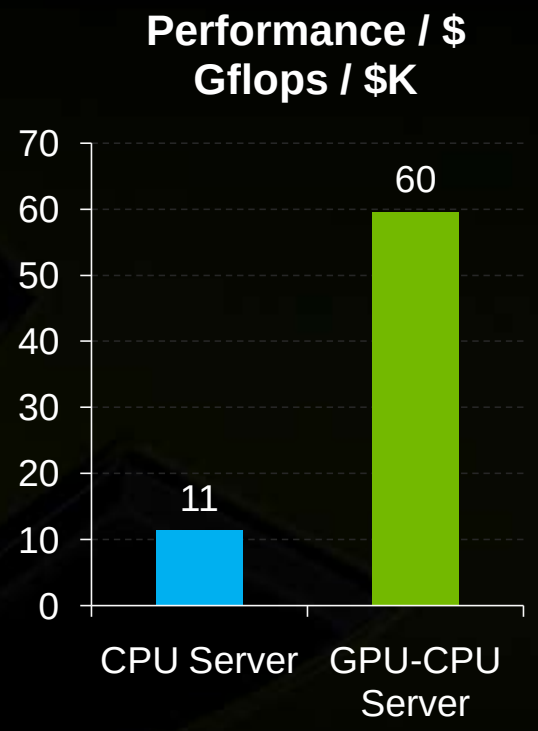
Welcome

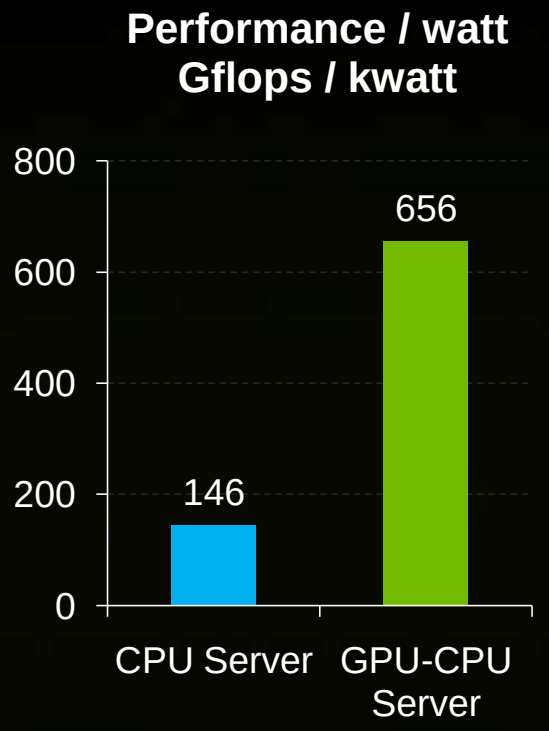
- Goals:
  - An introduction to High Performance Computing with CUDA
  - Help you get started developing and optimizing CUDA applications

- Outline
  - Motivation and introduction
  - CUDA C/C++ basics
  - CUDA libraries and CUDA Fortran
  - Analysis and optimization
  - Lessons learned in production codes

GPUs are Fast!

8x Higher Linpack





CPU 1U Server: 2x Intel Xeon X5550 (Nehalem) 2.66 GHz, 48 GB memory, $7K, 0.55 kw

GPU-CPU 1U Server: 2x Tesla C2050 + 2x Intel Xeon X5550, 48 GB memory, $11K, 1.0 kw

Tesla GPUs Power 3 of Top 5 Supercomputers

#1 : Tianhe-1A
7168 Tesla GPU's 2.5 PFLOPS

#3 : Nebulae
4650 Tesla GPU's 1.2 PFLOPS

#4 : Tsubame 2.0
4224 Tesla GPU's 1.194 PFLOPS

"We not only created the world's fastest computer, but also implemented a heterogeneous computing architecture incorporating CPU and GPU, this is a new innovation."

