NVIDIA

LIFE AND MATERIAL SCIENCES

Mark Berger; mberger@nvidia.com


NVIDIA
Founded 1993
Invented GPU 1999 – Computer Graphics
Visual Computing, Supercomputing,
Cloud & Mobile Computing

NVIDIA - Core Technologies and Brands

GPU

GeForce®
Quadro® , Tesla®

Mobile

Tegra®

Cloud

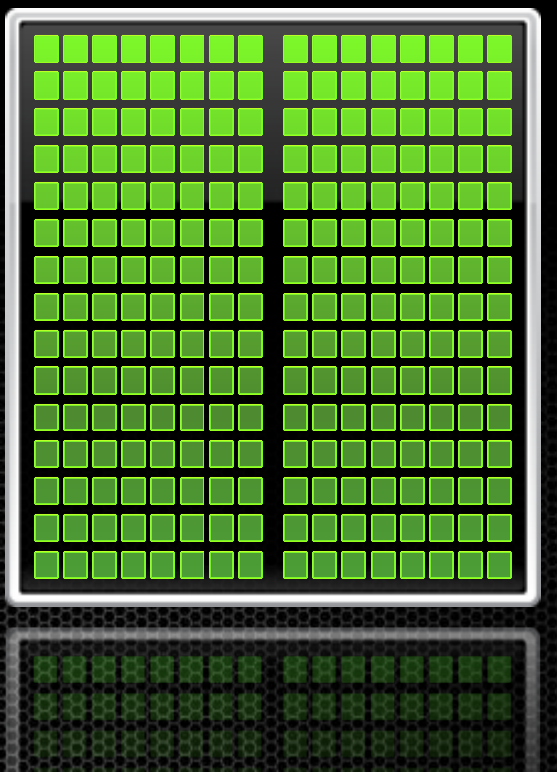
GRID

Accelerated Computing

Multi-core plus Many-cores

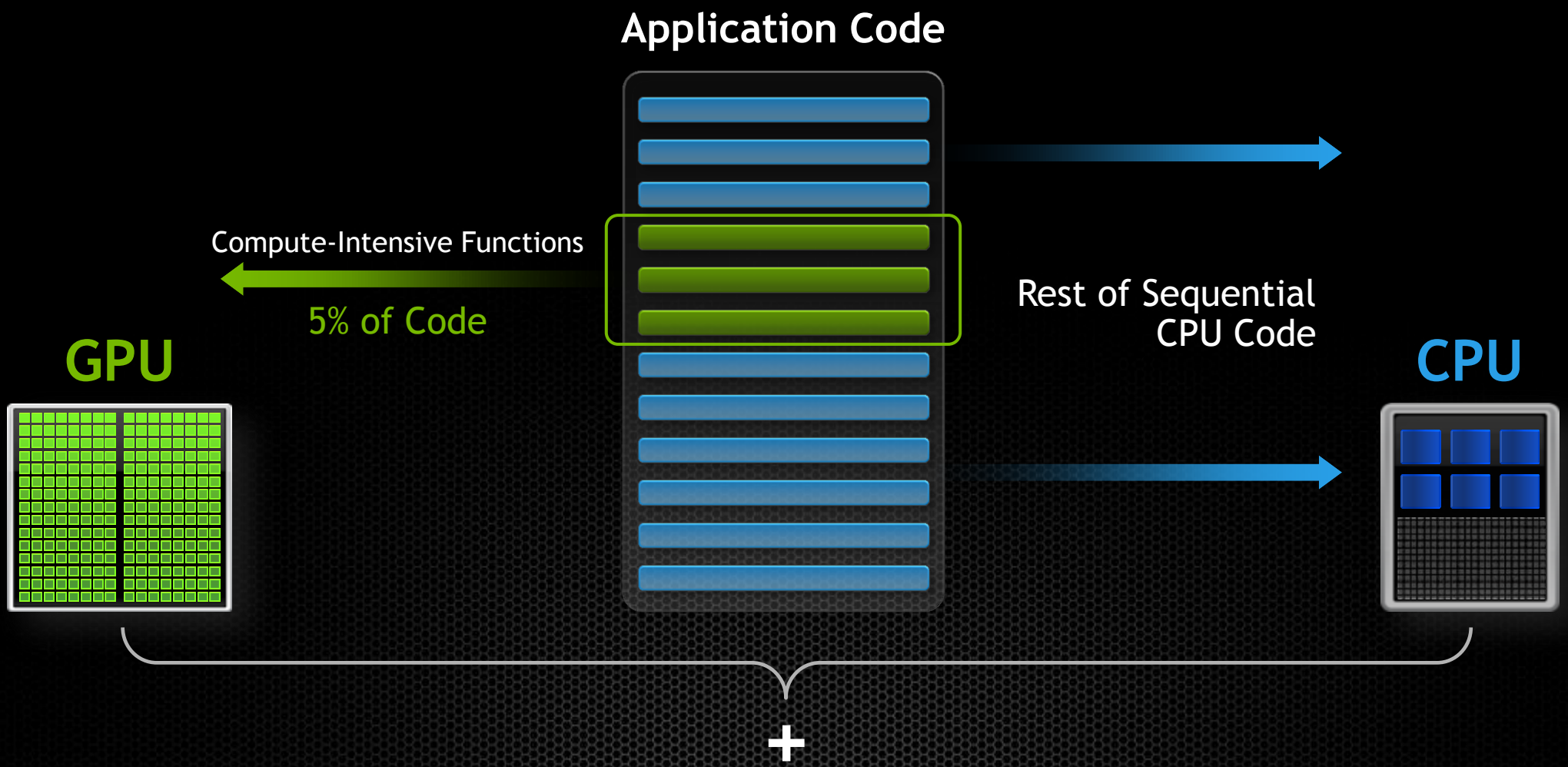
CPU
Optimized for
Serial Tasks

GPU Accelerator
Optimized for Many
Parallel Tasks

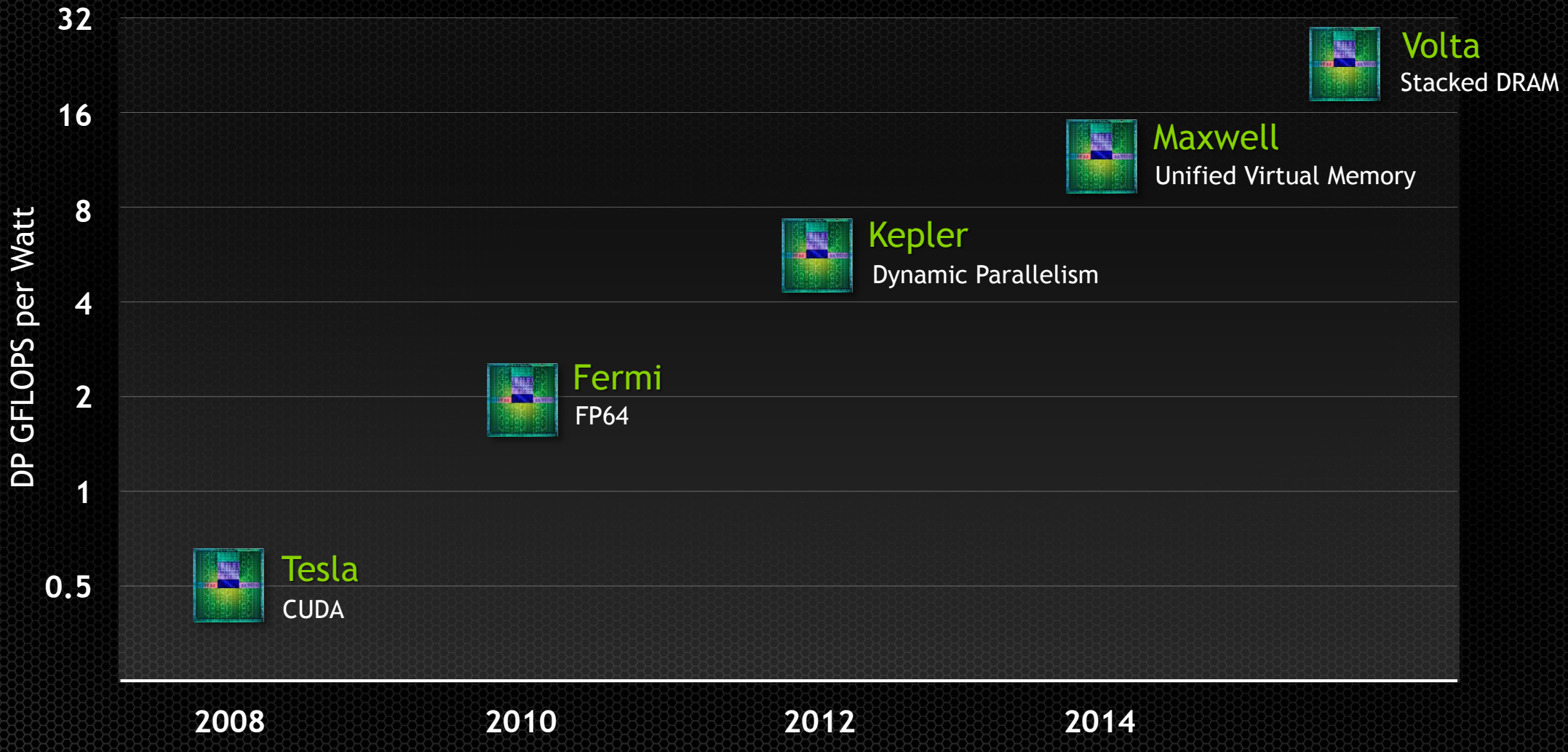


3-10X+ Comp Thruput
7X Memory Bandwidth
