



GPUs Revolutionize Graphics and
Parallel Computing;
Now, Let's go after Science!

David B. Kirk NVIDIA Fellow

NVIDIA Corporation
The Visual Computing Company





Fallou



CO

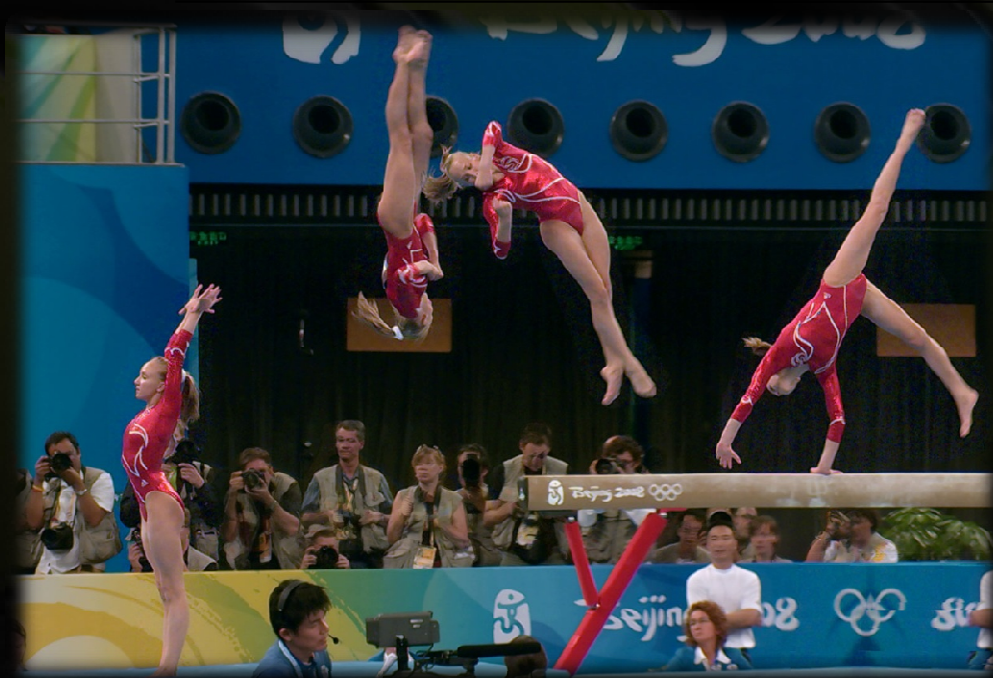
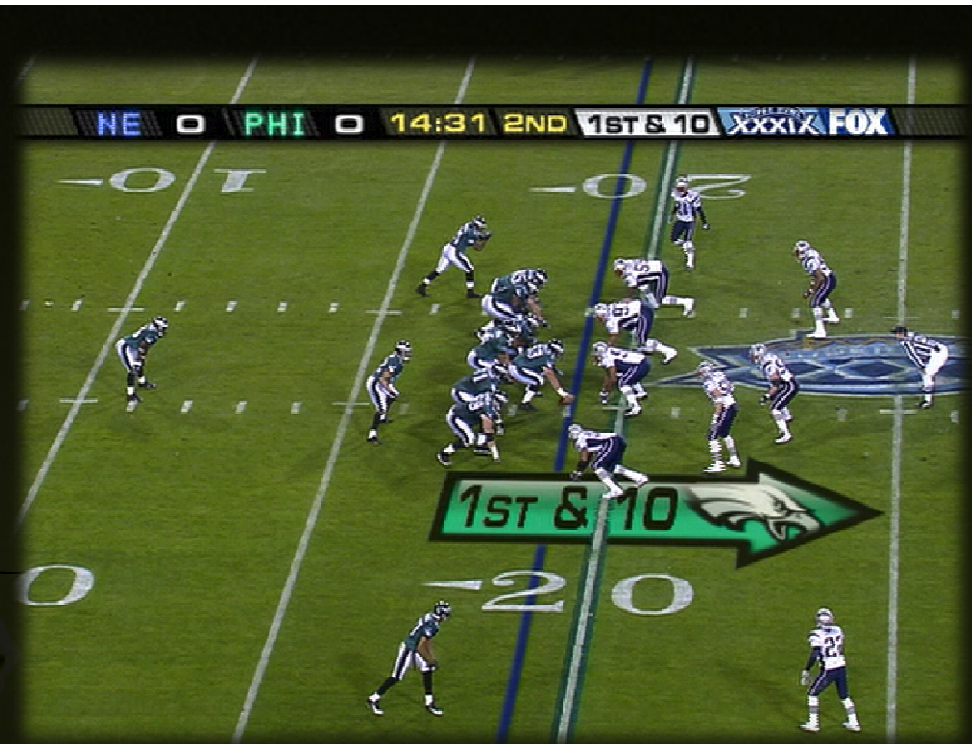
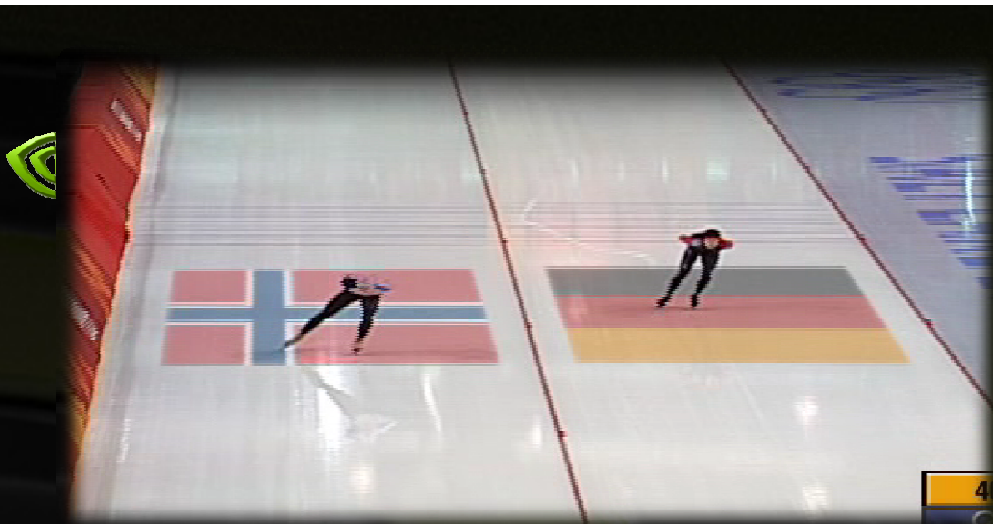


