GPU Progress and Directions for Earth System Modeling

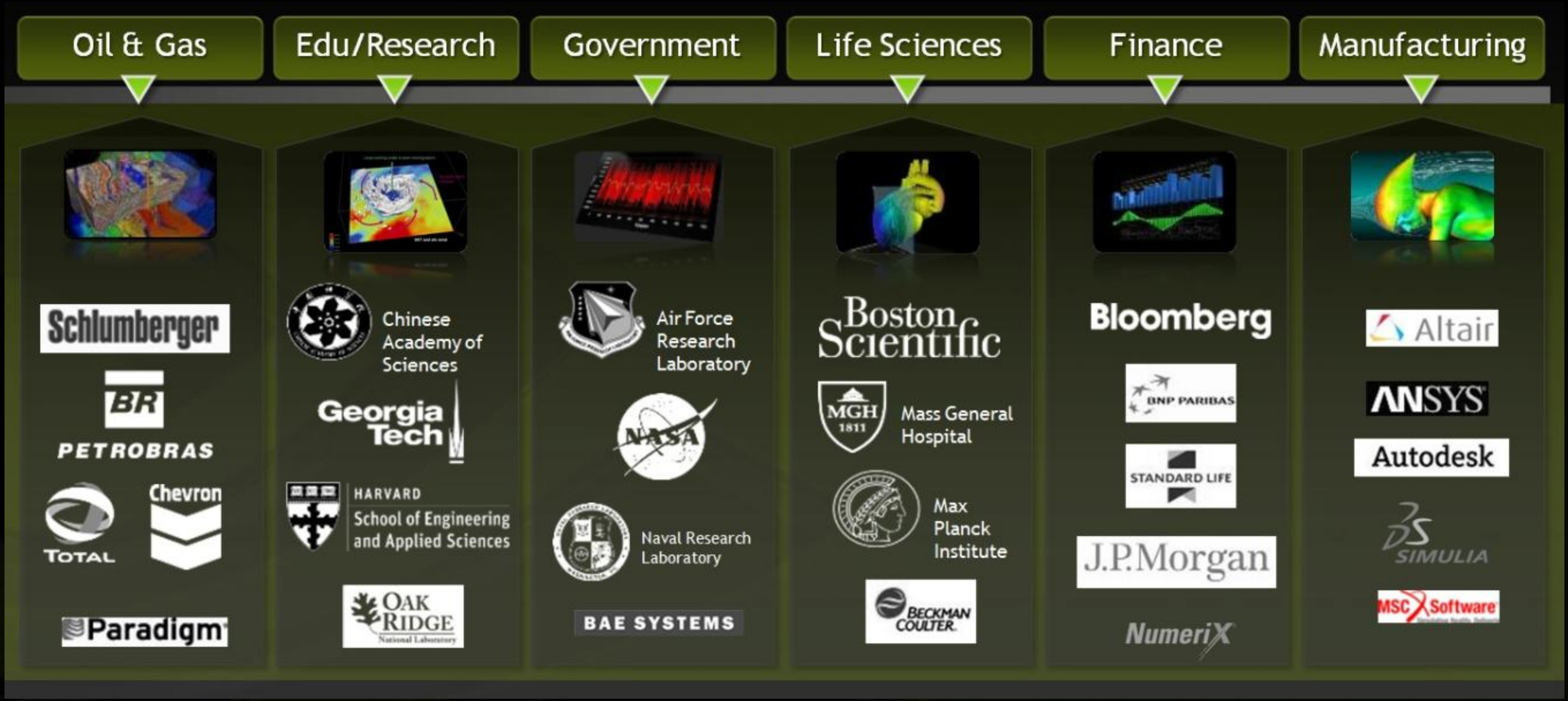
Stan Posey, sposey@nvidia.com
NVIDIA, Santa Clara, CA, USA

Agenda: GPU Progress and Directions for Earth System Modeling

- Introduction of GPUs in HPC
- NVIDIA Application Strategy
- GPU Progress in ES Modeling
- NVIDIA Technology Directions

GPU Growth as Mainstream HPC Technology

NVIDIA



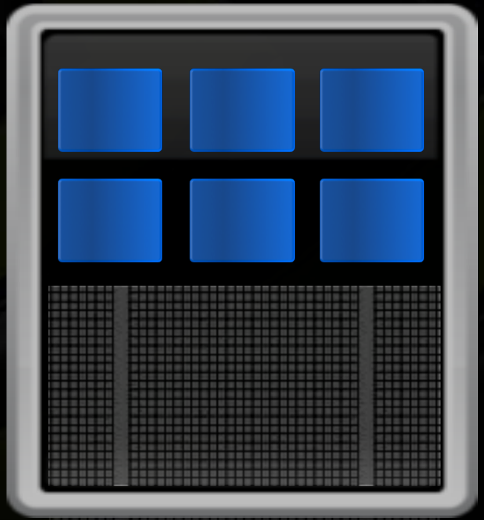
The buyer plans for including accelerators in their next technical computing server purchase has more than doubled from 29% to over 65% in last 20 months.