5x Energy Efficiency

How GPU Acceleration Works



GPUs : Two Year Heart Beat



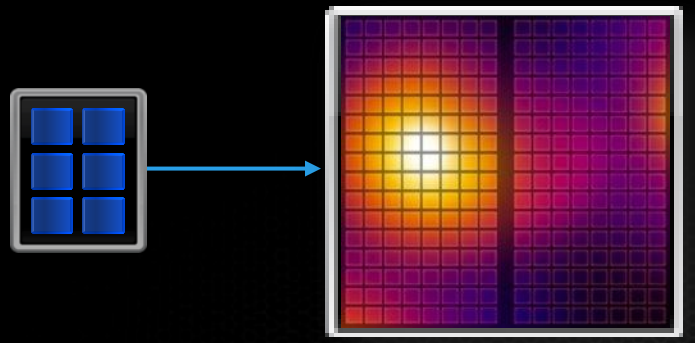
Kepler Features Make GPU Coding Easier

Hyper-Q

Speedup Legacy MPI Apps

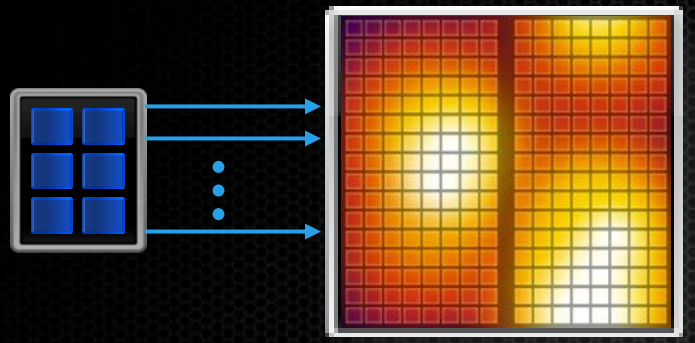
FERMI

1 Work Queue



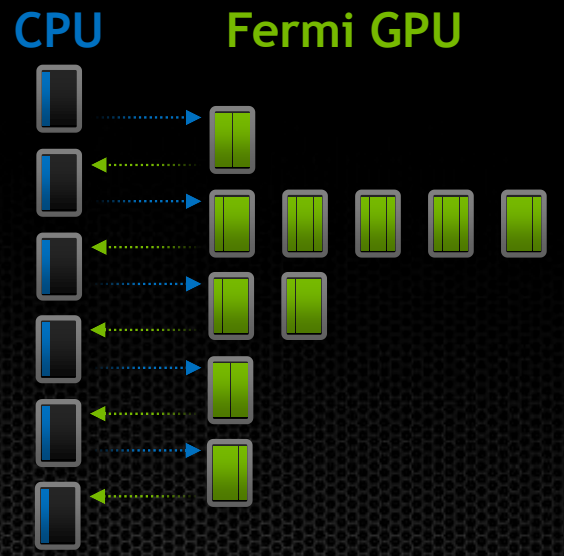
KEPLER

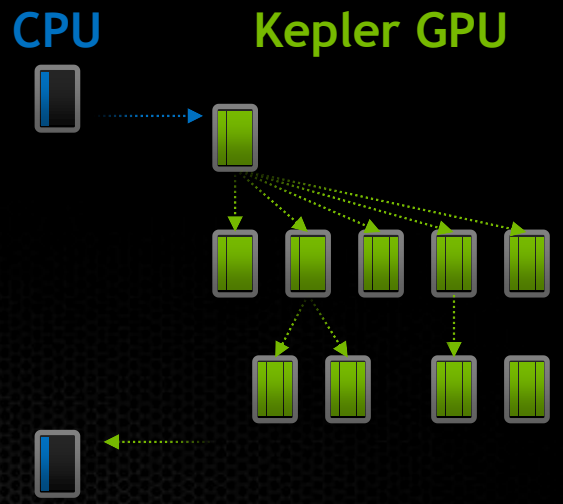
32 Concurrent Work Queues



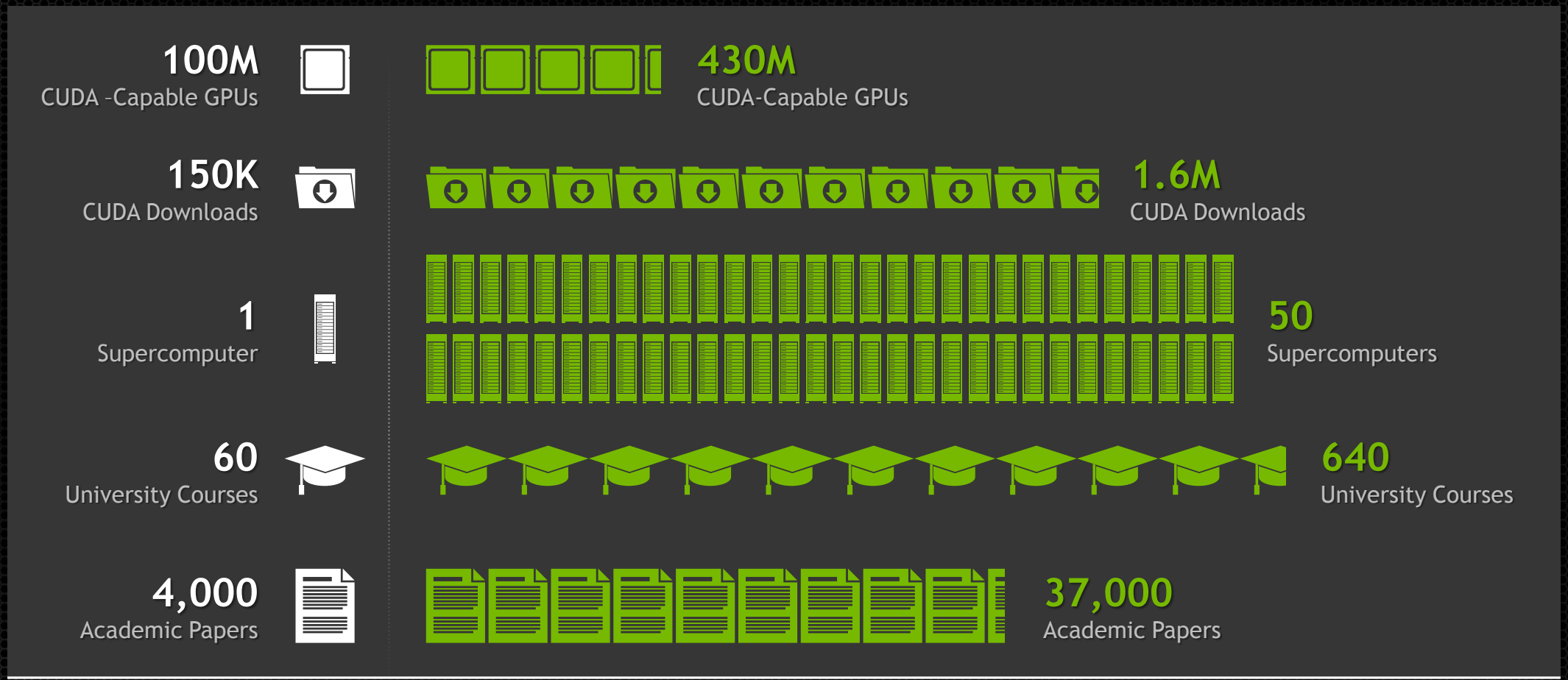
Dynamic Parallelism

Less Back-Forth, Simpler Code

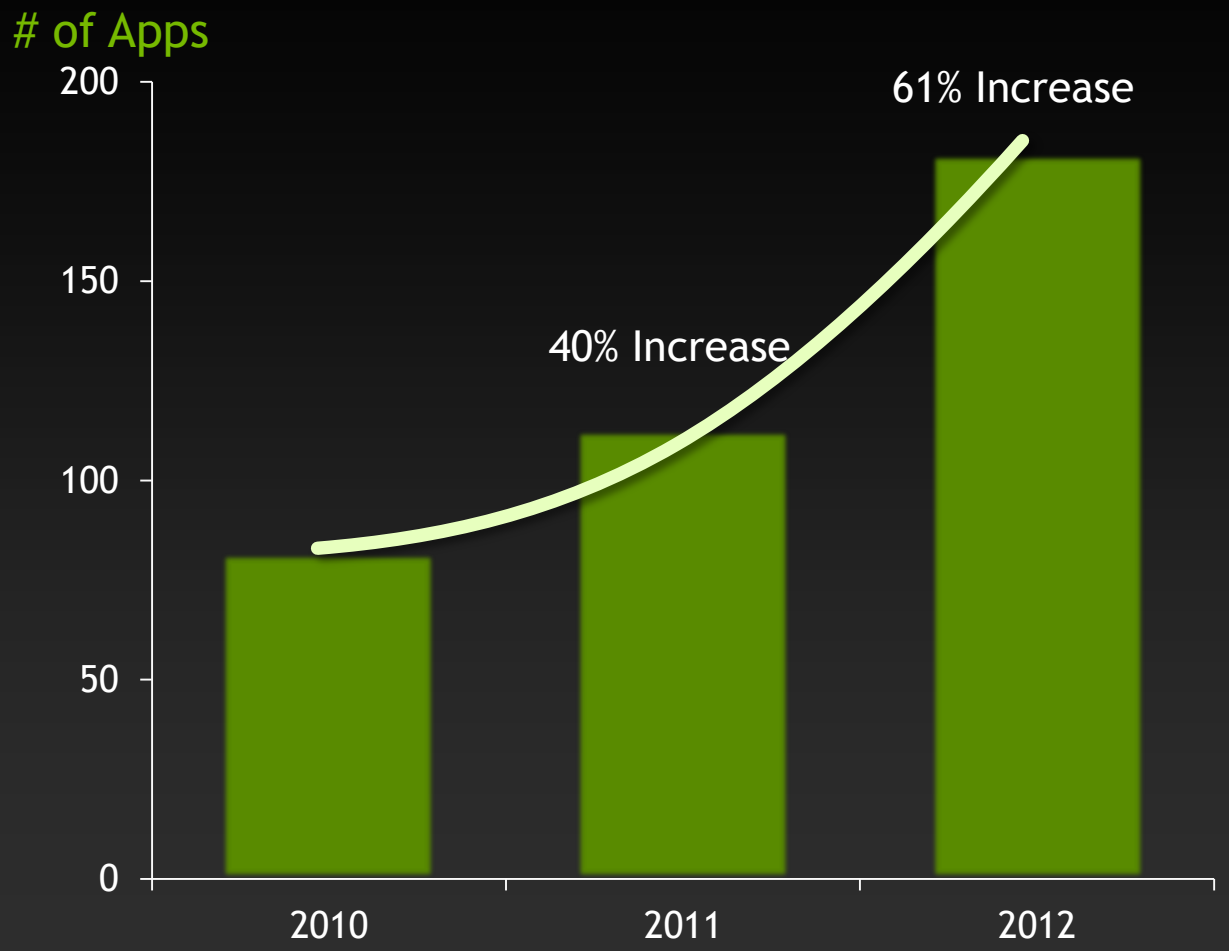




Developer Momentum Continues to Grow



Explosive Growth of GPU Accelerated Apps