GRID

Need For Speed Undercover



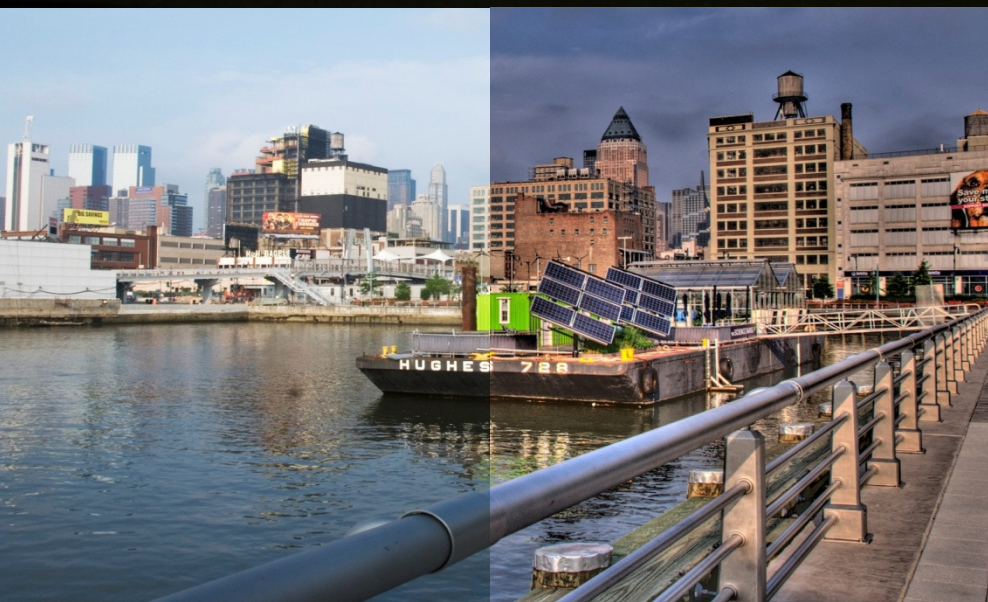




Nastia Liukin



NBC Virtual Set
Ann Curry



SD

HD

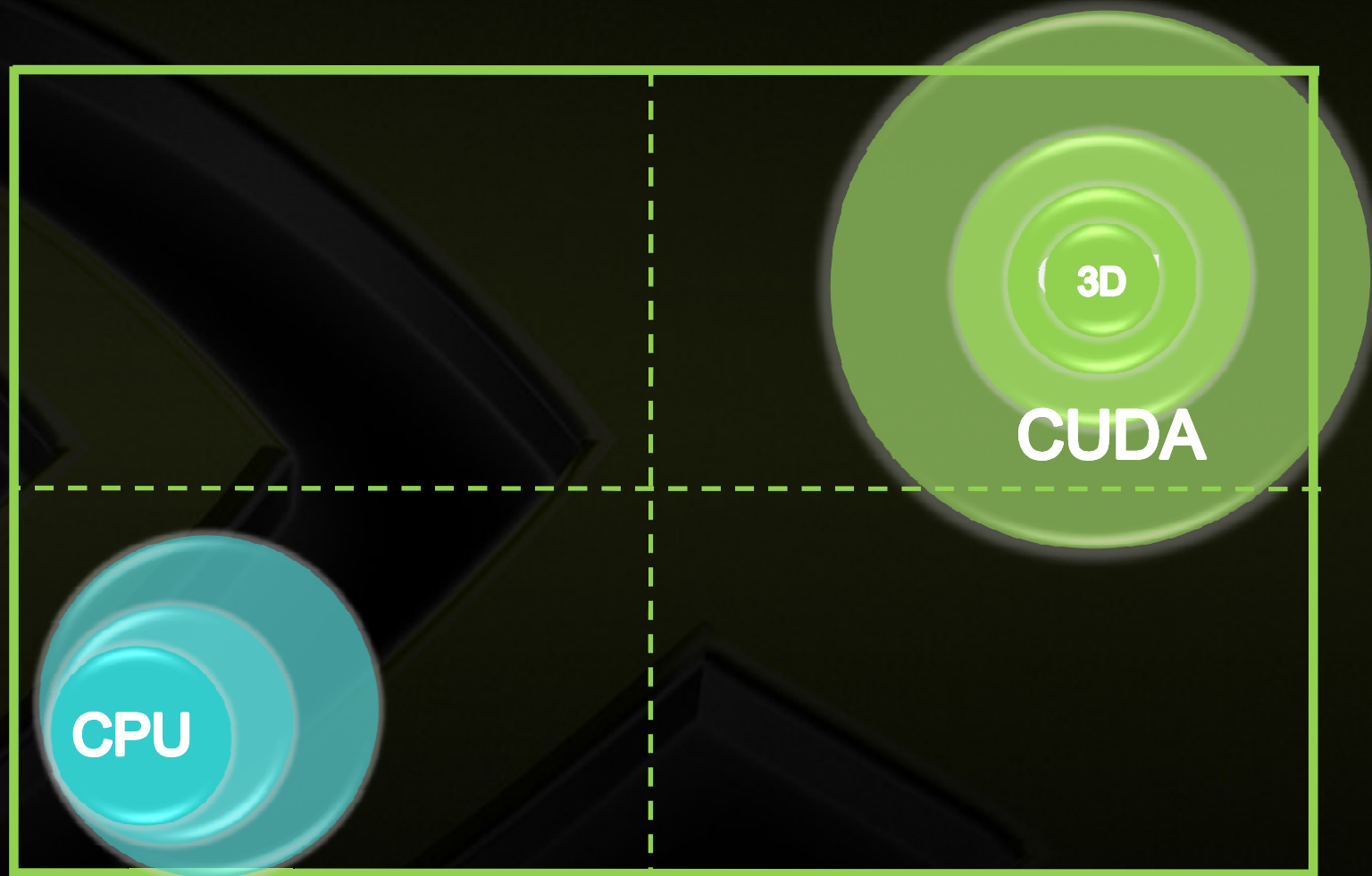


Evolution of Processors



**Massive
Data
Parallelism**

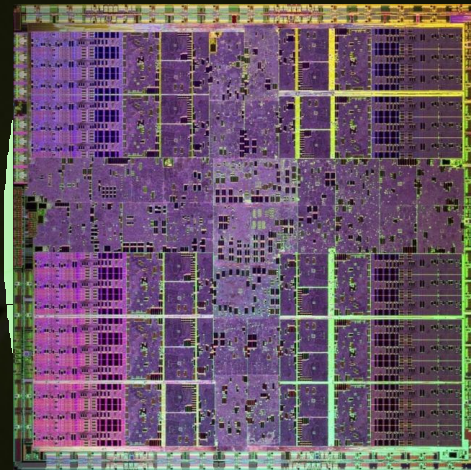
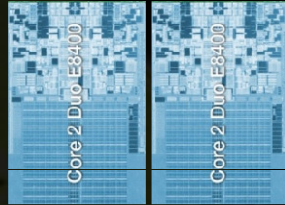
**Instruction
Level
Parallelism**



