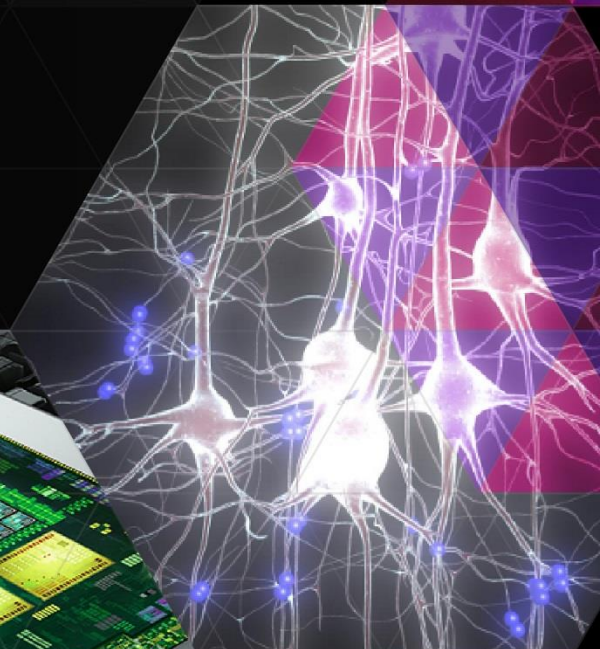


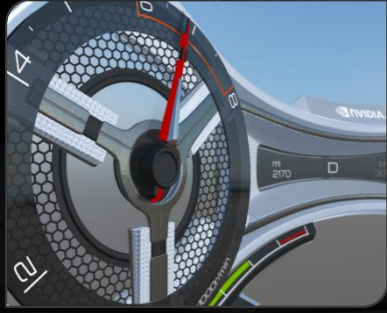


# THE VISUAL COMPUTING COMPANY

Edmondo Orloff, Channel Sales Director Europe  
Professional Solutions Group



**GTC14 VIDEO**



# NVIDIA

Two Decades of Visual Computing



GAMING



PROFESSIONAL  
VISUALIZATION



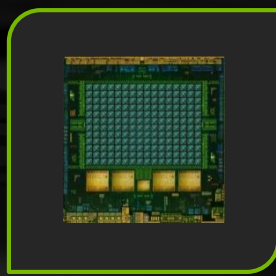
HPC and  
BIG DATA



MOBILE COMPUTING



IP



GPUs & SOCs



GRAPHICS CARDS



SYSTEMS

**WORLD LEADER IN VISUAL COMPUTING**



## PC GAMING IS THRIVING

600M PC gamers worldwide

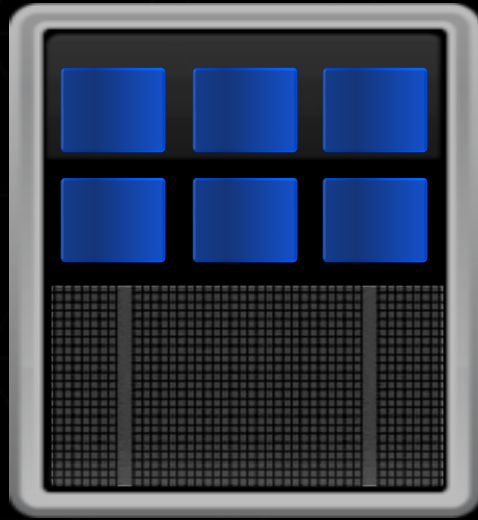
\$26B PC gaming market in 2014

GeForce GPUs: Choice of gamers

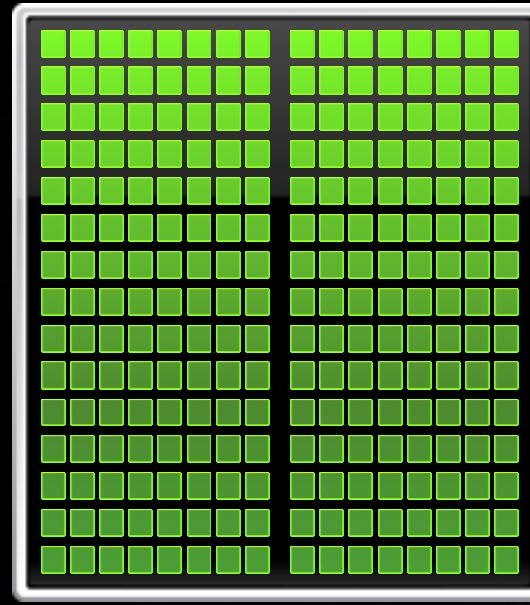
# ACCELERATED COMPUTING

*10x Performance & 5x Energy Efficiency for HPC*

**CPU**  
Optimized for  
Serial Tasks

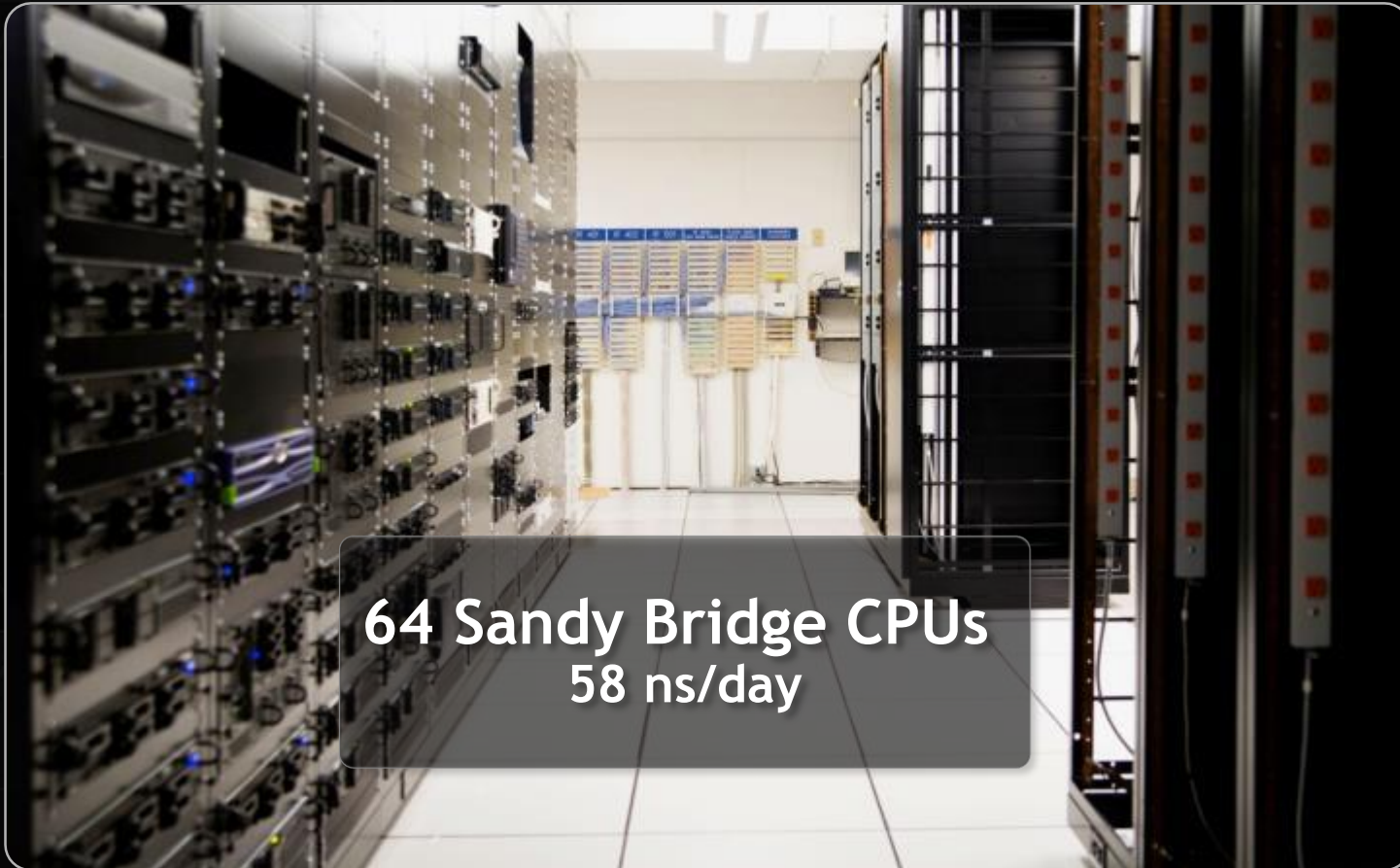


**GPU Accelerator**  
Optimized for  
Parallel Tasks

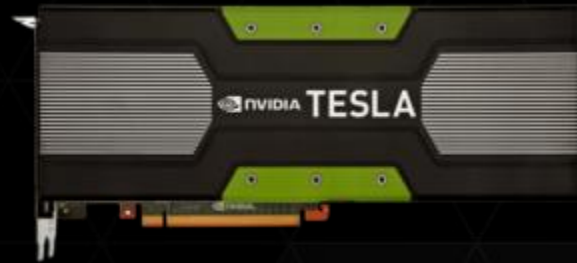


# REVOLUTIONIZING SCIENTIFIC COMPUTING

AMBER Molecular Dynamics Simulation  
DHFR NVE Benchmark



**64 Sandy Bridge CPUs**  
58 ns/day



**1 Tesla K40 GPU**  
102 ns/day

Power for CPU-only  
Exaflop Supercomputer



=

Power for the Bay Area, CA  
(*San Francisco + San Jose*)