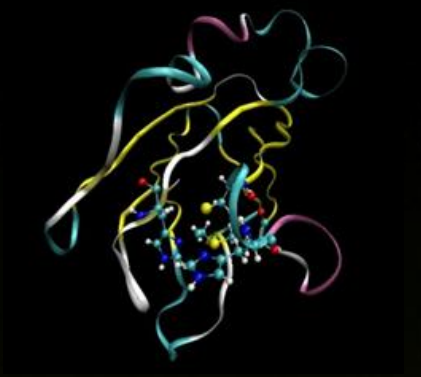
Top Scientific Apps

Molecular Dynamics	AMBER CHARMM GROMACS	LAMMPS NAMD DL_POLY
Quantum Chemistry	QMCPACK Quantum Espresso GAMESS-US	Gaussian NWChem VASP
Climate & Weather	COSMO GEOS-5	CAM-SE NIM WRF
Physics	Chroma Denovo GTC	GTS ENZO MILC
CAE	ANSYS Mechanical MSC Nastran SIMULIA Abaqus	ANSYS Fluent OpenFOAM LS-DYNA

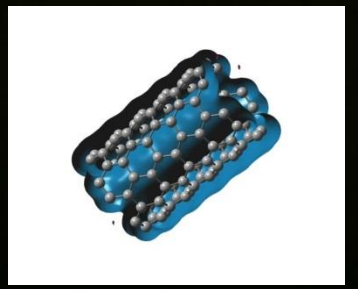
Accelerated, In Development

NVIDIA GPU Life Science Focus

- Molecular Dynamics: All codes are available
  - AMBER, CHARMM, DESMOND, DL_POLY, GROMACS, LAMMPS, NAMD
  - Great multi-GPU performance
  - GPU codes: ACEMD, HOOMD-Blue
  - Focus: scaling to large numbers of GPUs



- Quantum Chemistry: key codes ported or optimizing
  - Active GPU acceleration projects:
    - VASP, NWChem, Gaussian, GAMESS, ABINIT, Quantum Espresso, BigDFT, CP2K, GPAW, etc.
  - GPU code: TeraChem



- Analytical and Medical Imaging Instruments
- Genomics/Bioinformatics

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 4.04 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 4.6.2; 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/day 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