IDC Market Research
April, 2013

All Major OEM Servers Offer NVIDIA GPUs

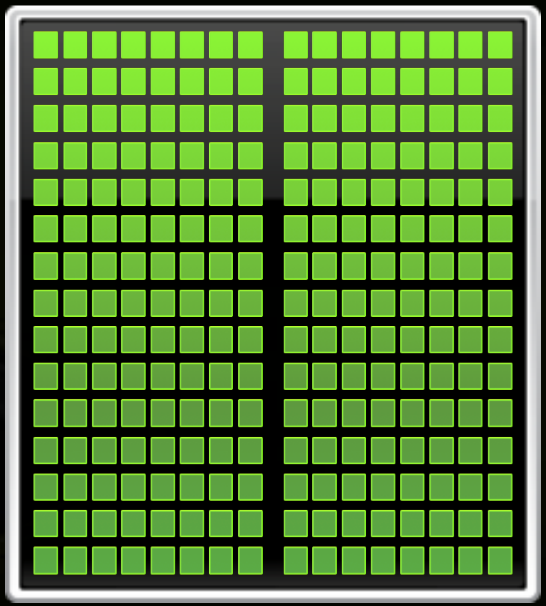
Schematic of GPU Accelerated Computing

CPU
Optimized for
Serial Tasks



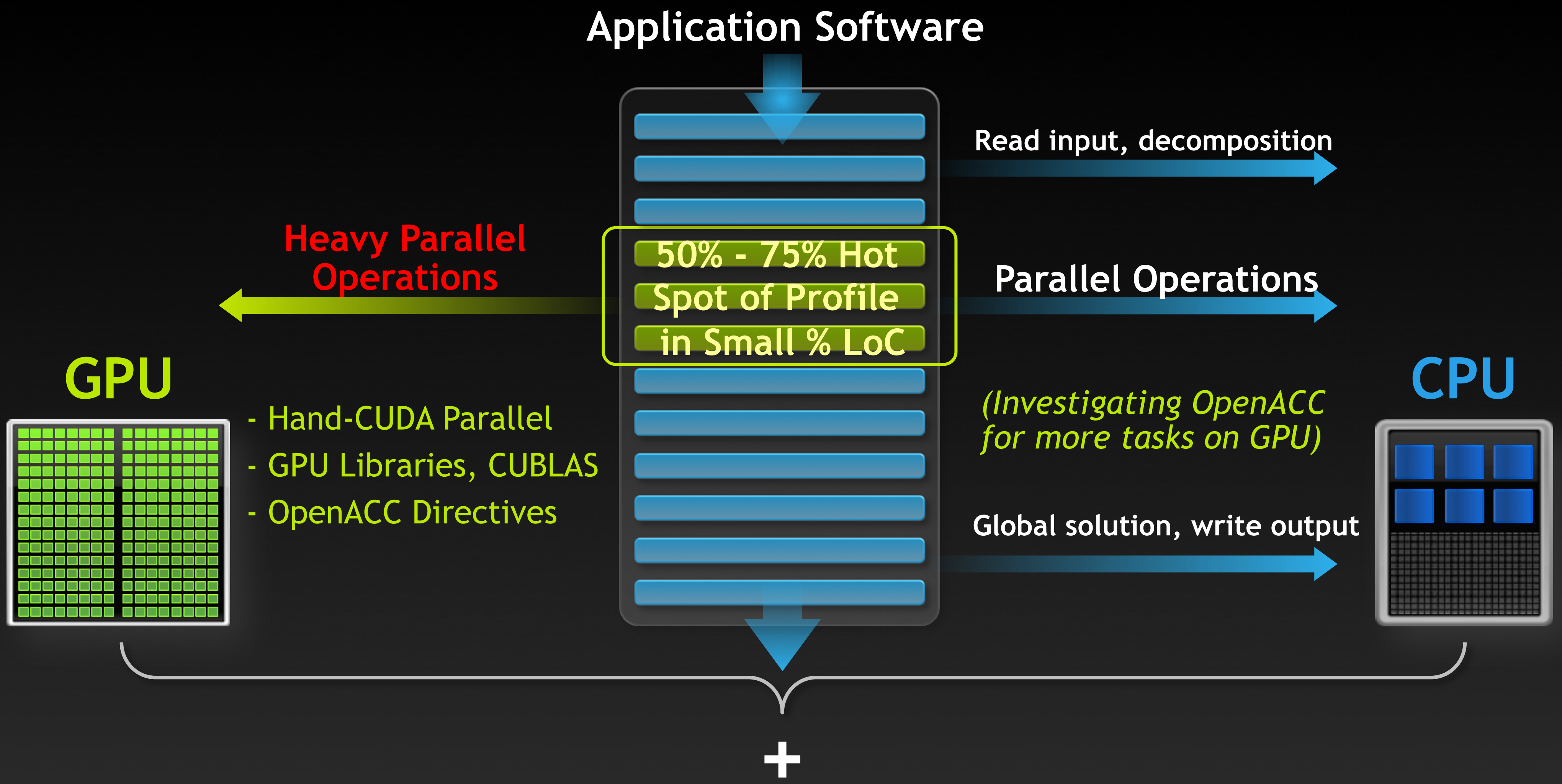
\+

GPU Accelerator
Optimized for
Parallel Tasks



- 10x Peak Performance
- 5x Energy Efficiency

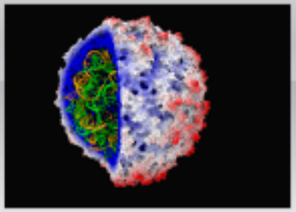
How Applications Are Usually Accelerated



NVIDIA HPC Marketing ~2009 . . .

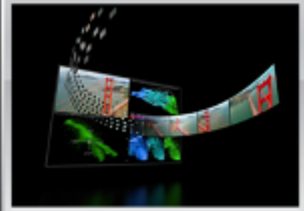
146X

Medical Imaging
U of Utah



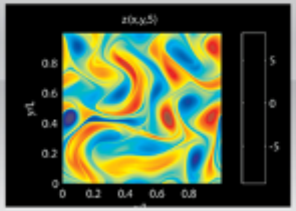
36X

Molecular Dynamics
U of Illinois, Urbana



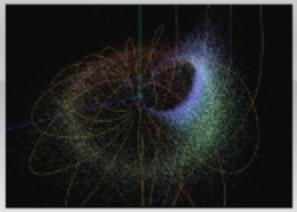
18X

Video Transcoding
Elemental Tech



50X

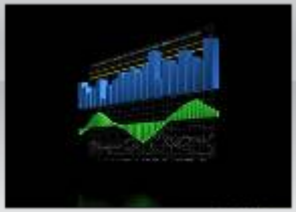
Matlab Computing
AccelerEyes



100X

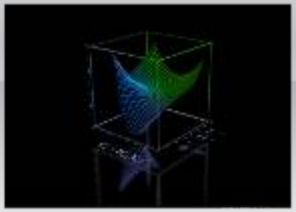
Astrophysics
RIKEN

Real Application Speedups



149X

Financial Simulation
Oxford



47X

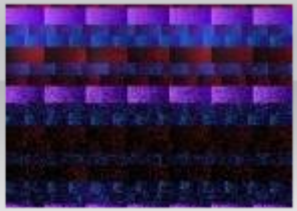
Linear Algebra
Universidad Jaime

20X

3D Ultrasound
Techniscan

130X

Quantum Chemistry
U of Illinois, Urbana



30X

Gene Sequencing
U of Maryland

Context!?

- Full application? *Often kernel only without data transfer . . .*

- What is the reference CPU? *Often old and dusty x86 . . .*

- How many CPU cores in the comparison?

