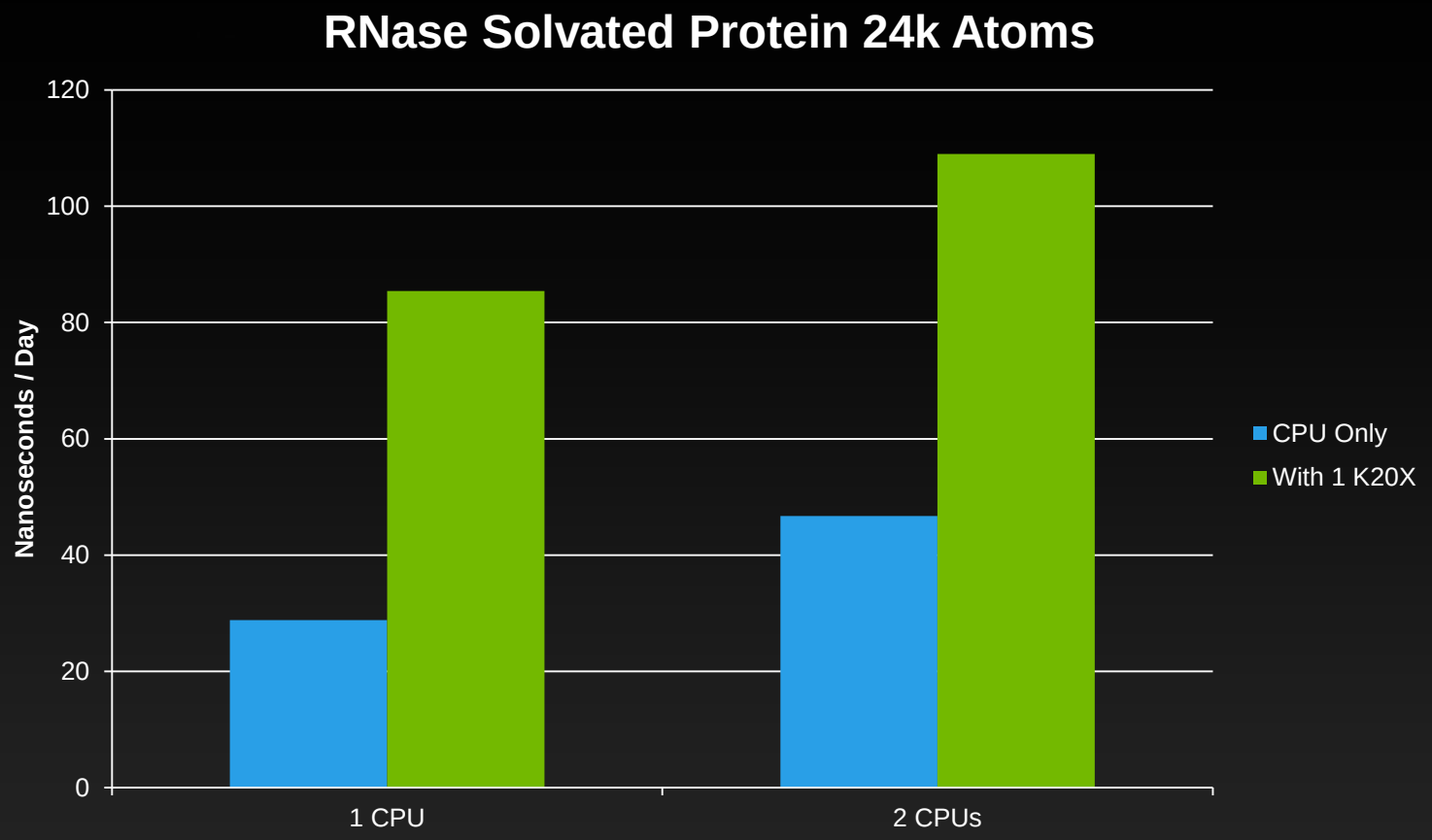
The green nodes contain 1 or 2 E5-2687W CPUs (8 Cores per CPU) and 1 or 2 NVIDIA K20X GPUs (235W each).



Ribonuclease

Upgrading an M2090 to a K20X increases performance 10-45%

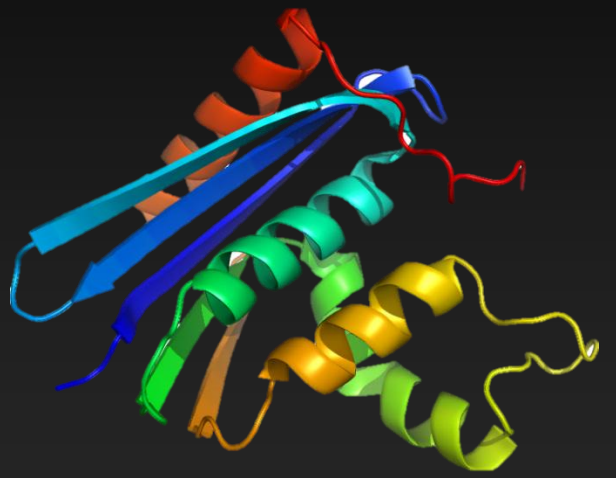
K20X - Fast



Running GROMACS version 4.6 beta

The blue nodes contain 1 or 2 E5-2687W CPUs (150W each, 8 Cores per CPU).

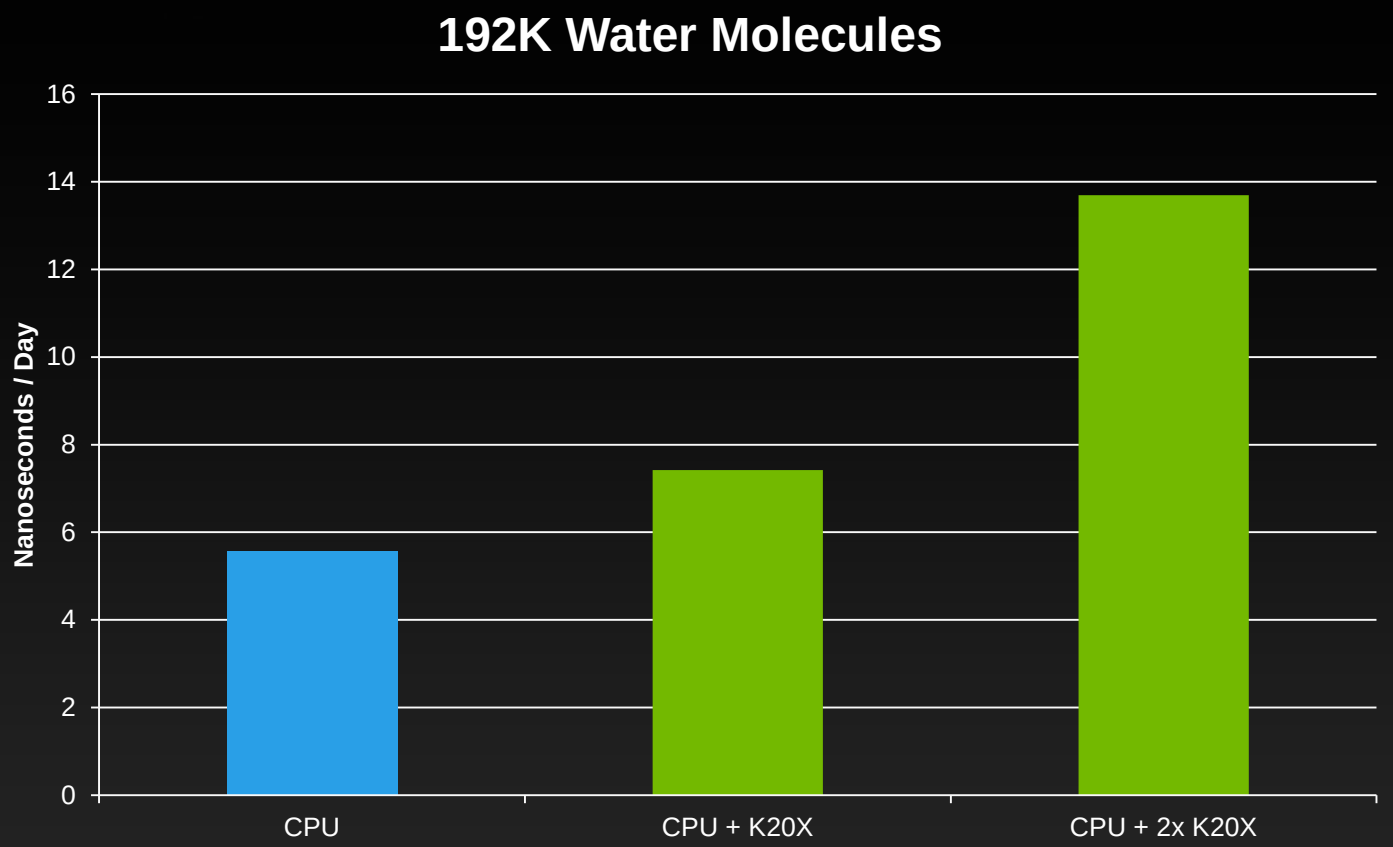
The green nodes contain 1 or 2 E5-2687W CPUs (8 Cores per CPU) and 1 or 2 NVIDIA K20X GPUs (235W each).



Ribonuclease

Adding a K20X increases performance by up to 3x

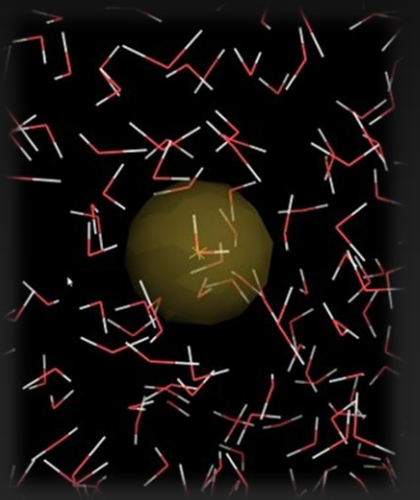
K20X, the Fastest Yet



Running GROMACS version 4.6-beta2 and CUDA 5.0.35

The blue node contains 2 E5-2687W CPUs (150W each, 8 Cores per CPU).

The green nodes contain 2 E5-2687W CPUs (8 Cores per CPU) and 1 or 2 NVIDIA K20X GPUs (235W each).



Water

Using K20X nodes increases performance by 2.5x

Try GPU accelerated GROMACS 4.6 for free – www.nvidia.com/GPUTestDrive

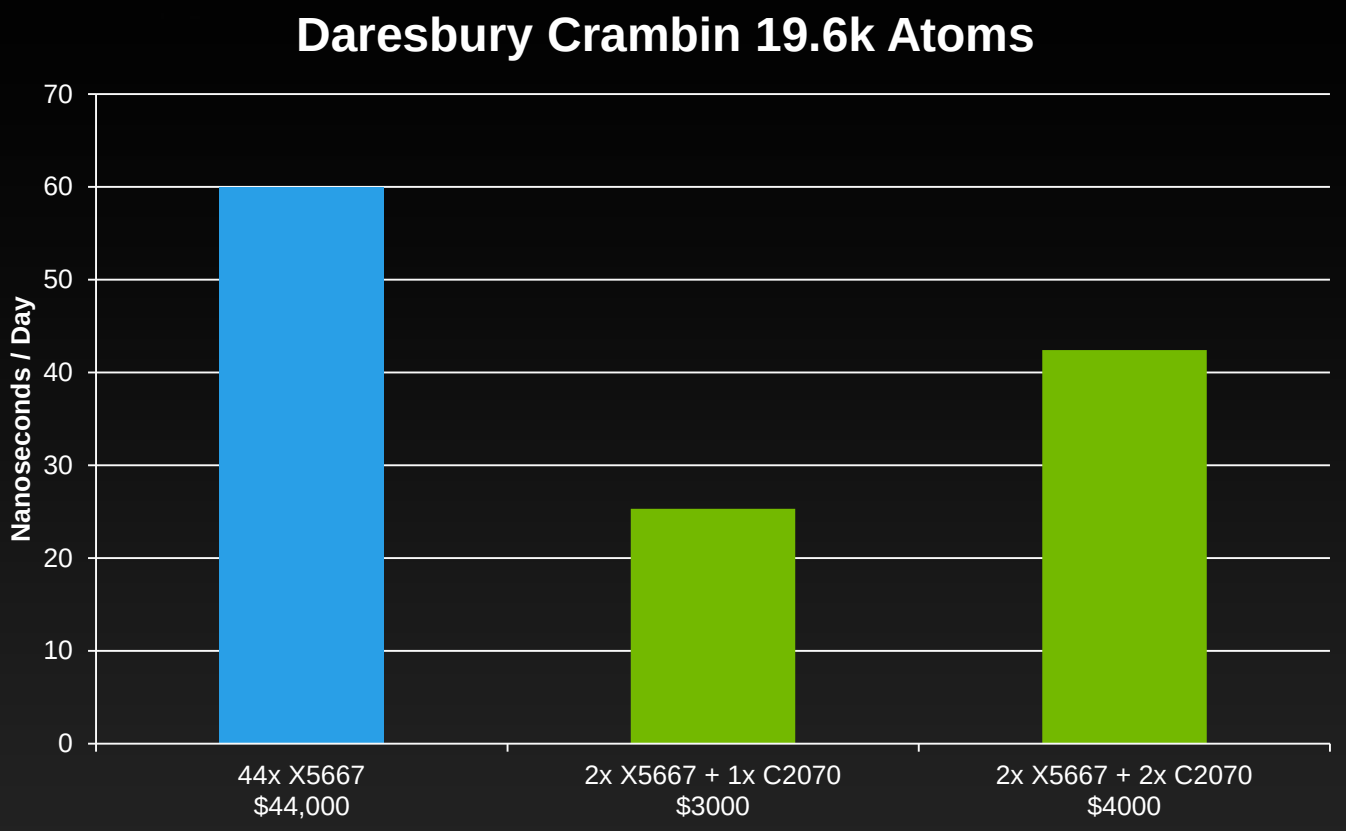
Recommended GPU Node Configuration for GROMACS Computational Chemistry

Workstation or Single Node Configuration	
# of CPU sockets	2
Cores per CPU socket	6+
CPU speed (Ghz)	2.66+
System memory per socket (GB)	32
GPUs	Kepler K10, K20, K20X Fermi M2090, M2075, C2075
# of GPUs per CPU socket	1x Kepler-based GPUs (K20X, K20 or K10): need fast Sandy Bridge or perhaps the very fastest Westmeres, or high-end AMD Opterons
GPU memory preference (GB)	6
GPU to CPU connection	PCIe 2.0 or higher
Server storage	500 GB or higher
Network configuration	Gemini, InfiniBand

Scale to multiple nodes with same single node configuration

CHARMM Release C37b1

GPUs Outperform CPUs



1 GPU = 15 CPUs

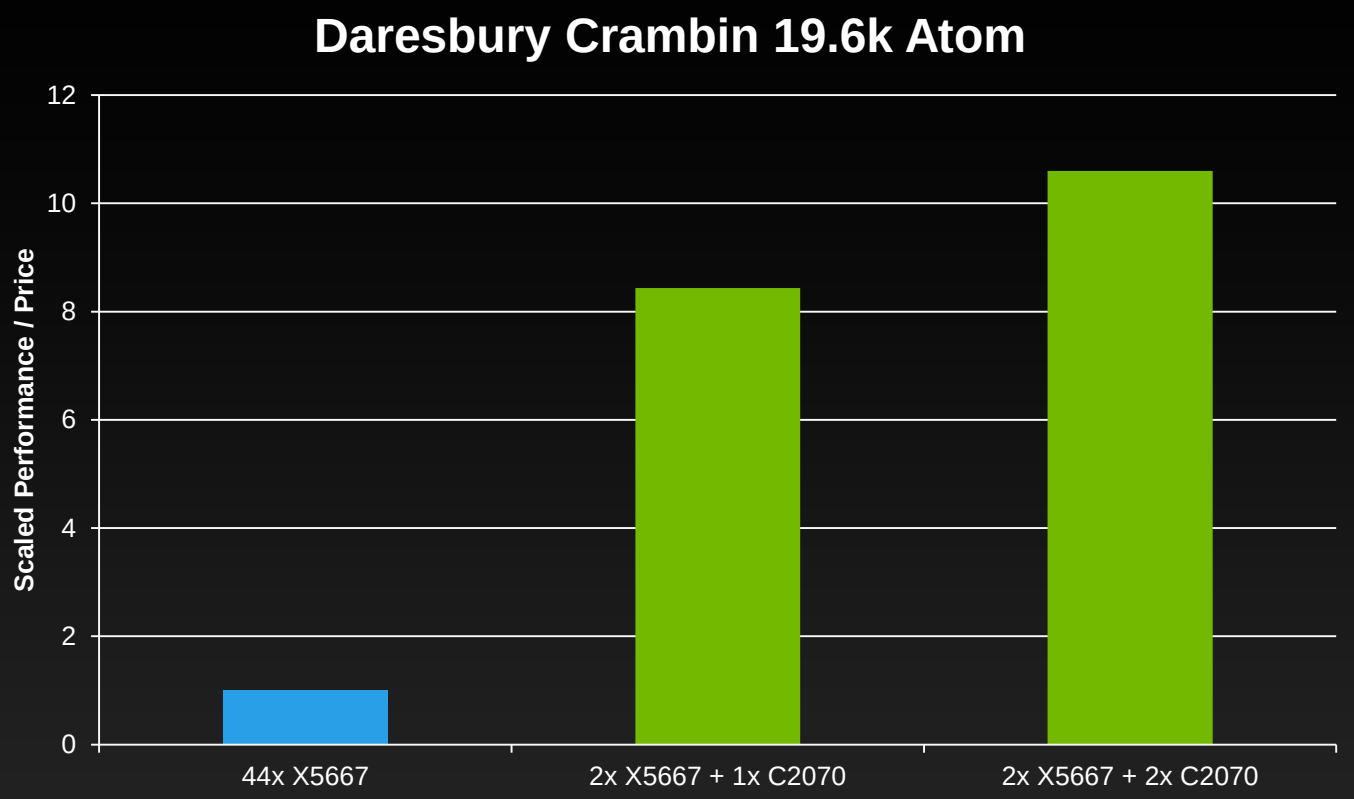
Running CHARMM release C37b1

The blue nodes contains 44 X5667 CPUs (95W, 4 Cores per CPU).

