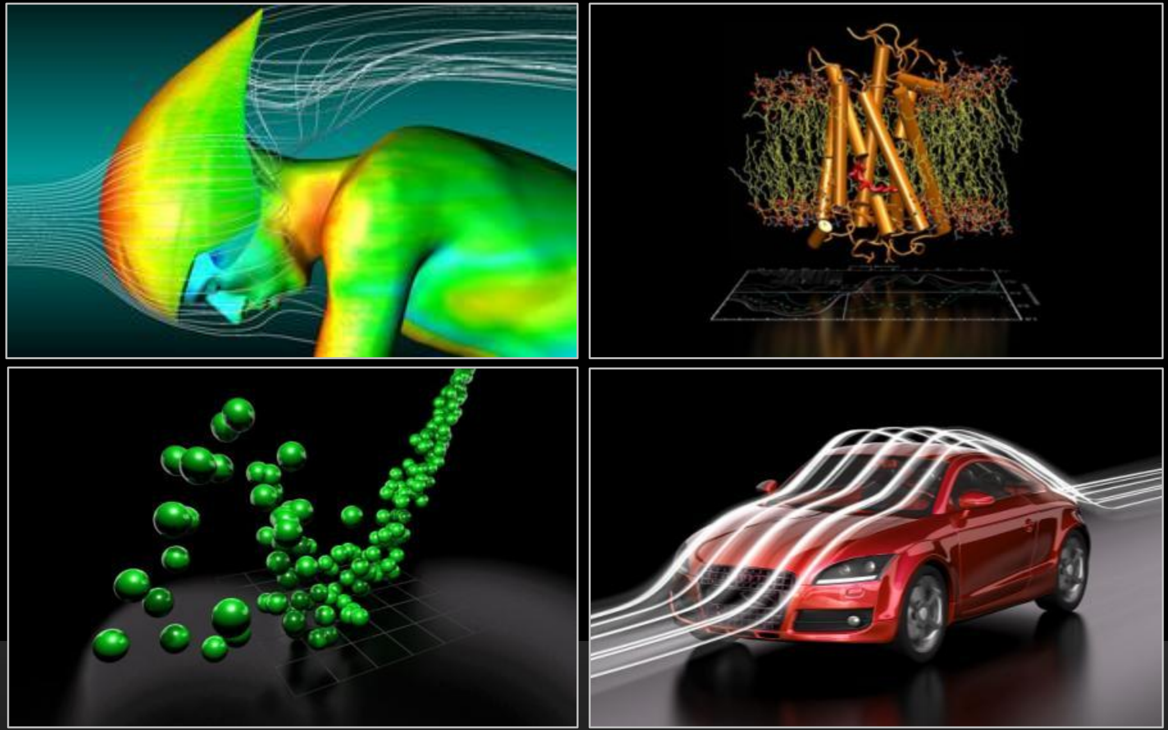
AGENDA 8:00 - 9:45

- “Tesla Supercomputing”
  - Andy Keane, NVIDIA
- “Multi-Scale HPC System for Multi-Scale Discrete Simulation”
  - Prof. Wei Ge, Chinese Academy of Sciences
- “Divide & Conquer Molecular Sim: GROMACS, OpenMM, and CUDA”
  - Prof. Eric Lindahl, Stockholm University
- "Extreme scale simulations of high-temperature superconductivity"
  - Dr. Thomas Schulthess, Director Swiss National Supercomputing Center & Chair of Computational Physics ETH-Zurich

TESLA

HPC Computing



Andy Keane

General Manager Tesla Computing
NVIDIA Corporation

CS192g

Jeff Wong - Spectral Matrix Decomposition

Peter P. - Radially Restricted Geodesic Distance

Jorge - CUDA Bench

Mustafa G. - Ultrasound Image Computation

Hoi Wong - Macro Adaptive Filter

Matt Culbreth - 3D Navier-Stokes

Stephen Russell & Peter Kimball - Range Ray Tracing for Particle Filters

Pete S. & Eugene R. - Convex solver

Matt L. & David K. - Local receptive fields

Jeffery Wang - searching straight line programs

Patrice C. & David W. - Navier Stokes

Andrew H. & Vadim O. - "CUDA Creatures"

Sanjeeb B. - Navier-Stokes Solver

Kiran M. & Sean A.- Image feature detection

Chunjing X. & Xiang Z. - ARPES (spectroscopy)

Changhan R. - Monte Carlo Stochastic Diff Eq

Jingyu C. - Image Reconstruction for PET

Sungpack H. & Tayo O. - Breadth-first search

Divij G. - Sudoku solver

Wen Z. - Particle level set

Kranthi K. - Graph partitioning

Hongli T. - Quantum chemistry

Steve L. - ILM Harry Potter fire

Why is HPC so important to NVIDIA?

Do GPUs apply to my work?

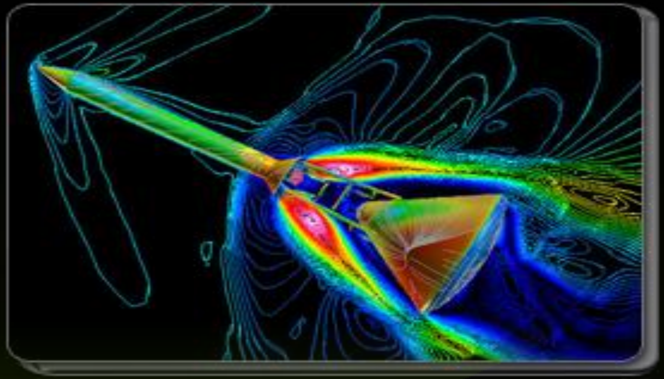
Why do we need something new?