**HPC's Biggest Challenge: Power**

# WORLD'S 15 GREENEST SUPERCOMPUTERS



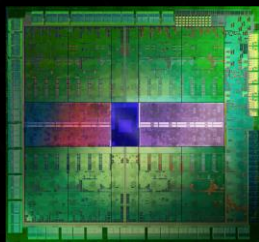
Green500 Rank	MFLOPS/W	Site
1	4,390	TSUBAME-KFC - GSIC Center, Tokyo Institute of Technology
2	3,632	Cambridge University
3	3,518	Center for Computational Sciences, University of Tsukuba
4	3,459	SURFsara - Netherlands
5	3,186	Swiss National Supercomputing Centre (CSCS)
6	3,131	ROMEO HPC Center - Champagne-Ardenne
7	3,020	CSIRO
8	2,952	TSUBAME 2.5 - GSIC Center, Tokyo Institute of Technology
9	2,813	ENI S.p.A. - Italian Energy Corporation
10 - 14	2,629	Financial Institutions
15	2,629	Max-Planck-Gesellschaft MPI/IPP

# KEPLER GENERATION OF GPUS

## Tesla K10



Dual GK104 GPUs



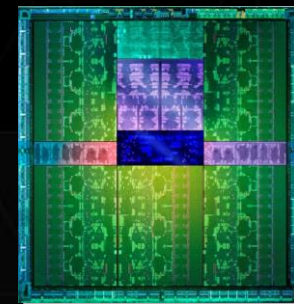
3x Single Precision

Video, Signal, Life Sciences, Seismic

## Tesla K20, K20X, K40



GK110 GPU



3x Double Precision

CFD, FEA, Finance, Physics, etc.

# Growth of GPU Computing

**100M**  
CUDA -Capable GPUs



**150K**  
CUDA Downloads



**1**  
Supercomputer



**60**  
University Courses



**4,000**  
Academic Papers



**2008**

# Growth of GPU Computing

**100M**  
CUDA -Capable GPUs



**430M**  
CUDA-Capable GPUs

**150K**  
CUDA Downloads



**1.6M**  
CUDA Downloads

**1**  
Supercomputer



**50**  
Supercomputers

**60**  
University Courses



**640**  
University Courses

**4,000**  
Academic Papers



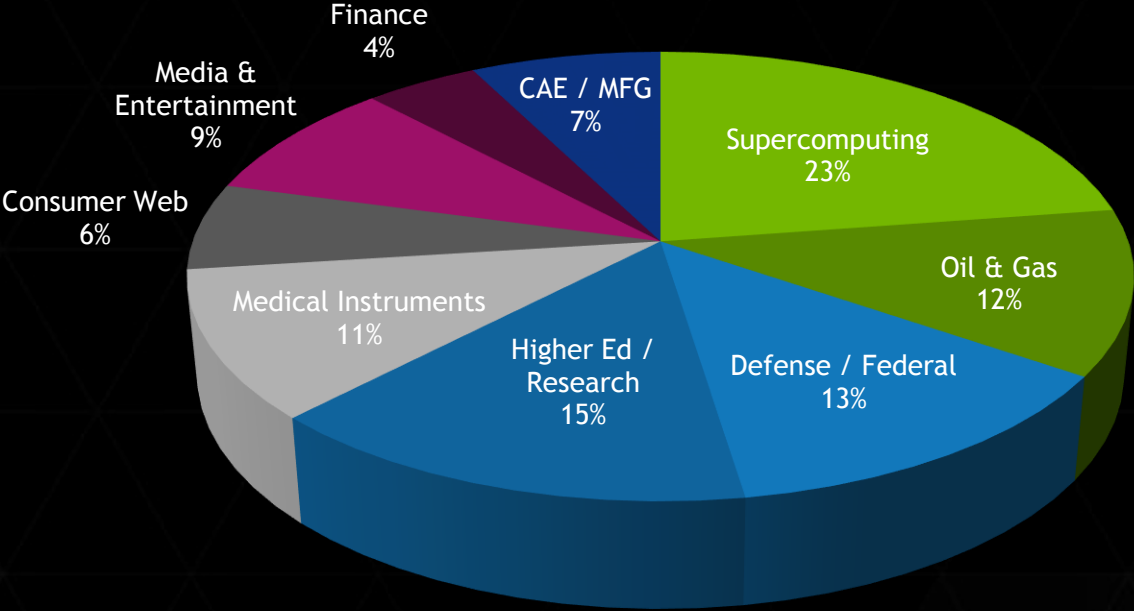
**37,000**  
Academic Papers

**2008**

**2013**

# DIVERSE MARKETS

## FY14 Segments



NVIDIA estimates

# ACCELERATING DATACENTERS

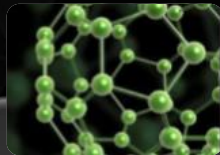


Oil & Gas



PETROBRAS

Chevron



Higher Ed



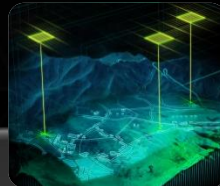
Chinese Academy of Sciences



HARVARD School of Engineering and Applied Sciences



STANFORD UNIVERSITY



Government



Air Force Research Laboratory



Naval Research Laboratory



Supercomputing



CSCS Swiss National Supercomputing Centre



NCSA

Tokyo Institute of Technology



Finance



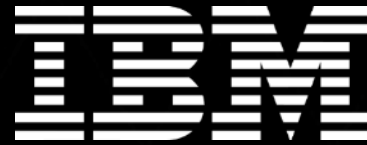
Web 2.0



# MACHINE LEARNING & DATA ANALYTICS



Speech/Image Recognition



Hadoop-based Clustering



Recommendation Engine



Auto Tagging in Creative Cloud



Database Queries

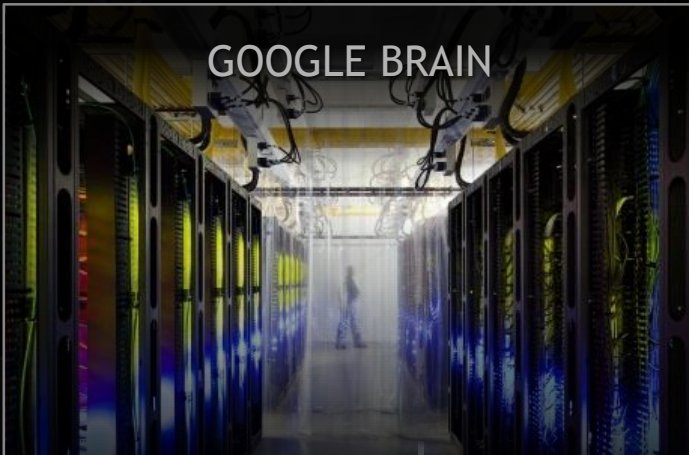


Search Ranking

# MACHINE LEARNING

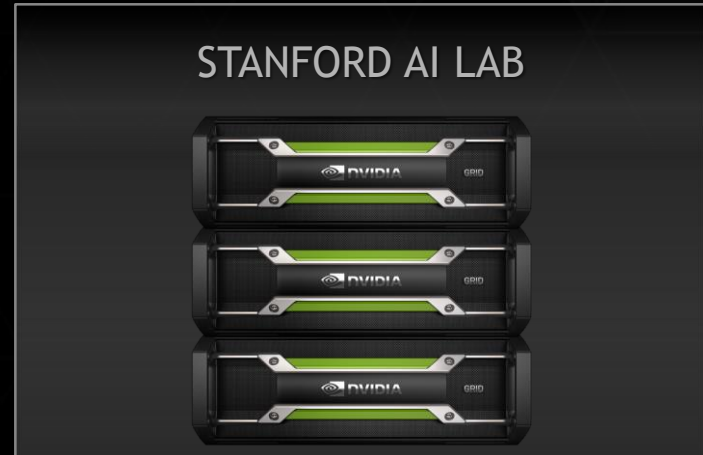
Artificial neural network at a fraction of the cost with GPUs