The green nodes contain 2 X5667 CPUs and 1 or 2 NVIDIA C2070 GPUs (238W each).

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

More Bang for your Buck



Running CHARMM release C37b1

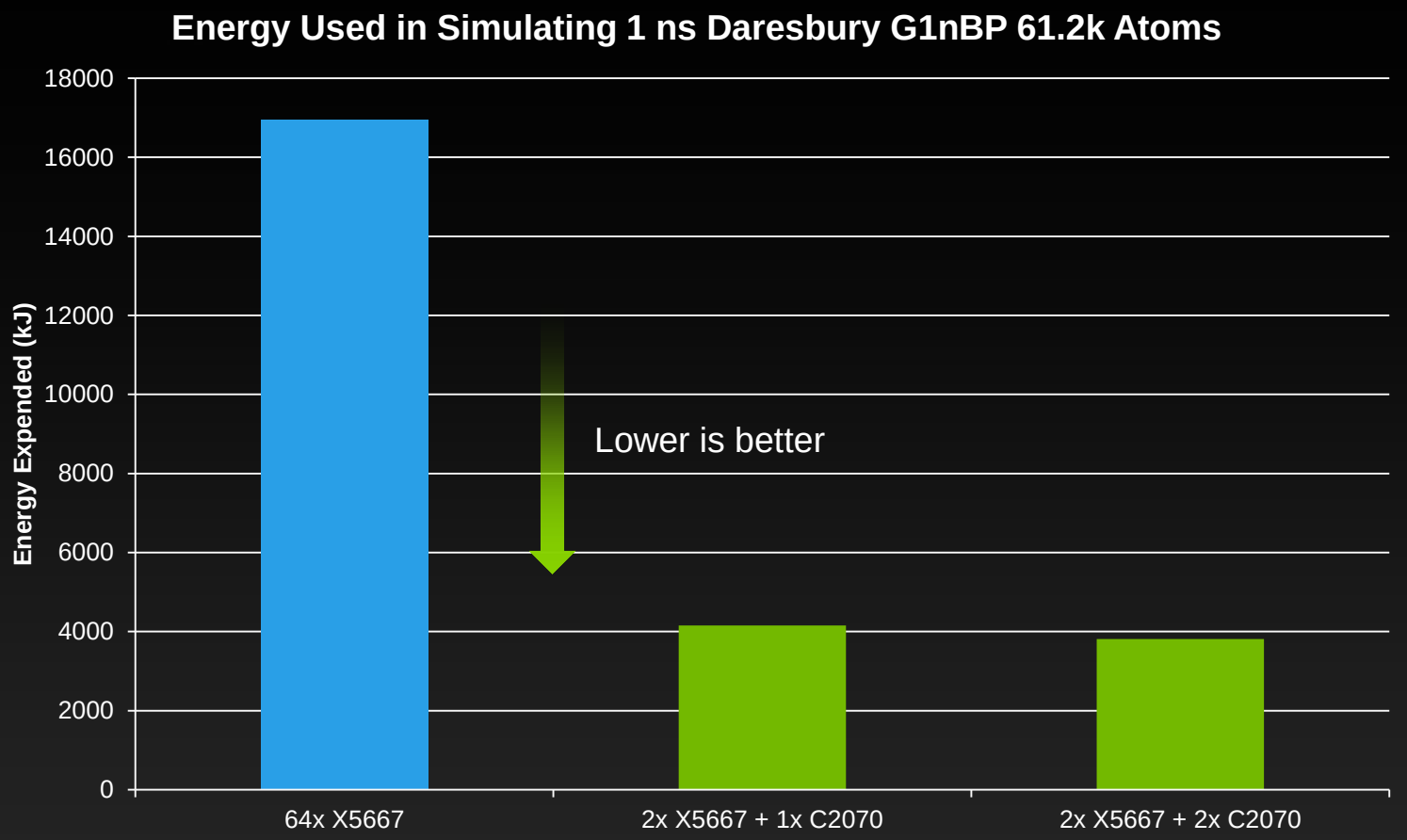
The blue nodes contains 44 X5667 CPUs (95W, 4 Cores per CPU).

The green nodes contain 2 X5667 CPUs and 1 or 2 NVIDIA C2070 GPUs (238W).

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

Using GPUs delivers 10.6x the performance for the same cost

Greener Science with NVIDIA



Using GPUs will decrease energy use by 75%

Running CHARMM release C37b1

The blue nodes contains 64 X5667 CPUs (95W, 4 Cores per CPU).

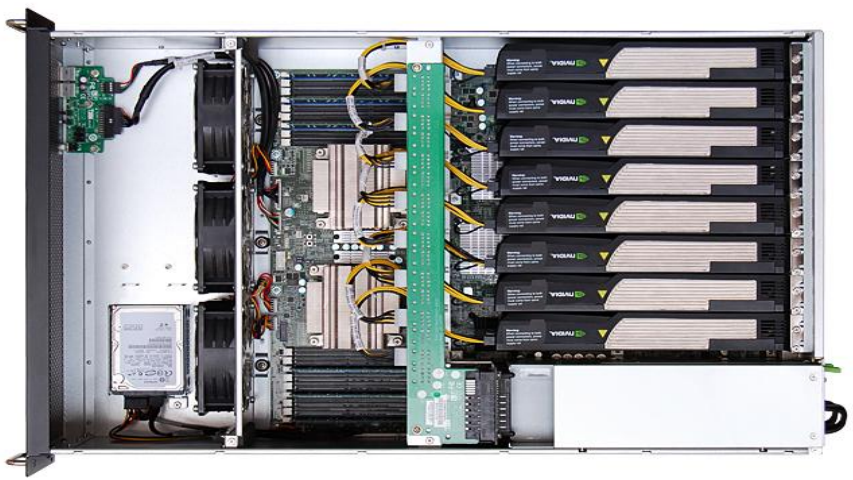
The green nodes contain 2 X5667 CPUs and 1 or 2 NVIDIA C2070 GPUs (238W each).

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

Energy Expended
= Power x Time

www.acellera.com

ACEMD: Extremely efficient and robust MD software built on GPUs



470 ns/day on 1 GPU for L-Iduronic acid (1362 atoms)

116 ns/day on 1 GPU for DHFR (23K atoms)

M. Harvey, G. Giupponi and G. De Fabritiis, *ACEMD: Accelerated molecular dynamics simulations in the microseconds timescale*, J. Chem. Theory and Comput. 5, 1632 (2009)

www.acellera.com

ACEMD

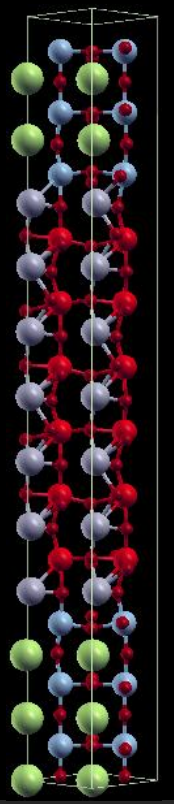
- **Standardised and easy to use**: ACEMD reads CHARMM/NAMD and AMBER input files and uses similar syntax to other MD software.
- **Fully featured:** NVT, NPT, PME, TCL, PLUMED, CAMSHIFT[1]
- **Robust**: ACEMD is a proven computational engine and is used in one of the largest distributed computing projects worldwide: GPUGRID.
- **Compatible**: ACEMD works with CUDA and OpenCL, the new standard framework for parallel and high-performance computing.
- **Validated**: ACEMD is used in reputable academic and industrial institutions. Results describing its applications have appeared in peer-reviewed journals of high impact such as PNAS, PLoS and JACS.[2]

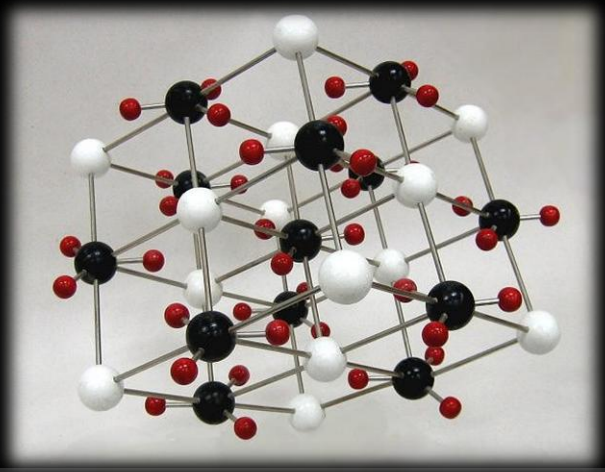
1 M. J. Harvey and G. De Fabritiis, *An implementation of the smooth particle-mesh Ewald (PME) method on GPU hardware*, J. Chem. Theory Comput., 5, 2371–2377 (2009)

2 For a list of selected references see http://www.acellera.com/acemd/publications

TESLA

Quantum Chemistry Module





Quantum Chemistry Applications

Application	Features Supported	GPU Perf	Release Status	Notes
Abinit	Local Hamiltonian, non-local Hamiltonian, LOBPCG algorithm, diagonalization / orthogonalization	1.3-2.7X	Released since Version 6.12 Multi-GPU support	www.abinit.org
ACES III	Integrating scheduling GPU into SIAL programming language and SIP runtime environment	10X on kernels	Under development Multi-GPU support	http://www.olcf.ornl.gov/wp-content/training/electronic-structure-2012/deumens_ESaccel_2012.pdf
ADF	Fock Matrix, Hessians	TBD	Pilot project completed, Under development Multi-GPU support	www.scm.com
BigDFT	DFT; Daubechies wavelets, part of Abinit	5-25X (1 CPU core to GPU kernel)	Released June 2009, current release 1.6 Multi-GPU support	http://inac.cea.fr/L_Sim/BigDFT/news.html, http://www.olcf.ornl.gov/wp-content/training/electronic-structure-2012/BigDFT-Formalism.pdf and http://www.olcf.ornl.gov/wp-content/training/electronic-structure-2012/BigDFT-HPC-tues.pdf
Casino	TBD	TBD	Under development, Spring 2013 release Multi-GPU support	http://www.tcm.phy.cam.ac.uk/~mdt26/casino.html
CP2K	DBCSR (spare matrix multiply library)	2-7X	Under development Multi-GPU support	http://www.olcf.ornl.gov/wp-content/training/ascc_2012/friday/ACSS_2012_VandeVondele_s.pdf
GAMESS-US	Libqc with Rys Quadrature Algorithm, Hartree-Fock, MP2 and CCSD in Q4 2012	1.3-1.6X, 2.3-2.9x HF	Released Multi-GPU support	Next release Q4 2012. http://www.msg.ameslab.gov/gamess/index.html

GPU Perf compared against Multi-core x86 CPU socket.
GPU Perf benchmarked on GPU supported features and may be a kernel to kernel perf comparison

Quantum Chemistry Applications

Application	Features Supported	GPU Perf	Release Status	Notes
GAMESS-UK	(ss|ss) type integrals within calculations using Hartree Fock *ab initio* methods and density functional theory. Supports organics & inorganics.	8x	Release in Summer 2012 Multi-GPU support	http://www.ncbi.nlm.nih.gov/pubmed/21541963
Gaussian	Joint PGI, NVIDIA & Gaussian Collaboration	TBD	Under development Multi-GPU support	Announced Aug. 29, 2011 http://www.gaussian.com/g_press/nvidia_press.htm
GPAW	Electrostatic poisson equation, orthonormalizing of vectors, residual minimization method (rmm-diis)	8x	Released Multi-GPU support	https://wiki.fysik.dtu.dk/gpaw/devel/projects/gpu.html, Samuli Hakala (CSC Finland) & Chris O'Grady (SLAC)
Jaguar	Investigating GPU acceleration	TBD	Under development Multi-GPU support	Schrodinger, Inc. http://www.schrodinger.com/kb/278
LSMS	Generalized Wang-Landau method	3x with 32 GPUs vs. 32 (16-core) CPUs	Under development Multi-GPU support	NICS Electronic Structure Determination Workshop 2012: http://www.olcf.ornl.gov/wp-content/training/electronic-structure-2012/Eisenbach_OakRidge_February.pdf
MOLCAS	CU_BLAS support	1.1x	Released, Version 7.8 Single GPU. Additional GPU support coming in Version 8	www.molcas.org
MOLPRO	Density-fitted MP2 (DF-MP2), density fitted local correlation methods (DF-RHF, DF-KS), DFT	1.7-2.3X projected	Under development Multiple GPU	www.molpro.net Hans-Joachim Werner

GPU Perf compared against Multi-core x86 CPU socket.
GPU Perf benchmarked on GPU supported features
and may be a kernel to kernel perf comparison

Quantum Chemistry Applications

Application	Features Supported	GPU Perf	Release Status	Notes
MOPAC2009	pseudodiagonalization, full diagonalization, and density matrix assembling	3.8-14X	Under Development Single GPU	Academic port. http://openmopac.net
NWChem	Triples part of Reg-CCSD(T), CCSD & EOMCCSD task schedulers	3-10X projected	Release targeting end of 2012 Multiple GPUs	Development GPGPU benchmarks: www.nwchem-sw.org And http://www.olcf.ornl.gov/wp-content/training/electronic-structure-2012/Krishnamoorthy-ESCMA12.pdf
Octopus	DFT and TDDFT	TBD	Released	http://www.tddft.org/programs/octopus/
PEtot	Density functional theory (DFT) plane wave pseudopotential calculations	6-10X	Released Multi-GPU	First principles materials code that computes the behavior of the electron structures of materials
Q-CHEM	RI-MP2	8x-14x	Released, Version 4.0	http://www.q-chem.com/doc_for_web/qchem_manual_4.0.pdf

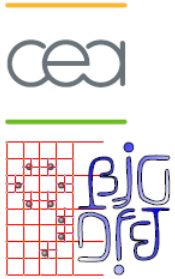
GPU Perf compared against Multi-core x86 CPU socket.
GPU Perf benchmarked on GPU supported features
and may be a kernel to kernel perf comparison

Quantum Chemistry Applications

Application	Features Supported	GPU Perf	Release Status	Notes
QMCPACK	Main features	3-4x	Released Multiple GPUs	NCSA University of Illinois at Urbana-Champaign http://cms.mcc.uiuc.edu/qmcpack/index.php/GPU_version_of_QMCPACK
Quantum Espresso/PWscf	PWscf package: linear algebra (matrix multiply), explicit computational kernels, 3D FFTs	2.5-3.5x	Released Version 5.0 Multiple GPUs	Created by Irish Centre for High-End Computing http://www.quantum-espresso.org/index.php and http://www.quantum-espresso.org/
TeraChem	"Full GPU-based solution"	44-650X vs. GAMESS CPU version	Released Version 1.5 Multi-GPU/single node	Completely redesigned to exploit GPU parallelism. YouTube: http://youtu.be/EJODzk6RFxE?hd=1 and http://www.olcf.ornl.gov/wp-content/training/electronic-structure-2012/Luehr-ESCMA.pdf
VASP	Hybrid Hartree-Fock DFT functionals including exact exchange	2x 2 GPUs comparable to 128 CPU cores	Available on request Multiple GPUs	By Carnegie Mellon University http://arxiv.org/pdf/1111.0716.pdf

GPU Perf compared against Multi-core x86 CPU socket.
GPU Perf benchmarked on GPU supported features and may be a kernel to kernel perf comparison

BigDFT



HPC and BigDFT

Electronic Structure Calculation Methods on Accelerators

ORNL/NICS – Oak Ridge, Tennessee

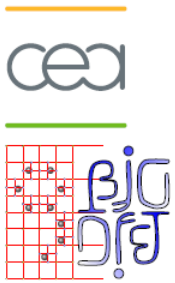
Wavelet-Based DFT calculations on Massively Parallel Hybrid Architectures

Luigi Genovese

L_Sim – CEA Grenoble

February 7, 2012

BigDFT version 1.6.0: (ABINIT-related) capabilities



http://inac.cea.fr/L_Sim/BigDFT

- Isolated, surfaces and 3D-periodic boundary conditions (k-points, symmetries)
- All XC functionals of the ABINIT package (libXC library)
- Hybrid functionals, Fock exchange operator
- Direct Minimisation and Mixing routines (metals)
- Local geometry optimizations (with constraints)
- External electric fields (surfaces BC)
- Born-Oppenheimer MD, ESTF-IO
- Vibrations
- Unoccupied states
- Empirical van der Waals interactions
- Saddle point searches (NEB, Granot & Bear)
- All these functionalities are GPU-compatible

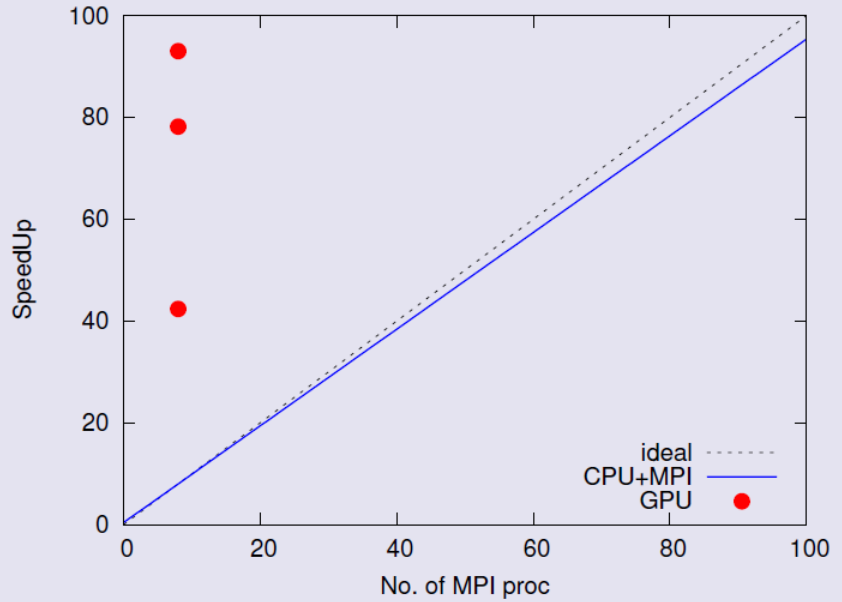
The time-to-solution problem I: Efficiency

Good example: 4 C at, surface BC, 113 Kpts

Parallel efficiency of 98%, convolutions largely dominate.