Data Fits in Cache

Huge Data Sets

CUDA GPU Enables Heterogeneous Parallel Computing



4 cores

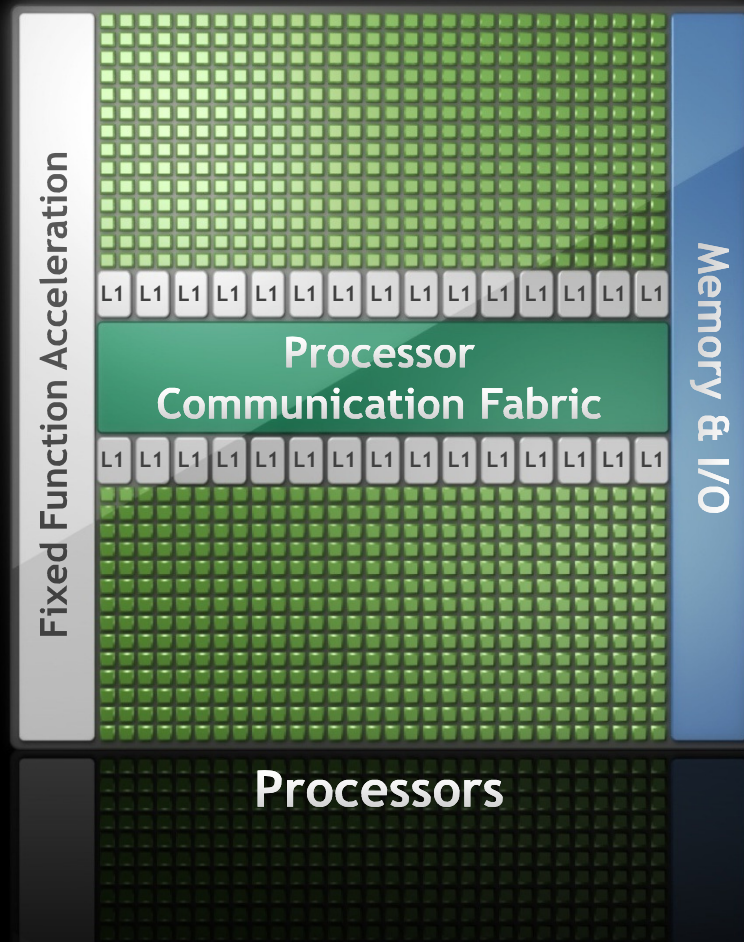
**Multi-Core
CPU**

240 cores

**Parallel-Core
GPU**



Processors



NVIDIA Tesla 10-Series GPU

Massively parallel, many core architecture

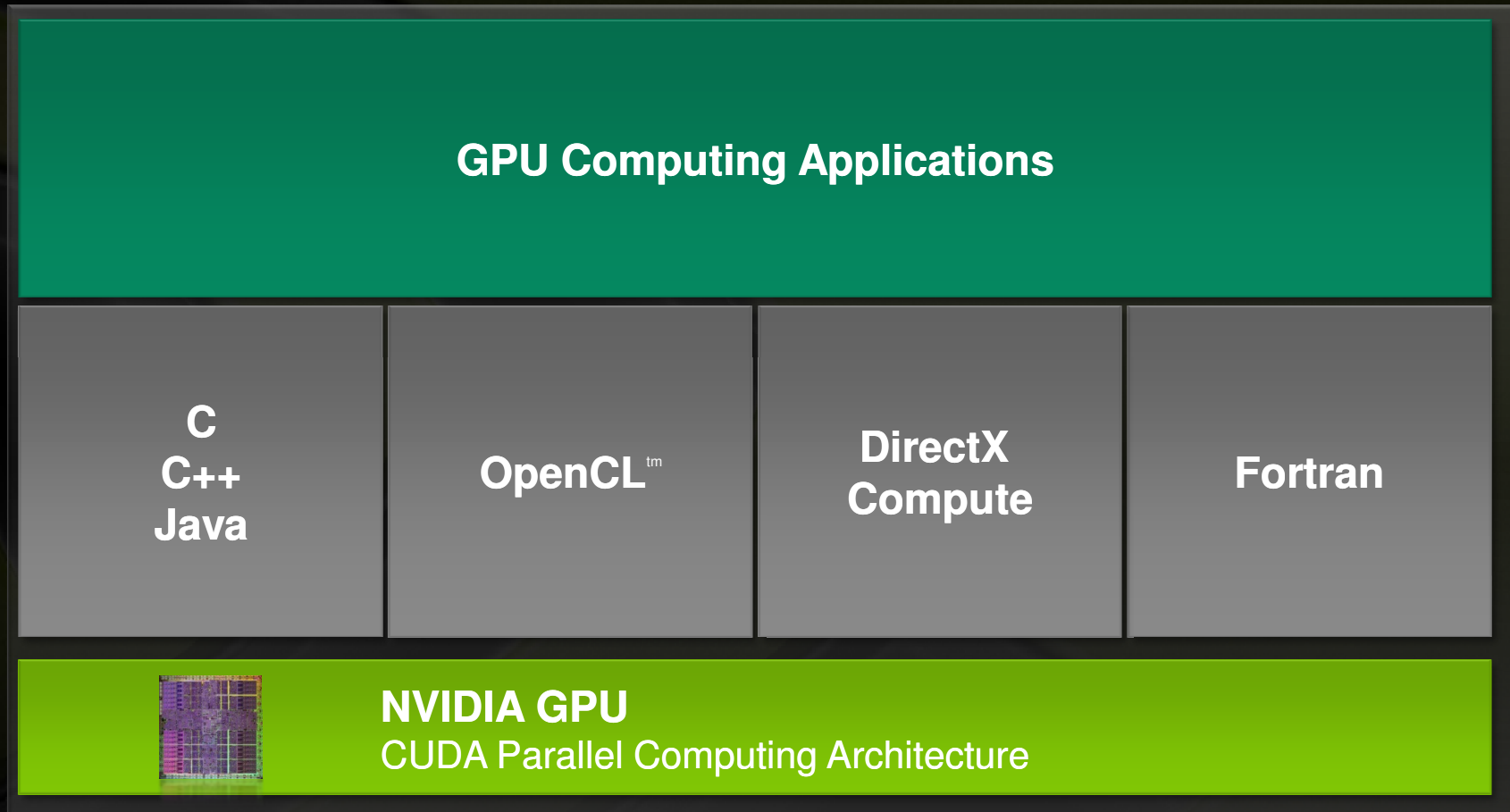
240 Processor Cores

1 Teraflops - 1,000 times Cray X-MP

IEEE Compliant Double Precision Floating Point

Designed for Scientific Computing

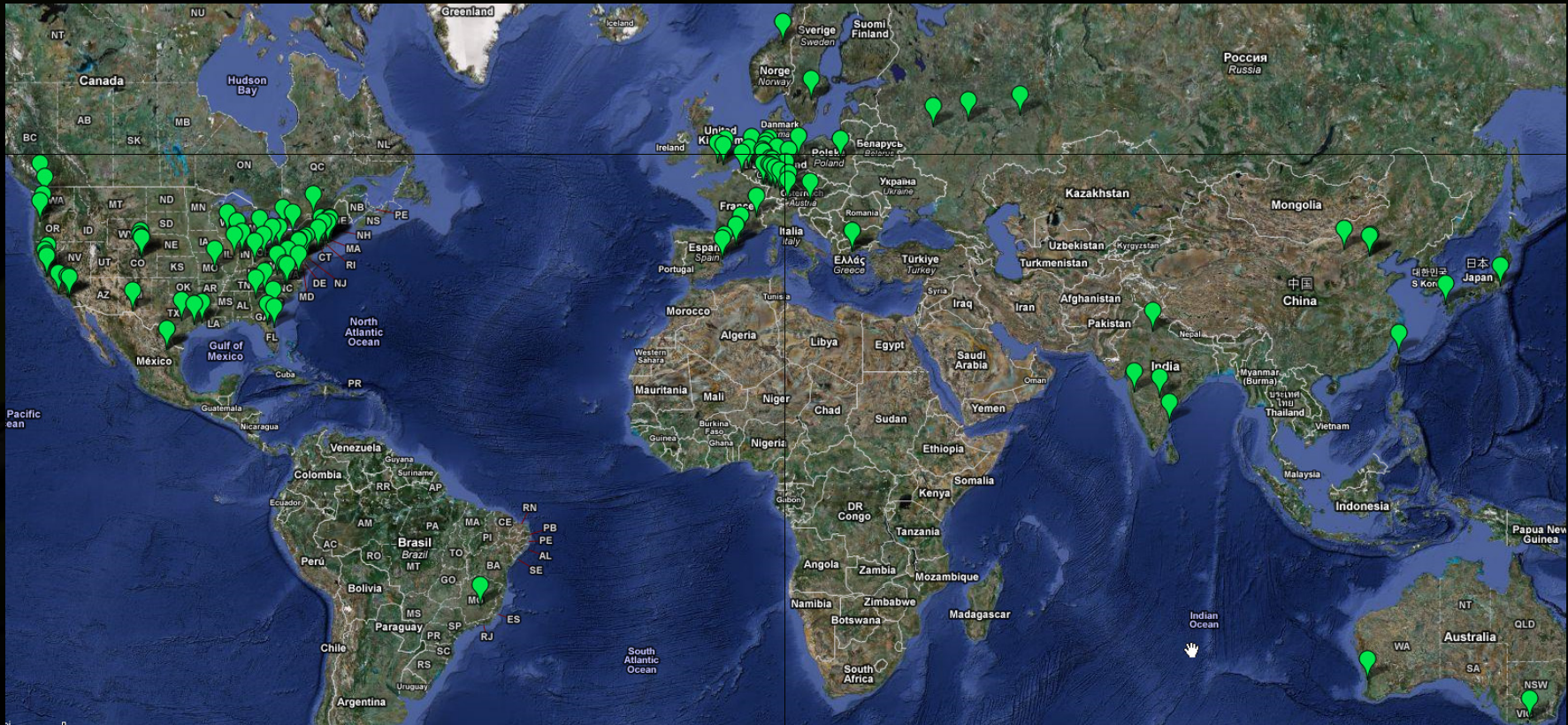
CUDA Parallel Computing Architecture



CUDA: Most Widely Adopted Parallel Programming Model



- 1000+ Research Papers
- 200+ universities teaching CUDA
- 120 Million CUDA GPUs
- 60,000+ Active Developers