GOOGLE BRAIN



1,000 CPU Servers 2,000 CPUs • 16,000 cores	<b>600 kWatts</b> <b>\$5,000,000</b>
--	---

STANFORD AI LAB

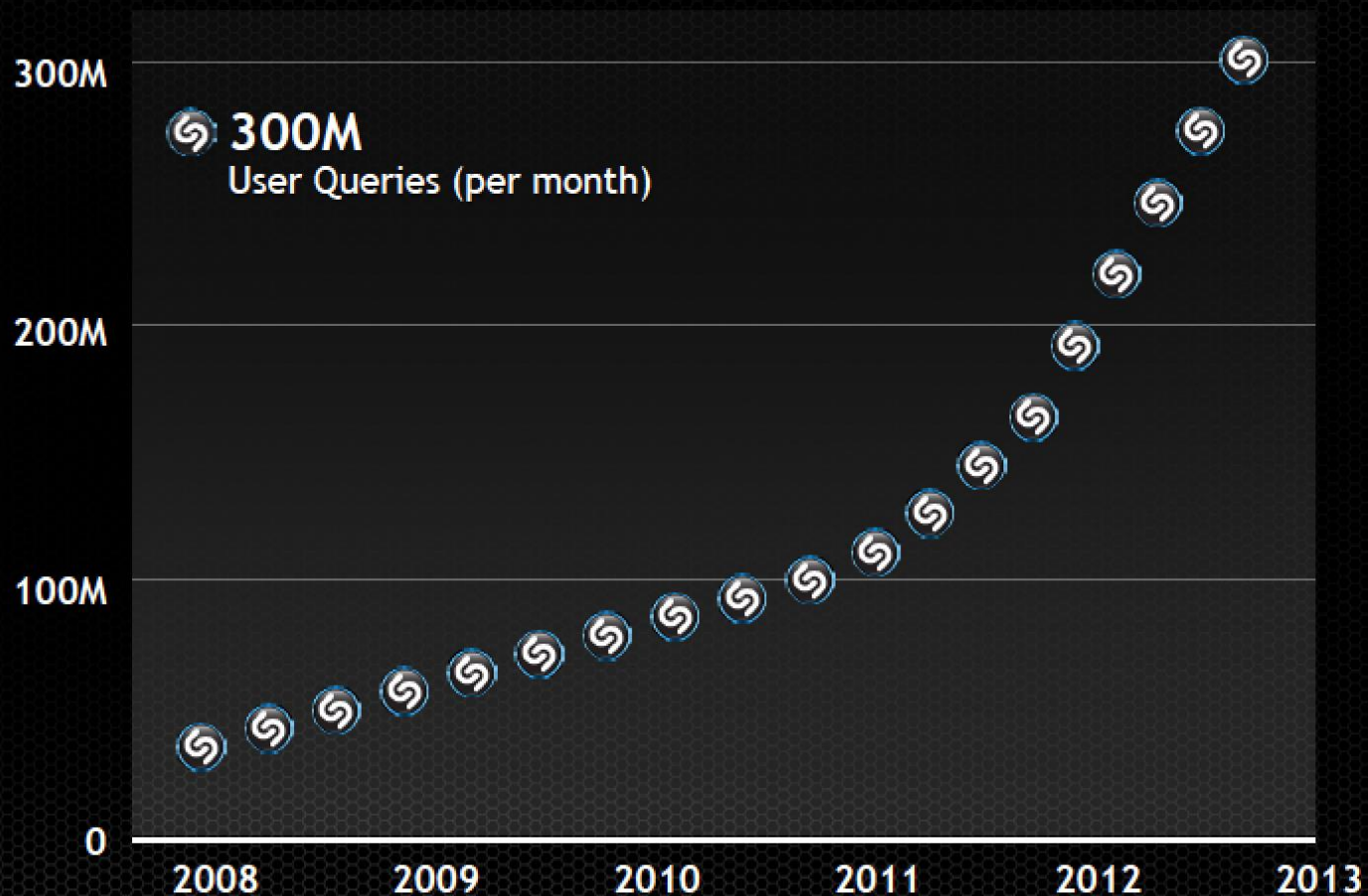


3 GPU-Accelerated Servers 12 GPUs • 18,432 cores	<b>4 kWatts</b> <b>\$33,000</b>
---	------------------------------------

# GPUs and Audio Search

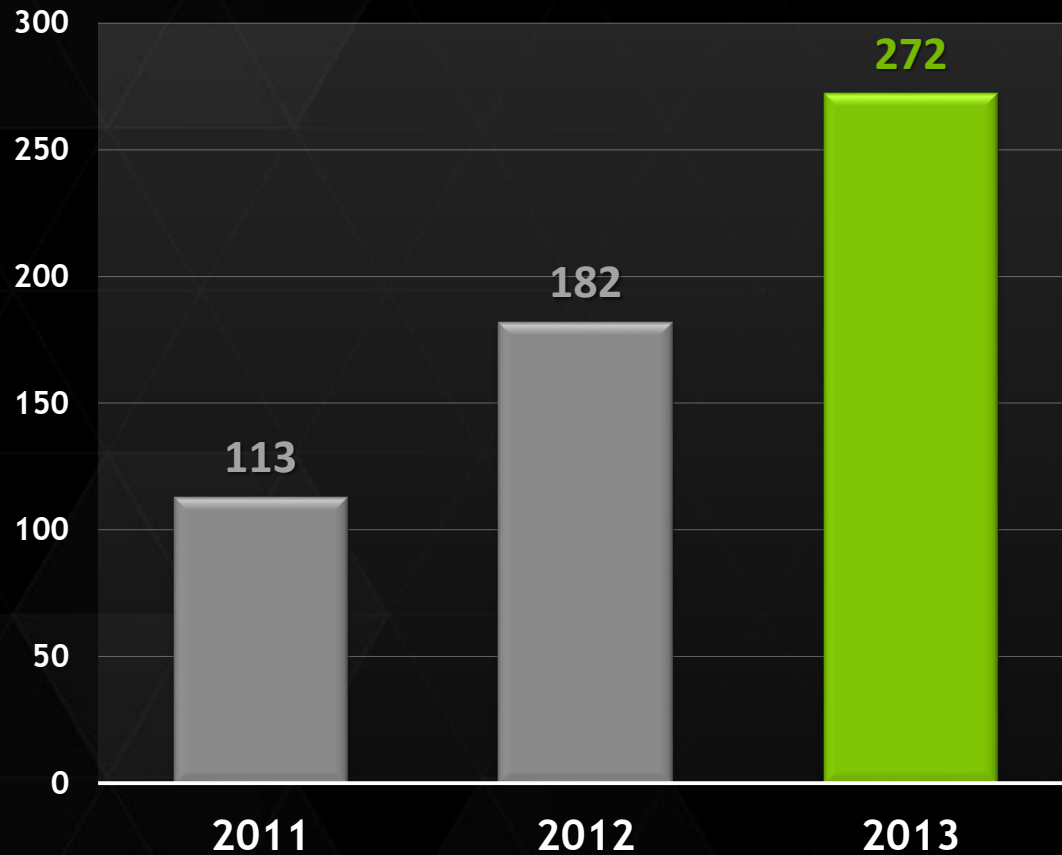


## SHAZAM



# SOLID GROWTH OF GPU ACCELERATED APPS

# of GPU-Accelerated Apps



## Top HPC Applications

Molecular Dynamics	AMBER CHARMM DESMOND	GROMACS LAMMPS NAMD
Quantum Chemistry	Abinit Gaussian	GAMESS NWChem
Material Science	CP2K QMCPACK	Quantum Espresso VASP
Weather & Climate	COSMO GEOS-5 HOMME	CAM-SE NEMO NIM WRF
Lattice QCD	Chroma	MILC
Plasma Physics	GTC	GTS
Structural Mechanics	ANSYS Mechanical LS-DYNA Implicit MSC Nastran	OptiStruct Abaqus/Standard
Fluid Dynamics	ANSYS Fluent	Culises (OpenFOAM)