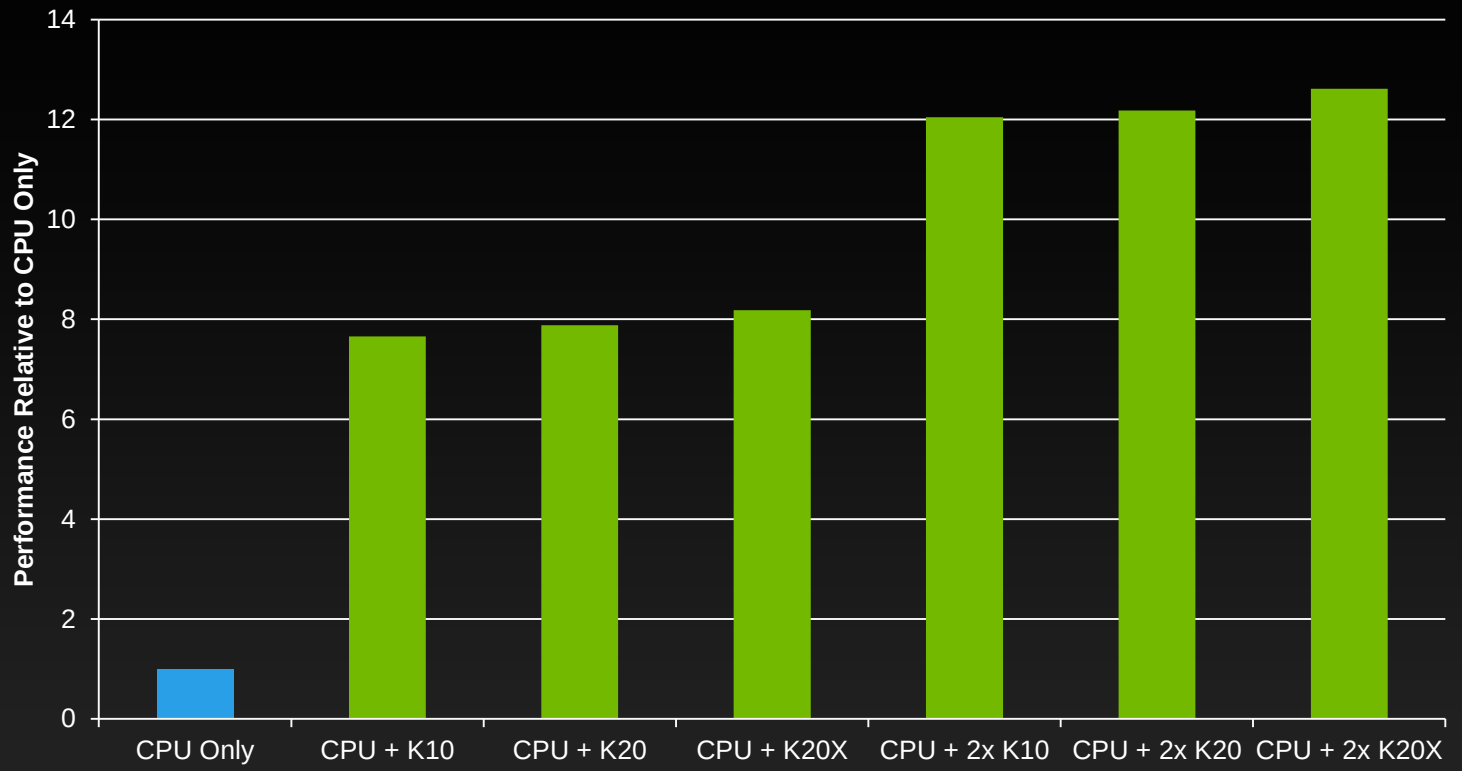
Node:
2× Fermi + 8 × Westmere
8 MPI processes

# GPU added	2	4	8
SpeedUp (SU)	5.3	9.8	11.6
# MPI equiv.	44	80	96
Acceler. Eff.	1	.94	.56



CP2K

Kepler, it's faster



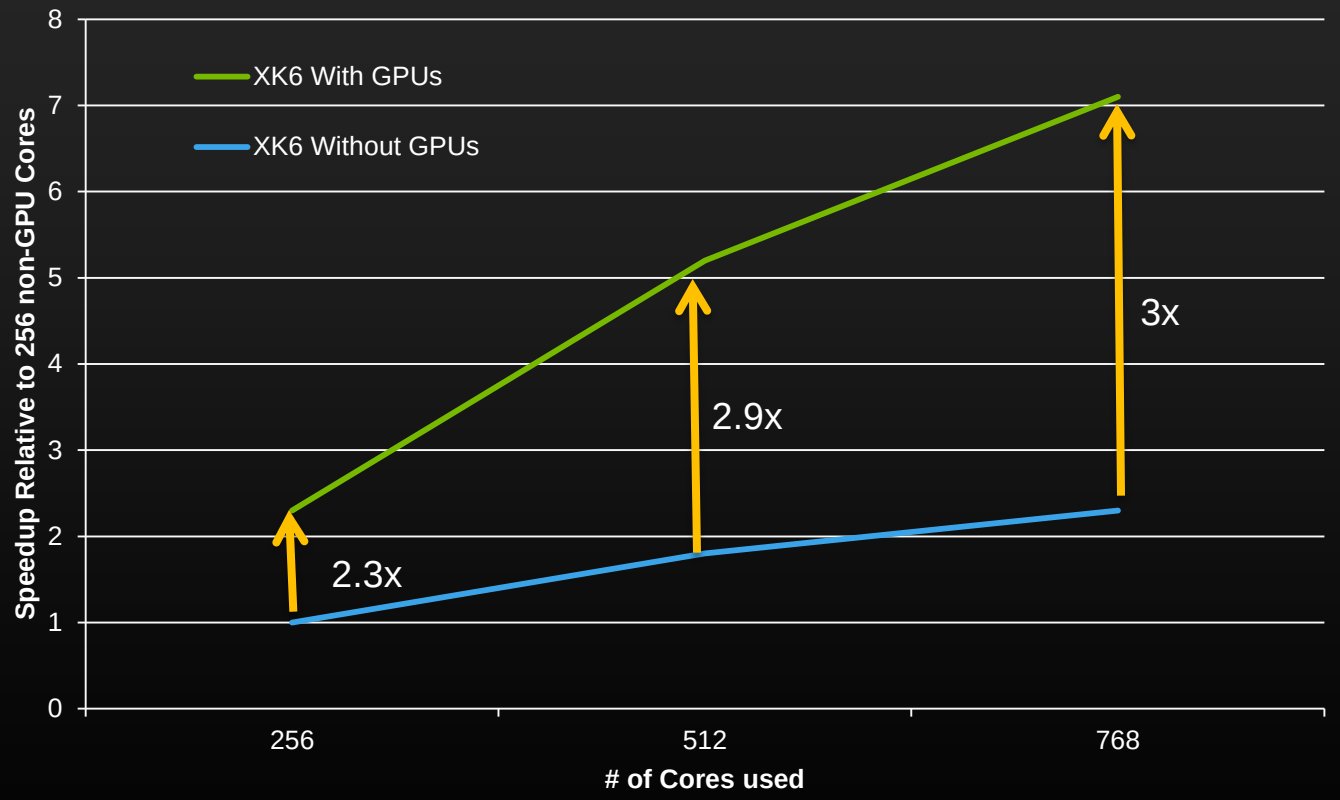
Running CP2K version 12413-trunk on CUDA 5.0.36

The blue node contains 2 E5-2687W CPUs (150W, 8 Cores per CPU).

The green nodes contain 2 E5-2687W CPUs and 1 or 2 NVIDIA K10, K20, or K20X GPUs (235W each).

Using GPUs delivers up to 12.6x the performance per node

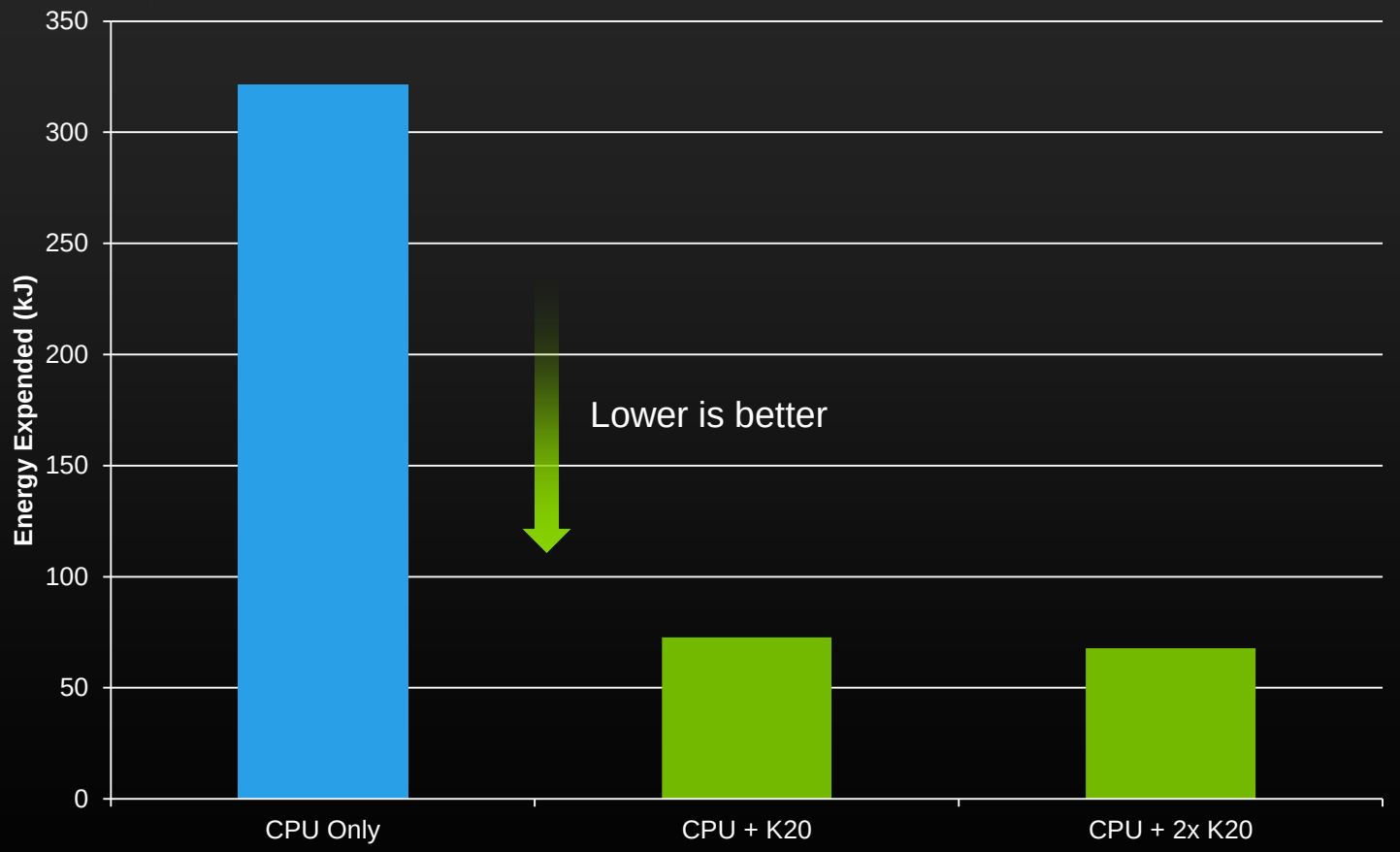
Strong Scaling



Conducted on Cray XK6
Using matrix-matrix multiplication
NREP=6 and N=159,000 with 50% occupation

Speedups increase as more nodes are added, up to 3x at 768 nodes

Kepler, keeping the planet Green



Running CP2K version 12413-trunk on CUDA 5.0.36

The blue node contains 2 E5-2687W CPUs (150W, 8 Cores per CPU).

The green nodes contain 2 E5-2687W CPUs and 1 or 2 NVIDIA K20 GPUs (235W each).

Energy Expended
= Power x Time

Using K20s will lower energy use by over 75% for the same simulation

GAUSSIAN

Gaussian

- Key quantum chemistry code
- ACS Fall 2011 press release
  - Joint collaboration between Gaussian, NVDA and PGI for GPU acceleration: http://www.gaussian.com/g_press/nvidia_press.htm
  - No such release exists for Intel MIC or AMD GPUs
- Mike Frisch quote:
  - *"Calculations using Gaussian are limited primarily by the available computing resources," said Dr. Michael Frisch, president of Gaussian, Inc. "By coordinating the development of hardware, compiler technology and application software among the three companies, the new application will bring the speed and cost-effectiveness of GPUs to the challenging problems and applications that Gaussian's customers need to address."*

GAMESS

GAMESS Partnership Overview

- **Mark Gordon and Andrey Asadchev, key developers of GAMESS, in collaboration with NVIDIA. Mark Gordon is a recipient of a NVIDIA Professor Partnership Award.**

- **Quantum Chemistry one of major consumers of CPU cycles at national supercomputer centers**

- **NVIDIA developer resources fully allocated to GAMESS code**

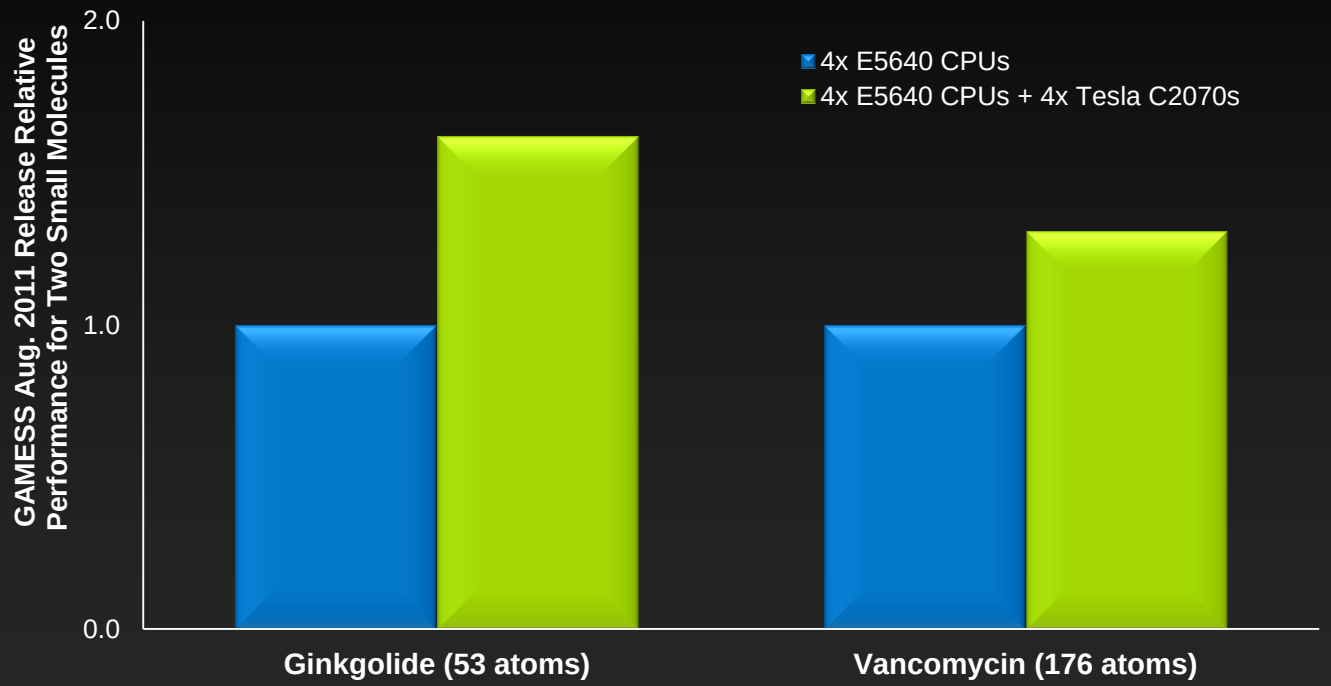
"We like to push the envelope as much as we can in the direction of highly scalable efficient codes. GPU technology seems like a good way to achieve this goal. Also, since we are associated with a DOE Laboratory, energy efficiency is important, and this is another reason to explore quantum chemistry on GPUs."