CUDA Ecosystem

Over 200 Universities Teaching CUDA



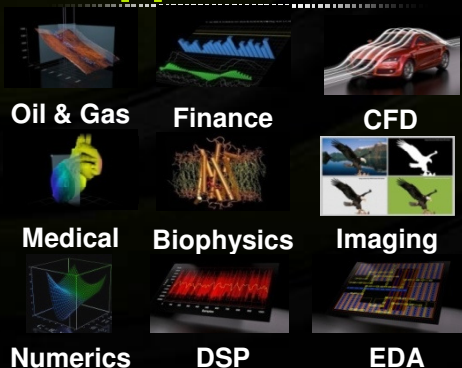
Languages

C, C++
DirectX
Fortran
Java
OpenCL
Python

Compilers

PGI Fortran
CAPs HMPP
MCUDA
MPI
NOAA Fortran2C
OpenMP

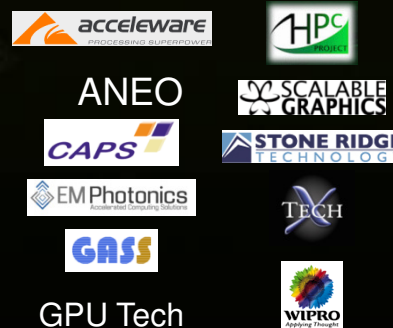
Applications



Libraries

FFT
BLAS
LAPACK
Image processing
Video processing
Signal processing
Vision

Consultants



OEMs

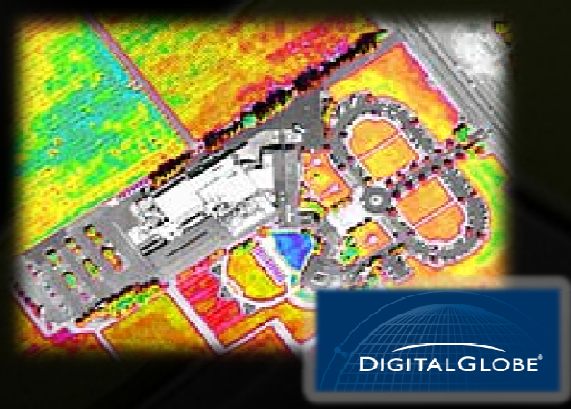
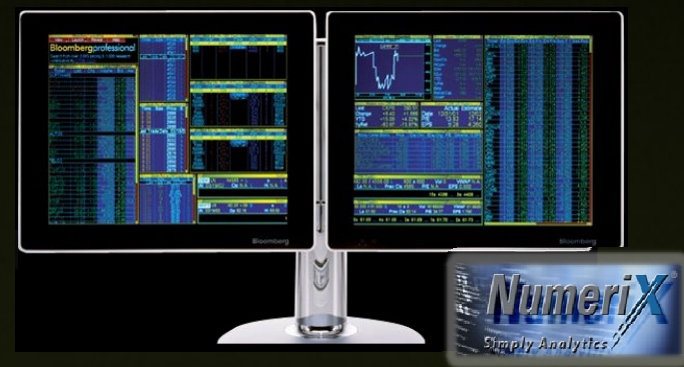
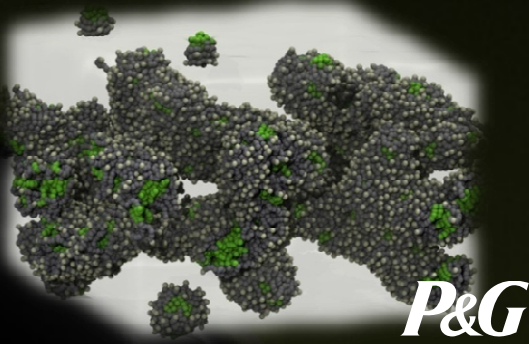
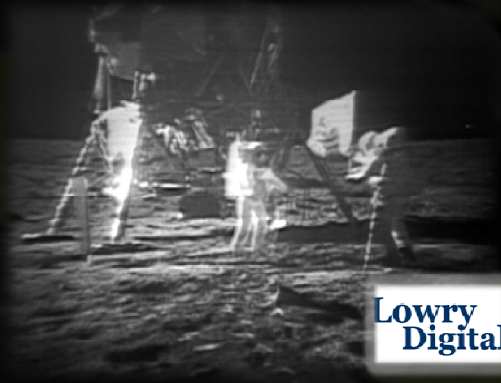


Released Applications

CUDA



Bio-Sciences	Bio-Informatics	Medical Imaging	Defense	Oil and Gas
GROMACS using OpenMM NAMD alpha VMD, 1.8.7 beta HOOMD	<ul style="list-style-type: none">• GPU HMMER• MUMmerGPU: Sequence Alignment• Accelereyes: MATLAB plugin	<ul style="list-style-type: none">• GPULib: IDL acceleration• Acceleware CT Recon• Digisens CT Recon• Accelereyes: MATLAB plugin	<ul style="list-style-type: none">• GPU VSIPL: Signal Processing• GPULib: IDL acceleration• Ikena: Imagery Analysis, Video Forensics• GIS: Manifold• Accelereyes: MATLAB plugin	<ul style="list-style-type: none">• Acceleware: Time Migration• SeismicCity: Prestack• Headwave: Prestack• OpenGeoSolutions: Spectral Decomp• Mercury: 3D viz• ffA: 3D Seismic proces• GIS: Manifold
EDA	Weather & Ocean Modeling	Finance	Electro-magnetics	
CST: 3D EM Agilent: ADS SPICE Synopsys: TCAD	<ul style="list-style-type: none">• WRF beta release• Particle simulation Boltzmann solver• Tsunami simulation: Tokyo Tech• NOAA new model being developed	<ul style="list-style-type: none">• Numerix: Counterparty• Scicomp: Derivative Pricing• Hanweck: Options Pricing• Exegy: Risk Analysis• Aqumin: 3D Viz	<ul style="list-style-type: none">• Acceleware: FDTD Solver• Quantum electrodynamics library• CST Microwave Studio• GPMAD : Particle beam dynamics simulator	



More Information

<http://www.nvidia.com/tesla>

Products

Vertical Solutions

CUDA GPU Programming Training

GPU Developer Conference

Sept 30 – Oct 2, 2009

San Jose, CA

<http://www.nvidia.com/gtc>

“ GPUs have evolved to the point where many real world applications are easily implemented on them and run significantly faster than on multi-core systems.

Future computing architectures will be hybrid systems with parallel-core GPUs working in tandem with multi-core CPUs.

”