# POPULAR GPU-ACCELERATED APPLICATIONS

## CONTENTS

- 02 Research: Higher Education and Supercomputing
  - COMPUTATIONAL CHEMISTRY AND BIOLOGY
  - NUMERICAL ANALYTICS
  - PHYSICS
  - WEATHER AND CLIMATE FORECASTING
- 06 Defense and Intelligence
- 07 Computational Finance
- 08 Manufacturing: CAD and CAE
  - COMPUTER AIDED DESIGN
  - COMPUTATIONAL FLUID DYNAMICS
  - COMPUTATIONAL STRUCTURAL MECHANICS
  - ELECTRONIC DESIGN AUTOMATION
- 10 Media and Entertainment
  - ANIMATION, MODELING AND RENDERING
  - COLOR CORRECTION AND GRAIN MANAGEMENT
  - COMPOSITING, FINISHING AND EFFECTS
  - EDITING
  - ENCODING AND DIGITAL DISTRIBUTION
  - ON-AIR GRAPHICS
  - ON-SET, REVIEW AND STEREO TOOLS
  - SIMULATION
  - WEATHER GRAPHICS
- 14 Oil and Gas

## Research: Higher Education and Supercomputing

### COMPUTATIONAL CHEMISTRY AND BIOLOGY

#### Bioinformatics

Application	Description	Supported Platforms	Accelerated Speed (x)	Downloaded GPU <sup>SM</sup>	Multi-GPU Support	Release Status
BarraCUDA	Sequence mapping software	Alignment of short sequencing reads	6-10x	T 2075, 2090, K10, K20, K20X	Yes	Available now Version 0.6.2
CUDASW++	Open source software for Smith-Waterman protein database searches on GPUs	Parallel search of Smith-Waterman database	10-50x	T 2075, 2090, K10, K20, K20X	Yes	Available now Version 2.0.8
CUSHAW	Parallelized short read aligner	Parallel, accurate long read aligner - gapped alignments to large genomes	10x	T 2075, 2090, K10, K20, K20X	Yes	Available now Version 1.0.40
GPU-BLAST	Local search with fast k-tuple heuristic	Protein alignment according to blastp, multi cpu threads	3-4x	T 2075, 2090, K10, K20, K20X	Single only	Available now Version 2.2.26
GPU-HMMER	Parallelized local and global search with profile Hidden Markov models	Parallel local and global search of Hidden Markov Models	60-100x	T 2075, 2090, K10, K20, K20X	Yes	Available now Version 2.3.2
mCUDA-MEME	Ultrafast scalable motif discovery algorithm based on MEME	Scalable motif discovery algorithm based on MEME	4-10x	T 2075, 2090, K10, K20, K20X	Yes	Available now Version 3.0.12
SeqFind	A GPU Accelerated Sequence Analysis Toolset	Reference assembly, blast, smith-waterman, hmm, de novo assembly	400x	T 2075, 2090, K10, K20, K20X	Yes	Available now
UGENE	Opensource Smith-Waterman for SSE/CUDA, Suffix array based repeats finder and dotplot	Fast short read alignment	6-8x	T 2075, 2090, K10, K20, K20X	Yes	Available now Version 1.11
WideLM	Fits numerous linear models to a fixed design and response	Parallel linear regression on multiple similarly-shaped models	150x	T 2075, 2090, K10, K20, K20X	Yes	Available now Version 0.1-1

#### Molecular Dynamics

Application	Description	Supported Platforms	Accelerated Speed (x)	Downloaded GPU <sup>SM</sup>	Multi-GPU Support	Release Status
Abalone	Models molecular dynamics of biopolymers for simulations of proteins, DNA and ligands	Simulations (on 1060 GPU)	4-29x	T 2075, 2090, K10, K20, K20X	Single Only	Available now Version 1.8.48
ACEMD	GPU simulation of molecular mechanics force fields, implicit and explicit solvent	Written for use on GPUs	160 ns/day GPU version only	T 2075, 2090, K10, K20, K20X	Yes	Available now
AMBER	Suite of programs to simulate molecular dynamics on biomolecule	PMEMD: explicit and implicit solvent	89.44 ns/day JAC NVE	T 2075, 2090, K10, K20, K20X	Yes	Available now Version 12 + bugfix9
DL-POLY	Simulate macromolecules, polymers, ionic systems, etc on a distributed memory parallel computer	Two-body forces, Link-cell pairs, Ewald SPME forces, Shake W	4x	T 2075, 2090, K10, K20, K20X	Yes	Available now, Version 4.0 Source only
CHARMM	MD package to simulate molecular dynamics on biomolecule.	Implicit (Sx), Explicit (2x) Solvent via DgeMM	TBD	T 2075, 2090, K10, K20, K20X	Yes	In Development Q4/12
GRMOMACS	Simulation of biochemical molecules with complicated bond interactions	Implicit (Sx), Explicit(2x) solvent	165 ns/Day DHFR	T 2075, 2090, K10, K20, K20X	Single only	Available now Version 4.6 in Q4/12
HOOMD-Blue	Particle dynamics package written grounds up for GPUs	Written for GPUs	2x	T 2075, 2090, K10, K20, K20X	Yes	Available now
LAMMPS	Classical molecular dynamics package	Lennard-Jones, Morse, Buckingham, CHARMM, Tabulated, Course grain SDK, Anisotropic Gay-Bern, RE-squared, "hybrid" combinations	3-18x	T 2075, 2090, K10, K20, K20X	Yes	Available now

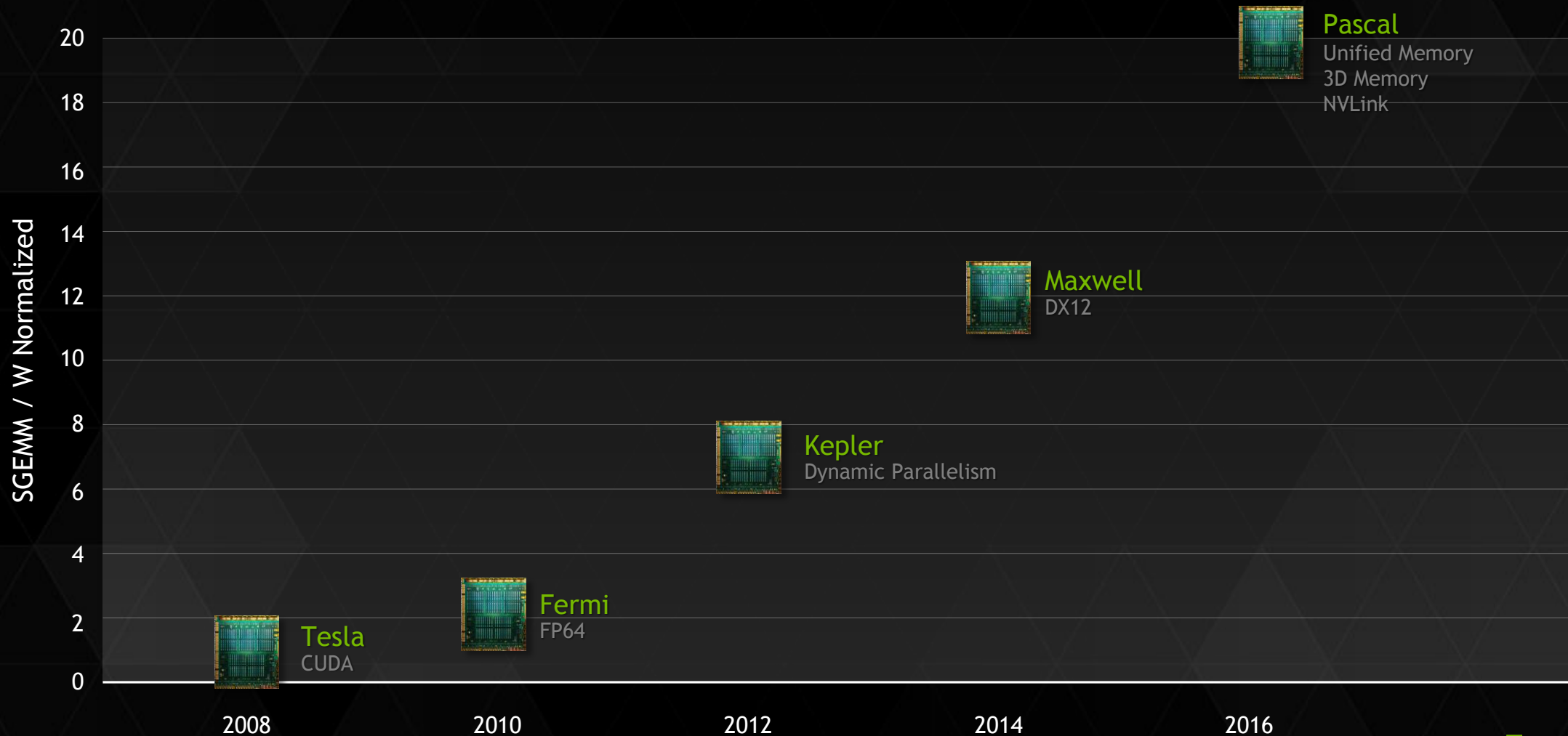
# Hundreds of GPU-Accelerated Applications

[www.nvidia.com/appscatalog](http://www.nvidia.com/appscatalog)

The background of the slide features a dark green, grid-like pattern that appears to be a distorted or warped mesh, possibly representing a GPU's internal structure or a 3D rendering technique. The grid lines are thin and dark, creating a complex, interconnected web of lines. The overall color palette is a range of dark greens, from deep black to a slightly lighter, vibrant green. The text is centered and rendered in a bright, neon-like green color, which stands out sharply against the dark background.