NVIDIA Business Model

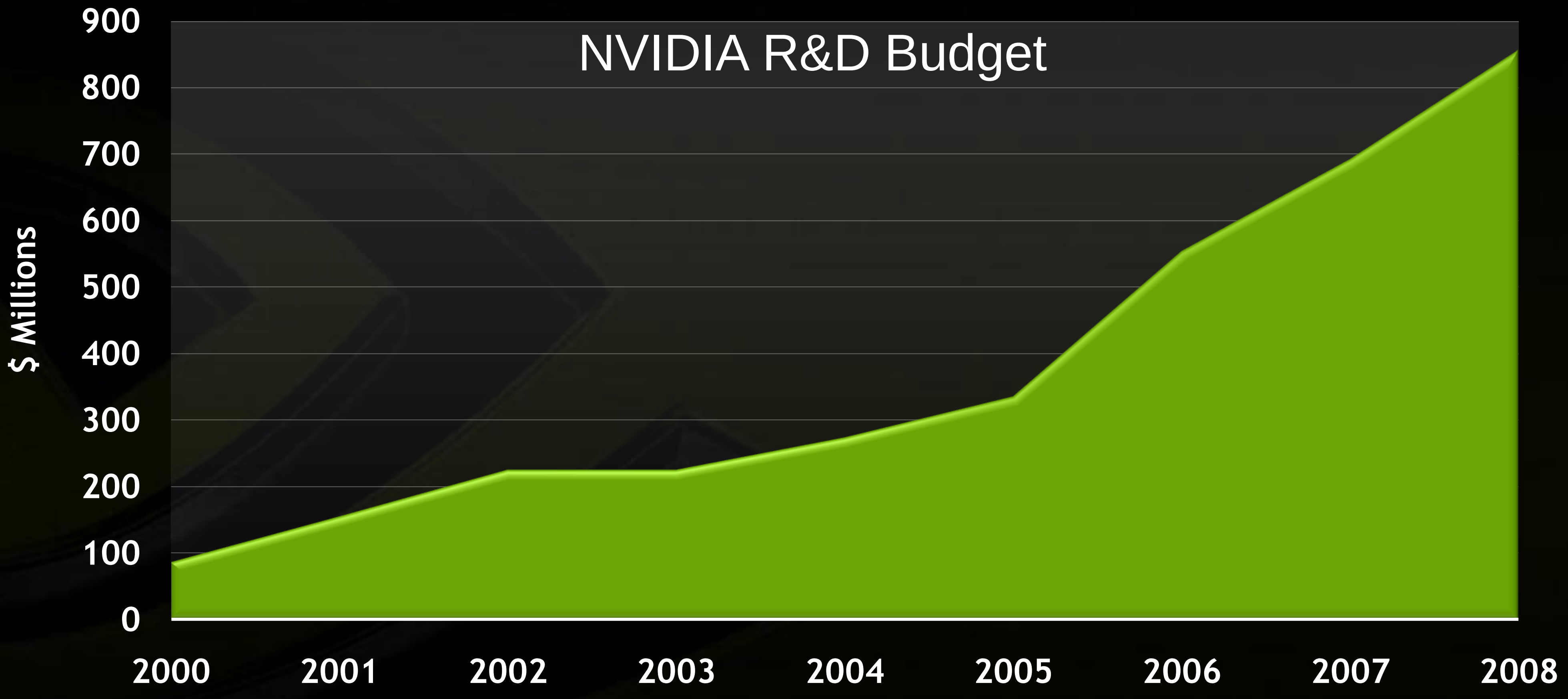
GeForce®
Entertainment

Quadro®
Design & Creation

Tesla™
High Performance Computing



GPU Innovation Accelerating



Mainstream Applications Going Parallel

CUDA Accelerates Adobe Mercury Playback Engine

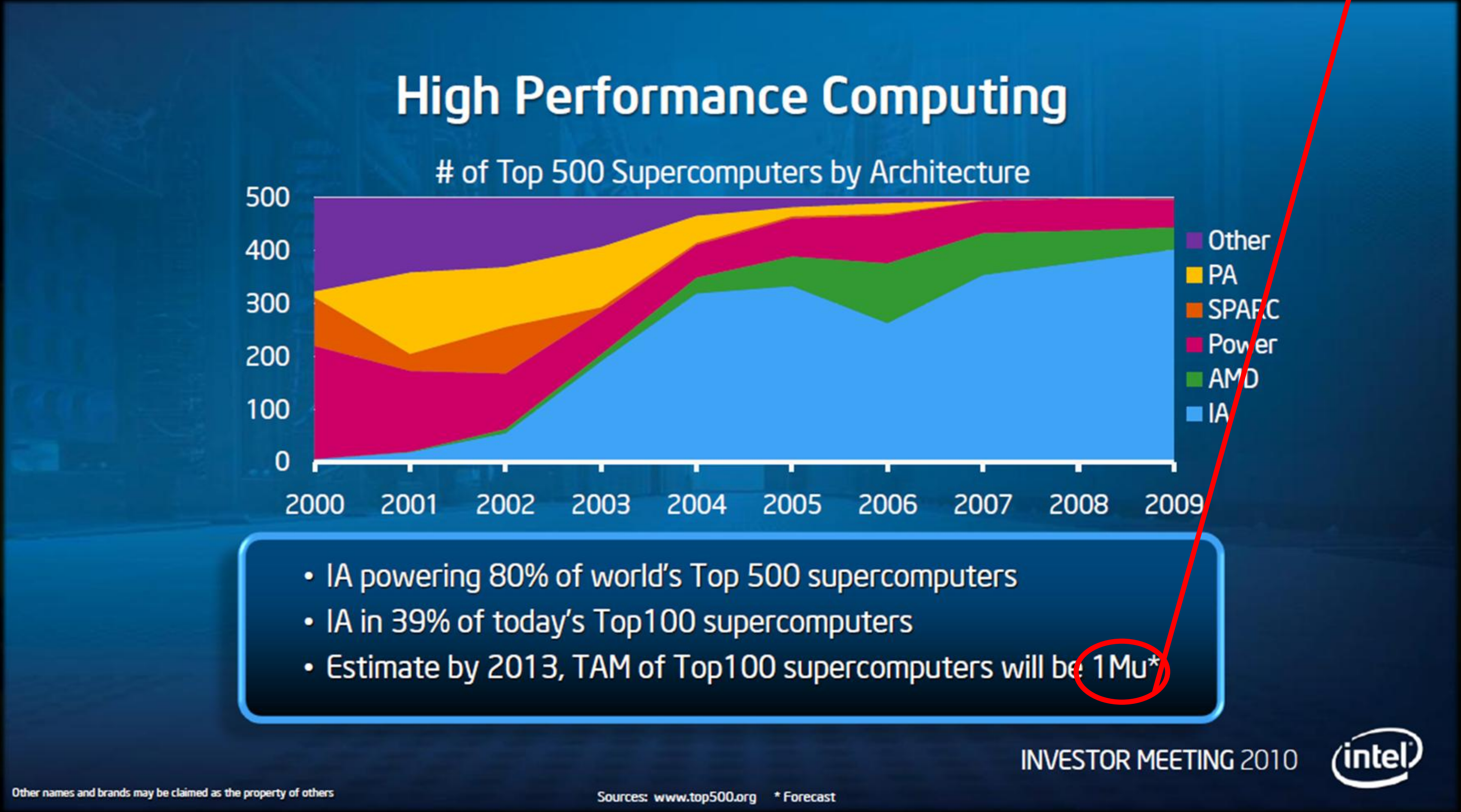
Amazingly fluid, real-time video editing