Jack Dongarra

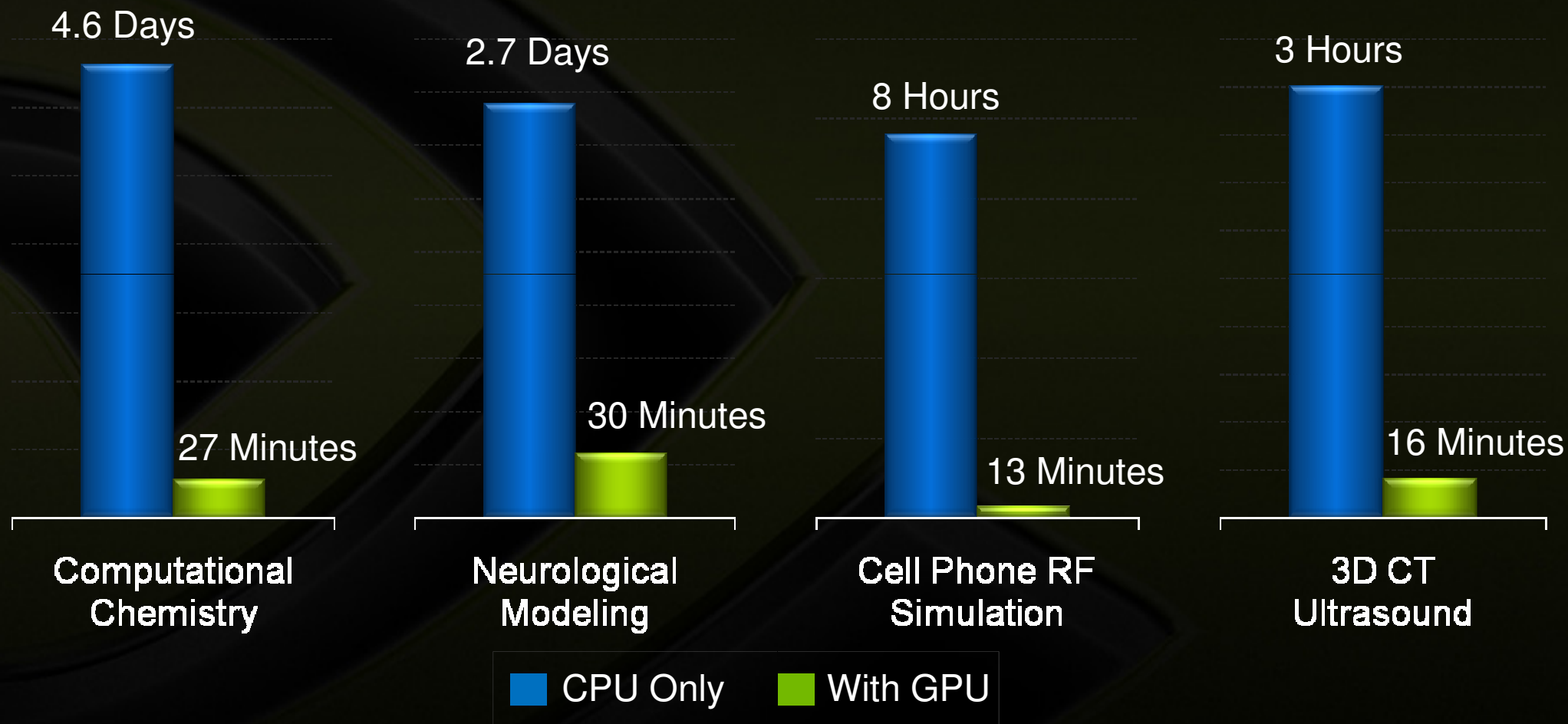
University of Tennessee

Developer of Linpack benchmark

Huge Speed-Ups from GPU Computing

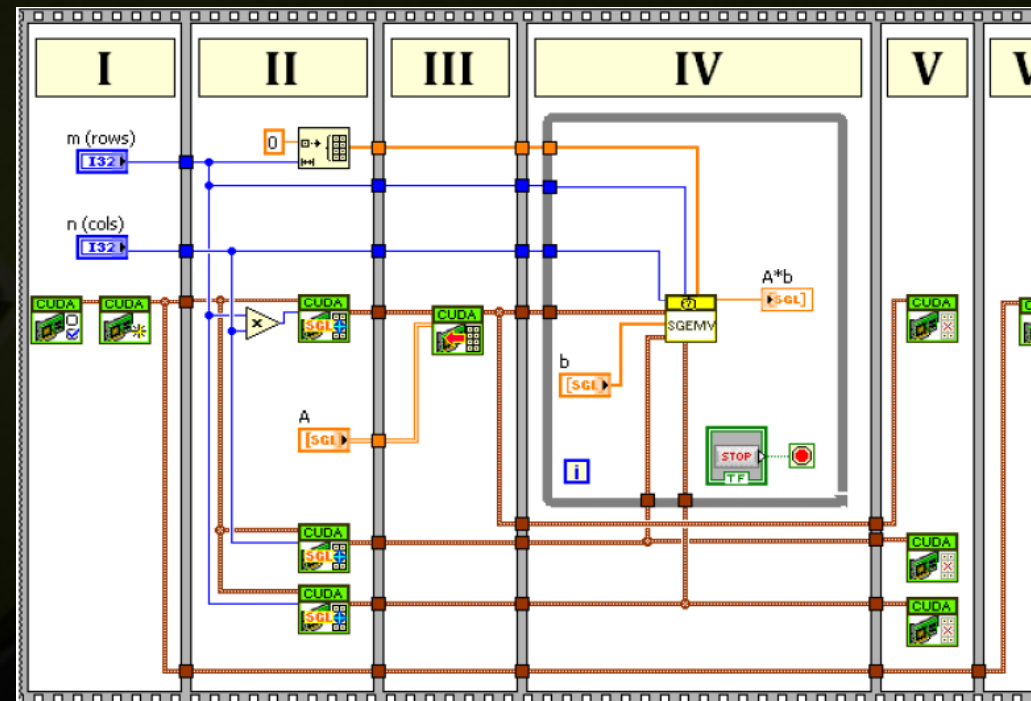
Algorithm	Field	Speedup
2-Electron Repulsion Integral	Quantum Chemistry	130X
Lattice Boltzmann	CFD	123X
Euler Solver	CFD	16X
Gromacs	Molecular Dynamics	137X
Lattice QCD	Physics	30X
Multifrontal Solver	FEA	20X
nbody	Astrophysics	100X
Simultaneous Iterative Reconstruction Technique	Computed Tomography	32X

Accelerating Time to Discovery



LabVIEW with CUDA

- Native interface for CUDA functions and libraries
 - Call cuBLAS functions
 - Call user-defined functions
- Support for Multiple GPUs
- 10-50x Speedups using GPUs



Tesla in Tsubame Supercomputer

“ the Tesla GPUs delivered speed-ups that we had never seen before – this will be a tremendous boost for our scientists and engineers