Prof. Mark Gordon

Distinguished Professor, Department of Chemistry, Iowa State University and Director, Applied Mathematical Sciences Program, AMES Laboratory

GAMESS August 2011 GPU Performance

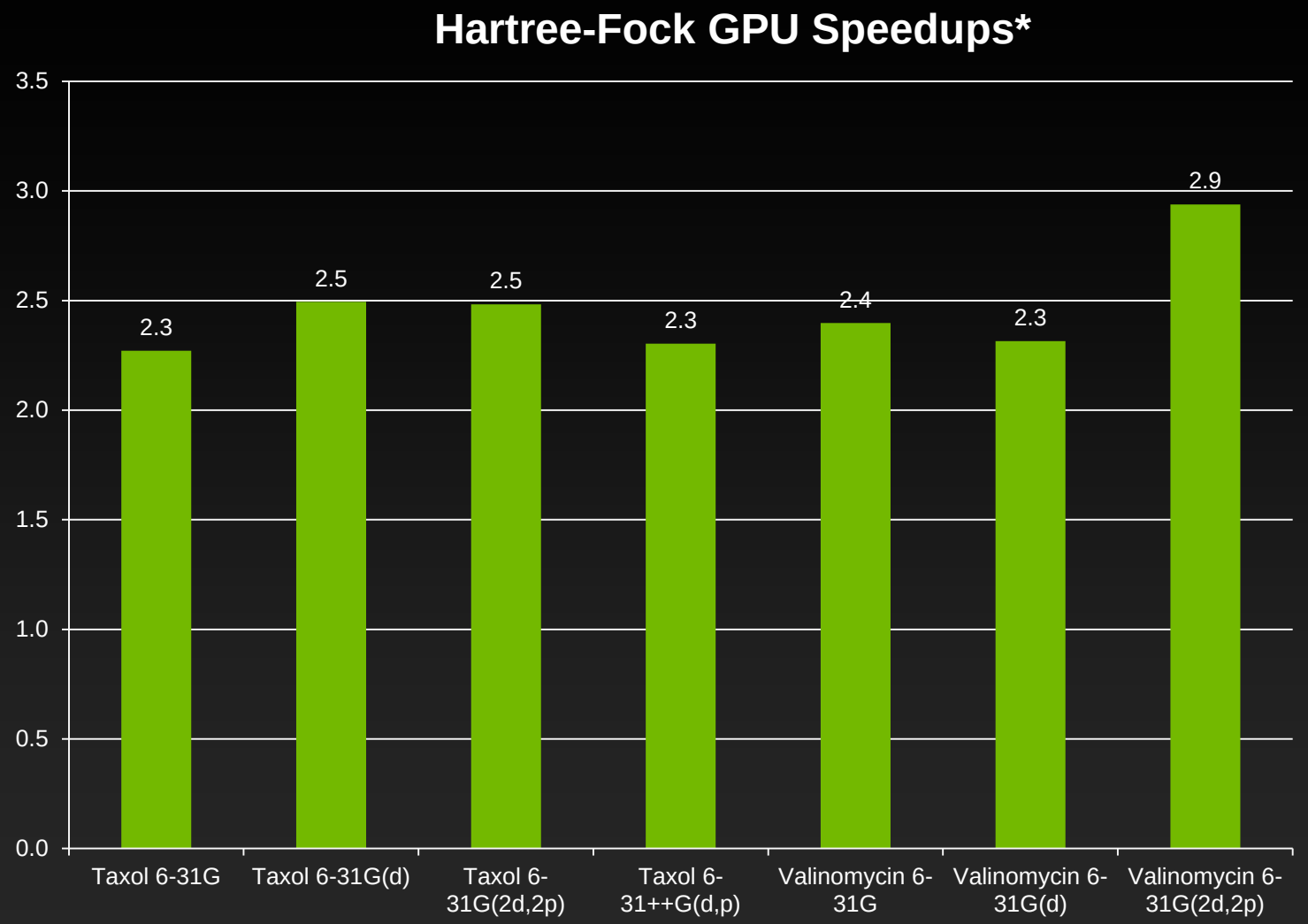
- First GPU supported GAMESS release via "libqc", a library for fast quantum chemistry on multiple NVIDIA GPUs in multiple nodes, with CUDA software
- 2e- AO integrals and their assembly into a closed shell Fock matrix



Upcoming GAMESS Q4 2012 Release

- Multi-nodes with multi-GPUs supported
- Rys Quadrature
- Hartree-Fock
  - 8 CPU cores: 8 CPU cores + M2070 yields 2.3-2.9x speedup. See 2012 publication
- Møller–Plesset perturbation theory (MP2): Preliminary code completed
  - Paper in development
- Coupled Cluster SD(T): CCSD code completed, (T) in progress

GAMESS - New Multithreaded Hybrid CPU/GPU Approach to H-F



Adding 1x 2070 GPU speeds up computations by 2.3x to 2.9x

* A. Asadchev, M.S. Gordon, "New Multithreaded Hybrid CPU/GPU Approach to Hartree-Fock," Journal of Chemical Theory and Computation (2012)

GPAW

Parallel Electronic Structure Calculations Using Multiple GPUs

Samuli Hakala
COMP/Department of Applied Physics, Aalto University
P.O. Box 11100, 00076 Aalto, Espoo, Finland
Email: samuli.hakala@aalto.fi

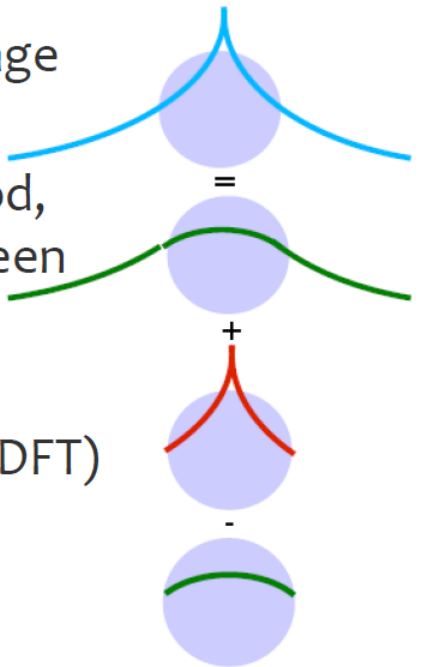
Aalto University
School of Science

NVIDIA

Used with permission from Samuli Hakala

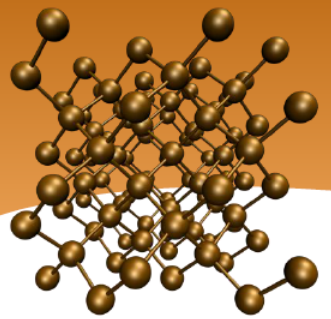
GPAW

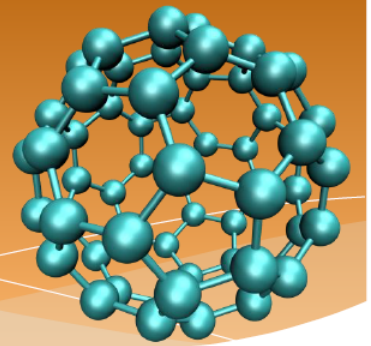
- Density Functional Theory (DFT) program package for electronic structure calculations
- Uses Projector Augmented Wave (PAW) method, which is based on a linear transformation between smooth valence pseudo wave-functions and all electron wave-functions
- Time-Dependent Density Functional Theory (TDDFT) is implemented in the linear response and time propagation schemes



J. Enkovaara et al., E J. Phys.: Condens. Matter 22, 253202 (2010)

DFT performance





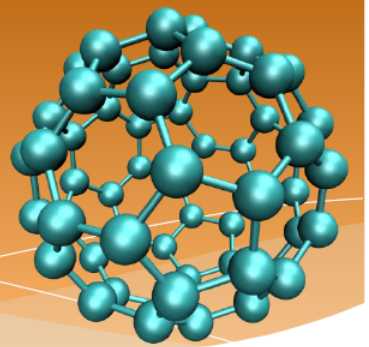
* Bulk silicon with 95 atoms in a supercell with periodic boundary conditions, 380 bands and 1 k-point. Grid size: 56x56x80. Time is in seconds

* Fullerene molecule C60 with 240 valence electrons. Grid size: 84x84x84

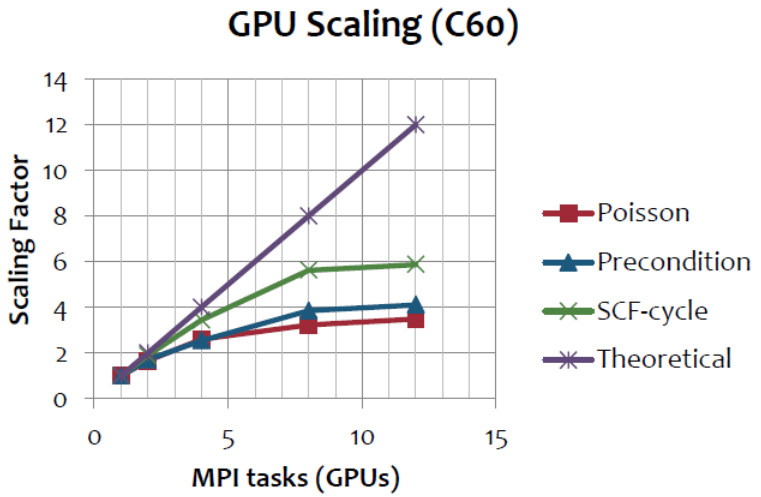
Si95	CPU	GPU	S-Up
Poisson Solver	1.8	0.13	**14**
Orthonormalization	23	3.0	**7.7**
Precondition	9.4	0.77	**12**
RMM-DIIS other	32	3.2	**10**
Subspace Diag	23	2.1	**11**
Other	2.7	2.7	**1.0**
Total (SCF-Iter)	93	13	9.7/7.7

C60	CPU	GPU	S-Up
	13	0.64	**20**
	11	1.2	**9.2**
	16	0.99	**16**
	8.1	0.6	**13**
	22	2.1	**10**
	3.5	3.2	**1.1**
	76	9.1	13/8.3

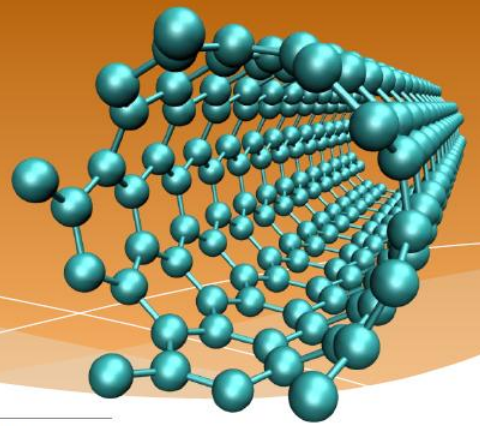
Strong Scaling

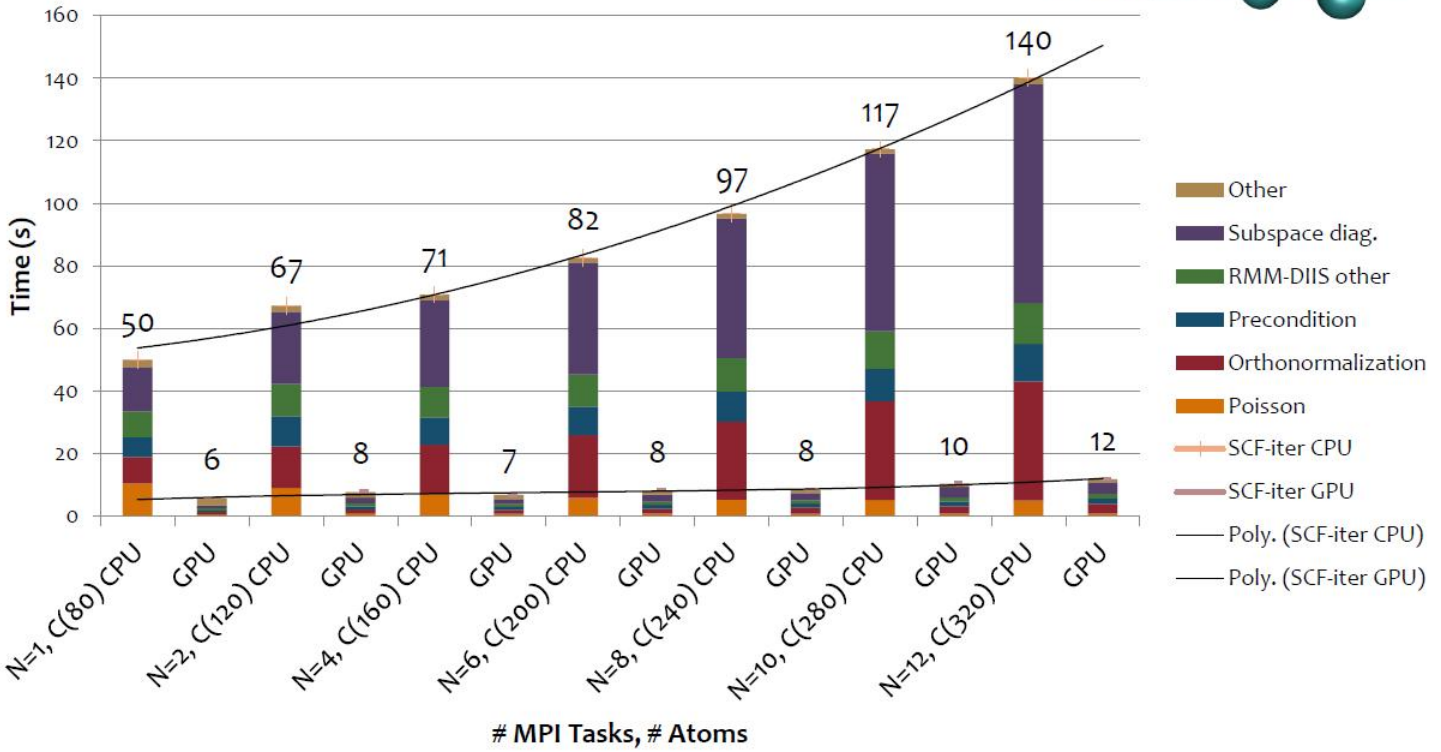


- Fullerene molecule C60 with 240 electronic states. Grid size: 84x84x84. Using domain decomposition
- 7 node GPU cluster.
- Nvidia Tesla M2070 and M2050