Premier Wen Jiabao
Public comments acknowledging Tianhe-1A

World's Greenest Petaflop Supercomputer

Tsubame 2.0

Tokyo Institute of Technology

- 1.19 Petaflops
- 4,224 Tesla M2050 GPUs

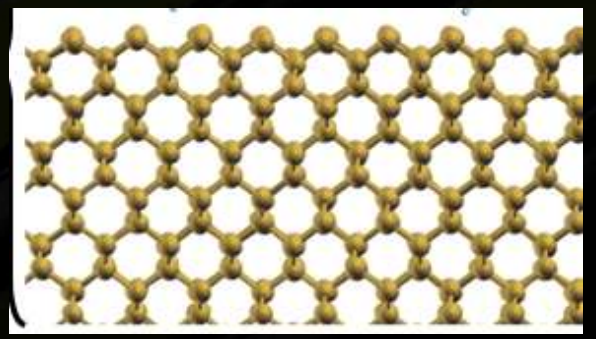
World's Fastest MD Simulation

Sustained Performance of 1.87 Petaflops/s

Institute of Process Engineering (IPE)

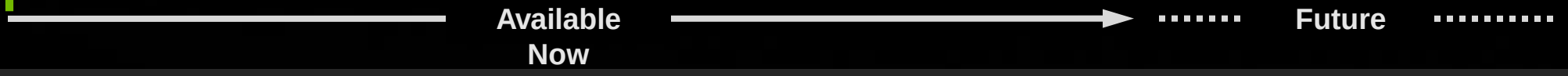
Chinese Academy of Sciences (CAS)

MD Simulation for Crystalline Silicon



Used all 7168 Tesla GPUs on
Tianhe-1A GPU Supercomputer

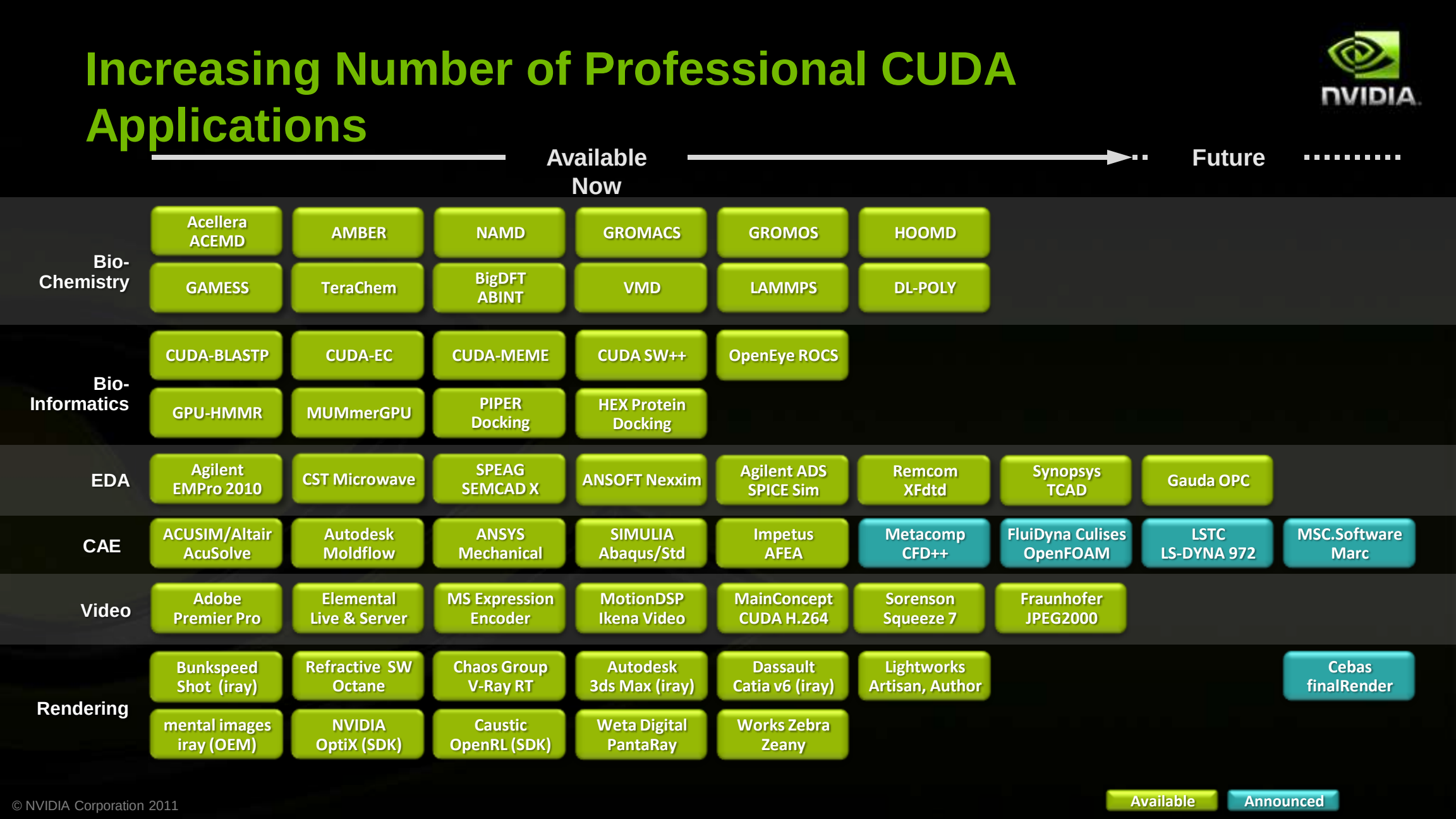
Increasing Number of Professional CUDA Applications