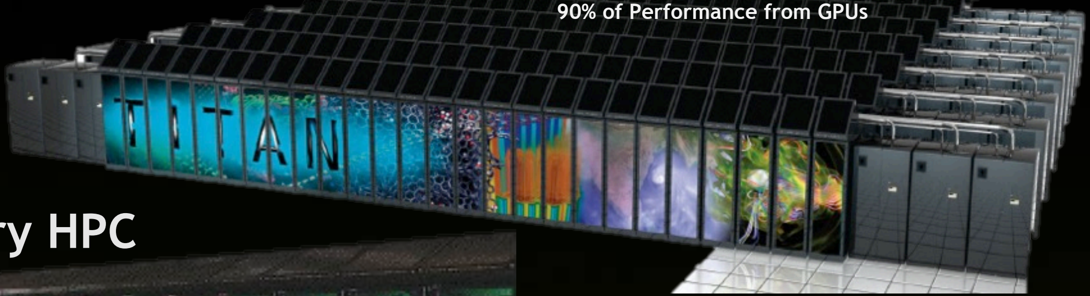
Often 1 core . . . but who uses only 1 core nowadays?

NOTE: Missing context often fault of NVIDIA and not the organizations referenced

Next Migration Underway: Accelerated HPC

Accelerated HPC

Titan: 18,688 Tesla K20X GPUs
27 Petaflops Peak, 17.59 Petaflops on Linpack
90% of Performance from GPUs



Distributed Memory HPC

Shared Memory HPC

Vector HPC

1980 1990 2000 2010 2020

GPU-Driven Fast and Energy Efficient HPC

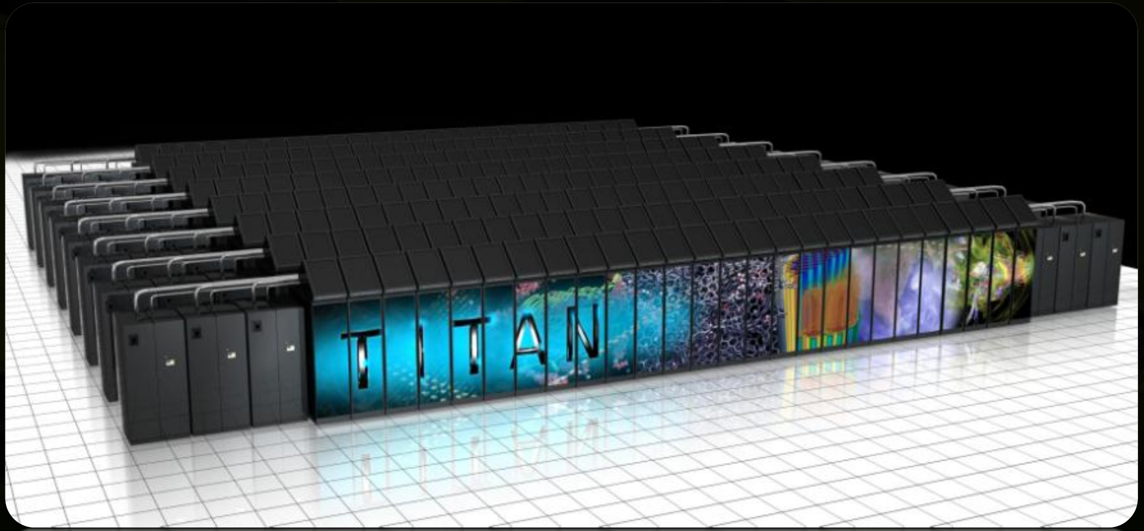
TITAN at ORNL

World's Fastest Open Science Supercomputer

18,688 Tesla K20X GPU Accelerators

27 Petaflops Peak

90% of Performance from GPUs



Eurora at CINECA

World's Most Energy Efficient Supercomputer

128 Tesla K20 GPU Accelerators

3150 MFLOPS/Watt

\$100k Energy & 300 Tons of CO_2 Saving Per Year

GPU Acceleration at Leadership HPC Sites

CRAY THE SUPERCOMPUTER COMPANY
DELL
hp
IBM
sgi

United States
Oak Ridge National Labs
Lawrence Livermore National Labs
Sandia National Labs
NOAA Gaea (ORNL)
NCSA Blue Waters
NCAR Yellowstone (Geyser & Caldera)
NASA Pleadies; Discover