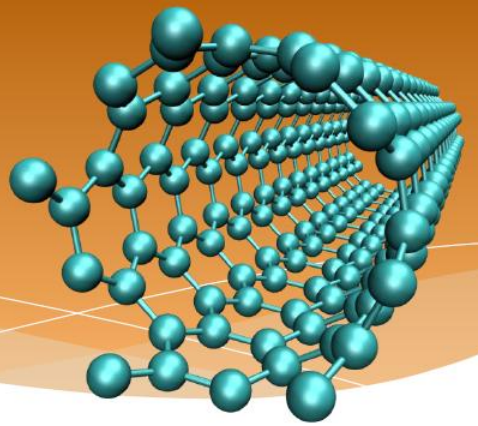


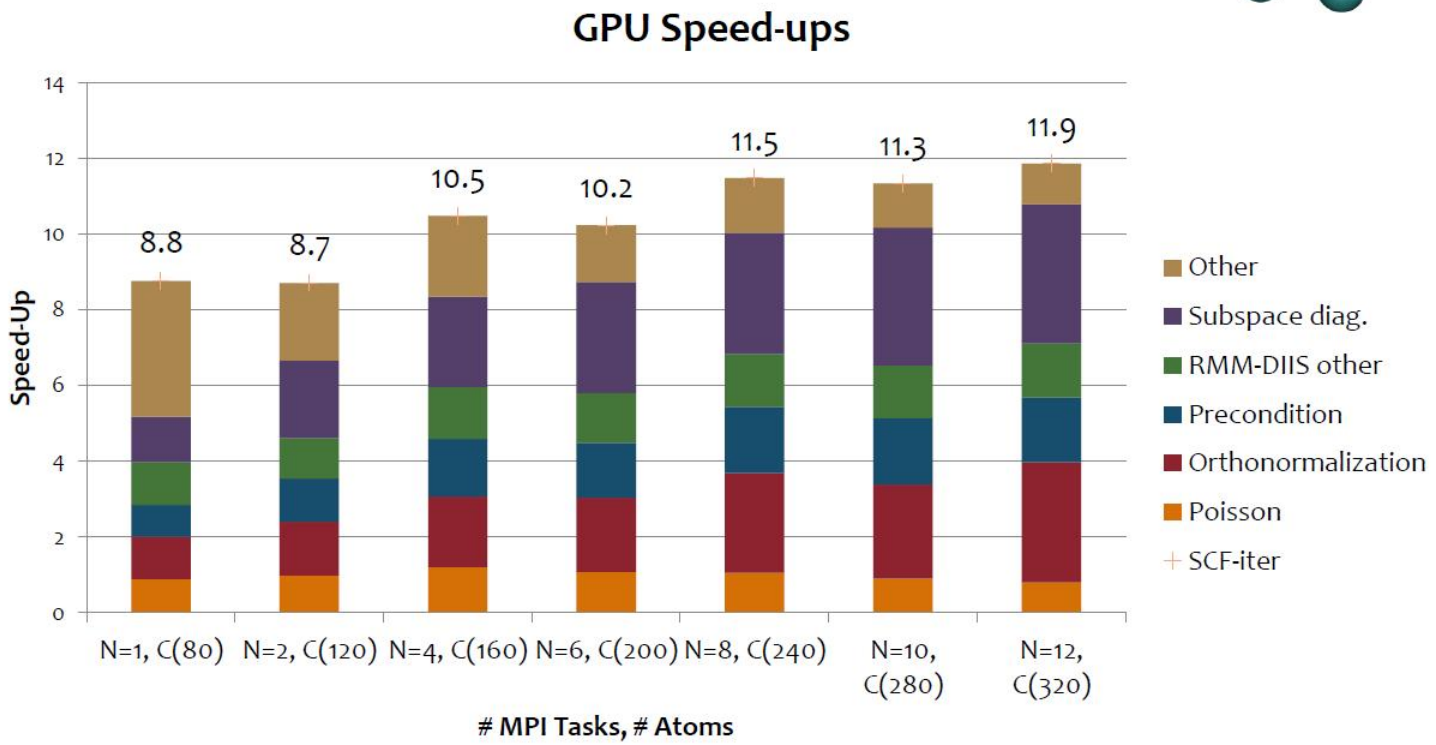
Carbon Nanotube



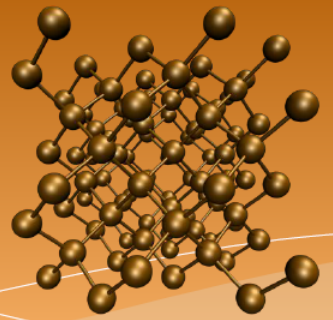


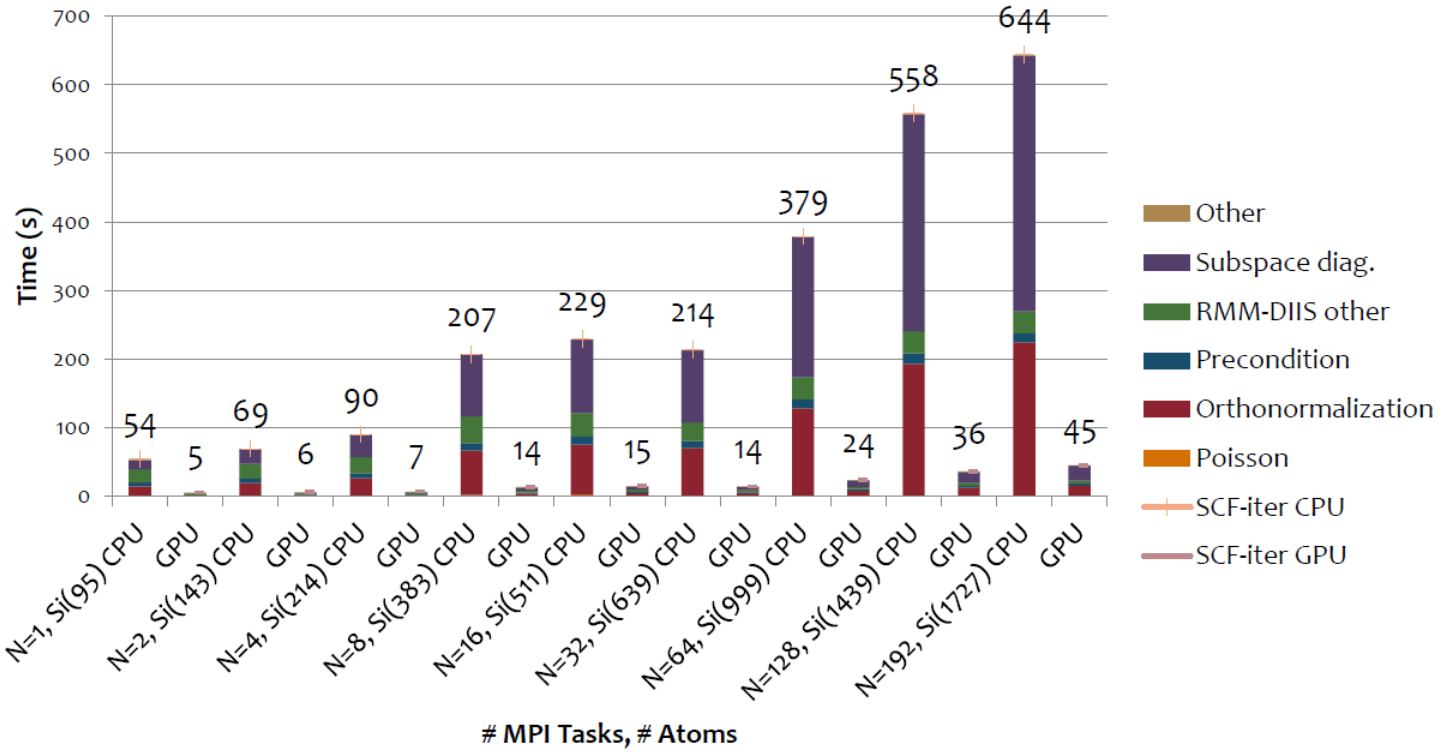
Nanotube Speed-ups



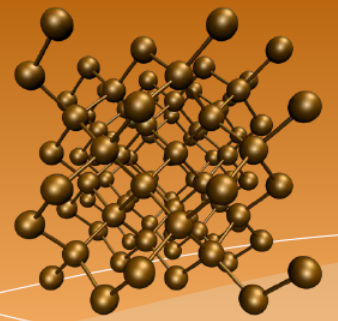


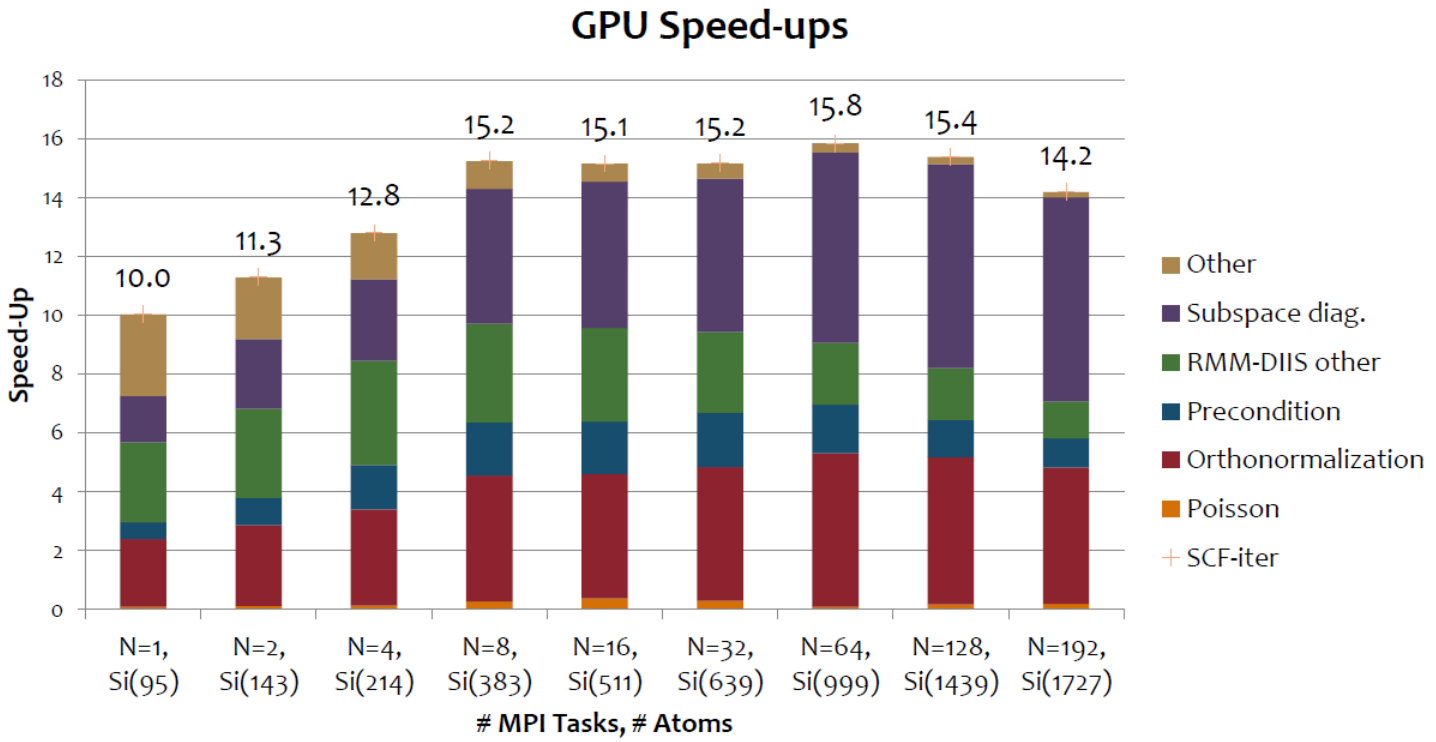
Bulk Silicon





Bulk Silicon Speed-ups



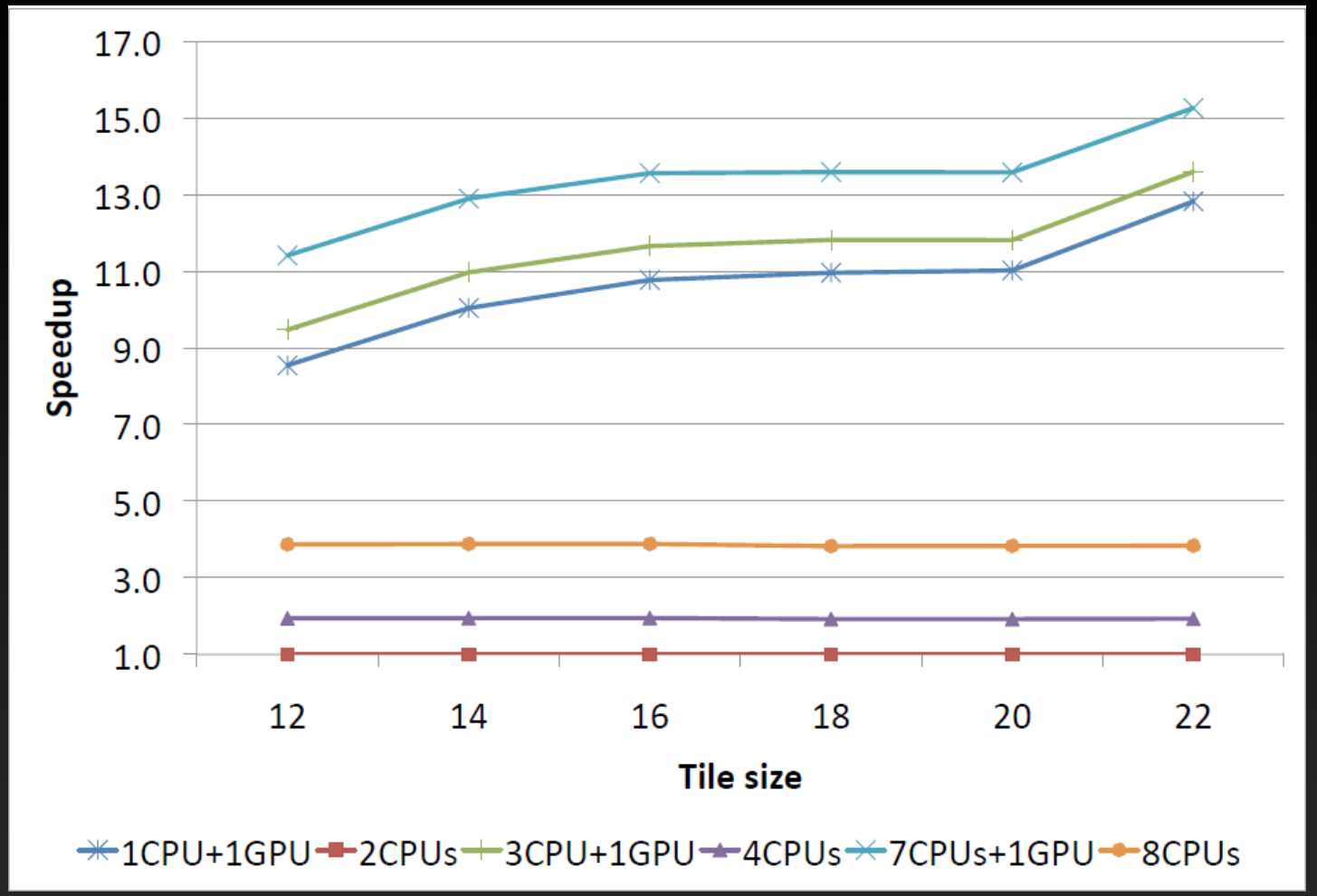


Conclusions

* GPUs can provide significant speed-ups in practical electronic structure calculations
* Good performance requires lots of data and/or algorithmic parallelism and careful programming
* Multi-GPU implementation scales well.
* CSC and PRACE provided the computing resources used in this work

NWChem

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



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

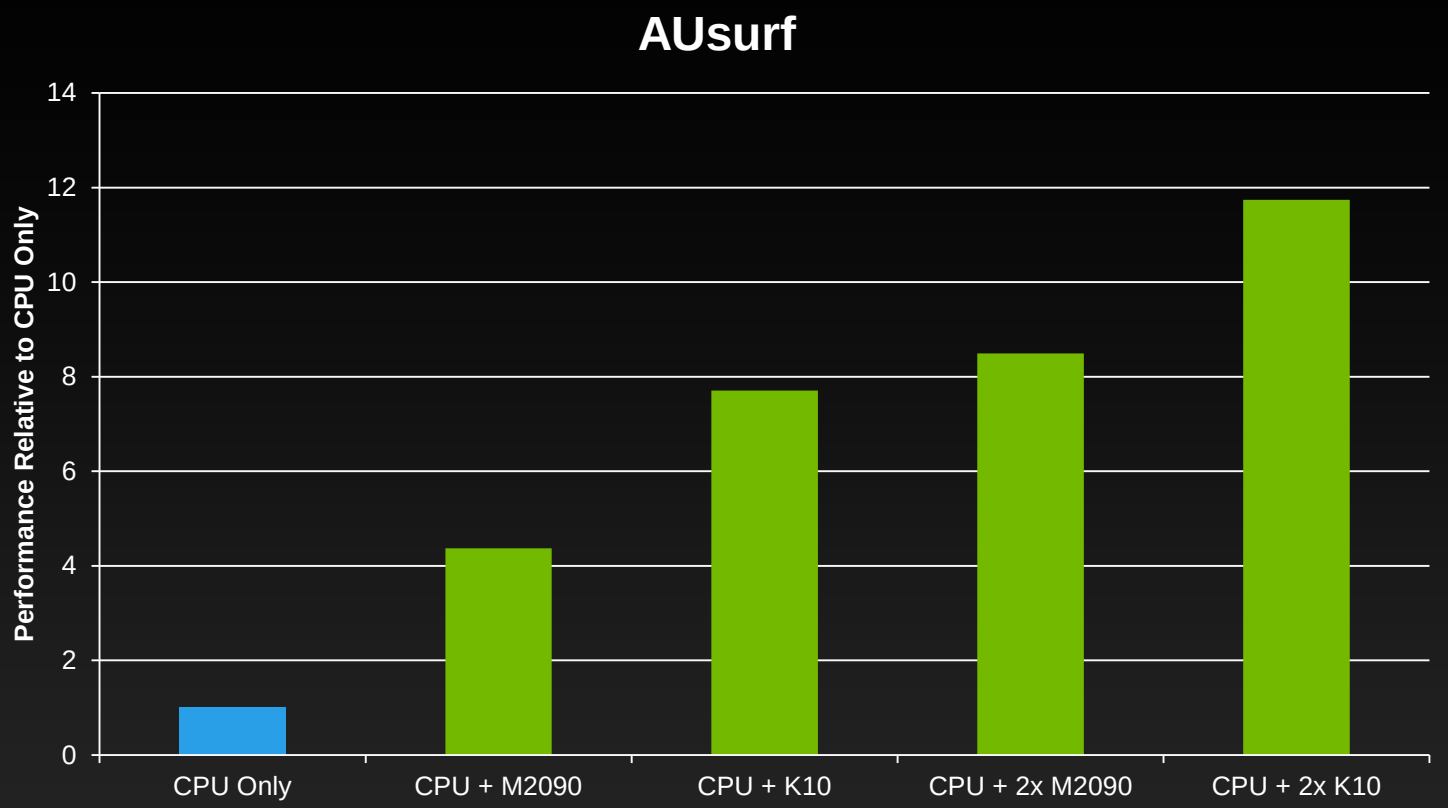
- 8-core AMD Interlagos processors
- 64 GB of memory
- Tesla M2090 (Fermi) GPUs

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

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

Quantum Espresso/PWscf

Kepler, fast science



Running Quantum Espresso version 5.0-build7 on CUDA 5.0.36

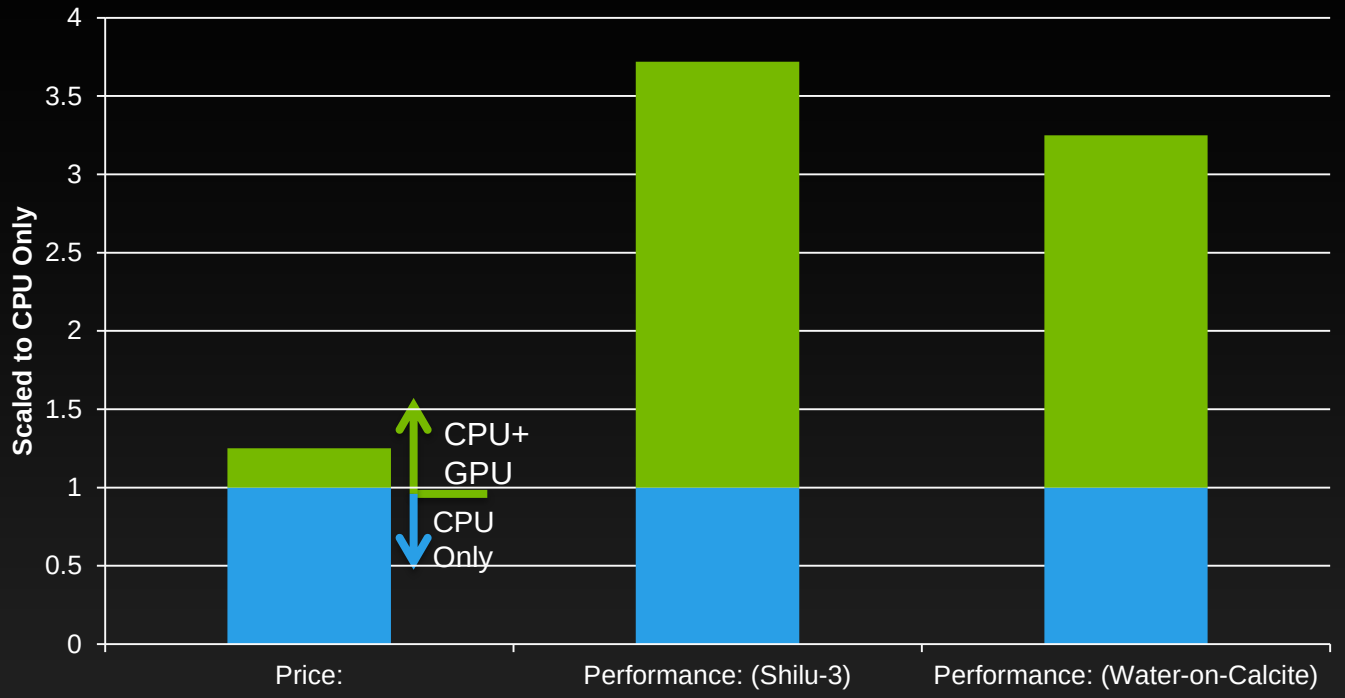
The blue node contains 2 E5-2687W CPUs (150W, 8 Cores per CPU).

The green nodes contain 2 E5-2687W CPUs and 1 or 2 NVIDIA M2090 or K10 GPUs (225W and 235W respectively).