	Available Now							Future	
Tools & Libraries	CUDA C/C++	Parallel Nsight Vis Studio IDE	NVIDIA Video Libraries	ParaTools VampirTrace	PGI Accelerators	EMPhotonics CULAPACK	Allinea DDT Debugger	TauCUDA Perf Tools	PGI CUDA-X86
	NVIDIA NPP Perf Primitives	PGI Fortran	Thrust C++ Template Lib	Bright Cluster Manager	CAPS HMPP	MAGMA	GPU Packages For R Stats Pkg	Platform LSF Cluster Mgr	GPU.net
	pyCUDA	R-Stream Reservoir Labs	PBSWorks	MOAB Adaptive Comp	Torque Adaptive Comp	TotalView Debugger	IMSL		
Oil & Gas	Headwave Suite	OpenGeo Solns OpenSEIS	GeoStar Seismic	Acceleware RTM Solver	StoneRidge RTM	Seismic City RTM	Tsunami RTM		Schlumberger Petrel
	ffA SVI Pro	Paradigm SKUA	VSG Open Inventor	Paradigm GeoDepth RTM	VSG Avizo	SVI Pro	SEA 3D Pro 2010	Schlumberger Omega	Paradigm VoxelGeo
Numerical Analytics	LabVIEW Libraries	AccelerEyes Jacket: MATLAB	MATLAB	Mathematica					
Finance	NAG RNG	Numerix CounterpartyRisk	SciComp SciFinance	Aquimin AlphaVision	Hanweck Volera Options Analysi	Murex MACS			
Other	Siemens 4D Ultrasound	Digisens CT	Schrodinger Core Hopping	Useful Prog Medical Imag	ASUCA Weather Model				
	Manifold GIS	MVTech Mach Vision	Dalsa Mach Vision	WRF Weather					

Available | Announced

Increasing Number of Professional CUDA Applications



CUDA by the Numbers

300,000,000	CUDA Capable GPUs
500,000	CUDA Toolkit Downloads
100,000	Active CUDA Developers
400	Universities Teaching CUDA
100	% OEMs offer CUDA GPU PCs

GPU Computing Applications

Libraries & Middleware

CUBLAS	CUFFT	CULAPACK	NPP & CUDPP	Video	PhysX Physics	OptiX Ray tracing	mental ray iray Rendering	Reality Server 3D web services