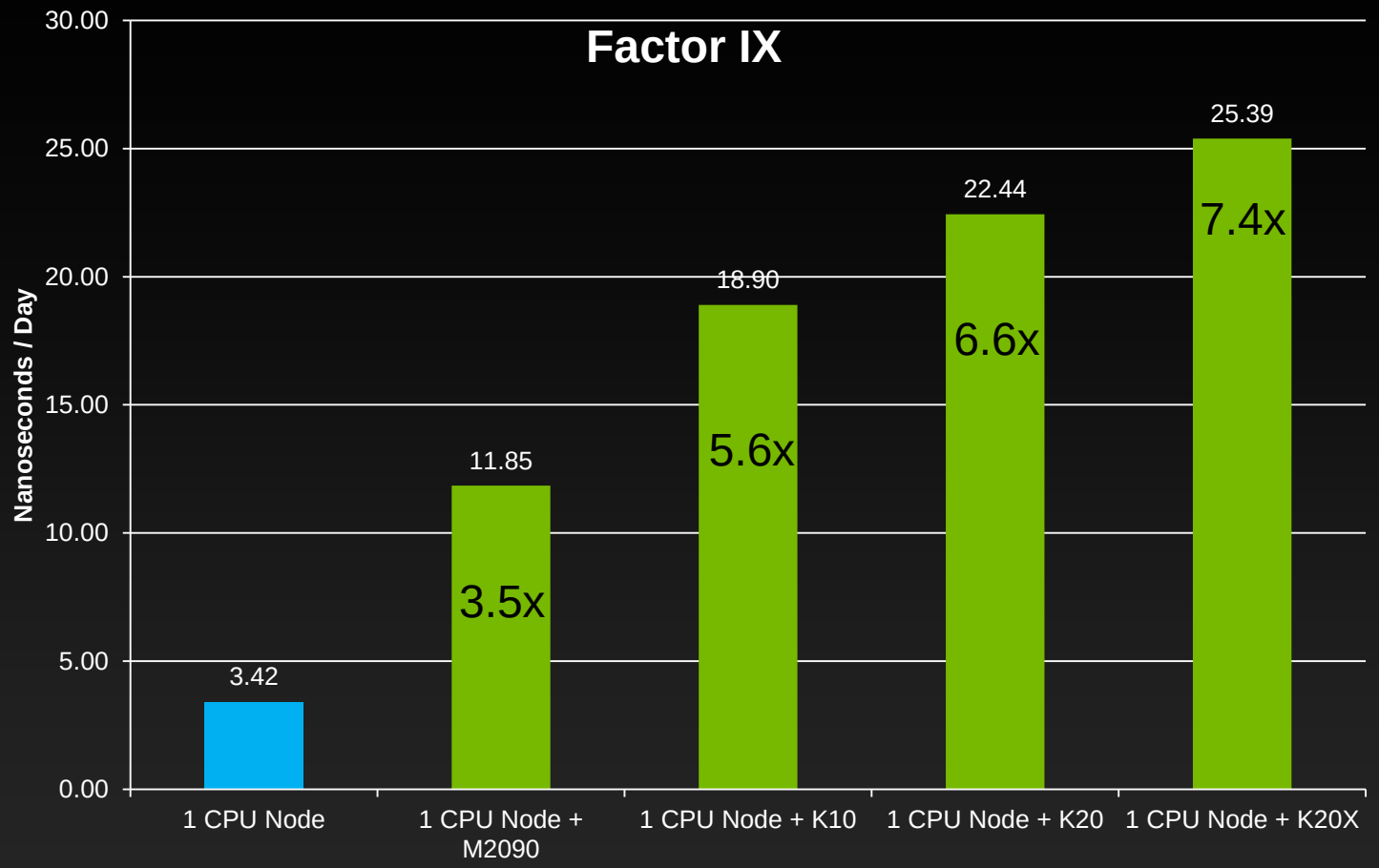
New/Additional MD Applications Ramping

Application	Features Supported	GPU Perf	Release Status	Notes
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
DESMOND	TBD	TBD	Under development	http://www.deshawresearch.com/resources_whatdesmonddoes.html
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 0.11.3 Single and Multi-GPU on 1 node	http://codeblue.umich.edu/hoomd-blue/ Multi-GPU w/ MPI in September 2013
mdcore	TBD	TBD	Released, Version 0.1.7	http://mdcore.sourceforge.net/download.html
OpenMM	Implicit and explicit solvent, custom forces	Implicit: 127-213 ns/day Explicit: 18-55 ns/day DHFR	Released, Version 5.1 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