Germany
Juelich
HLRS
Max Planck
TU Dresden

UK
Cambridge
EPCC
Oxford
STFC

Rest of Europe
BSC, Spain
CINECA, Italy
CEA, France
CSCS, Switzerland

Japan
Tokyo Tech
RIKEN
Tsukuba

China
NSC, Shenzhen
NSC, Tianjin
CAS IPE

Rest of World
MSU, Russia
RAS, Russia
IITs, India

Important OEM Collaborations in ES Modeling

Collaboration on large deployments; OpenACC development

- TITAN —ORNL; Blue Waters —NCSA; Gaea —NOAA/ORNL; Piz Daint —CSCS

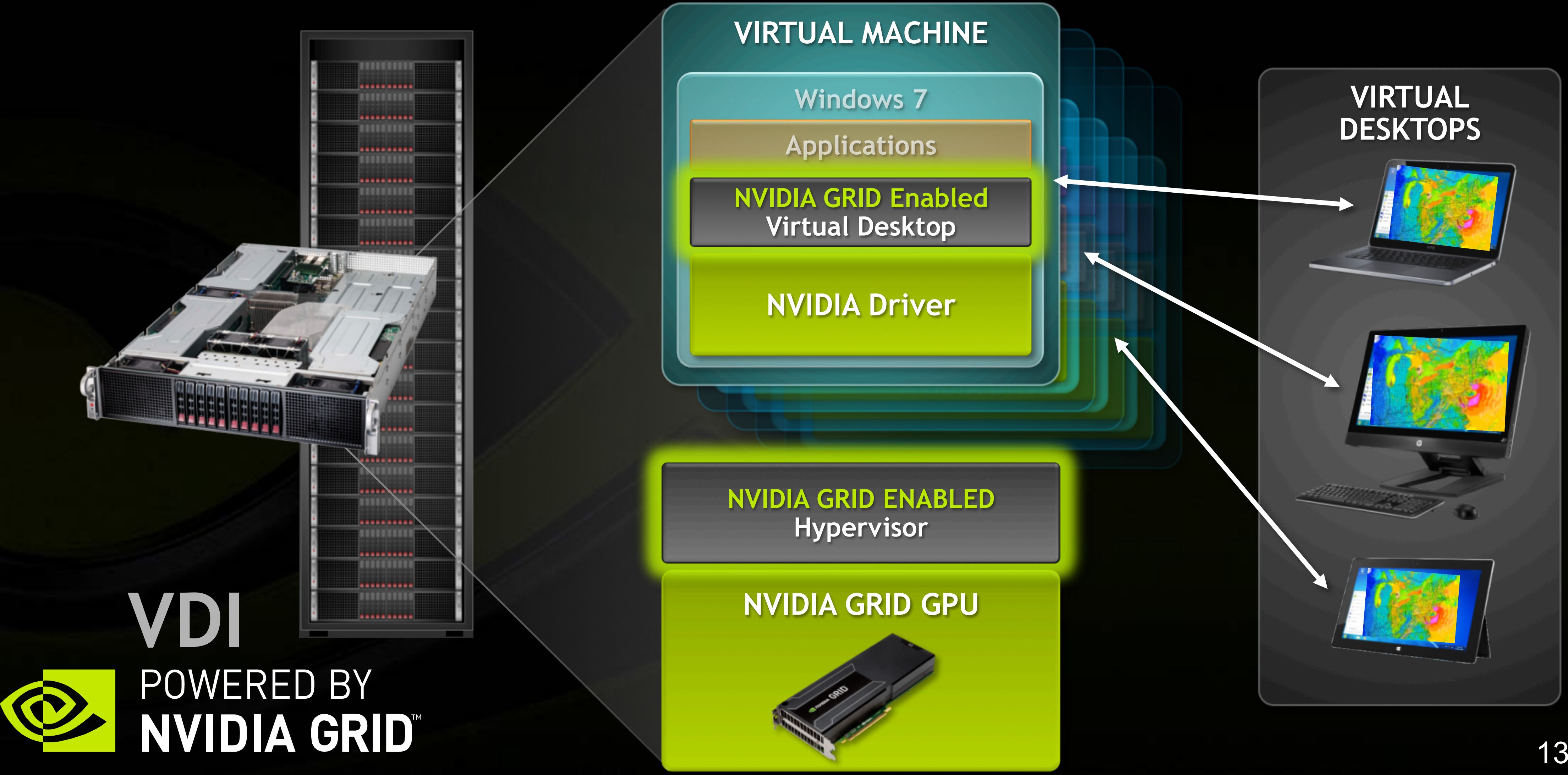
Collaboration on strategic deployments; Member Openpower

- Yellowstone — NCAR; Discover — NASA GSFC
- Openpower Consortium: IBM, NVIDIA, Google, Mellanox, Tyan, others

Strategic large deployments including Tsubame – Tokyo Inst of Tech

Strategic deployments including Pleadies – NASA ARC

Remote Visualization With GPU-Driven VDI



GPU Adoption Underway in Data Analytics

NVIDIA

Beyond HPC

GPU-Driven Big Data

Analyzing Twitter	salesforce, Twitter
Searching Audio	Shazam
Visual Shopping	CORTEXICA vision systems, ebay
Real-time Video Delivery	ELEMENTAL, ESPN

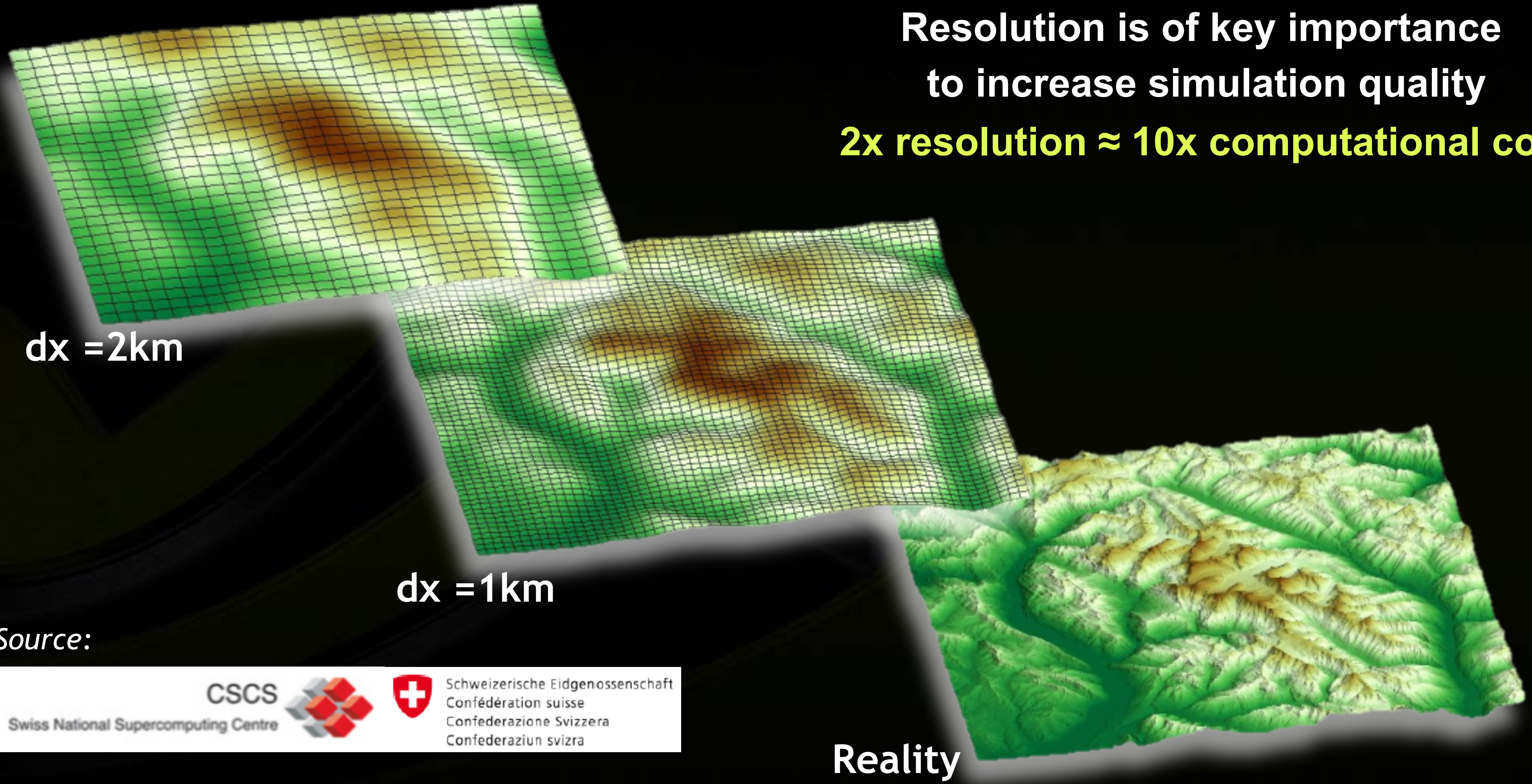
Agenda: GPU Progress and Directions for Earth System Modeling