C	C++	OpenCL™	Direct Compute	Fortran	Java & Python

NVIDIA GPU with CUDA Parallel Computing Architecture

Fermi architecture (compute capability 2.x)	GeForce 500 series GeForce 400 series	Quadro Fermi series	Tesla 20 series
Tesla architecture (compute capability 1.x)	GeForce 200 series GeForce 9 series GeForce 8 series	Quadro FX series QuadroPlex series Quadro NVS series	Tesla 10 series

NVIDIA Developer Ecosystem

NVIDIA

Numerical Packages

MATLAB
Mathematica
NI LabView
pyCUDA

Debuggers & Profilers

cuda-gdb
NV Visual Profiler
Parallel Nsight
Visual Studio
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
GPULib

GPGPU Consultants & Training

acceleware · GASS · STONE RIDGE TECHNOLOGY · ANEO · GPU Tech · TECH-X · CAPS · HPC · EM Photonics · SCALABLE GRAPHICS · WIPRO

OEM Solution Providers

DELL · hp · IBM · CRAY · ASUS · SUPERMICRO · sgi · FUJITSU · BULL · APPRO · lenovo 联想 · NEC

GPU TECHNOLOGY CONFERENCE

GPU Technology Conference Worldwide Events

GTC Workshop Japan, Tokyo, July 22, 2011

Co-hosted with the Tokyo Institute of Technology and bringing together top researchers, scientists and industry leaders to focus on critical research, trends and opportunities in GPU computing.

GTC China, Beijing, December 15-16, 2011

Focusing on the very latest scientific research and commercial applications in GPU computing.

GTC 2012, San Jose, CA, May 14-17, 2012

Advancing awareness of High Performance Computing and the transformational impact of GPUs.

www.gputechconf.com

INPAR2012

INNOVATIVE PARALLEL COMPUTING

Foundations & Applications of GPU, Manycore, and Heterogeneous Systems

San Jose, CA / May 13-14, 2012

- *InPar provides a academic venue for peer-reviewed, archival publication in the emerging fields of parallel computing*
- *Call for Papers*

  Seeking papers involving current GPU/manycore architectures, new or emerging commodity parallel architectures (such as Intel "MIC" products), and hybrid or heterogeneous systems.
- *Join the InPar 2012 Mailing List at innovativeparallel.org*
- *InPar 2012 is co-located with NVIDIA's* ***GPU Technology Conference***.

gpucomputing.net is a research and development community that fosters collaborative domain-focused GPU research across disciplines.

- 5,175 Papers, Events, Forums, & Job Postings
- In 43 Communities

gpucomputing.net

Connect ▪ Communicate ▪ Collaborate

NVIDIA at ISC”11

- NVIDIA Booth #630
- GPU Debate – The Fast Lane on the Road to Better Science: Tuesday, June 21, 2011