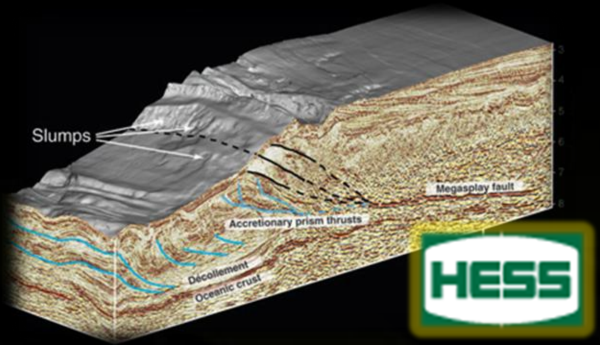
Quick preview of real time edits and effects

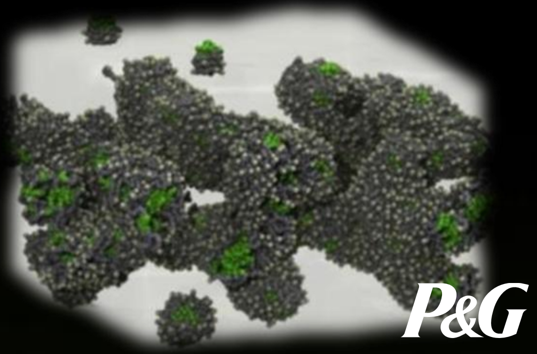
Realistic preview of final content

Faster encoding

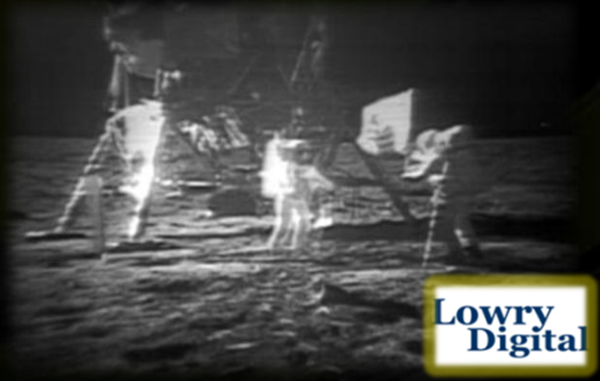
1 Million Units by 2013




Slumps
Megasplay fault
Accretionary prism thrusts
Décollement
Oceanic crust
HESS


P&G


Bloomberg

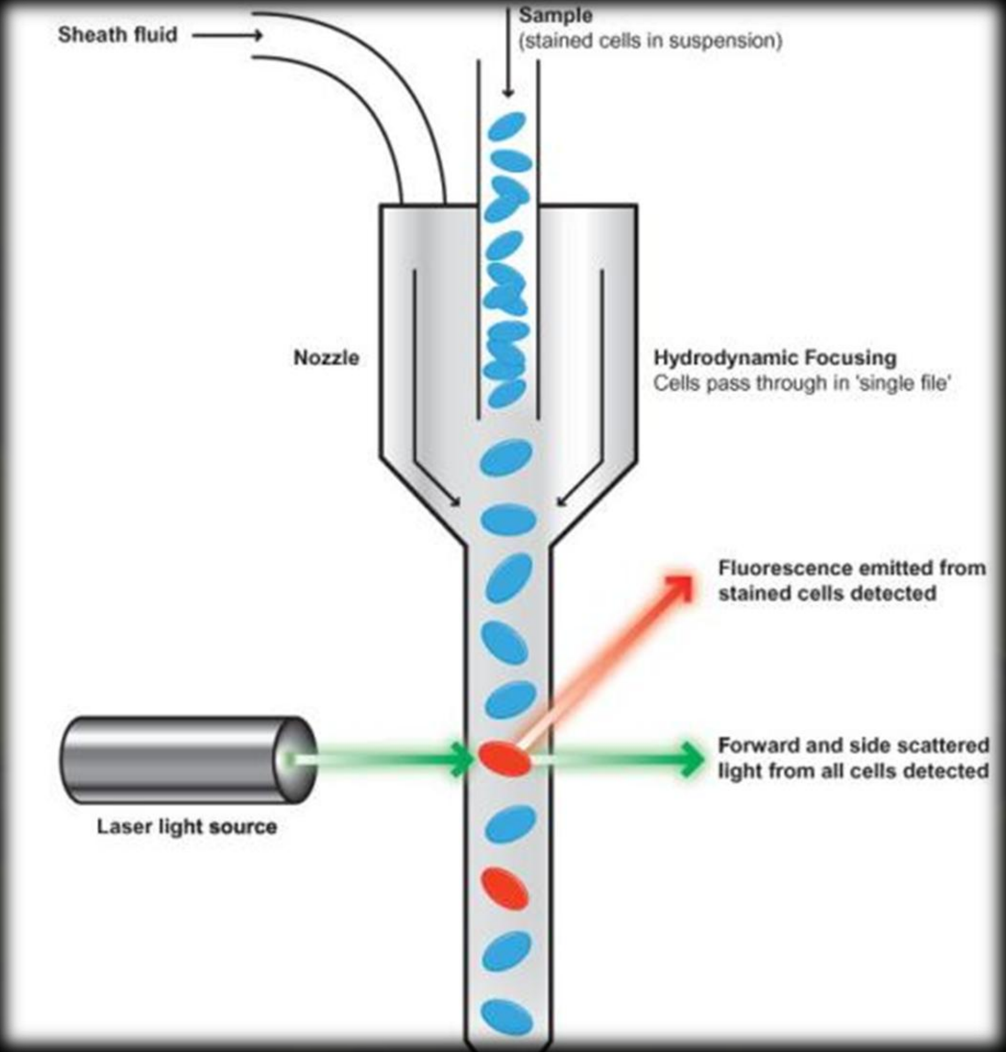

Lowry Digital


DIGITALGLOBE


BAE SYSTEMS

Reducing Radiation in CT

Flow Cytometry : Finding Cancer Cells



Flow Cytometer

Post-Katrina Hurricane: What Google Showed

What Was Really There : Digital Globe

Why is HPC important to NVIDIA?

Drives innovation in the GPU

Incredible growth opportunity

Solving really important problems

Max Planck Institute

National Center for Supercomputing Applications

Tokyo Institute of Technology

CSIRO

Developing With GPUs

Numerical Packages

MATLAB
Mathematica
NI LabView
pyCUDA

Debuggers & Profilers

cuda-gdb
NV Visual Profiler
“Nexus” VS 2008
Allinea
TotalView

GPU Compilers

C
C++
Fortran
OpenCL
DirectCompute
Java
Python

Parallelizing Compilers

PGI Accelerator
CAPS HMPP
mCUDA
OpenMP

Libraries

BLAS
FFT
LAPACK
NPP
Video
Imaging

CUDA C/C++ Leadership: 7 Releases in 3 Years

2007 | 2008 | 2009 | 2010