# *GPU HARDWARE ROADMAP*

# FAST PACED CUDA GPU ROADMAP





# Blaise Pascal

1623–1662

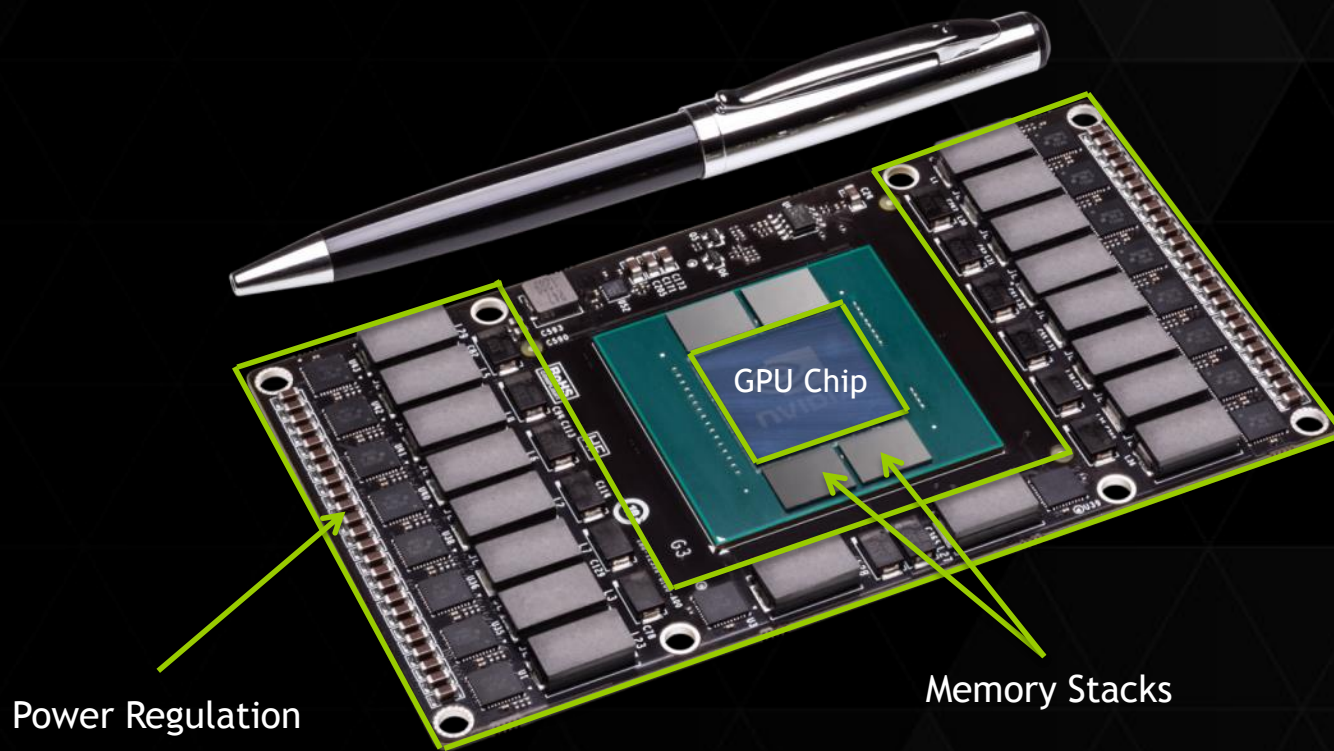
Mechanical Calculator  
Probability Theory  
Pascal's Theorem  
Pascal's Law

# PASCAL COMPUTING PLATFORM ...



# PASCAL

**NVLink** 5 to 12X PCIe 3.0  
**3D Memory** 2 to 4X memory BW & size  
**Module** 1/3 size of PCIe card



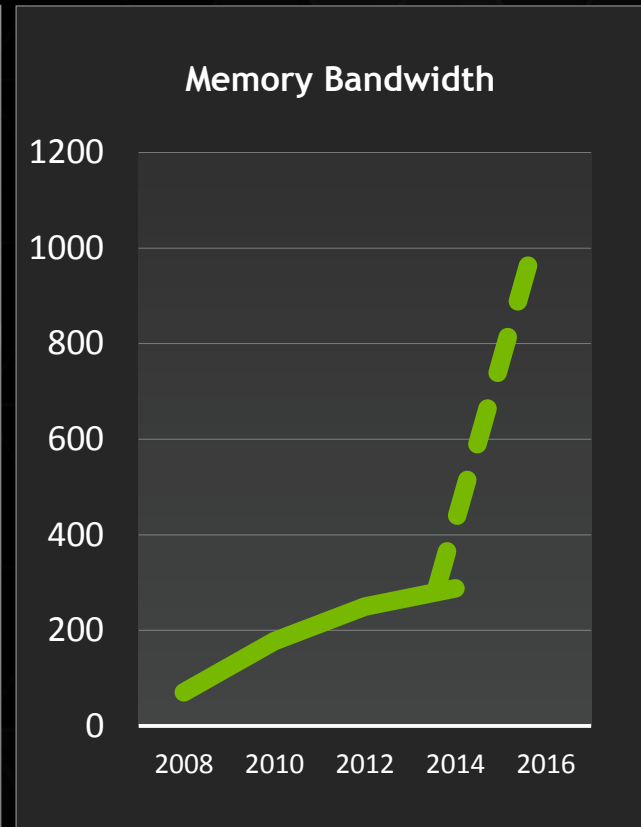
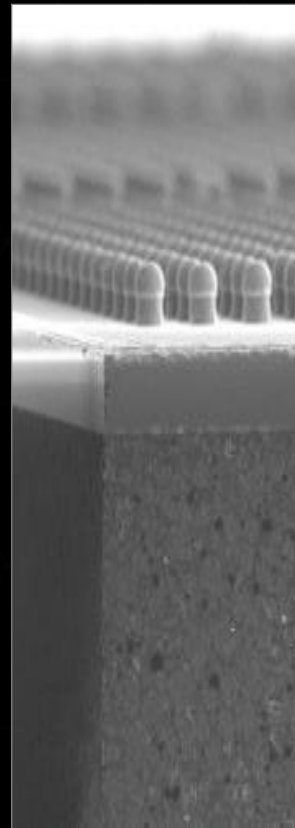
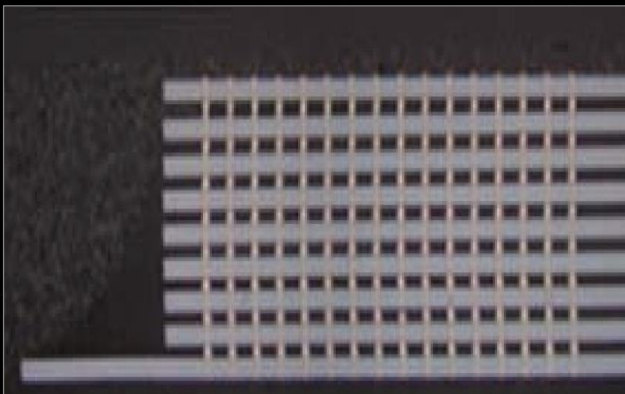
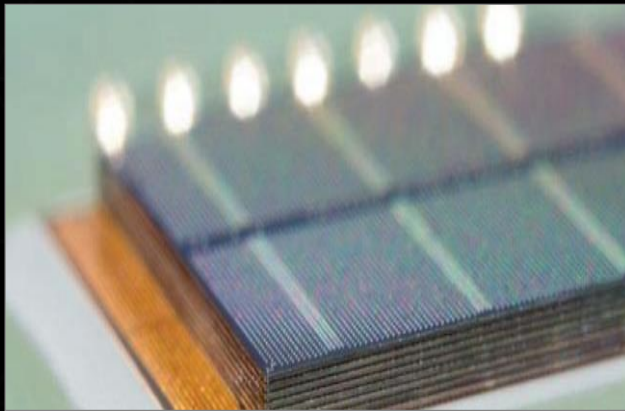
# 3D MEMORY