*Dr. Satoshi Matsuoka
Tokyo Institute of Technology*

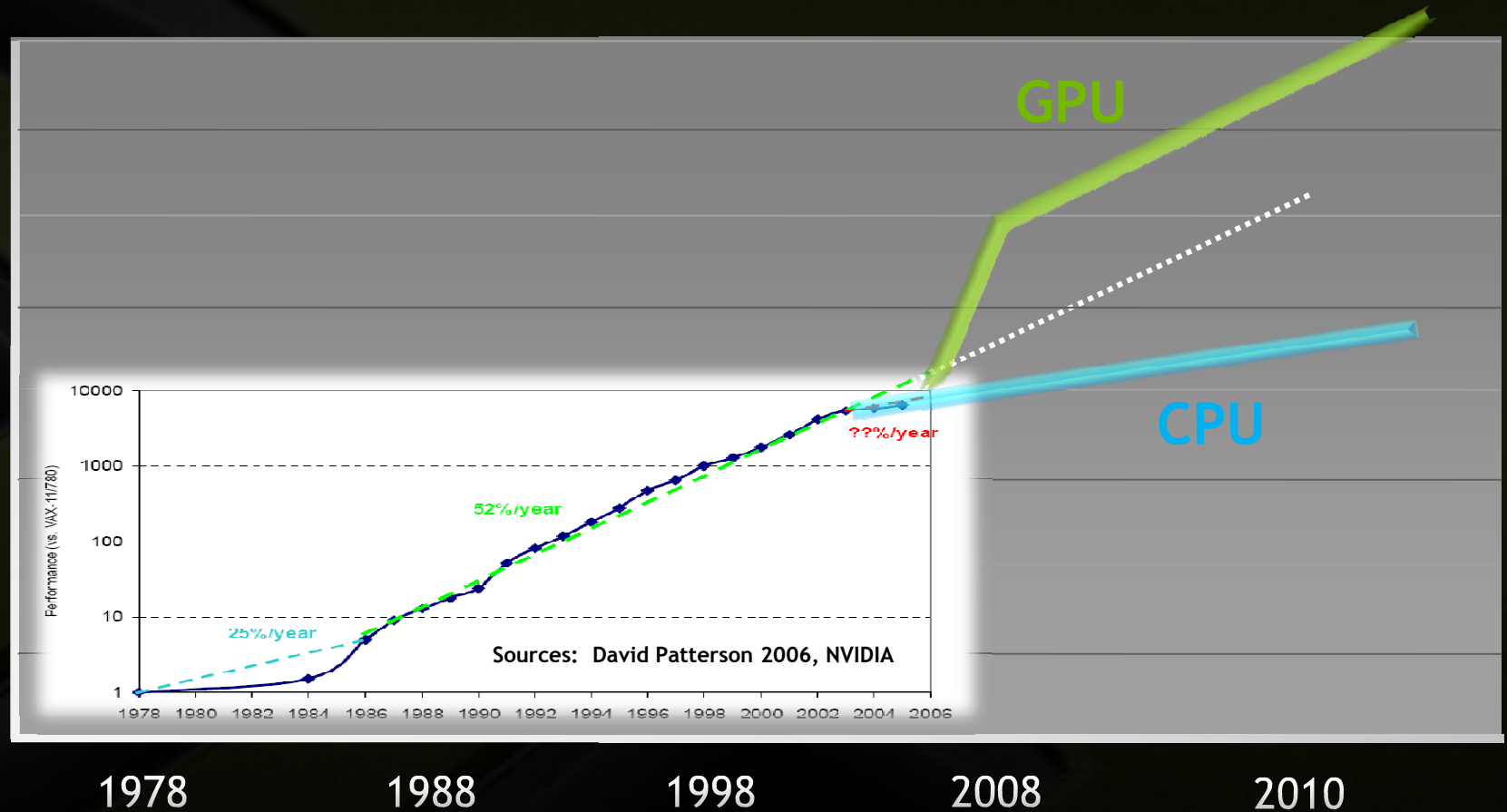


The GPU Computing Revolution

Performance
VAX 11/780 Equivalents

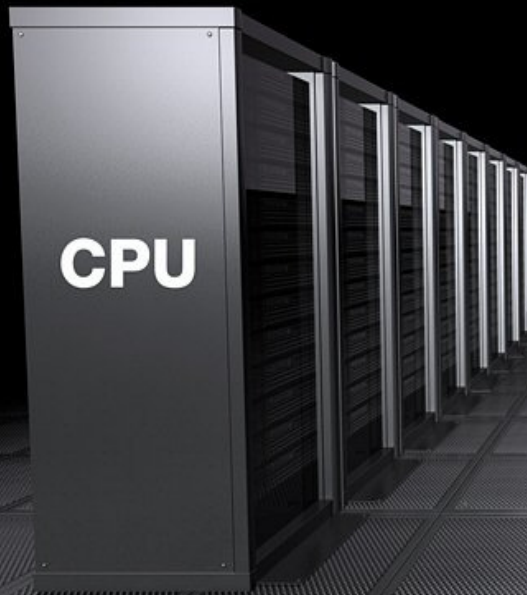
100,000,000
1,000,000
10,000
100
1

1

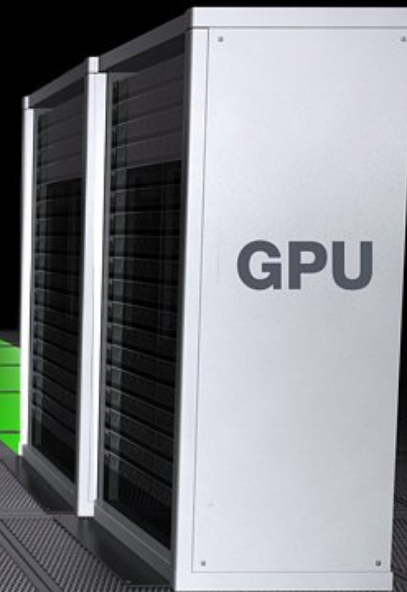


CUDA GPU Reduce Power 24X!

**3600 CPUs
60 Racks**



**120 GPUs + 60 CPUs
2 Racks**

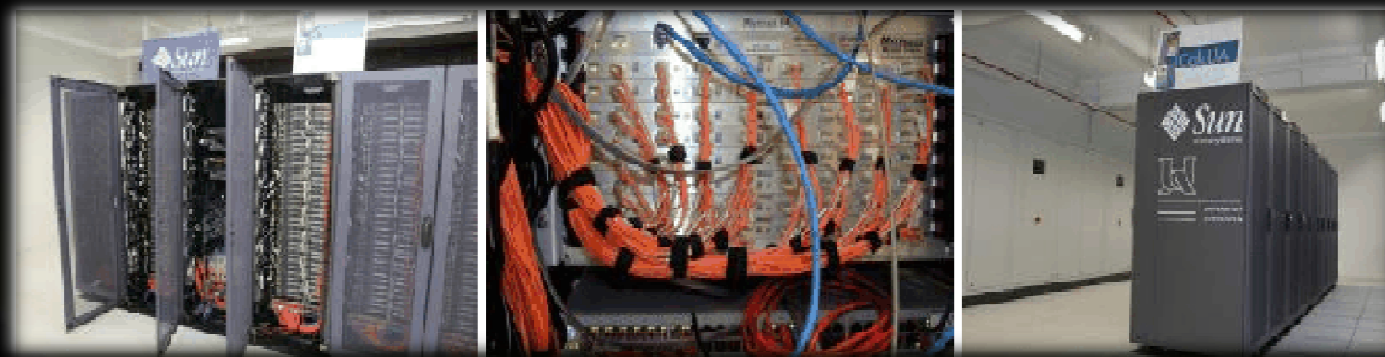


**60X Speed-Up (over 1 rack of CPU)
24X Less Power**

“Homemade” Supercomputer Revolution

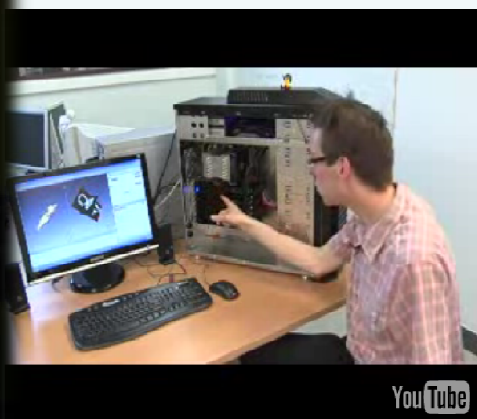


CalcUA
256 Nodes (512 cores)



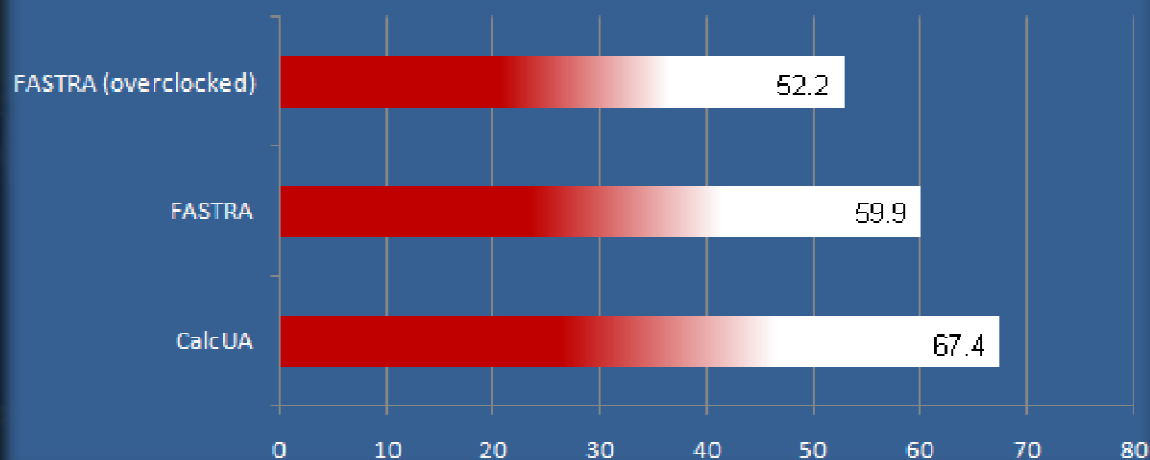
FASTRA
8 GPUs in a Desktop

FASTRA Desktop Supercomputer Built With 4
Nvidia 9800 GX2 Graphics Cards

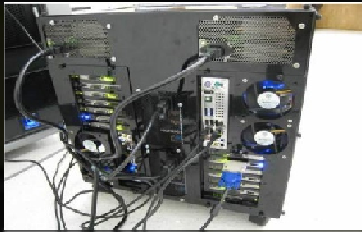


Looking

Reconstruction running times (secs)