- Introduction of GPUs in HPC
- **NVIDIA Application Strategy**
- GPU Progress in ES Modeling
- NVIDIA Technology Directions

GPU Motivation: Higher Resolution = High Cost

NVIDIA

Resolution is of key importance
to increase simulation quality
2x resolution ≈ 10x computational cost



Source:

CSCS
Swiss National Supercomputing Centre

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Example: NASA and Global Cloud Resolving GEOS-6

Programming weather, climate, and earth-system models on heterogeneous multi-core platforms

September 7-8, 2011 at the National Center for Atmospheric Research in Boulder, Colorado

The Finite-Volume Dynamical Core on GPUs within GEOS-5

- Dr. William Putman, Global Modeling and Assimilation Office, NASA GSFC

NASA targeting GEOS global model resolution at sub-10-km to 1-km range

Computational requirements for typical 5-day operational forecast:

Grid resolution	Westmere CPU cores	Comments
10 KM	12,000	Possible today
3 KM	300,000	Reasonable but not available
1 KM	10,000,000	Impractical, need acclerators

3.5-km GEOS-5 Simulated Clouds (CPU-Only)

Source: http://data1.gfdl.noaa.gov/multi-core/

NVIDIA Application Strategy Since 2010

- Initial focus on climate and atmospheric research
  - Opportunities to refactor code and use of CUDA
- Later focus on operational models with directives
  - User community imposed Fortran-only requirements
- Direct investments in applications engineering
  - Current collaborations in 6 models and growing
- Continued development in libraries and OpenACC
  - CUBLAS, CUSPARSE, collaborations with PGI, CAPS, Cray
- Collaborations with OEMs (development, benchmarks, etc.)
  - Cray, IBM, HP, SGI, etc.

Programming Strategies for GPU Acceleration

NVIDIA

Applications		
GPU Libraries	OpenACC Directives	Programming Languages
Provides Fast "Drop-In" Acceleration	GPU-acceleration in Standard Language (Fortran, C, C++)	Maximum GPU Architecture Flexibility

Increasing Development Effort →

NVIDIA Application Engineering Investments

Model	Focus	GPU Approach	Collaboration
WRF	NWP/Climate	CUDA C, OpenACC	NCAR, Cray, NVIDIA
COSMO	NWP/Climate	CUDA C, OpenACC	CSCS, SCS, MeteoSwiss, NVIDIA
CAM-SE	Climate	CUDA Ftn, OpenACC	ORNL, Cray, NVIDIA
NIM/FIM	NWP/Climate	F2C-ACC, OpenACC	NOAA, OACC Vendors, NVIDIA
GEOS-5	Climate	CUDA Ftn, OpenACC	NASA, NVIDIA
NEMO	Ocean Model	OpenACC	NVIDIA, STFC (future)

Other Evaluations: GFS, COAMPS, MPAS-A, ROMS; ICON, UKMO GungHo; GRAPES (CN), OLAM (BR)

Other Investments: Government and Research Institutes without direct NVIDIA collaboration

Example: NOPP/ONR Funding in Accelerated HPC

NVIDIA

NOPP NATIONAL OCEANOGRAPHIC PARTNERSHIP PROGRAM

DEPARTMENT OF THE NAVY ONR Science & Technology

http://www.onr.navy.mil/~/media/Files/Funding-Announcements/BAA/2013/13-011.ashx

ONR BAA13-011:

Advancing Air-Ocean-Land-Ice Global Coupled Prediction on Emerging Computational Architectures

Predictive simulations on heterogeneous architectures Central Processing Unit (CPU), MIC, GPU: identification of representative code patterns that either look particularly amenable or intractable to refactoring; establishment of pathways to maintain single source code compatible with multiple platforms; and determination of mechanisms to achieve optimal performance and portability.

- **Total of $3.75M funding distributed among 4 – 8 awards (closed Apr 2013)**
- **Models: GFS, HIRAM, NIM, MOM, CESM, HYCOM, CICE, Wavewatch3, NUMA**

Rapid OpenACC Growth Since 2011 Founding

NVIDIA

26+ Community Applications

2012

- Cloverleaf
- COSMO (Physics)
- HBM
- NIM
- S3D
- WRF

2013

Cloverleaf	MiniMD
COSMO (Physics)	NICAM
DNS	NIM
EMGS ELAN	NEMO GYRE
GAMESS	PALM-GPU
GENE	Quantum Espresso
GEOS	RAMSES
GTC	ROMS
Harmonie	S3D
HBM	Seismic CPML
ICON	SPECFEM-3D
LULESH	UPACS
MiniGhost	WRF

HPC Industry Support Grows 2x

2012

- Oak Ridge National Laboratory
- allinea
- University of Houston
- Rogue Wave Software
- PGI
- CAPS
- CRAY
- NVIDIA

2013

- Oak Ridge National Laboratory
- allinea
- University of Houston
- Rogue Wave Software
- PGI
- CAPS
- CRAY
- NVIDIA
- NOAA
- Indiana University
- 東京工業大学 Tokyo Institute of Technology
- epcc
- Georgia Institute of Technology
- Technische Universität Dresden
- CSCS

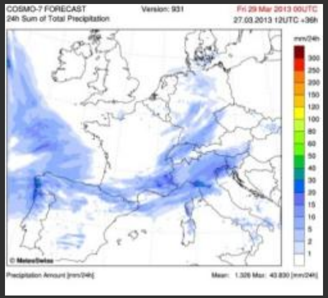
OpenACC in Practice for NWP and Climate

NVIDIA

Examples: Use of directives with less effort; ease in maintenance and flexibility