3D Chip-on-Wafer integration

Many X bandwidth

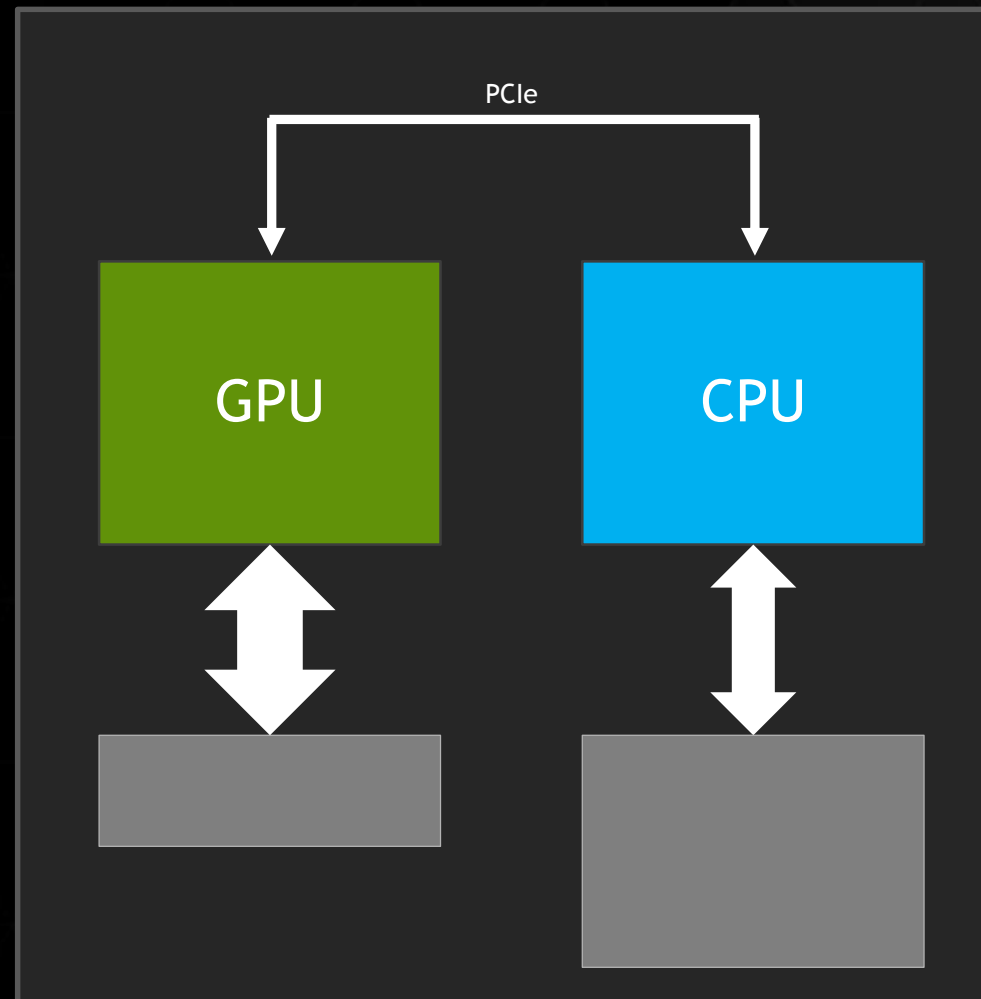
2.5X capacity

4X energy efficiency



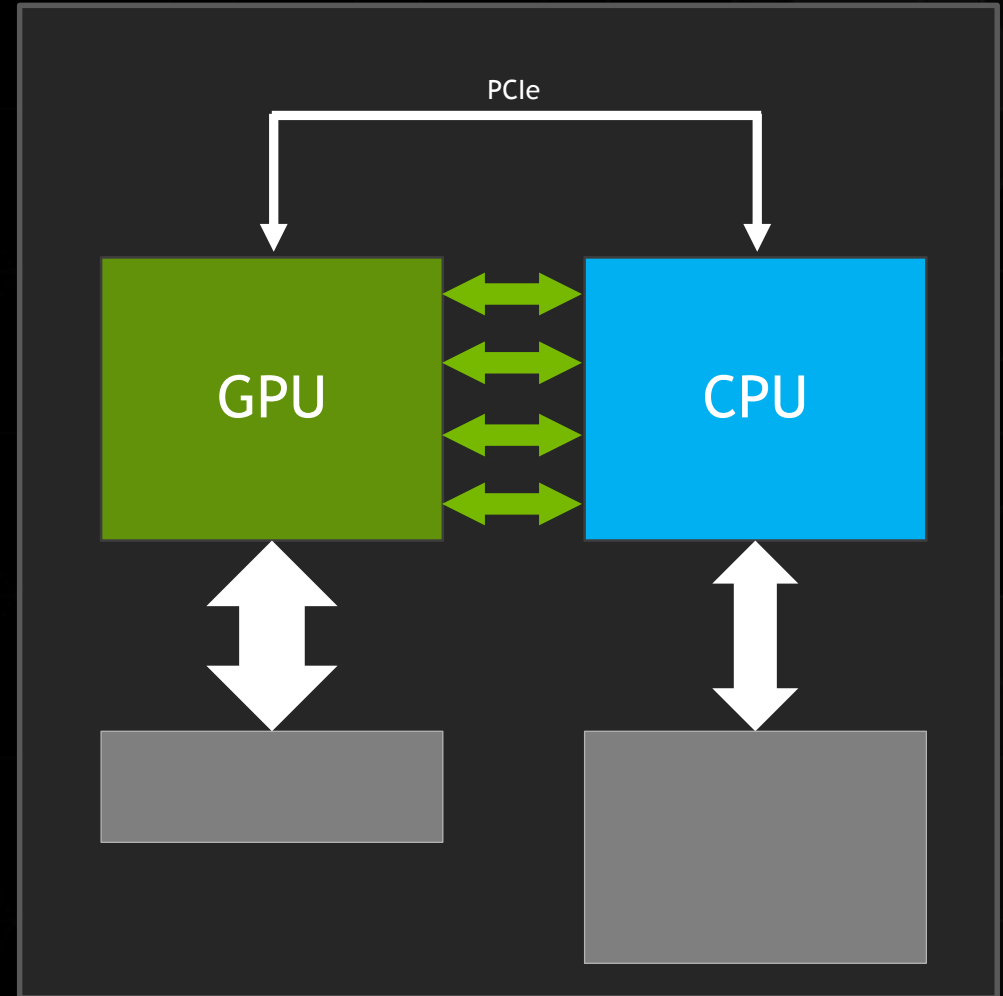
# BANDWIDTH BOTTLENECKS

PCI Express	16GB/sec
CPU Memory	60GB/sec
GPU Memory	288GB/sec

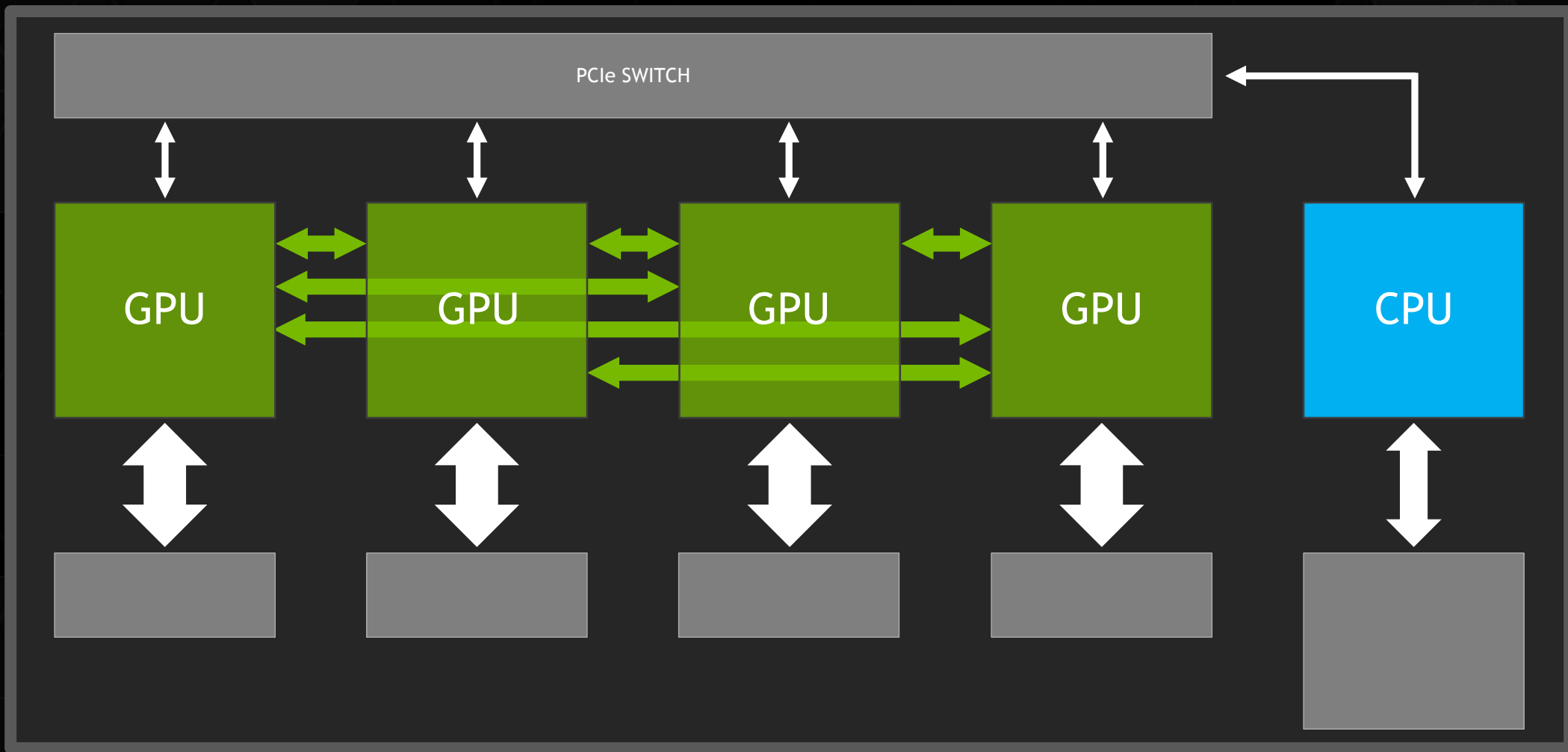


# INTRODUCING NVLINK

Differential with embedded clock  
PCIe programming model (w/ DMA+)  
Unified Memory  
Cache coherency in Gen 2.0  