July 07 — CUDA Toolkit 1.0
- C Compiler
- C Extensions
- Single Precision
- BLAS
- FFT
- SDK
  40 examples

Nov 07 — CUDA Toolkit 1.1
- Win XP 64
- Atomics support
- Multi-GPU support

April 08 — CUDA Visual Profiler 2.2

cuda-gdb
HW Debugger

Aug 08 — CUDA Toolkit 2.0
- Double Precision
- Compiler Optimizations
- Vista 32/64
- Mac OSX
- 3D Textures
- HW Interpolation

July 09 — CUDA Toolkit 2.3
- DP FFT
- 16-32 Conversion intrinsics
- Performance enhancements

Nov 09 — Parallel Nsight Beta

Mar 10 — CUDA Toolkit 3.0
- C++ inheritance
- Fermi arch support
- Tools updates
- Driver / RT interop

NVIDIA OpenCL

- Compute Capability
- Compiler Flags
- Byte Addr. Stores
- Images
- Global Atomics Base
- Global Atomics Extd.

NVIDIA released
R190 Driver
September 2009

- Compute Capability
- Compiler Flags
- Double Precision
- OpenGL Sharing
- Byte Addr. Stores
- Images
- Global Atomics Base
- Global Atomics Extd.
- ICD (beta)

NVIDIA released
R195 Driver
November 2009

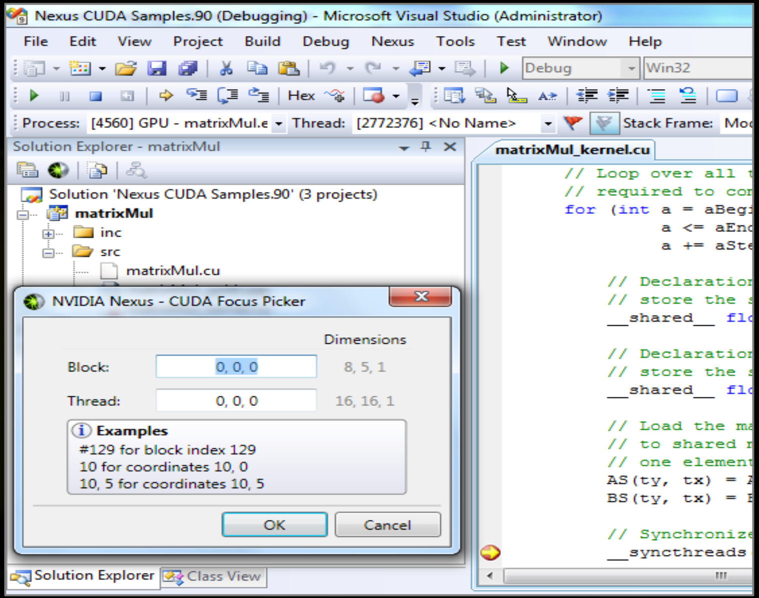
- D3D11 Sharing
- D3D10 Sharing
- D3D9 Sharing
- Pragma Unroll
- Compute Capability
- Compiler Flags
- Double Precision
- OpenGL Sharing
- Byte Addr. Stores
- Images
- Local Atomics Base
- Local Atomics Extd.
- Global Atomics Base
- Global Atomics Extd.
- ICD