“Homemade” Supercomputer Revolution



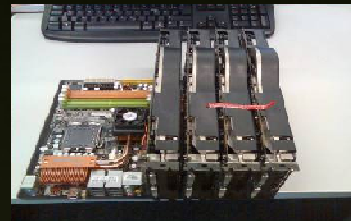
16 GPUs

MIT, Harvard



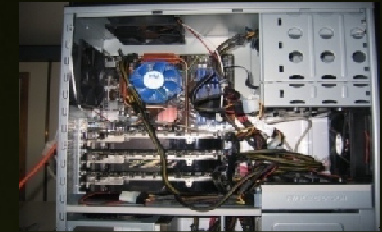
8 GPUs

University of Antwerp
Belgium



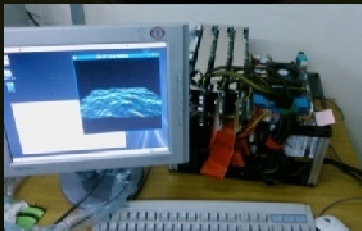
4 GPUs

TU Braunschweig,
Germany



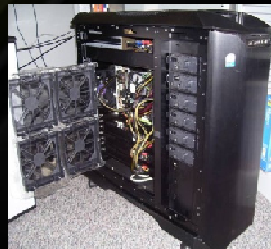
3 GPUs

University of Illinois



3 GPUs

Yonsei University,
Korea



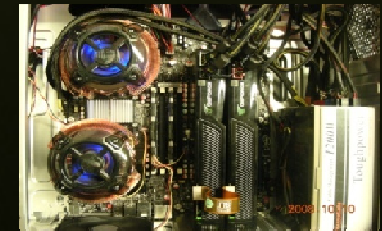
3 GPUs

Rice University



3 GPUs

University of
Cambridge, UK



2 GPUs

Georgia Tech

Tesla Personal Supercomputer



Supercomputing Performance

- Massively parallel CUDA Architecture
- 960 cores. 4 Teraflops
- 250x the performance of a desktop