CSCS Swiss National Supercomputing Centre

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

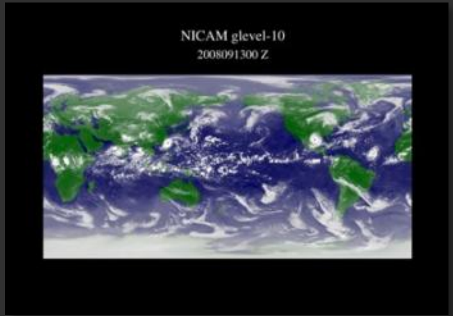


COSMO (Physics)

NWP

- **Goal: preserve *physics* code (22% of runtime), augmenting refactored *dynamics* in CUDA**
- **Physics scheme speedup 4.2X vs. multi-core Xeon**

東京工業大学
Tokyo Institute of Technology



NICAM

Climate

- **Hotspots using CUDA, then OpenACC**
- **CUDA: 3.1x faster on GPU vs. CPU node**
- **OpenACC: (*preliminary*)= 69-77% of CUDA**
  - **More portable, more maintainable**
  - **Full OpenACC port in progress**

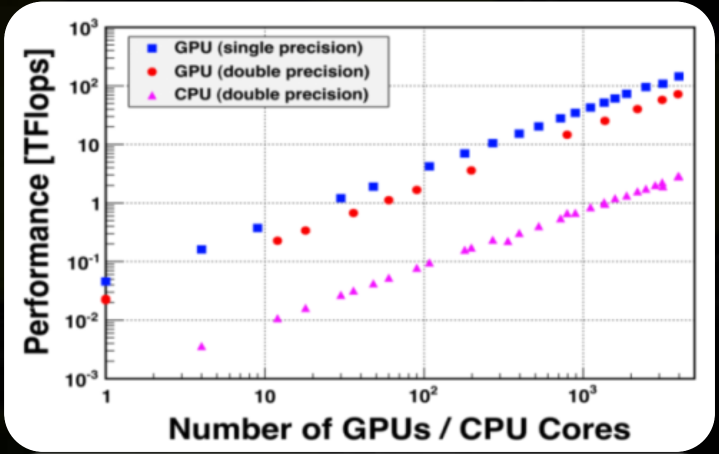
Results from MeteoSwiss/CSCS, and Tokyo Inst of Technology

TiTech NWP 2010 Achievement: ASUCA 145 TF

Tsubame 2.0
Tokyo Institute of Technology

東京工業大学 Tokyo Institute of Technology

- 1.19 Petaflops
- 4,224 Tesla M2050 GPUs

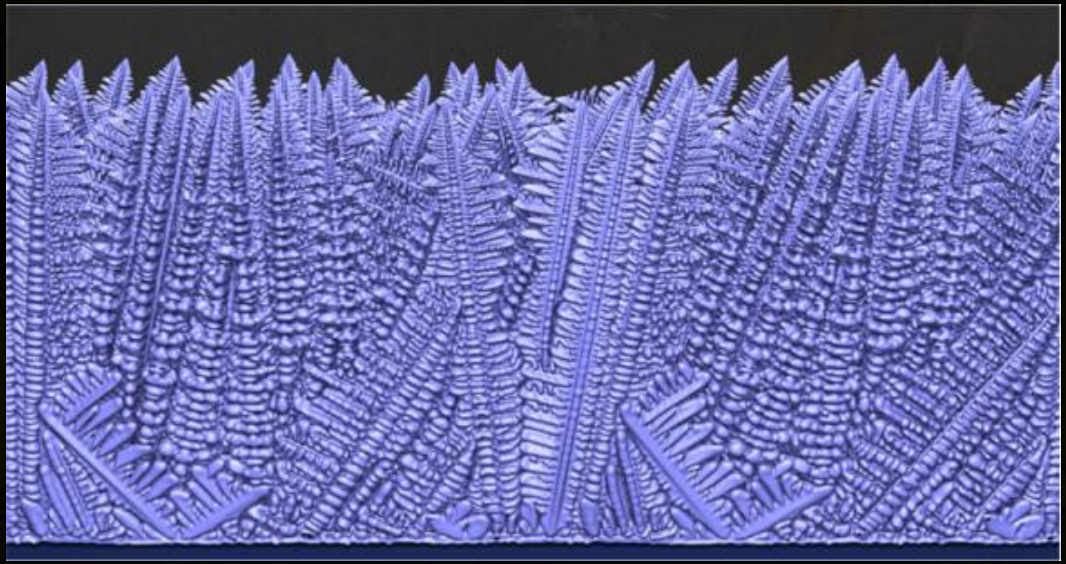


3990 Tesla M2050s
145.0 Tflops SP
76.1 Tflops DP

ASUCA and NWP Simulation on Tsubame 2.0, TiTech Supercomputer:
Dr. Takayuki Aoki, GSIC, Tokyo Institute of Technology, Tokyo Japan

TiTech Winner of 2011 Gordon Bell Prize

Special Achievement in Scalability and Time-to Solution



"Peta-scale Phase-Field Simulation for Dendritic Solidification on the TSUBAME 2.0 Supercomputer"
-- T. Shimokawabe, T. Aoki, et. al.

Tsubame 2.0
Tokyo Institute of Technology

4,224 Tesla GPUs + 2,816 x86 CPUs

Agenda: GPU Progress and Directions for Earth System Modeling

- Introduction of GPUs in HPC
- NVIDIA Application Strategy
- **GPU Progress in ES Modeling**
- NVIDIA Technology Directions

WRF

- http://irina.eas.gatech.edu/EAS8802_Spring2011/Lecture7.pdf
- http://www.mmm.ucar.edu/wrf/users/workshops/WS2010/presentations/Lectures/Microphysics10.pdf
- http://www.mmm.ucar.edu/wrf/users/docs/user_guide_V3.1/users_guide_chap5.htm#_Installing_WRF
- http://www.mmm.ucar.edu/wrf/WG2/GPU/WSM5.htm