NVIDIA released
R195 Refresh Driver
March 2010

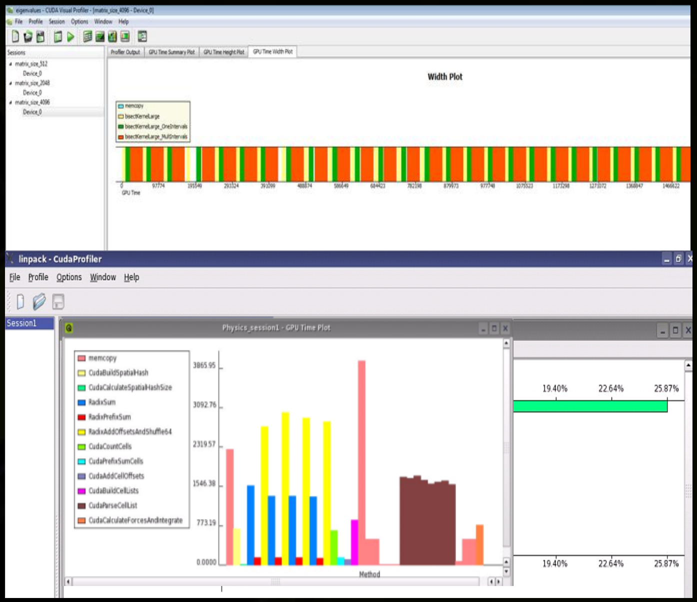
Vertical Extensions

Khronos Extensions

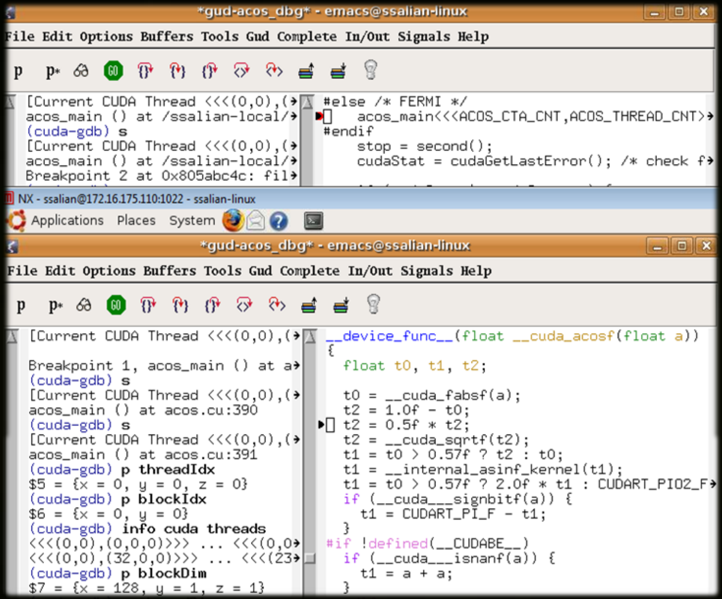
Parallel Nsight Visual Studio



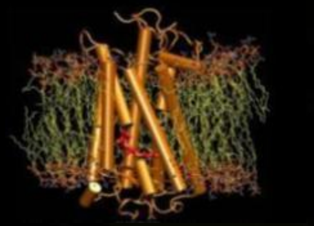
Visual Profiler For Linux



cuda-gdb For Linux

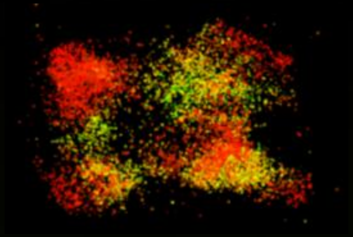


Key Research Applications Available on GPUs

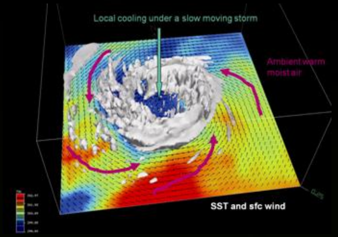


Molecular Dynamics / Quantum Chemistry

Computational Fluid Dynamics



Astrophysics



Weather & Climate Modeling

Many More

AMBER
ABINIT
DL_POLY
GROMACS
LAMMPS
NAMD
TeraChem

OpenCurrent
BAE Systems
Acusim
Euler Solvers
Lattice Boltzman
Navier Stokes

N-body
Chimera
GADGET2
Many published papers

ASUCA (Japan)
CO2 Modeling (Japan)
HOMME
Tsunami modeling
NOAA NIM
WRF

- **Materials Science**
  - **DCA++**
  - **gWL-LSMS**
- **Combustion**
  - **S3D**
- **Lattice QCD**
  - **Chroma (QUDA)**