Using K10s delivers up to 11.7x the performance per node over CPUs

And 1.7x the performance when compared to M2090s

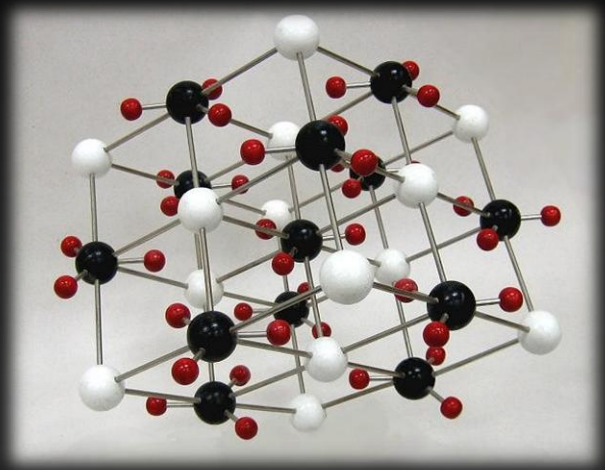
Extreme Performance/Price from 1 GPU



Simulations run on FERMI @ ICHEC.

A 6-Core 2.66 GHz Intel X5650 was used for the CPU

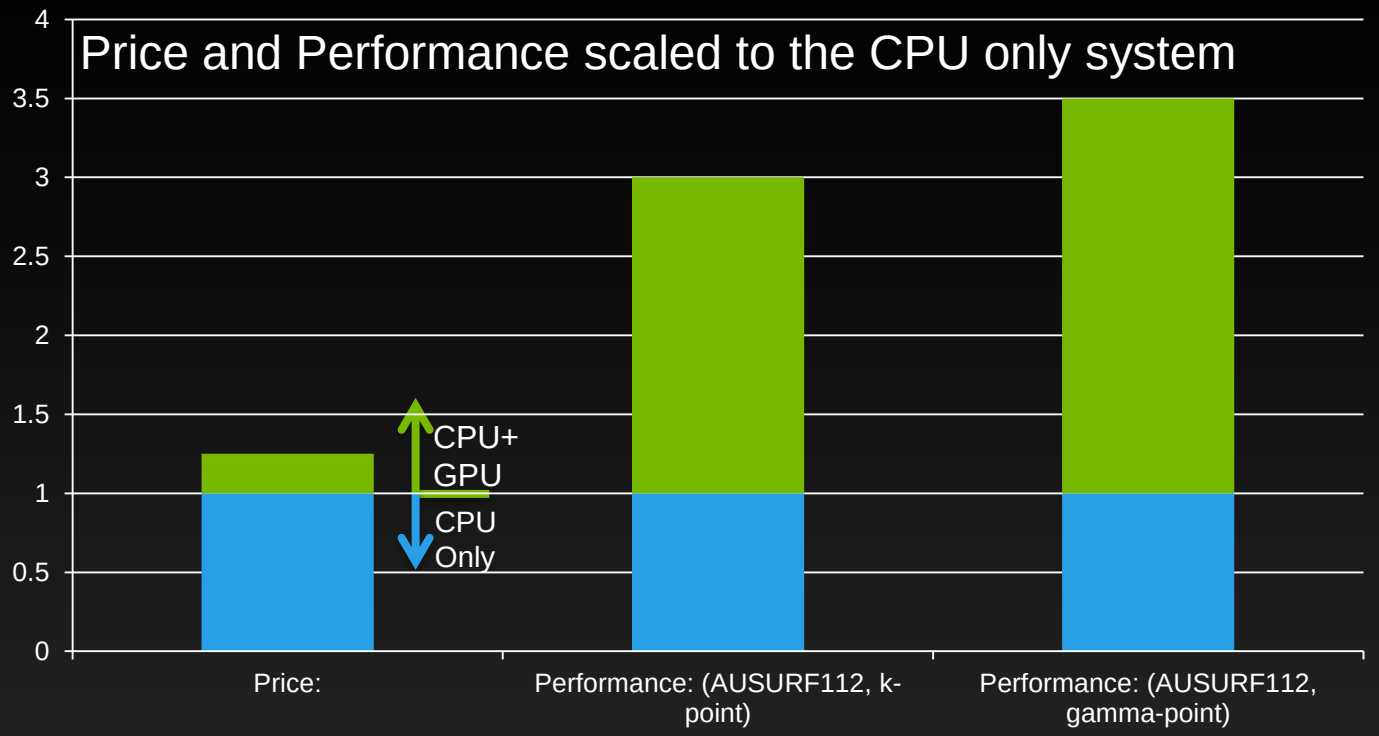
An NVIDIA C2050 was used for the GPU



Calcite structure

Adding a GPU can improve performance by 3.7x while only increasing price by 25%

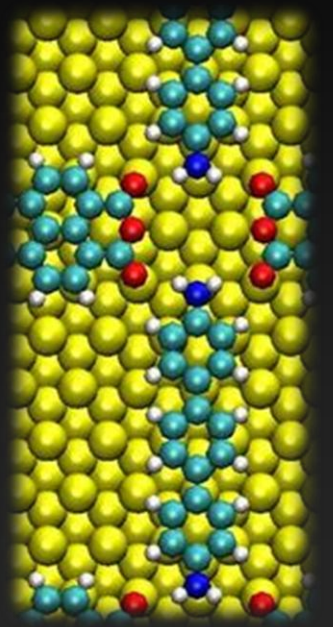
Extreme Performance/Price from 1 GPU



Simulations run on FERMI @ ICHEC.

A 6-Core 2.66 GHz Intel X5650 was used for the CPU

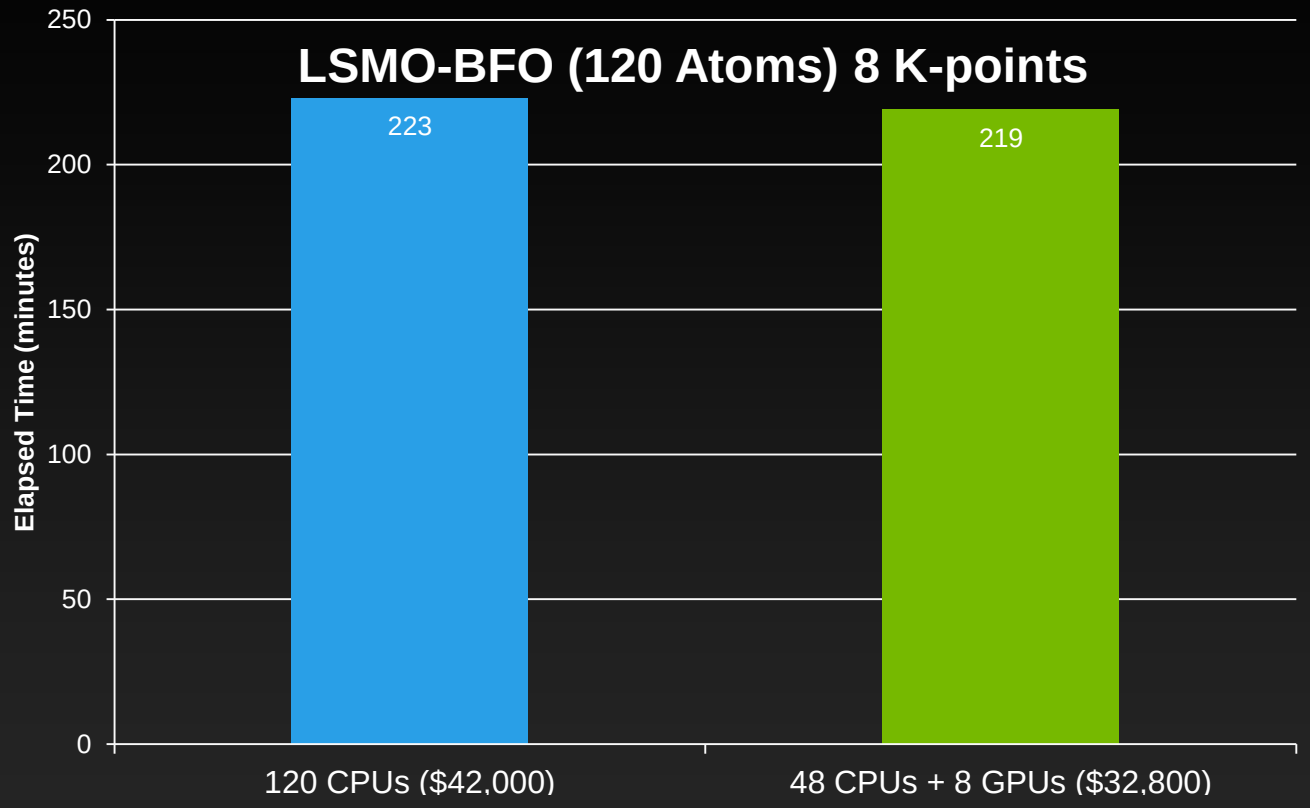
An NVIDIA C2050 was used for the GPU



Calculation done for a gold surface of 112 atoms

Adding a GPU can improve performance by 3.5x while only increasing price by 25%

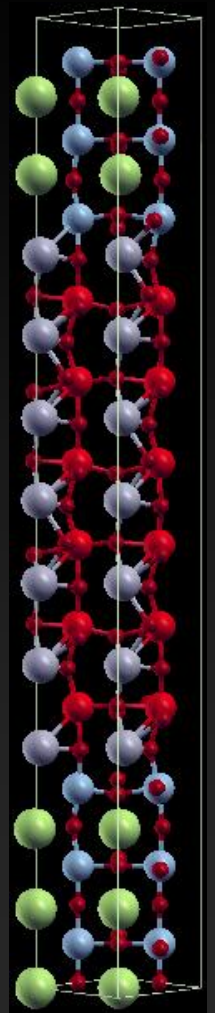
Replace 72 CPUs with 8 GPUs



Simulations run on PLX @ CINECA.

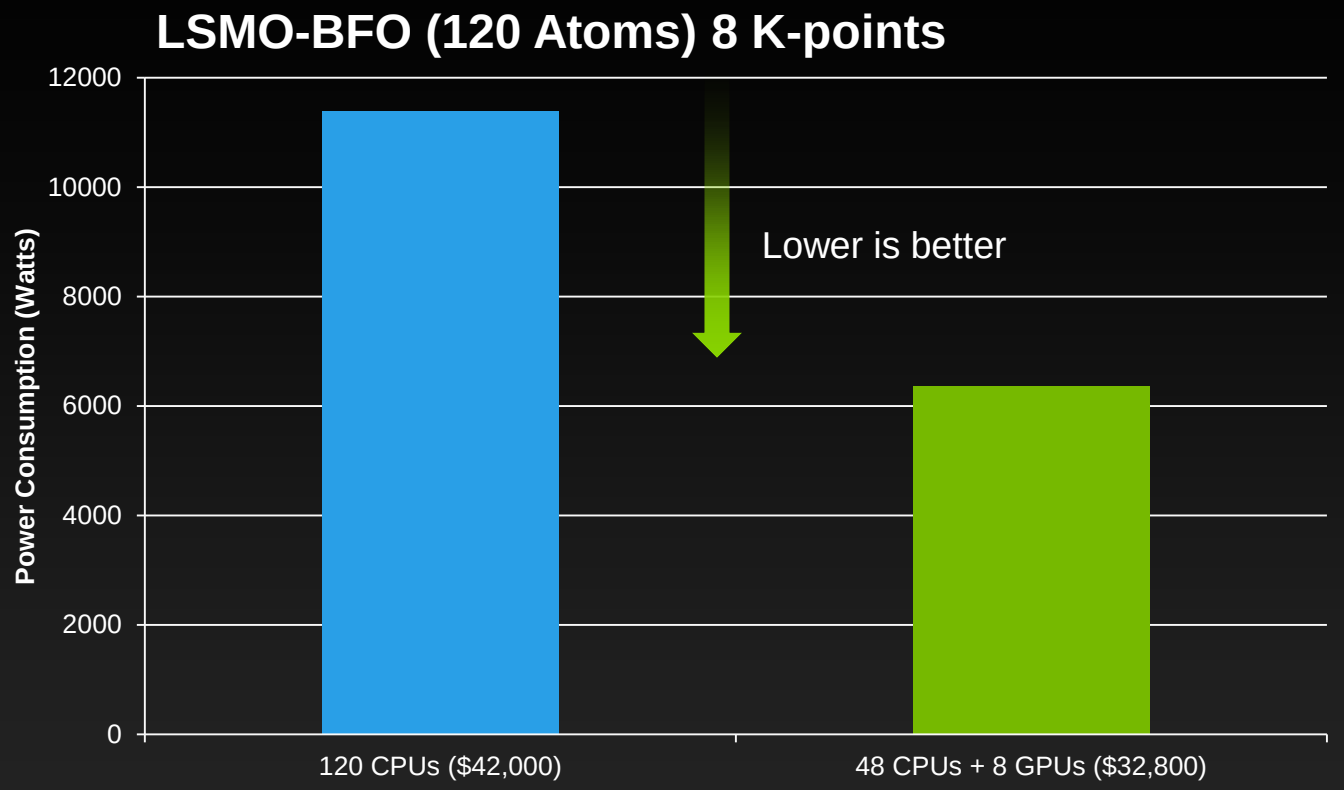
Intel 6-Core 2.66 GHz X5550 were used for the CPUs

NVIDIA M2070s were used for the GPUs



The GPU Accelerated setup performs faster and costs 24% less

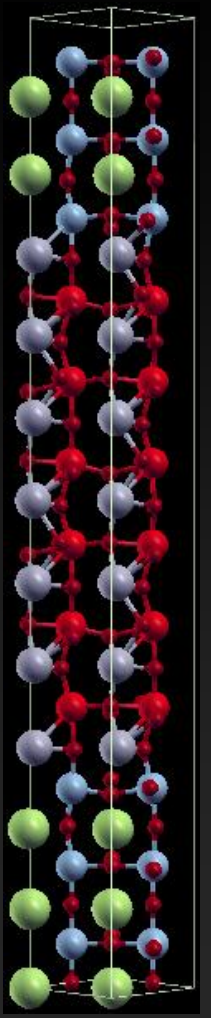
QE/PWscf - Green Science



Simulations run on PLX @ CINECA.

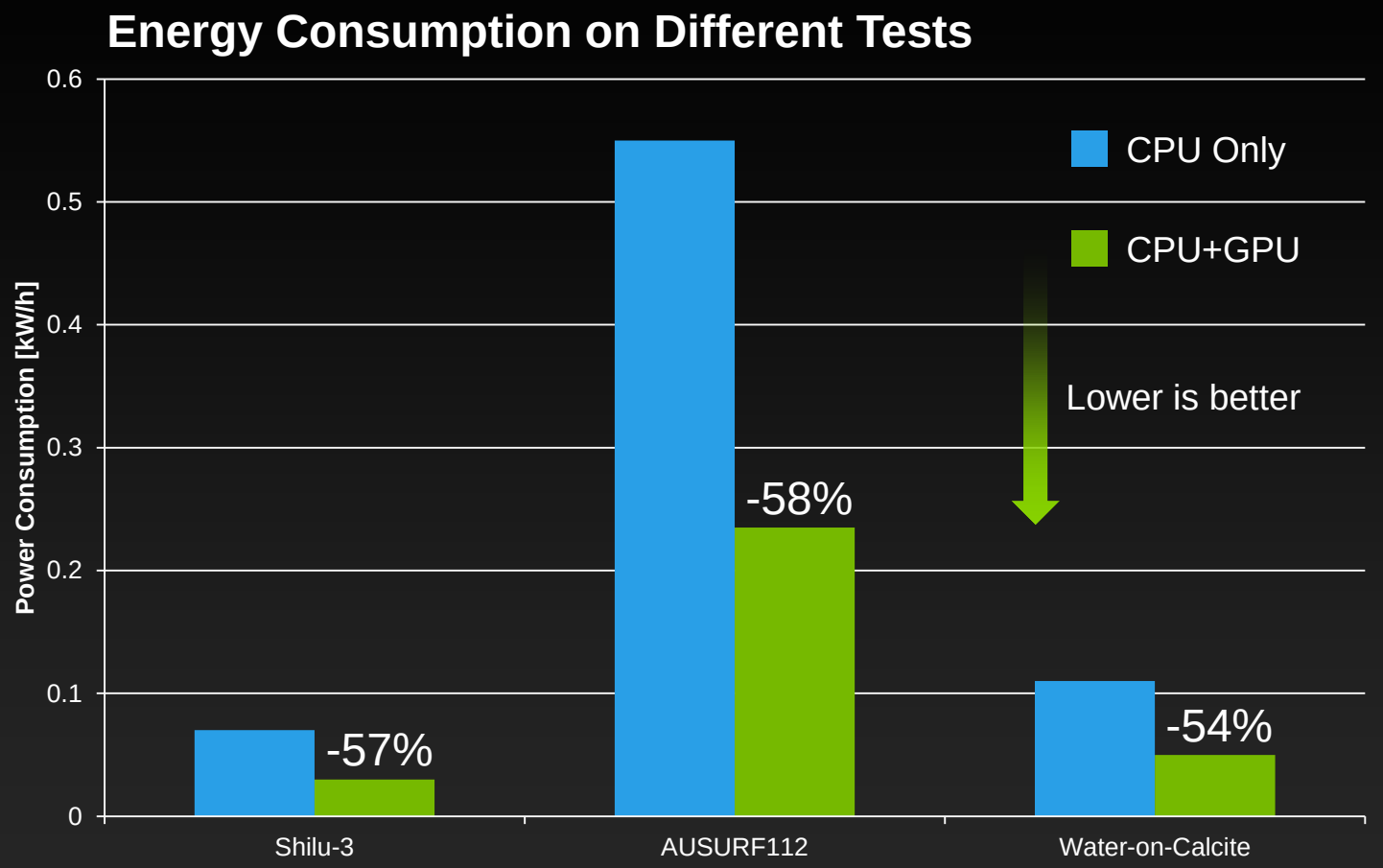
Intel 6-Core 2.66 GHz X5550 were used for the CPUs