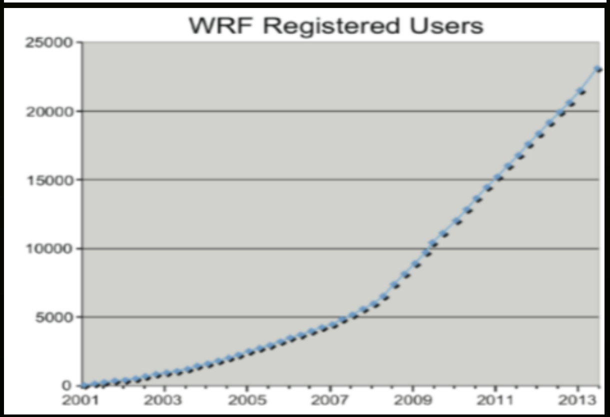
- Jarno Mielikainen, Bormin Huang, Hung-Lung Allen Huang, and Mitchell D, Goldberg, "Improved GPU/CUDA Based Parallel Weather and Research Forecast(WRF) Single Moment 5-Class (WSM5) Cloud Microphysics", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol 5, No.4, August 2012

WRF: Operational in 21 Countries; 153 Total

Registered Users

U.S. Universities	3,600
U. S. Gov't labs, Private sector	3,591
Foreign users	15,916

	23,107
Countries represented:	**153**



Source: Welcome Remarks, 14th Annual WRF Users' Workshop, 24-28 Jun 2013, Boulder, CO

GPU Status of WRF Developments

- Several non-trunk efforts at various stages:
  - 2009: Physics schemes by John Michalakes: www.mmm.ucar.edu/wrf/WG2/GPU/
  - 2010: Dynamics and some physics by Thomas Nipen at UBC – source at NVIDIA
  - 2011: Shortwave radiation model by NVIDIA (G. Ruetsch) and PGI (available)
  - 2012: C-DAC and HPC-FTE group working with NVIDIA India Developers
  - 2012: NOAA announced NIM dycore with WRF physics, but now GFS and YSU
  - 2012: Cray and OpenACC (Pete Johnsen) results at NCAR multi-core workshop
  - Ongoing: KernelGen project: www.kernelgen.org update at NVIDIA GTC 2013
  - Ongoing: 50% of physics schemes by Space Science Engineering Center, UW-M
- Trunk efforts at various stages:
  - WSM5 physics model (15% - 25%) in release 3.2 from 2009 (now dormant)
  - WRF 3.5 with OpenACC –NVIDIA and NCAR (MMM – Dave Gill) collaboration

OpenACC Developments for WRF 3.4/3.5

Programming weather, climate, and earth-system models on heterogeneous multi-core platforms

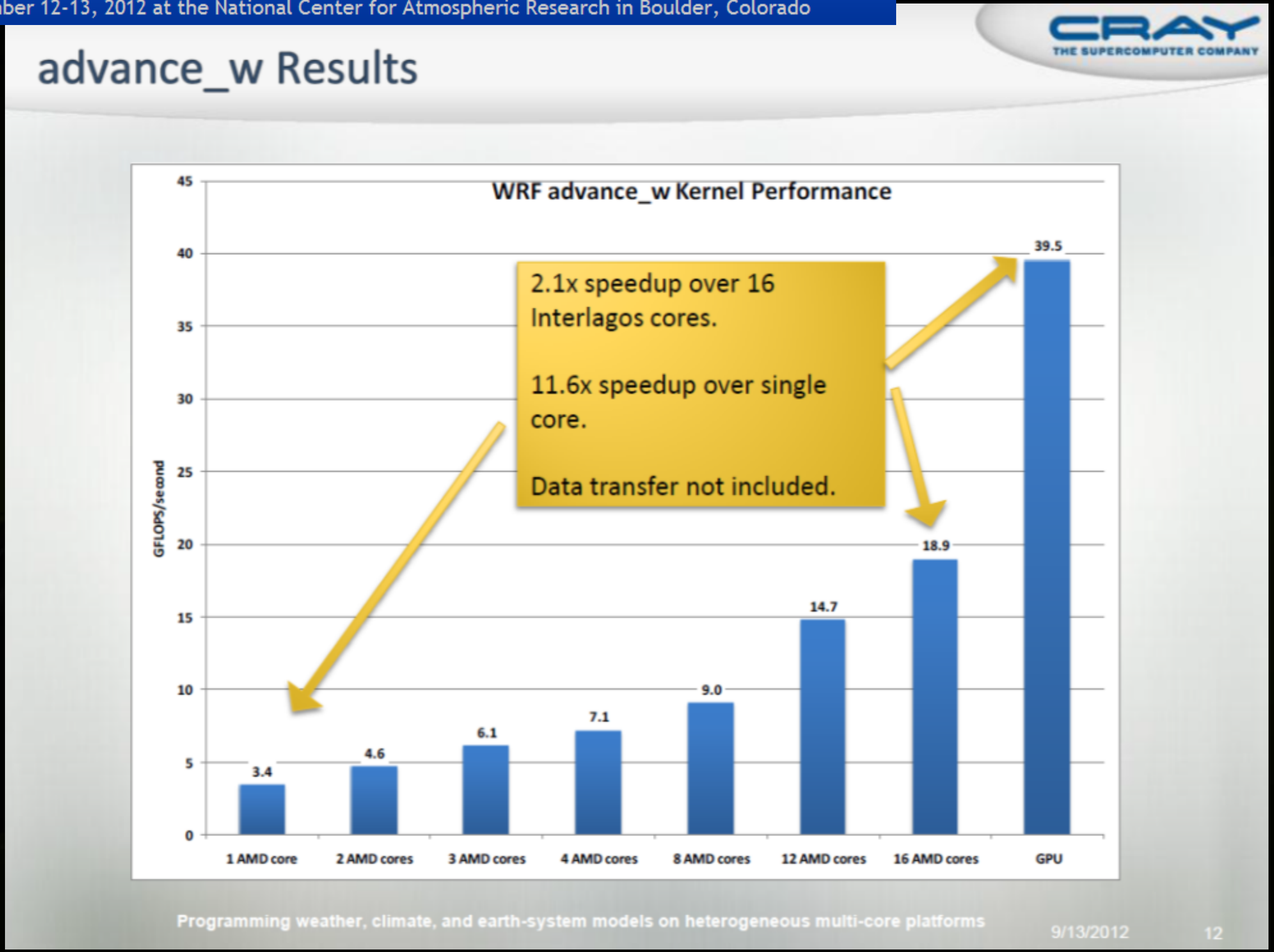
September 12-13, 2012 at the National Center for Atmospheric Research in Boulder, Colorado

WRF Experiments on GPU Accelerators using OpenACC

- Pete Johnsen, Cray, Inc.