GPU TECHNOLOGY CONFERENCE

3 days

133 hours of technical content

60 startups

Over 1,400 attendees from 40 countries

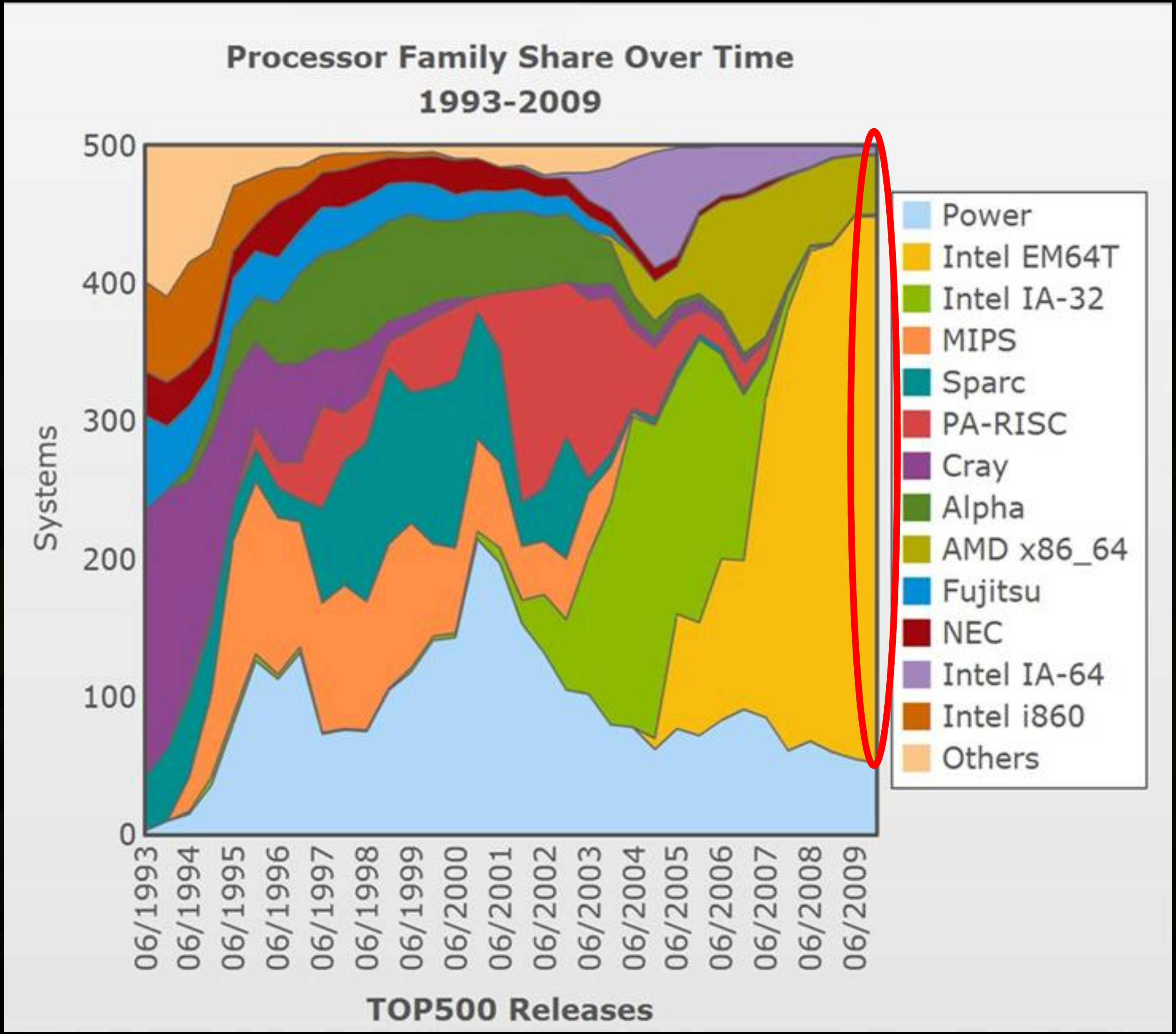
91 Research Posters

Do GPUs apply to my work?

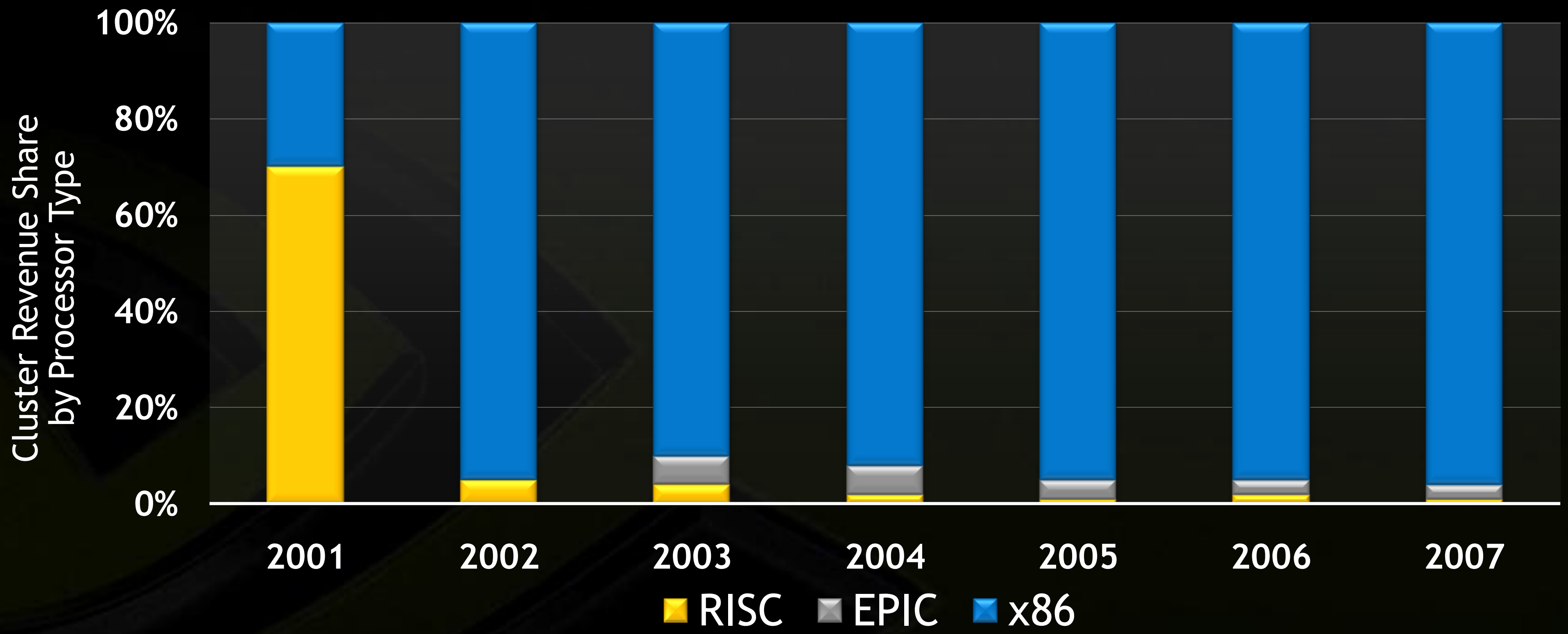
Successfully used in supercomputing

Research quality of software

Remarkable results from students


Processor Family Share Over Time
1993-2009
Systems
500
400
300
200
100
0
06/1993
06/1994
06/1995
06/1996
06/1997
06/1998
06/1999
06/2000
06/2001
06/2002
06/2003
06/2004
06/2005
06/2006
06/2007
06/2008
06/2009
TOP500 Releases
Power
Intel EM64T
Intel IA-32
MIPS
Sparc
PA-RISC
Cray
Alpha
AMD x86_64
Fujitsu
NEC
Intel IA-64
Intel i860
Others

