**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

NVIDIA R&D Budget

$ Millions


$ Millions

0 100 200 300 400 500 600 700 800 900

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

High Performance Computing

# of Top 500 Supercomputers by Architecture

- IA powering 80% of world’s Top 500 supercomputers
- IA in 39% of today’s Top100 supercomputers
- Estimate by 2013, TAM of Top100 supercomputers will be 1M units
Reducing Radiation in CT
Flow Cytometry: Finding Cancer Cells
What Was Really There: Digital Globe
Why is HPC important to NVIDIA?

- Drives innovation in the GPU
- Incredible growth opportunity
- Solving really important problems
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

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

- **CUDA Toolkit 1.1** (Nov 07)
  - Win XP 64
  - Atomics support
  - Multi-GPU support

- **CUDA Visual Profiler 2.2** (April 08)
  - cuda-gdb
  - HW Debugger

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

- **CUDA Toolkit 2.3** (July 09)
  - DP FFT
  - 16-32 Conversion intrinsics
  - Performance enhancements

- **Parallel Nsight Beta** (Nov 09)

- **CUDA Toolkit 3.0** (Mar 10)
  - 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 Extld.**
- **OpenGL Sharing**
- **Double Precision**

NVIDIA released
R190 Driver
September 2009

NVIDIA released
R195 Driver
November 2009

NVIDIA released
R195 Refresh Driver
March 2010

- **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 Extld.**
- **Global Atomics Base**
- **Global Atomics Extld.**
- **ICD**

- **Vendor Extensions**
- **Khronos Extensions**
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 Boltzmann
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)
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
Commodity CPUs Dominate HPC

Cluster Revenue Share by Processor Type

- RISC
- EPIC
- x86

2X every 18 months AND cheap!

Source: IDC 2008
In Supercomputing, Perf/\$ = Perf

- Cray Vector
- Floating Point Systems
- Thinking Machines
- Convex
- Cydrome
- Multiflow
- Supertek
- Ardent
- Kendall Square
- Maspar
- Intel SSE announced 1999
- Cray “clusters”
$500,000,000

$2,000,000,000,000

$13,000
Scaling to 5 Petaflop Cluster

- **Power (MegaWatts)**
  - 25
  - 20
  - 15
  - 10
  - 5
  - 0

- **Linpack Performance (Teraflops)**
  - 6000
  - 5000
  - 4000
  - 3000
  - 2000
  - 1000
  - 0

- **Systems**
  - JAGUAR x86 CPU
  - Roadrunner Cell
  - JUGENE BlueGene
  - 20 MWatt x86 CPU
  - 10 MWatt Tesla GPU
Why do we need something new?

- Innovation
- Performance Scaling
- Power Scaling
Mass market adoption
Maturing applications
Future scaling