5 to 12X PCIe



# 5X MORE BANDWIDTH FOR SCALING

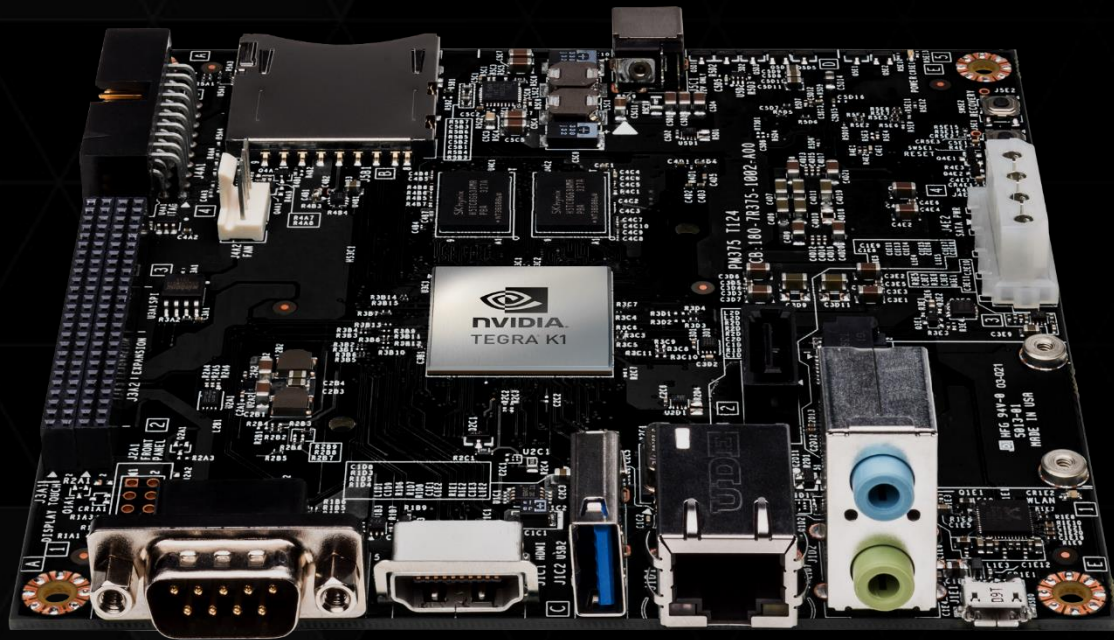


The background of the slide features a dark green, grid-like pattern that appears to be a distorted or warped mesh, possibly representing a GPU or a complex data structure. The grid lines are thin and dark, creating a subtle texture. The overall color palette is dark green and black, with the text in a bright, glowing green.