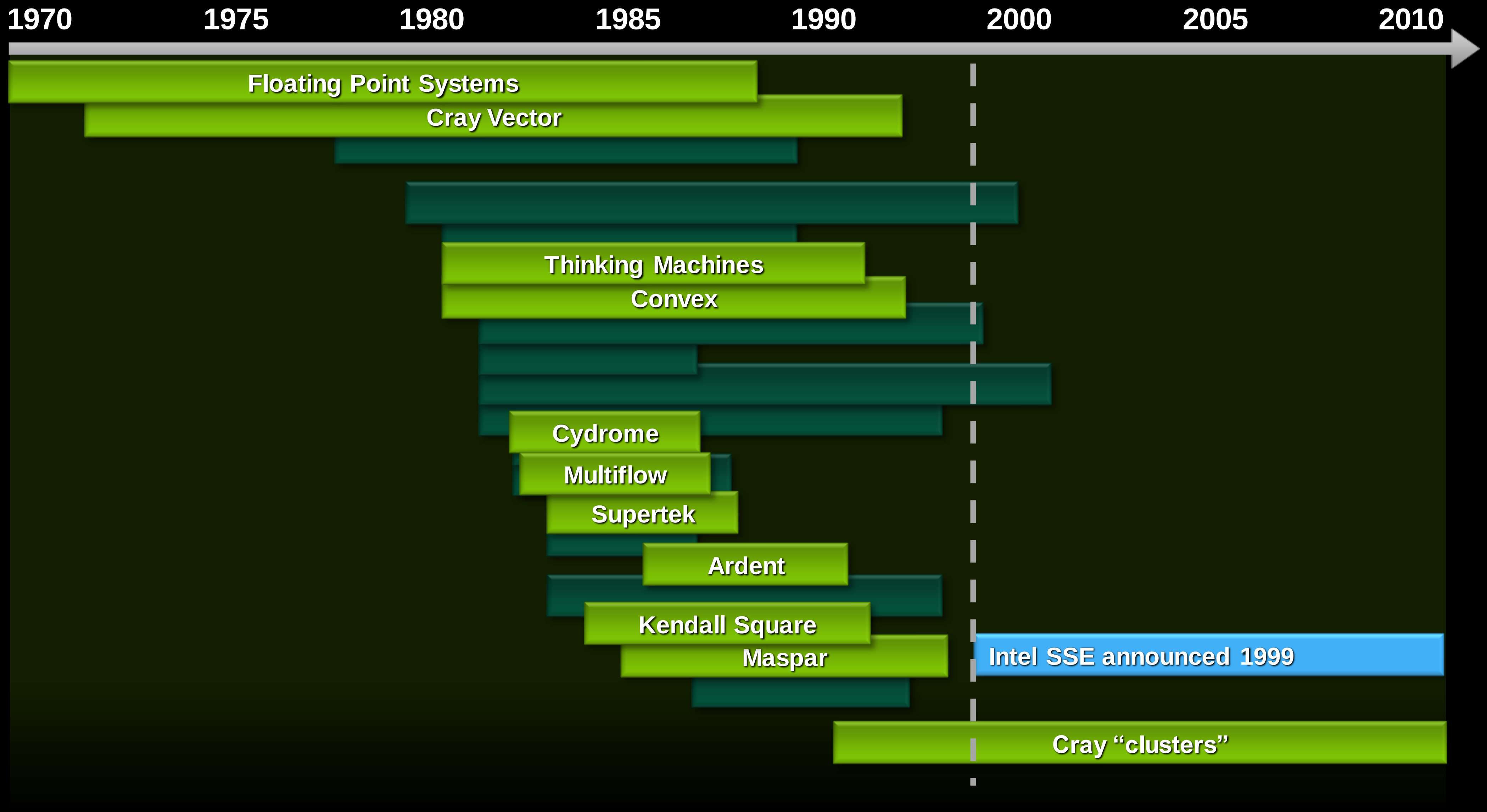
Commodity CPUs Dominate HPC



2X every 18 months AND cheap!