NVIDIA M2070s were used for the GPUs



Over a year, the lower power consumption would save $4300 on energy bills

NVIDIA GPUs Use Less Energy



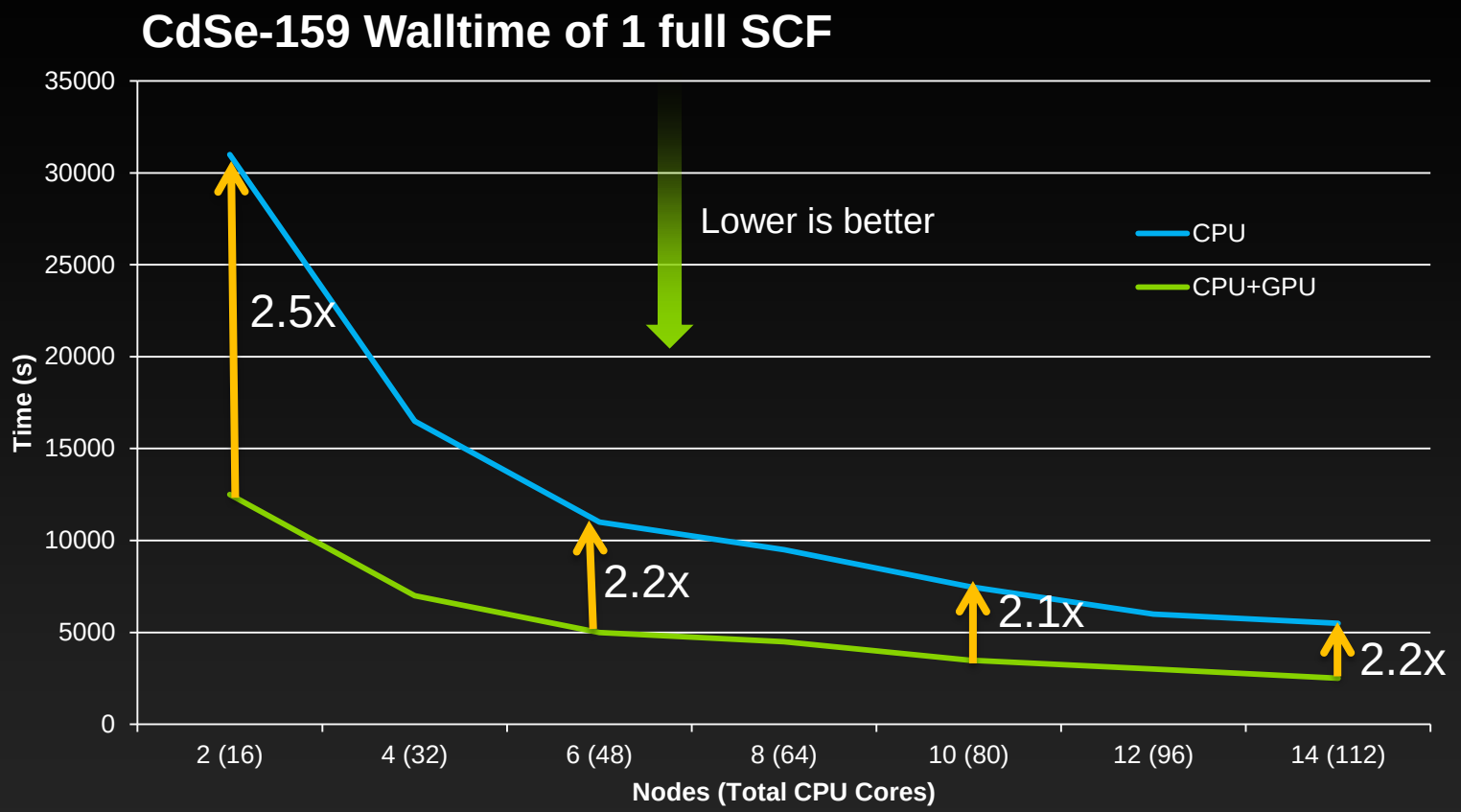
Simulations run on FERMI @ ICHEC.

A 6-Core 2.66 GHz Intel X5650 was used for the CPU

An NVIDIA C2050 was used for the GPU

In all tests, the GPU Accelerated system consumed less than half the power as the CPU Only

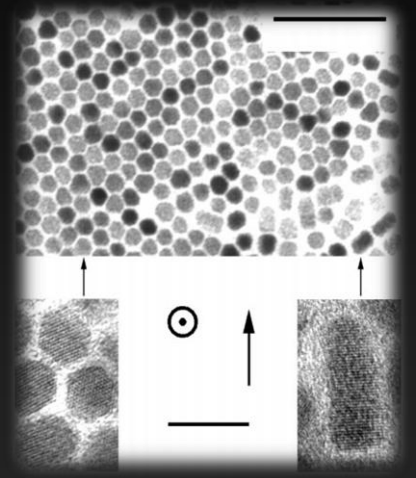
QE/PWscf - Great Strong Scaling in Parallel



Simulations run on STONEY @ ICHEC.

Two quad core 2.87 GHz Intel X5560s were used in each node

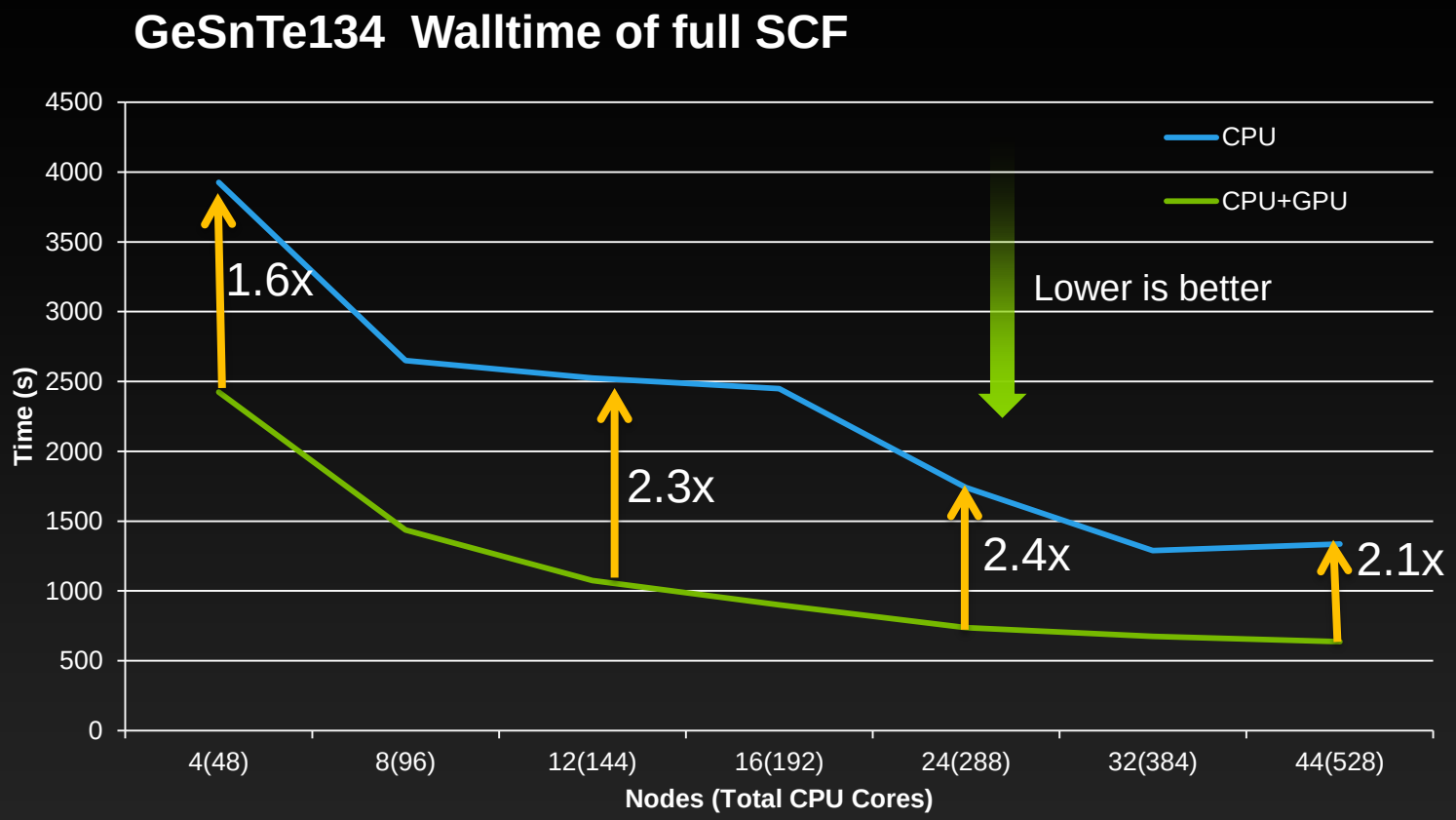
Two NVIDIA M2090s were used in each node for the CPU+GPU test



159 Cadmium Selenide nanodots

Speedups up to 2.5x with GPU Accelerations

QE/PWscf - More Powerful Strong Scaling



Simulations run on PLX @ CINECA.

Two 6-Core 2.4 GHz Intel E5645s were used in each node

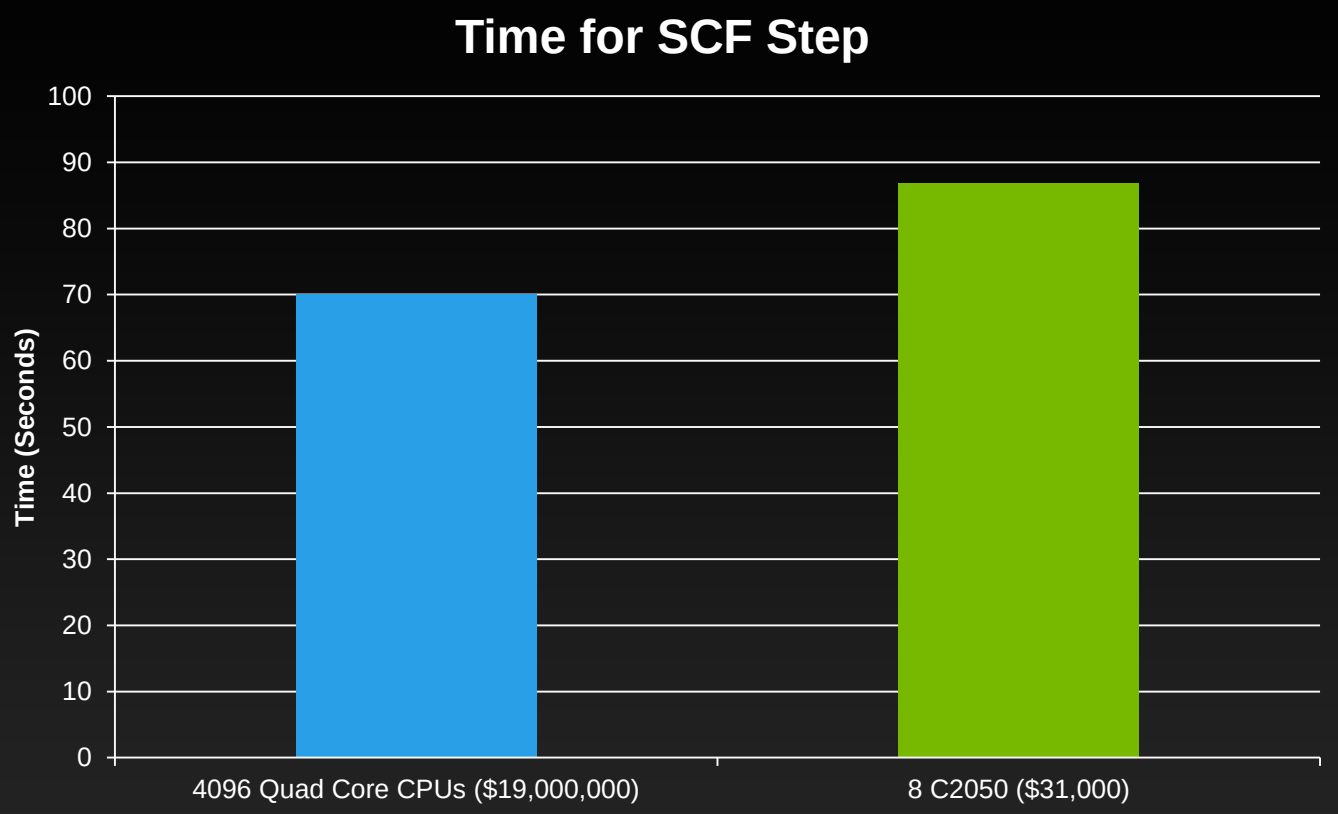
Two NVIDIA M2070s were used in each node for the CPU+GPU test

Accelerate your cluster by up to 2.1x with NVIDIA GPUs

Try GPU accelerated Quantum Espresso for free – www.nvidia.com/GPUTestDrive

TeraChem

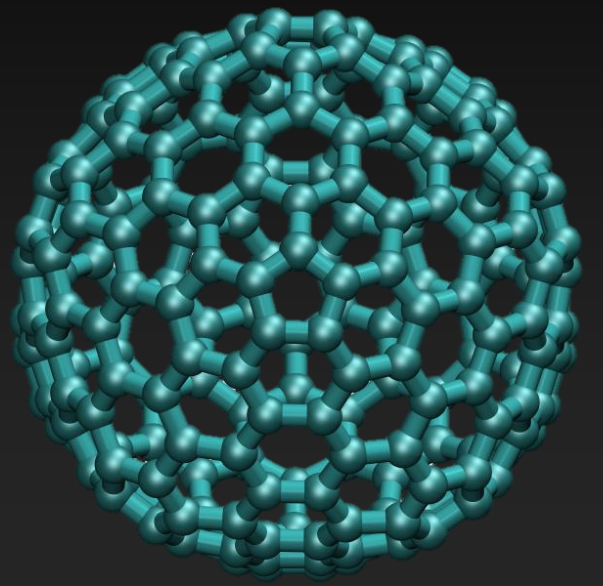
TeraChem
Supercomputer Speeds on GPUs



TeraChem running on 8 C2050s on 1 node

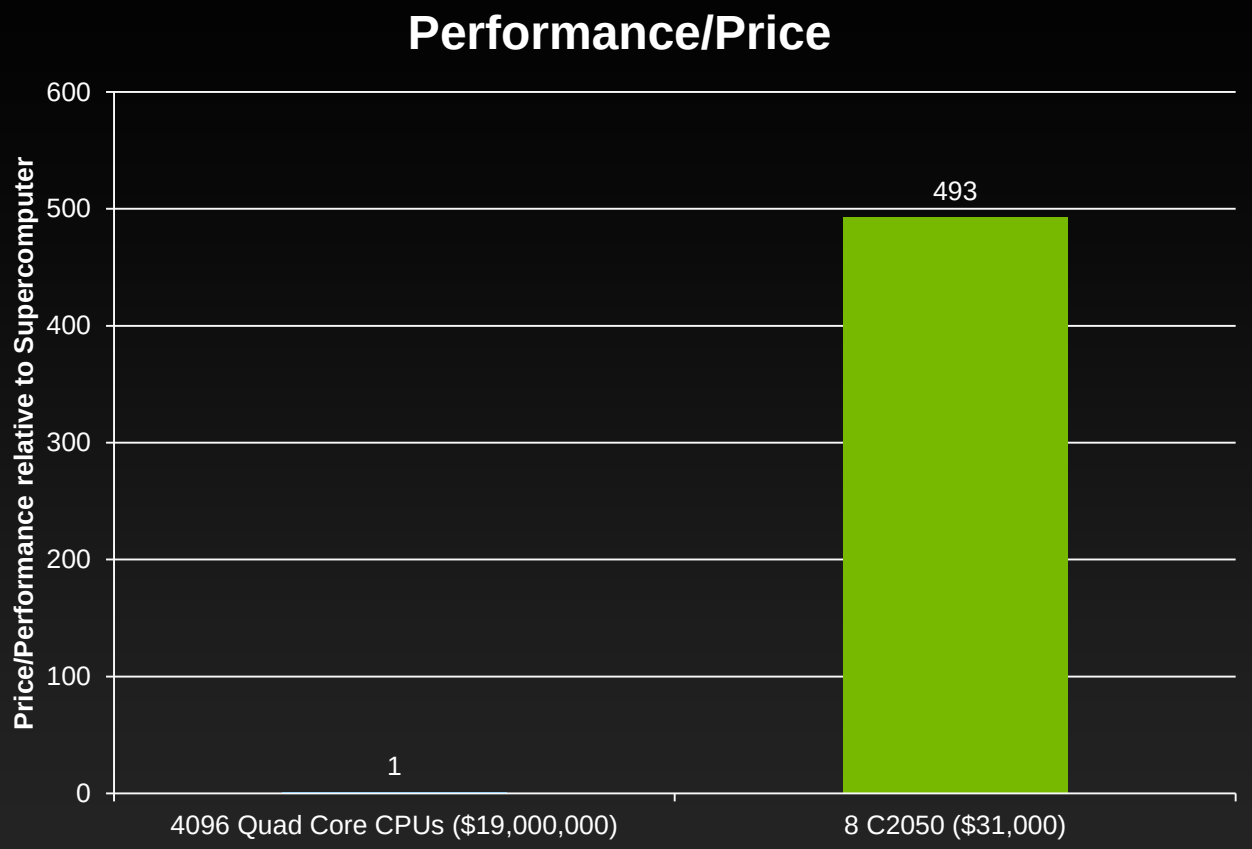
NWChem running on 4096 Quad Core CPUs
In the Chinook Supercomputer