AMBER 12
GPU Support Revision 12.2
1/22/2013

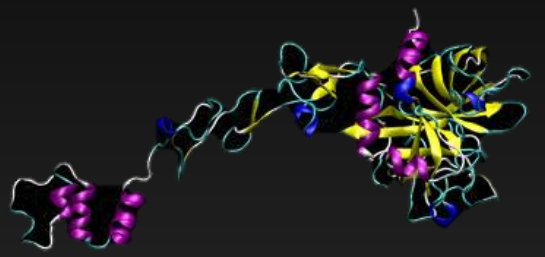
Kepler - Our Fastest Family of GPUs Yet



Running AMBER 12 GPU Support Revision 12.1

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

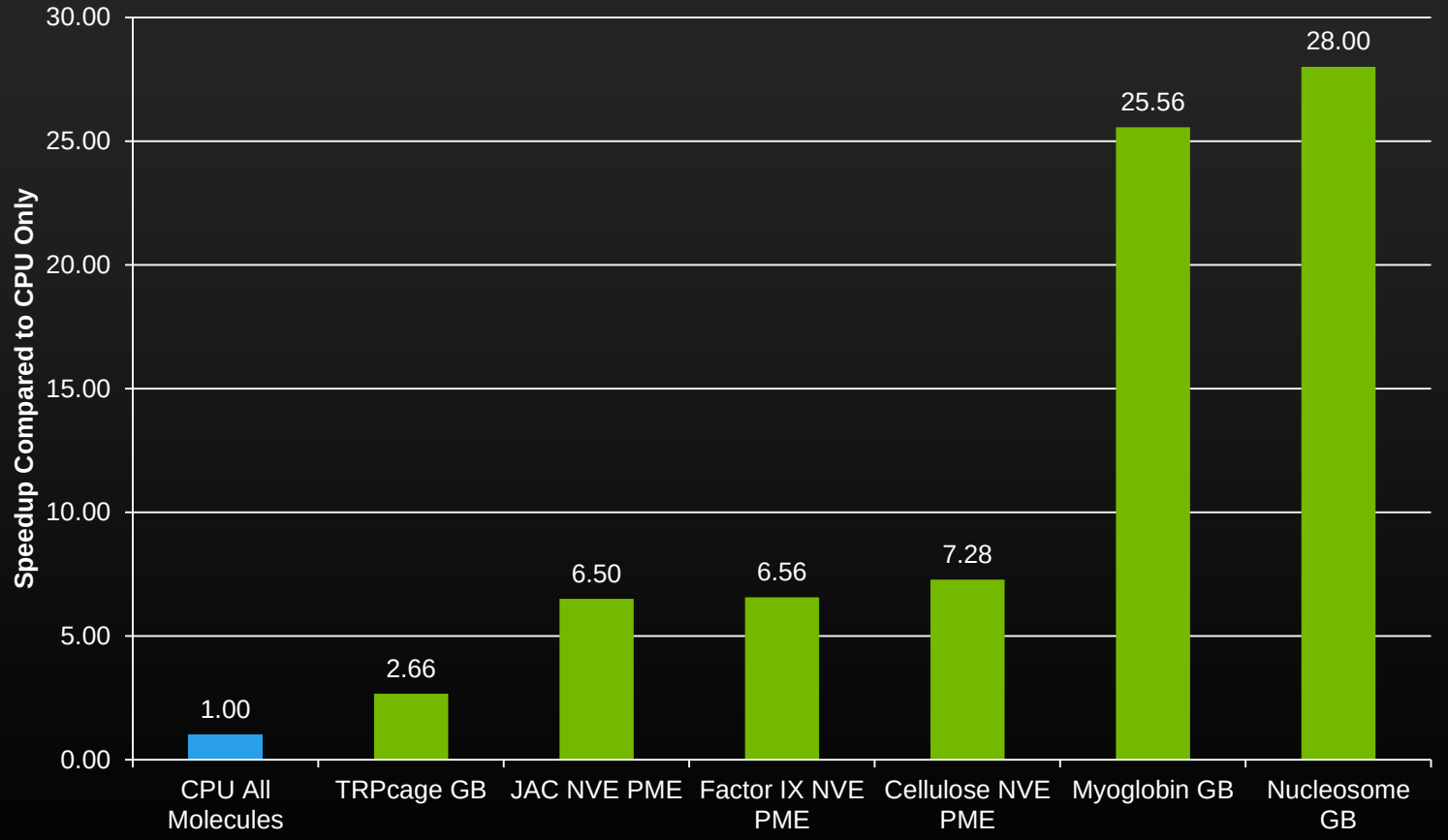
The green nodes contain Dual E5-2687W CPUs (8 Cores per CPU) and either 1x NVIDIA M2090, 1x K10 or 1x K20 for the GPU



Factor IX

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

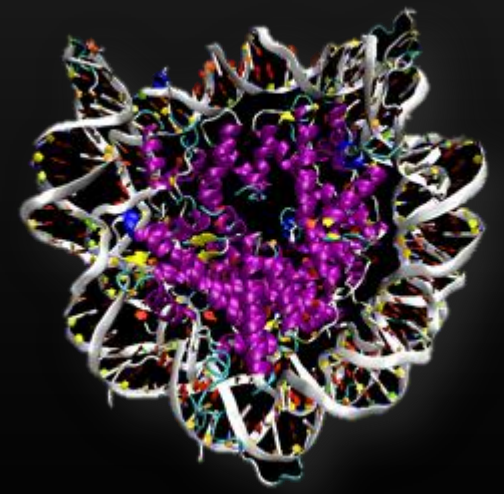
K20 Accelerates Simulations of All Sizes



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

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 7.2.2 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, Version 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	Code for performing quantum Monte Carlo (QMC) electronic structure calculations for finite and periodic systems	TBD	Under development Multi-GPU support	http://www.tcm.phy.cam.ac.uk/~mdt26/casino.html
CASTEP	TBD	TBD	Under development	http://www.castep.org/Main/HomePage
CP2K	DBCSR (spare matrix multiply	2.7X	Under development	http://www.olcf.ornl.gov/wp-

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	Released, Version 7.0 Multi-GPU support	http://www.ncbi.nlm.nih.gov/pubmed/21541963
Gaussian	Joint PGI, NVIDIA & Gaussian Collaboration	TBD	Under development Multi-GPU support	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 http://www.schrodinger.com/productpage/14/7/32/
LATTE	CU_BLAS, SP2 algorithm	TBD	Released Multi-GPU support	http://on-demand.gputechconf.com/gtc/2013/presentations/S3195-Fast-Quantum-Molecular-Dynamics-in-LATTE.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-	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
MOPAC2012	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	Released, Version 6.3 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	Full GPU support for ground-state, real-time calculations; Kohn-Sham Hamiltonian, orthogonalization, subspace diagonalization, poisson solver, time propagation	1.5-8X	Released, Version 4.1.0	http://www.tddft.org/programs/octopus/
ONETEP	TBD	TBD	Under development	http://www2.tcm.phy.cam.ac.uk/onetep/
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 4.0.1 Multi-GPU support	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 5.5 Single GPU	Visualization from Visage Imaging. 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 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