Source: IDC 2008

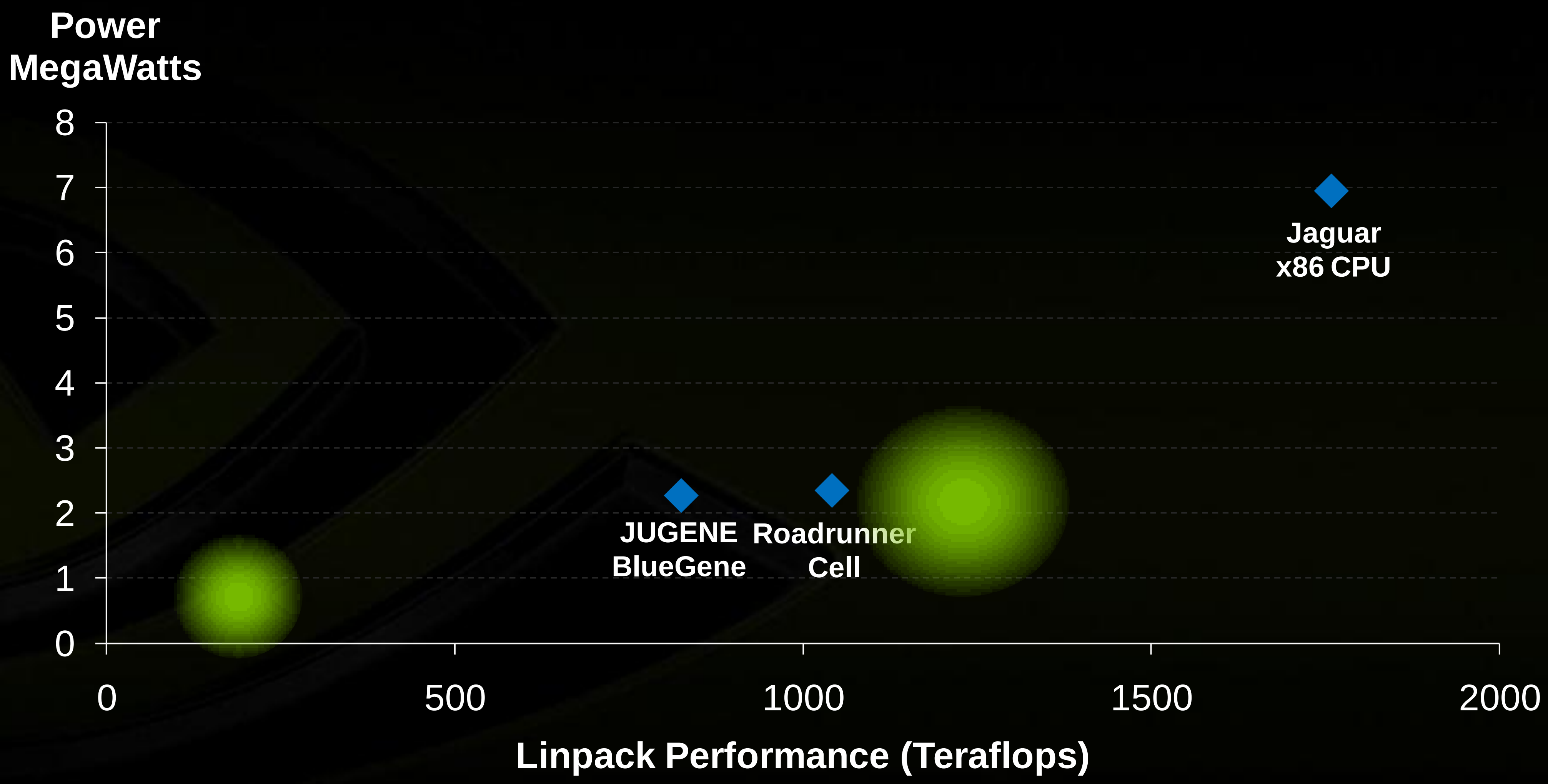
In Supercomputing, Perf/$ = Perf



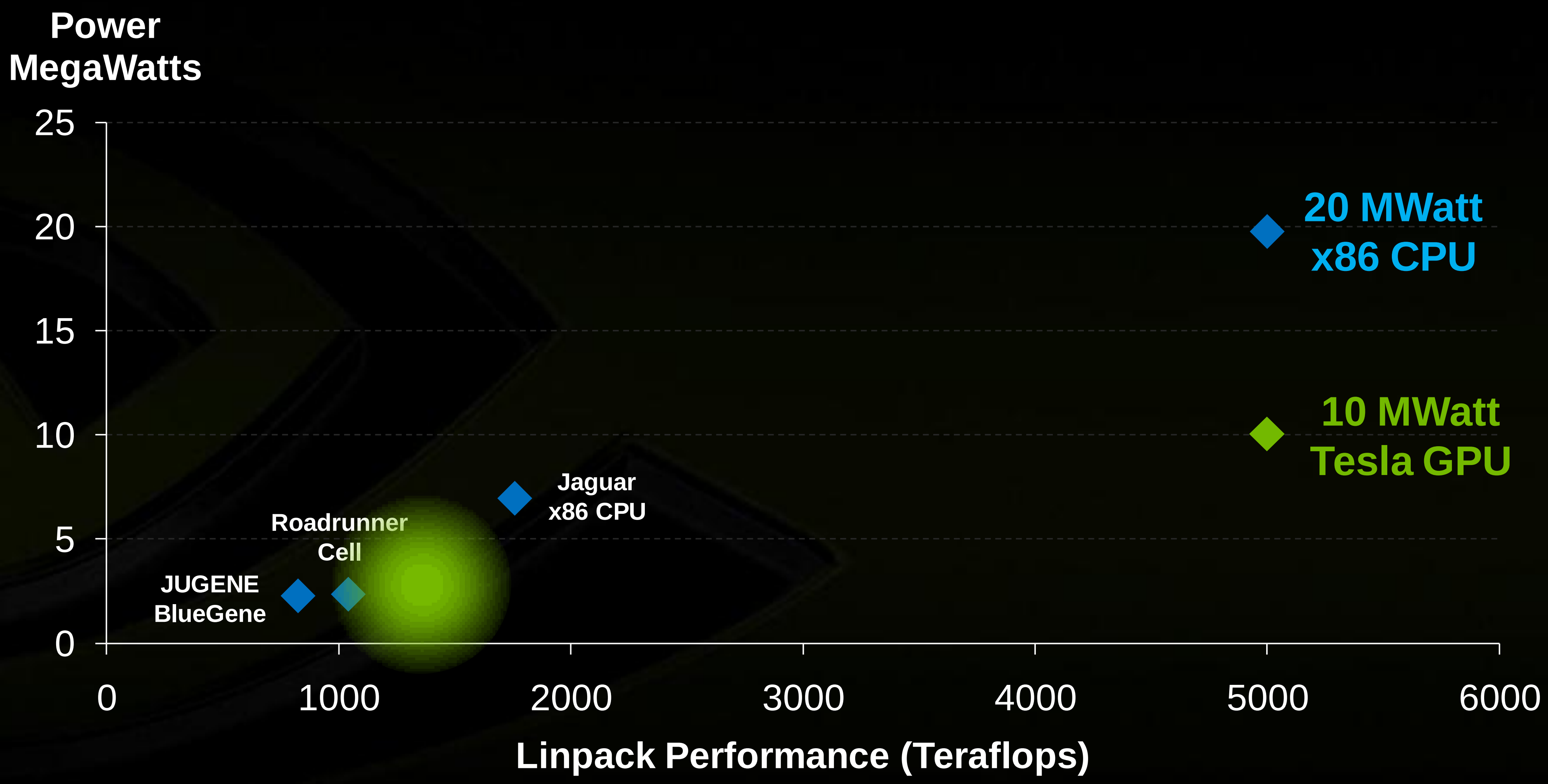
$500,000,000

$2,000,000,000

$13,000


Power
MegaWatts
8
7
6
5
4
3
2
1
0
Jaguar
x86 CPU
JUGENE
BlueGene
Roadrunner
Cell
0
500
1000
1500
2000
Linpack Performance (Teraflops)

Scaling to 5 Petaflop Cluster



Why do we need something new?

Innovation

Performance Scaling

Power Scaling

Mass market adoption

Maturing applications

Future scaling