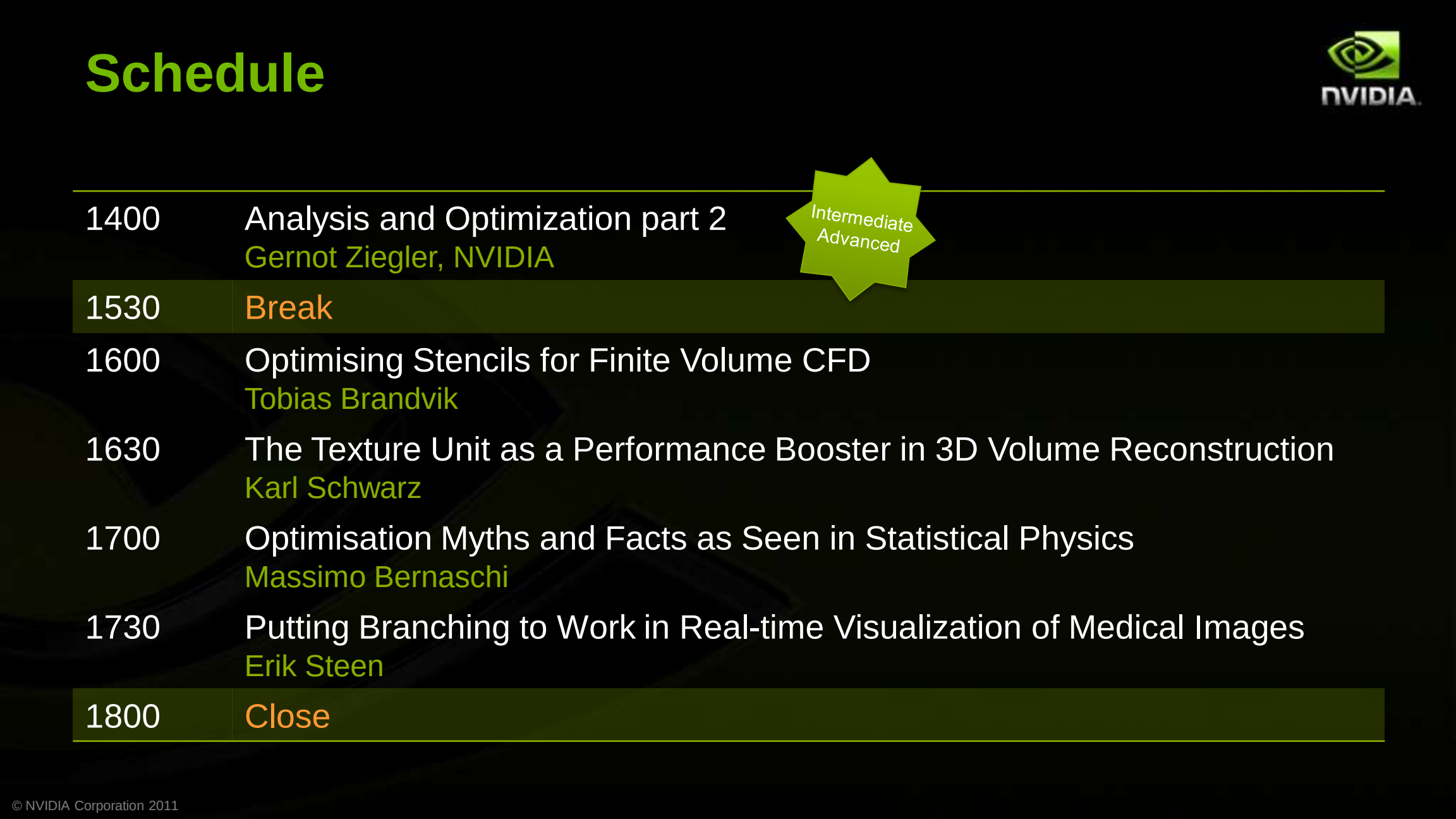
  Come and see Thomas Sterling from Louisiana State University and David Kirk from NVIDIA. The debate will be chaired by Horst Simon from Lawrence Berkeley National Laboratory.
- Presentations of the CUDA Tutorial talks available on Monday at

  http://www.nvidia.com/object/isc2011.html

Schedule

0900	Introduction
0915	CUDA C/C++ Basics Thomas Bradley, NVIDIA
1030	Break
1100	CUDA Libraries and CUDA Fortran Massimilano Fatica, NVIDIA
1145	Analysis and Optimization part 1 Tim Schröder, NVIDIA
1300	Lunch

Schedule



1400	Analysis and Optimization part 2 Gernot Ziegler, NVIDIA
1530	Break
1600	Optimising Stencils for Finite Volume CFD Tobias Brandvik
1630	The Texture Unit as a Performance Booster in 3D Volume Reconstruction Karl Schwarz
1700	Optimisation Myths and Facts as Seen in Statistical Physics Massimo Bernaschi
1730	Putting Branching to Work in Real-time Visualization of Medical Images Erik Steen
1800	Close