Personal

- One researcher, one supercomputer
- Plugs into standard power strip

Accessible

- Program in C for Windows, Linux
- Priced like a PC Workstation

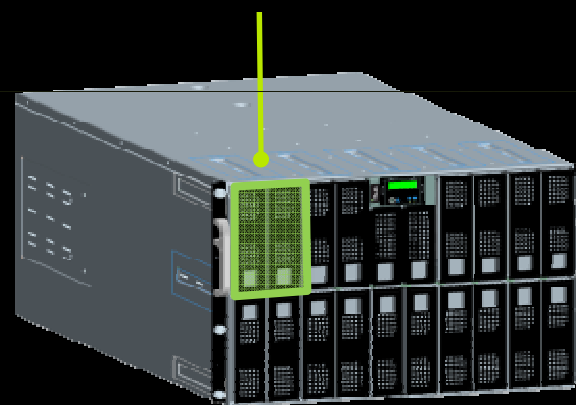
New Class of Hybrid CPU-GPU Servers

2 Tesla
M1060 GPUs



SuperMicro 1U
GPU Server

Upto 18 Tesla
M1060 GPUs



Bull Bullx
Blade Enclosure

Performance

10,000x

**Tesla
Co-processing
Cluster**

**Tesla
Personal
Supercomputer**

100x

**Traditional
CPU Cluster**

1x

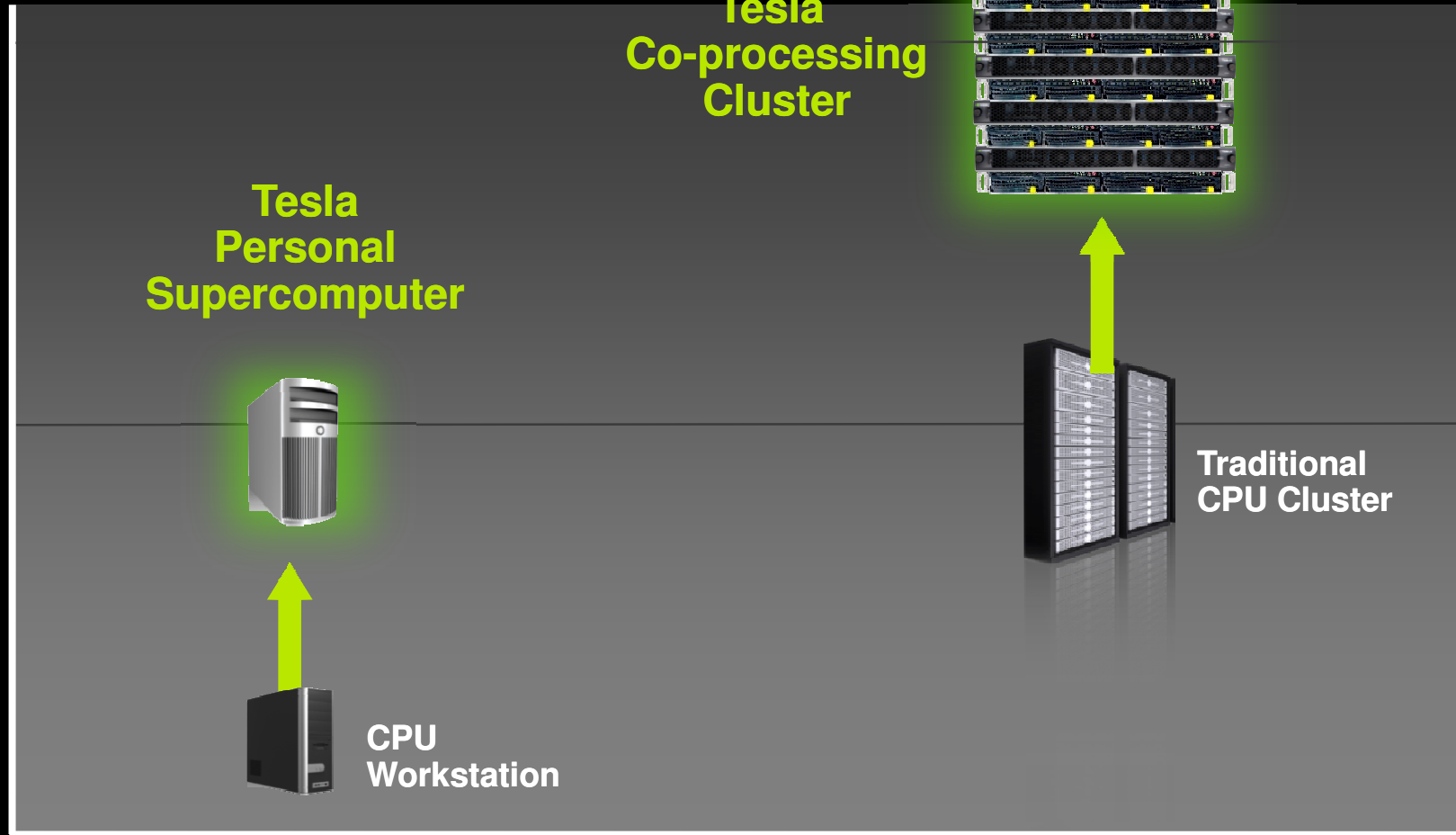


**CPU
Workstation**

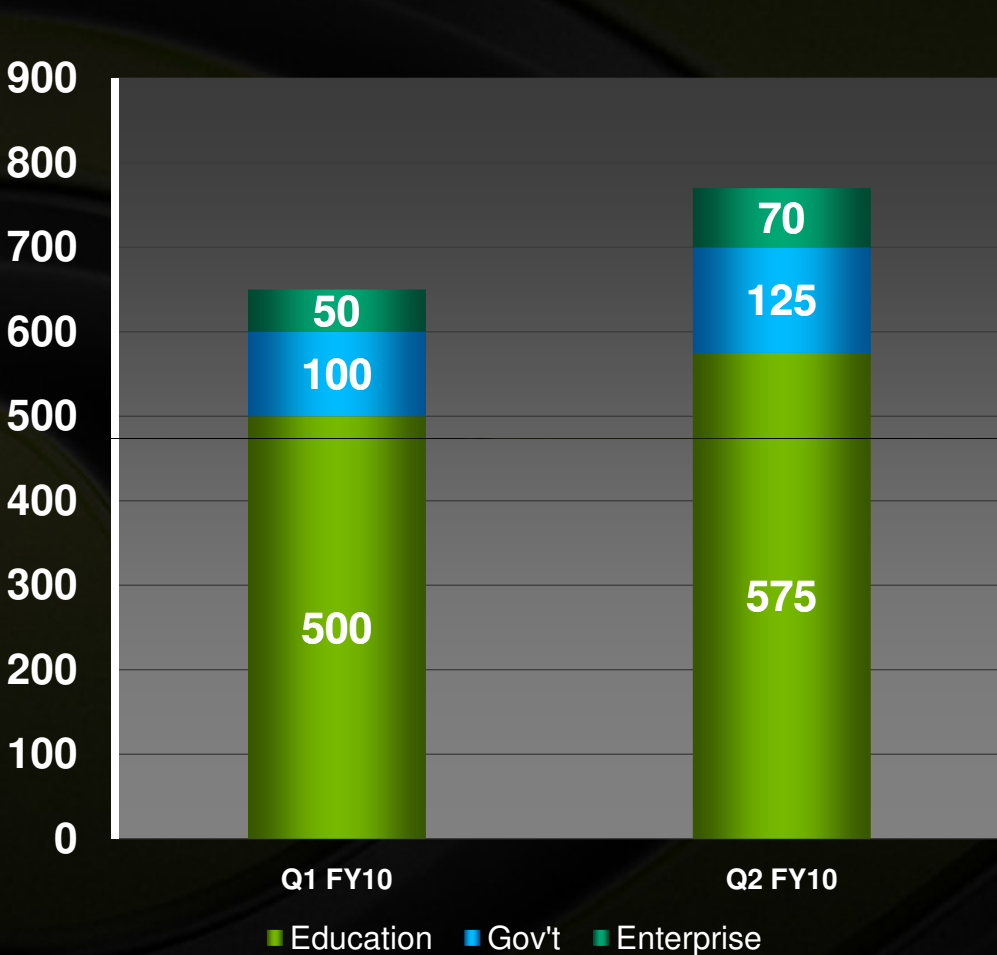


K\$

M\$



Tesla Server Installations



	# of GPUs
Commercial cluster	>3000
Government	>2000
Chinese Academy of Sciences - Industrial Process Institute	828
Tokyo Institute of Technology Supercomputing Center	680
NCSA – National Center for Supercomputing Applications	384
Seismic processing	256
Pacific Northwest National Labs – Biomedical research	256
CSIRO – Australian National Supercomputing Center	252
Riken – Japanese Astrophysical research	220
Seismic processing	200
Chinese Academy of Sciences – Institute of Modern Physics	200

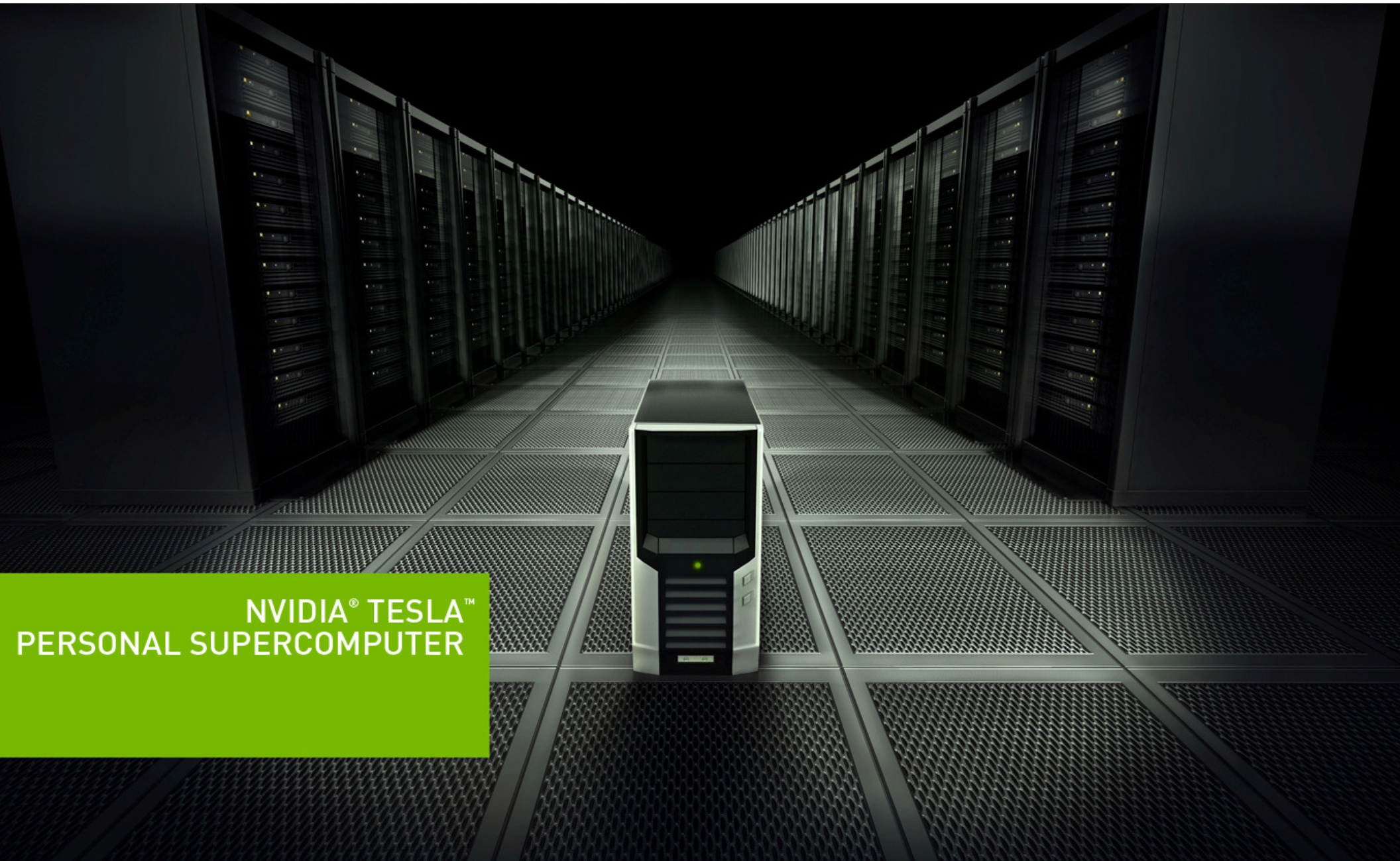
Final Thoughts

- GPU and heterogeneous parallel architecture will revolutionize computing
- Parallel computing key to solve some of the most interesting and important human challenges ahead
- Learning parallel programming is an imperative for students in computing and sciences

From Virtua Fighter to Tsubame

1995 – NV1	2008 – GT200
0.8M transistors	1,200M transistors
50MHz	1.3GHz
1M Bytes	4G Bytes
0 GFLOPS	1 TFLOPS
\$5M R&D	\$1B R&D

Another 1000x in 15 years?



NVIDIA® TESLA™
PERSONAL SUPERCOMPUTER