Giant Fullerene C240 Molecule



Similar performance from just a handful of GPUs

TeraChem
Bang for the Buck

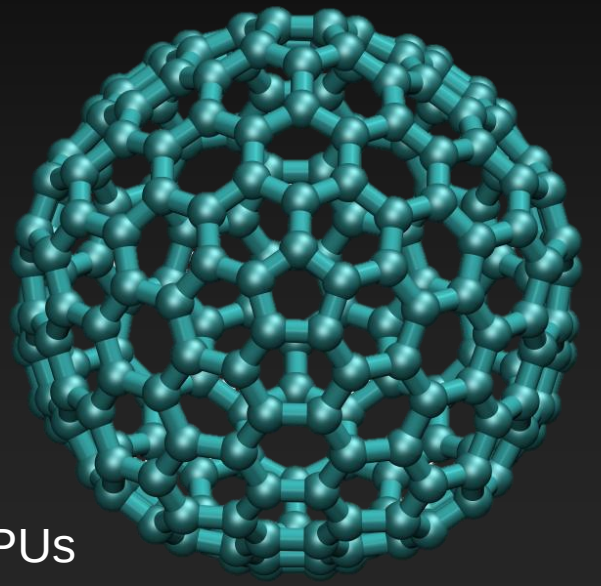


TeraChem running on 8 C2050s on 1 node

NWChem running on 4096 Quad Core CPUs
In the Chinook Supercomputer

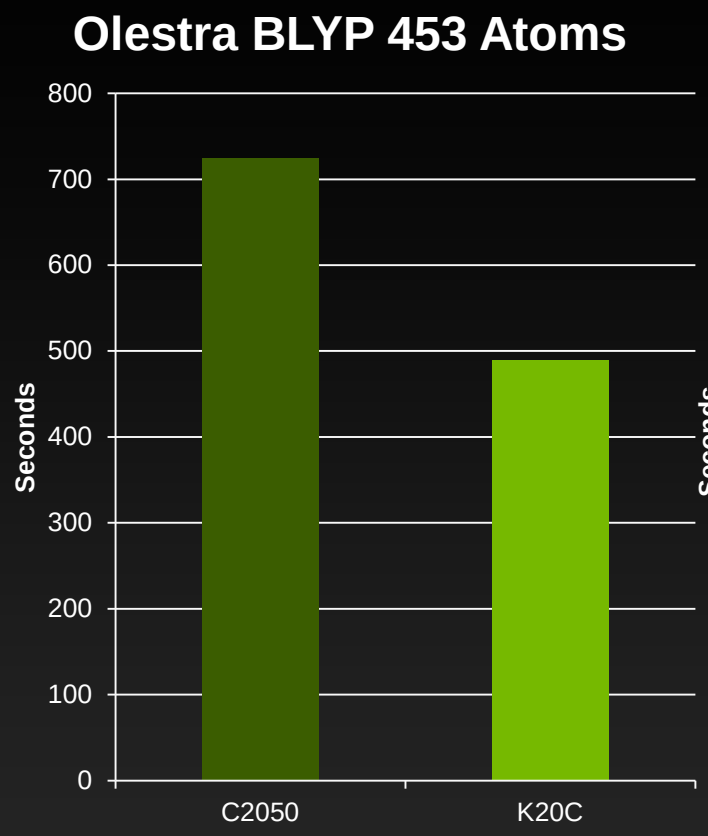
Giant Fullerene C240 Molecule

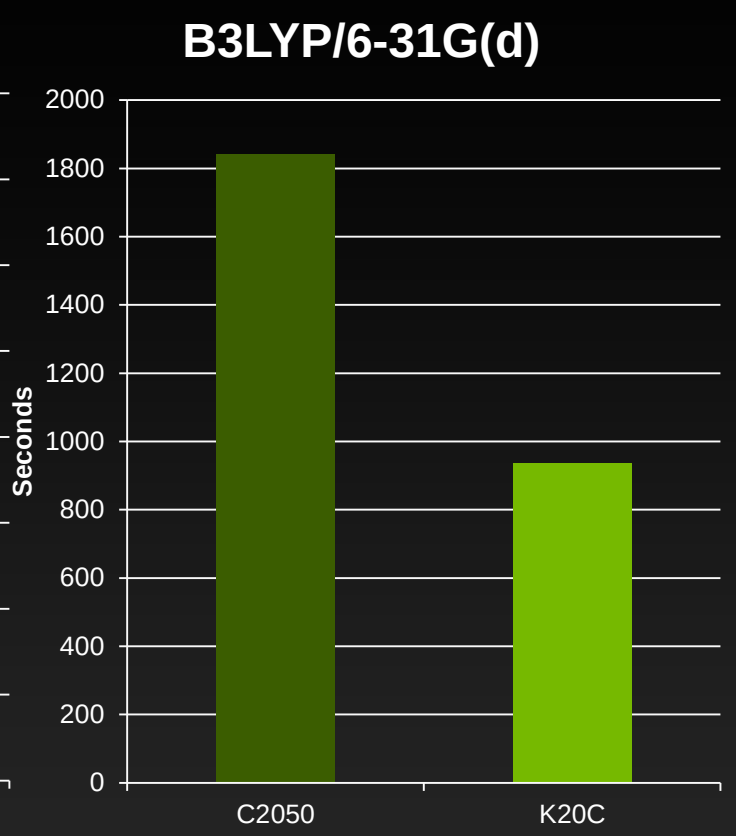
Note: Typical CPU and GPU node pricing used. Pricing may vary depending on node configuration. Contact your preferred HW vendor for actual pricing.



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

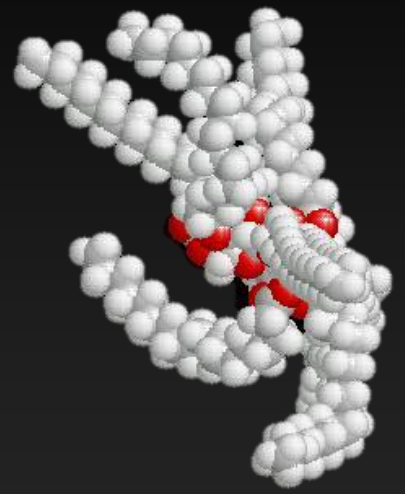
Kepler's Even Better





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

Viz, "Docking" and Related Applications Growing

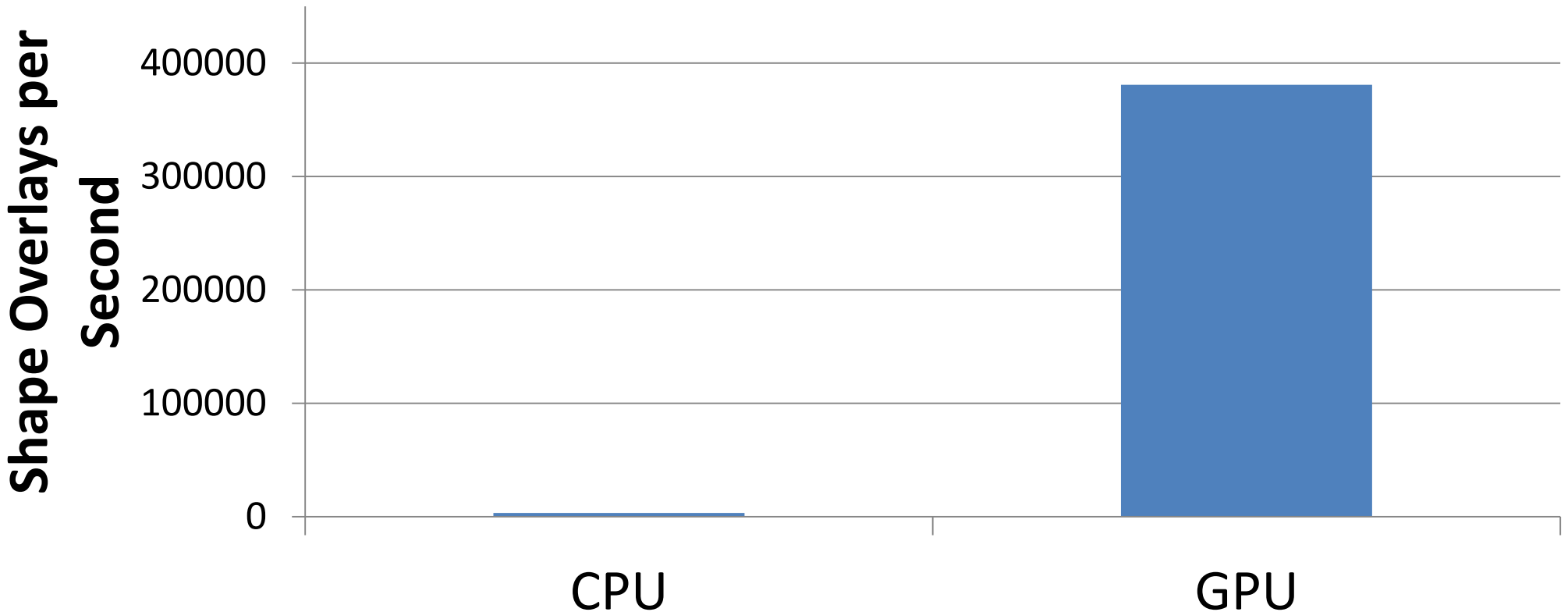
Related Applications	Features Supported	GPU Perf	Release Status	Notes
Amira 5®	3D visualization of volumetric data and surfaces	70x	Released, Version 5.3.3 Single GPU	Visualization from Visage Imaging. Next release, 5.4, will use GPU for general purpose processing in some functions http://www.visageimaging.com/overview.html
BINDSURF	Allows fast processing of large ligand databases	100X	Available upon request to authors; single GPU	High-Throughput parallel blind Virtual Screening, http://www.biomedcentral.com/1471-2105/13/S14/S13
BUDE	Empirical Free Energy Forcefield	6.5-13.4X	Released Single GPU	University of Bristol http://www.bris.ac.uk/biochemistry/cpfg/bude/bude.htm
Core Hopping	GPU accelerated application	3.75-5000X	Released, Suite 2011 Single and multi-GPUs.	Schrodinger, Inc. http://www.schrodinger.com/products/14/32/
FastROCS	Real-time shape similarity searching/comparison	800-3000X	Released Single and multi-GPUs.	Open Eyes Scientific Software http://www.eyesopen.com/fastrocs
PyMol	Lines: 460% increase Cartoons: 1246% increase Surface: 1746% increase Spheres: 753% increase Ribbon: 426% increase	1700x	Released, Version 1.5 Single GPUs	http://pymol.org/
VMD	High quality rendering, large structures (100 million atoms), analysis and visualization tasks, multiple GPU support for display of molecular orbitals	100-125X or greater on kernels	Released, Version 1.9	Visualization from University of Illinois at Urbana-Champaign http://www.ks.uiuc.edu/Research/vmd/

GPU Perf compared against Multi-core x86 CPU socket.
GPU Perf benchmarked on GPU supported features and may be a kernel to kernel perf comparison

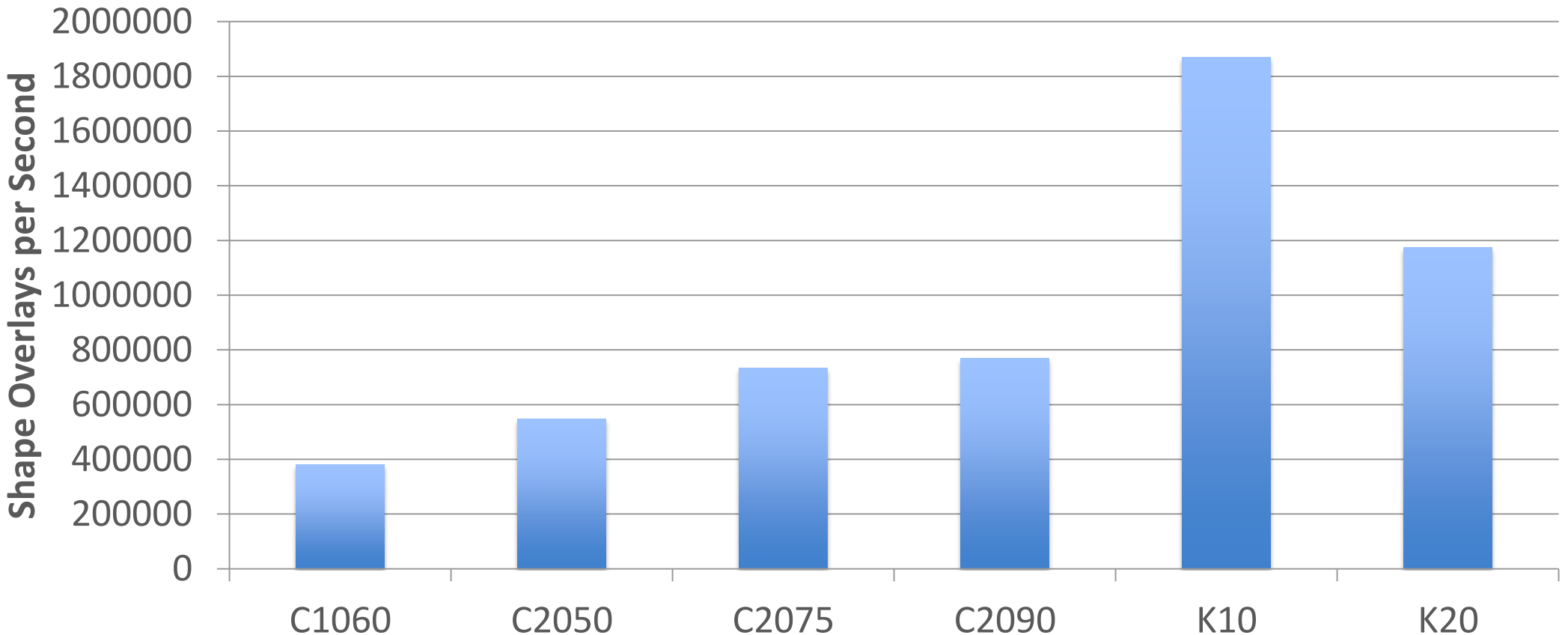
FastROCS

OpenEye Japan
Hideyuki Sato, Ph.D.

ROCS on the GPU: FastROCS

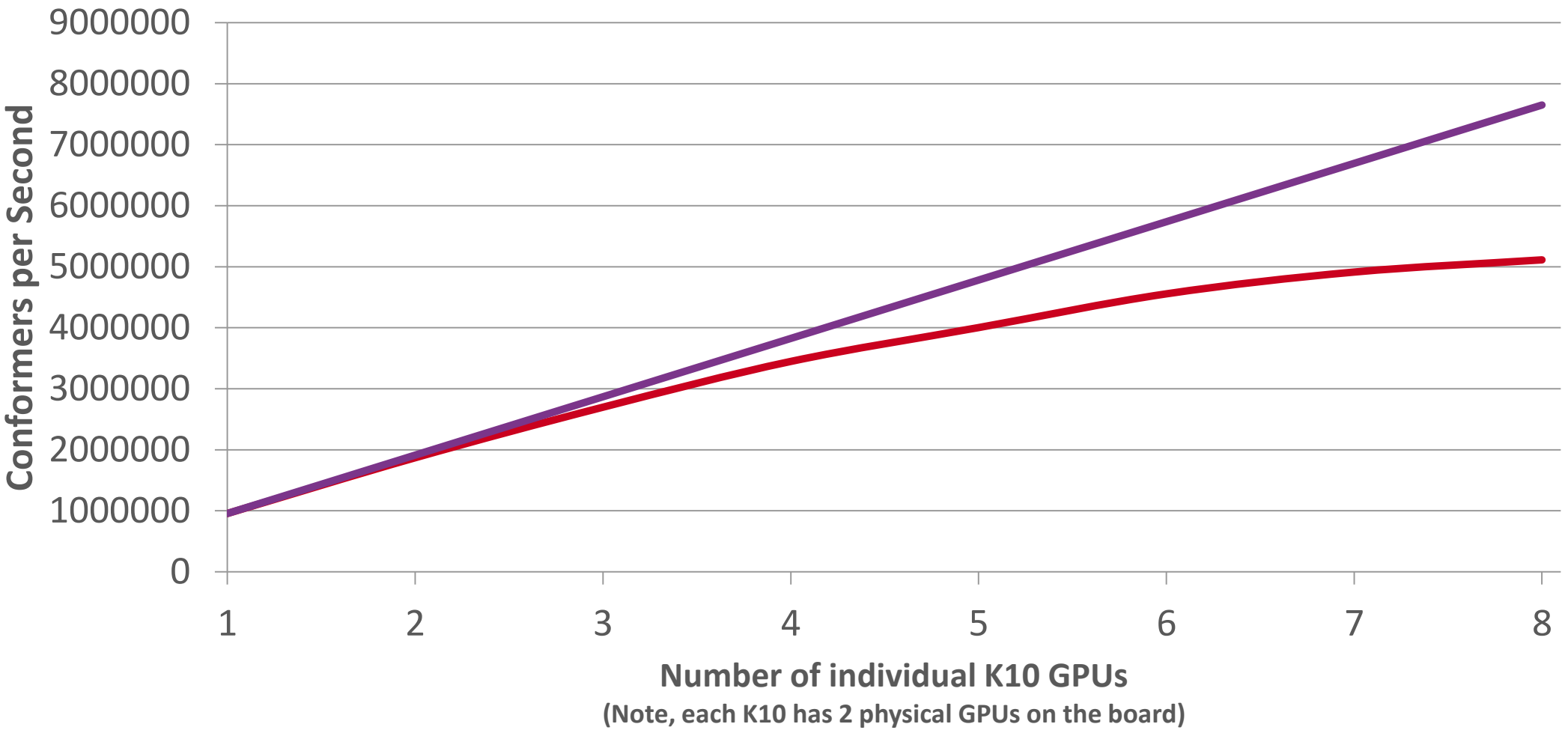


Riding Moore's Law



FastROCS scaling across 4x K10s (2 physical GPUs per K10)

53 million conformers (10.9 compounds of PubChem at 5 conformers per molecule)



OpenEye Scientific Software

Benefits of GPU Accelerated Computing

- Faster than CPU only systems in all tests
- Large performance boost with marginal price increase
- Energy usage cut by more than half
- GPUs scale well within a node and over multiple nodes
- K20 GPU is our fastest and lowest power high performance GPU yet

Try GPU accelerated TeraChem for free – www.nvidia.com/GPUTestDrive

GPU Test Drive
Experience GPU Acceleration

- For Computational Chemistry Researchers, Biophysicists
- Preconfigured with Molecular Dynamics Apps
- Remotely Hosted GPU Servers
- Free & Easy – Sign up, Log in and See Results

www.nvidia.com/gputestdrive