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
Interactive Molecule Visualizer	Written for use only on GPUs	TBD	Under development Multi-GPU support	http://www.molecular-visualization.com
Molegro Virtual Docker 6	Energy grid computation, pose evaluation and guided differential evolution	25-30x	Single GPU	http://www.molegro.com/gpu.php
PIPER Protein Docking	Molecule docking	3-8x	Released	http://www.bu.edu/caadlab/gpgpu09.pdf; http://www.schrodinger.com/productpage/14/40/
PyMol	Lines: 460% increase Cartoons: 1246% increase Surface: 1746% increase Spheres: 753% increase Ribbon: 426% increase	1700x	Released, Version 1.6 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 1.9.1 Multi-GPU support	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

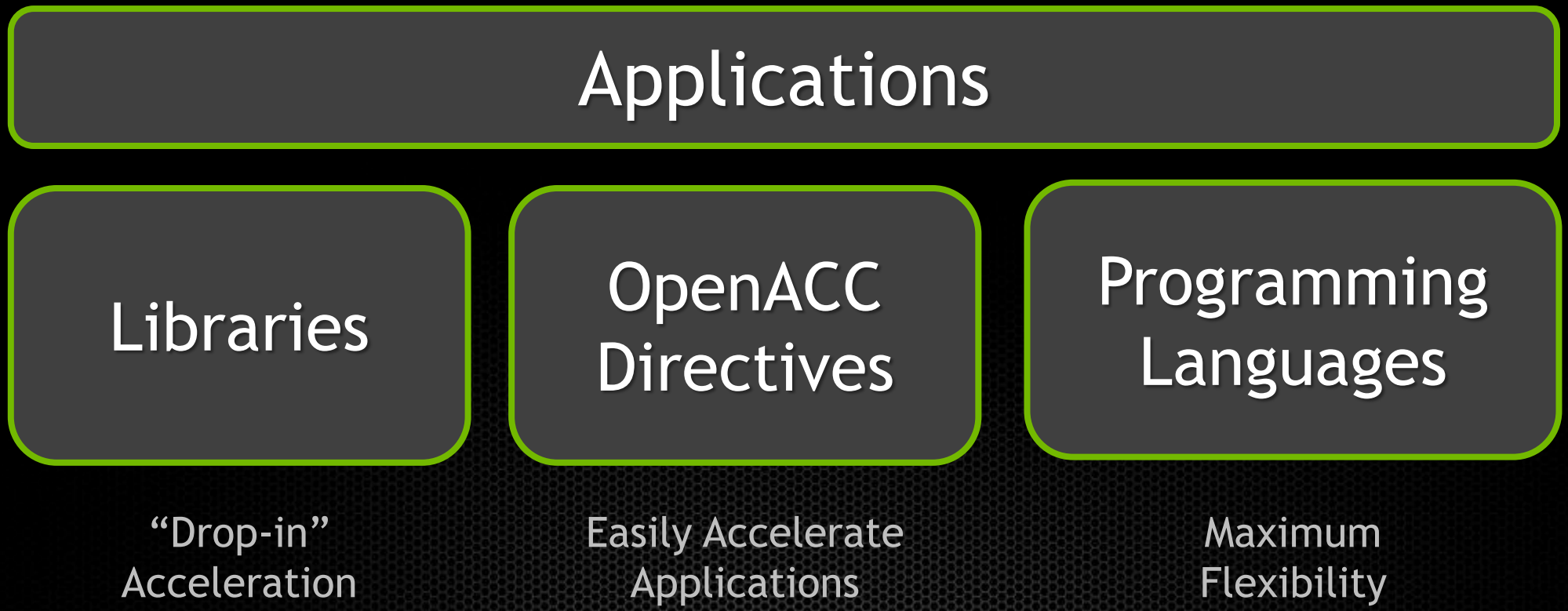
"Organic Growth" of Bioinformatics Applications

Application	Features Supported	GPU Speedup	Release Status	Website
BarraCUDA	Alignment of short sequencing reads	6-10x	Released, Version 0.6.2 Multi-GPU, multi-node	http://seqbarracuda.sourceforge.net/
CUDASW++	Parallel search of Smith-Waterman database	10-50x	Released, Version 2.0.10 Multi-GPU, multi-node	http://sourceforge.net/projects/cudasw/
CUSHAW	Parallel, accurate long read aligner for large genomes	10x	Released, Version 1.0.40 Multiple-GPU	http://cushaw.sourceforge.net/
GPU-BLAST	Protein alignment according to BLASTP	3-4x	Released, Version 2.2.26 Single GPU	http://eudoxus.cheme.cmu.edu/gpublast/gpublast.html
mCUDA-MEME	Scalable motif discovery algorithm based on MEME	4-10x	Released, Version 3.0.13 Multi-GPU, multi-node	https://sites.google.com/site/yongchaosoftware/mcuda-meme
MUMmer GPU	Aligns multiple query sequences against reference sequence in parallel	3-10x	Released, Version 3.0	http://sourceforge.net/apps/mediawiki/mummergpu/index.php?title=MUMmerGPU
SeqNFind	Hardware and software for reference assembly, blast, SW, HMM, de novo assembly	400x	Released Multi-GPU, multi-node	http://www.seqnfind.com/
SOAP3	Short read alignment tool that is not heuristic based; reports all answers	10x	Beta release, Version 0.01	http://soap.genomics.org.cn/soap3.html
UGENE	Fast short read alignment	6-8x	Released, Version 1.12	http://ugene.unipro.ru/