# *GPU EMBEDDED SOLUTIONS*

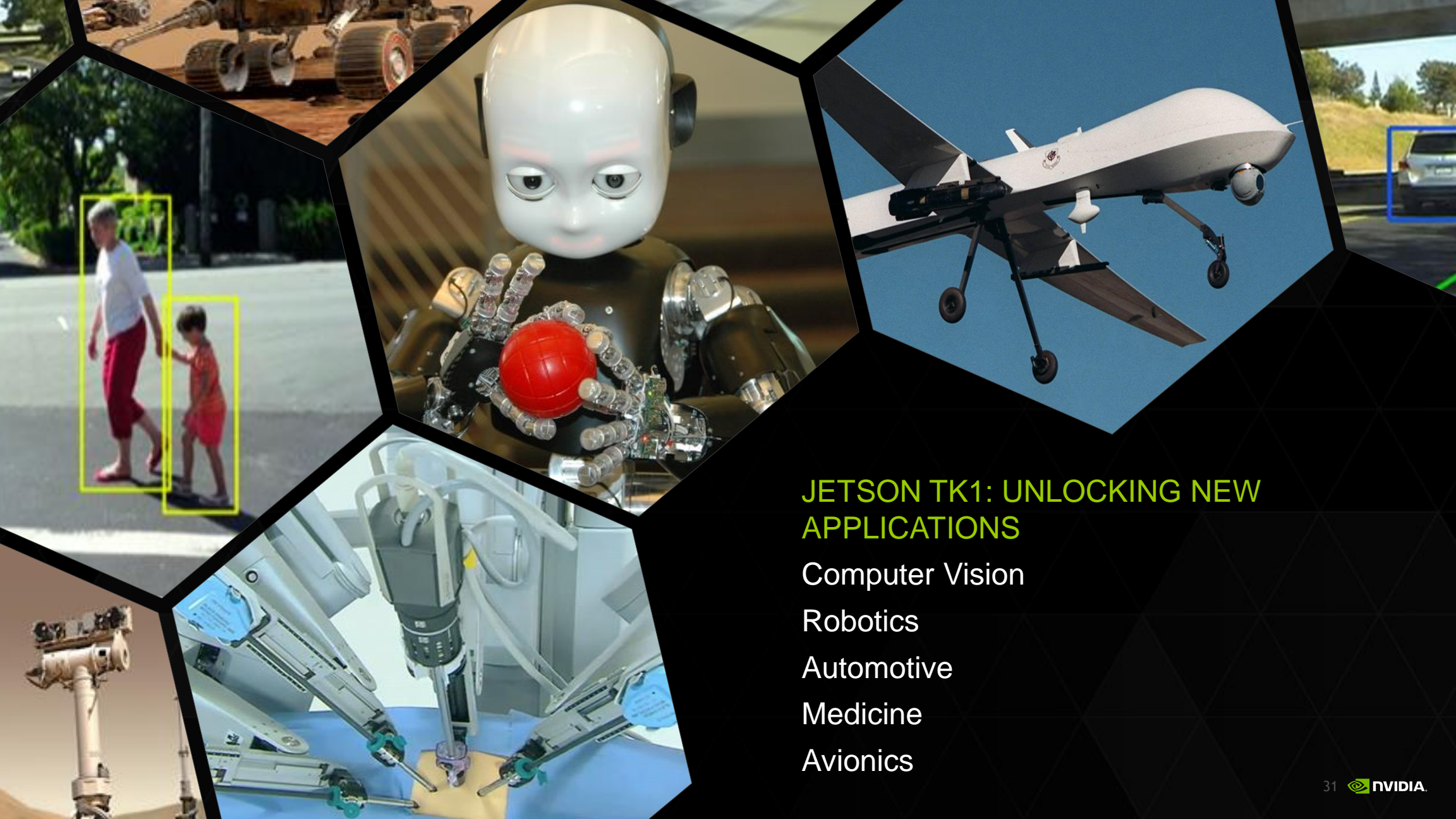
# JETSON TK1

THE WORLD'S 1st EMBEDDED SUPERCOMPUTER



Development Platform for Embedded  
Computer Vision, Robotics, Medical

Tegra K1 SoC  
CUDA Enabled  
\$192



## JETSON TK1: UNLOCKING NEW APPLICATIONS

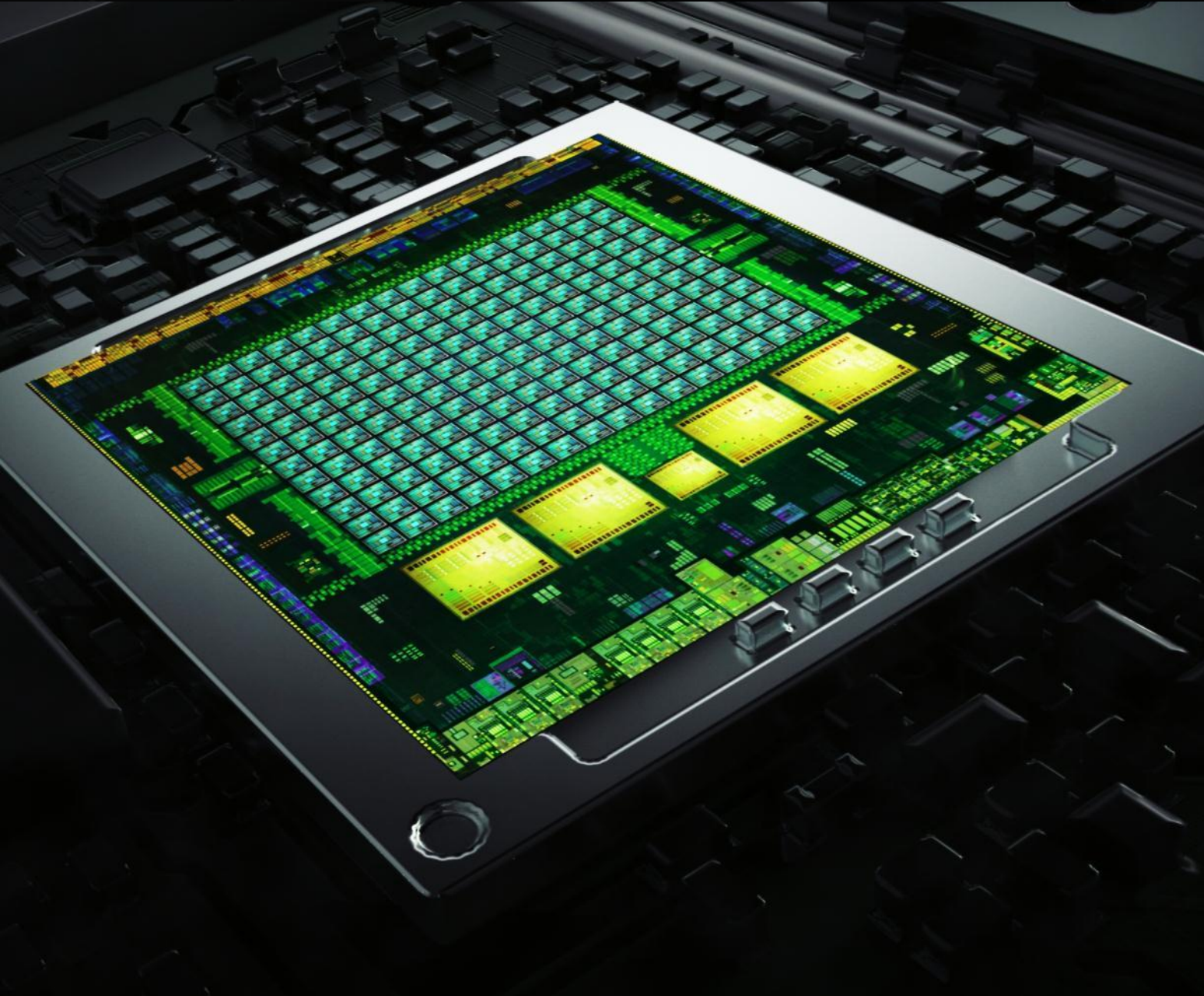
Computer Vision

Robotics

Automotive

Medicine

Avionics



# TEGRA K1

## IMPOSSIBLY ADVANCED

NVIDIA Kepler Architecture

4-Plus-1 Quad-Core A15

192 NVIDIA CUDA Cores

Compute Capability 3.2

326 GFLOPS

5 Watts

# CUDA 6 FOR EMBEDDED APPLICATIONS

- ▶ Tegra K1 Supports Full CUDA Toolkit v6.0
- ▶ Unified Memory
  - ▶ Memory physically unified, separate GPU and CPU caches
  - ▶ Same programming model as desktop and server
- ▶ OpenGL 4.4, DirectX 11 and OpenGL ES 3.0
- ▶ Jetson TK1 runs 32-bit Ubuntu 13.04 Linux for Tegra (L4T)

# REMOTE DEVELOPMENT TOOLS

- Remotely develop, debug and profile CUDA 6 applications with NVIDIA Nsight Eclipse Edition for Jetson TK1.
- Synchronise between x86 host and Jetson TK1, or cross-compile on x86 host.

The screenshot displays the NVIDIA Nsight Eclipse Edition interface for remote development on a Jetson TK1. The main window shows the source code of a CUDA kernel, `boofilter_kernel.cu`, with a debugger window open. The debugger window shows the execution state of the kernel, including the current thread, the device (boofilter\_arm on ubuntu), and the current block of execution. The right-hand side of the interface shows the 'Variables' and 'Registers' windows, which provide a detailed view of the kernel's state during execution. Below the source code, the 'Profile' window is visible, showing a timeline of the kernel's execution. The 'Analysis' window at the bottom provides a summary of the kernel's performance, including the start time, duration, grid size, block size, and throughput. The 'Console' window shows the output of the build process, including the compilation of the kernel and the generation of the executable. The 'Remote Connections' window is also visible, showing the connection to the Jetson TK1.

Name	Start Time	Duration	Grid Size	Block Size	Reps	Static Mem	Dynamic Mem	Size	Throughput
boofilter_arm [yml]	288.52ms	3.288ms							
d_boofilter_ghba_x[unsigned int, int, int]	297.244ms	13.303ms	[64, 1]	[64, 1]	43	0	0	164	164
d_boofilter_ghba_x[unsigned int, int, int]	298.823ms	2.265ms	[64, 1]	[64, 1]	32	0	0	164	164
d_boofilter_ghba_x[unsigned int, int, int]	298.934ms	13.308ms	[64, 1]	[64, 1]	43	0	0	164	164
d_boofilter_ghba_x[unsigned int, int, int]	300.846ms	1.732ms	[64, 1]	[64, 1]	32	0	0	164	164
d_boofilter_ghba_x[unsigned int, int, int]	304.532ms	13.316ms	[64, 1]	[64, 1]	43	0	0	164	164
d_boofilter_ghba_x[unsigned int, int, int]	347.658ms	1.832ms	[64, 1]	[64, 1]	32	0	0	164	164
d_boofilter_ghba_x[unsigned int, int, int]	382.329ms	13.771ms	[64, 1]	[64, 1]	43	0	0	164	164
d_boofilter_ghba_x[unsigned int, int, int]	396.065ms	1.965ms	[64, 1]	[64, 1]	32	0	0	164	164
d_boofilter_ghba_x[unsigned int, int, int]	423.167ms	13.476ms	[64, 1]	[64, 1]	43	0	0	164	164
d_boofilter_ghba_x[unsigned int, int, int]	448.819ms	2.221ms	[64, 1]	[64, 1]	32	0	0	164	164



Edmondo Orlotti  
eorlotti@nvidia.com