GPU Perf compared against same or similar code running on single CPU machine

Performance measured internally or independently

3 Ways to Accelerate Applications



GPU Accelerated Libraries

"Drop-in" Acceleration for your Applications

Linear Algebra
FFT, BLAS,
SPARSE, Matrix

CULA | tools

MAGMA
ICL UT

CUSP

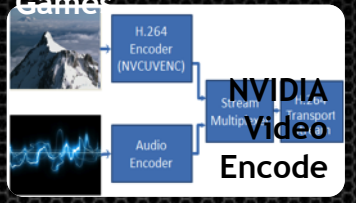
Numerical & Math
RAND, Statistics

ArrayFire

Data Struct. & AI
Sort, Scan, Zero Sum

Thrust

Visual Processing
Image & Video



Sundog™
Software

GPU Programming Languages

Numerical analytics	R, MATLAB, Mathematica, LabVIEW
Fortran	OpenACC, CUDA Fortran
C	OpenACC, CUDA C
C++	Thrust, CUDA C++
Python	PyCUDA, Anaconda Accelerate
C#	GPU.NET

Easiest Way to Learn CUDA

50K
Registered

127
Countries

Learn from the Best

Anywhere, Any Time

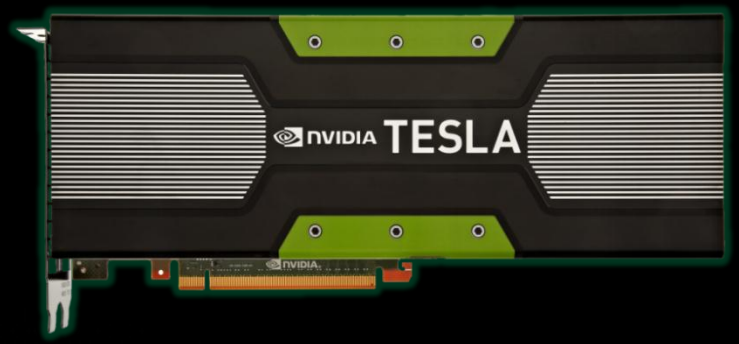
It's Free!

Engage with an Active Community

Develop on GeForce, Deploy on Tesla

GeForce GTX Titan

Tesla K20X/K20



Designed for Gamers & Developers

1+ Teraflop Double Precision Performance
Dynamic Parallelism
Hyper-Q for CUDA Streams

Available Everywhere!

Develop on any GPU released in the last 3 years;
Macs, PCs, Workstations, Clusters, Supercomputers

Designed for Cluster Deployment

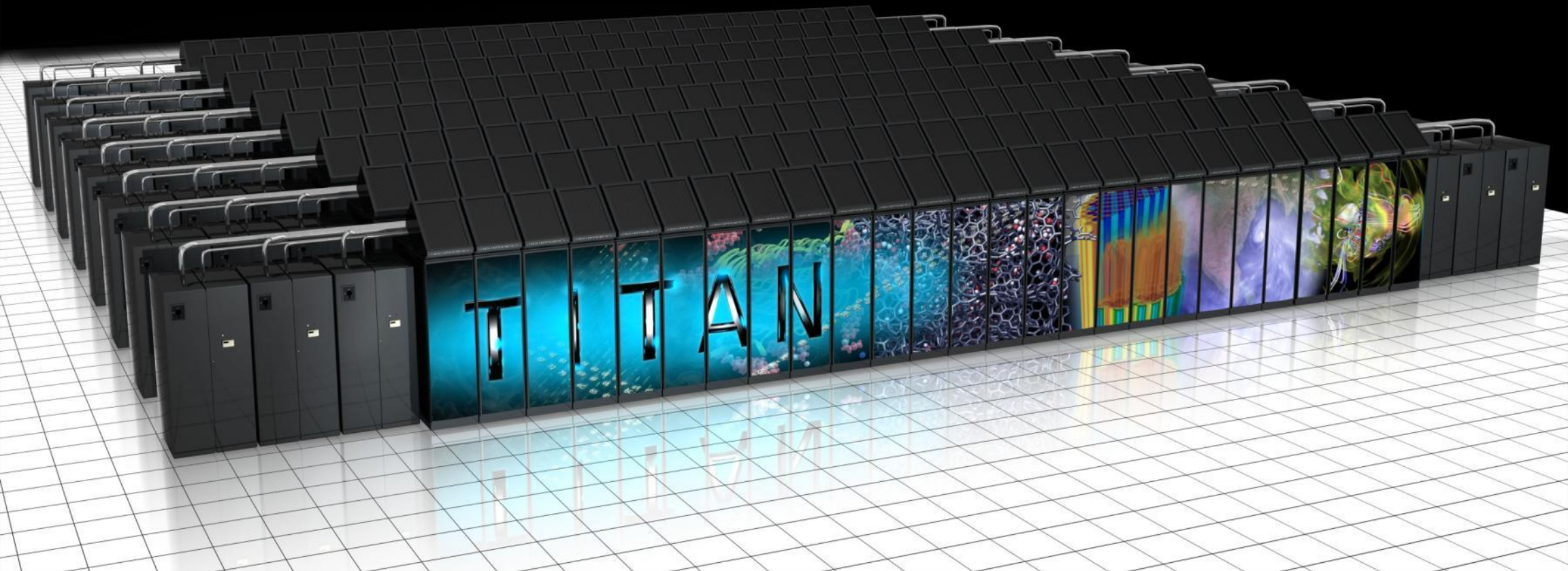
ECC
24x7 Runtime
GPU Monitoring
Cluster Management
GPUDirect-RDMA
Hyper-Q for MPI
3 Year Warranty
Integrated OEM Systems, Professional Support

TITAN: World's Fastest Supercomputer

18,688 Tesla K20X GPUs

27 Petaflops Peak, 17.59 Petaflops on Linpack

90% of Performance from GPUs



Test Drive K20 GPUs!

Experience The Acceleration

Run on Tesla K20 GPUs today

Sign up for FREE GPU Test Drive on remotely hosted clusters
www.nvidia.com/GPUTestDrive

NVIDIA

LIFE AND MATERIAL SCIENCES

Mark Berger; mberger@nvidia.com