WRF routine `advance_w`

- Dynamics routine to advance vertical velocity
- Standard Fortran use with OpenACC directives
- 2.1x speedup for 16 cores

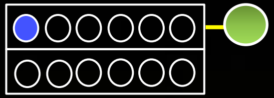


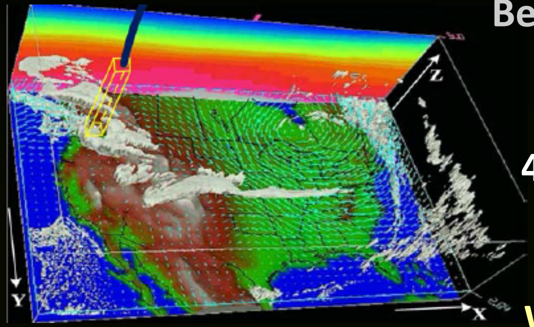
Source: http://data1.gfdl.noaa.gov/multi-core/

Published WRF Speedups from SSEC

WRF Module name	Speedup
Single moment 6-class microphysics	500x
Eta microphysics	272x
Purdue Lin microphysics	692x
Stony-Brook University 5-class microphysics	896x
Betts-Miller-Janjic convection	105x
Kessler microphysics	816x
New Goddard shortwave radiance	134x
Single moment 3-class microphysics	331x
New Thompson microphysics	153x
Double moment 6-class microphysics	206x
Dudhia shortwave radiance	409x
Goddard microphysics	1311x
Double moment 5-class microphysics	206x
Total Energy Mass Flux surface layer	214x
Mellor-Yamada Nakanishi Niino surface layer	113x
Single moment 5-class microphysics	350x
Yonsei University planetary boundary layer scheme	108x
5-Layer Thermal diffusion land surface layer	211x
Pleim-Xiu surface layer	665x

Hardware: Core-i7 3930K, 1 core use; GTX 590 GeForce





Benchmark: CONUS 12 km for 24 Oct 24

433 x 308, 35 levels

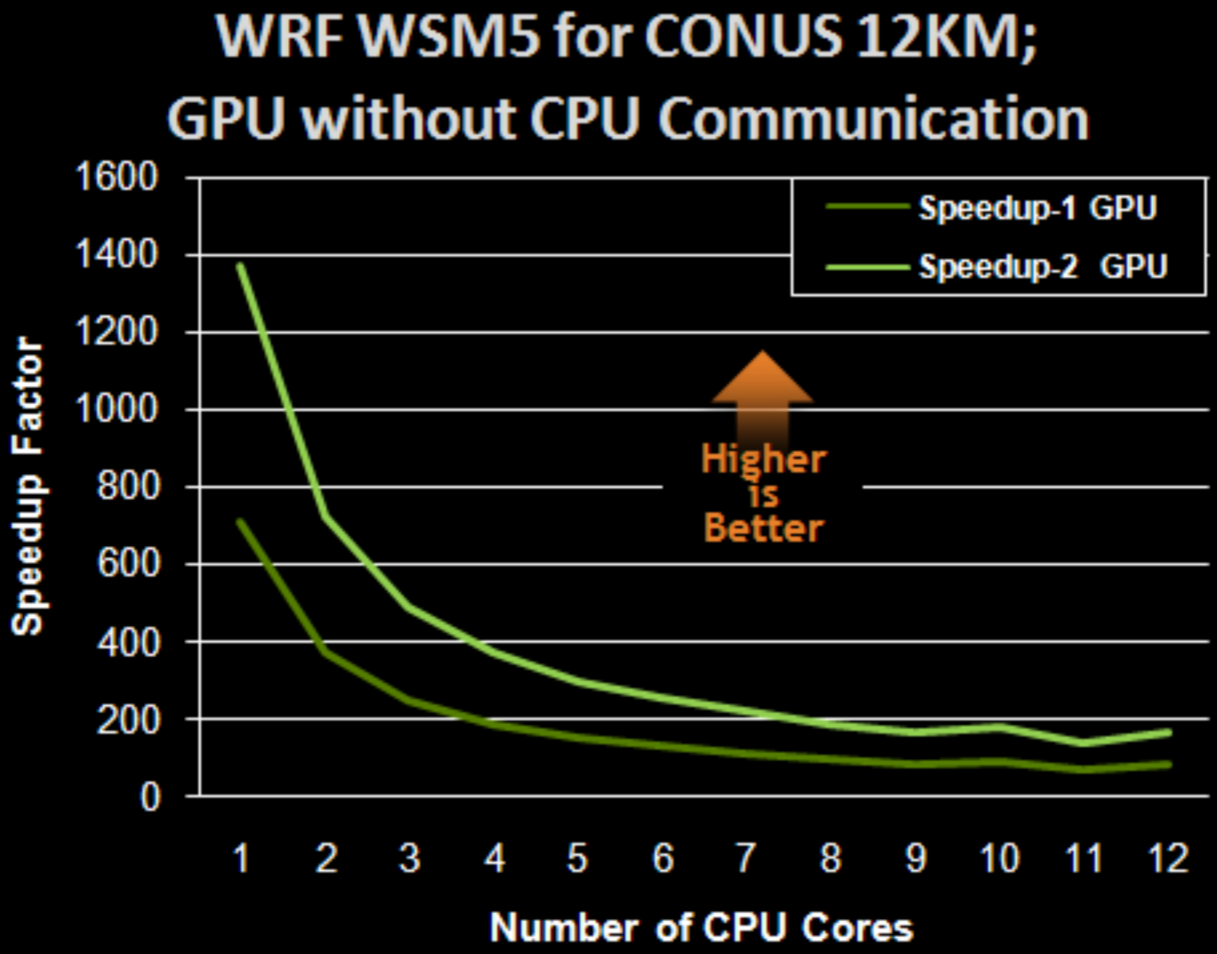
WRF V3.2 and V3.3

Verification: WSM5 by NREL (J. Michalakes) and NVIDIA Applications Engr [Next 2 slides]

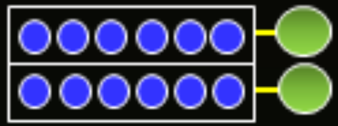
NOTE: All times without CPU data transfer

Source: Bormin Huang, Space Science and Engineering Center, UW-M

NVIDIA Verification of WSM5 from SSEC



2 x Core-i7 3930K,
Total of 12 Cores;
GPUs: 2 x Tesla K20X



Benchmark: CONUS 12 km resolution for October 24, 2001;

433 x 308 grid points, 35 vertical levels

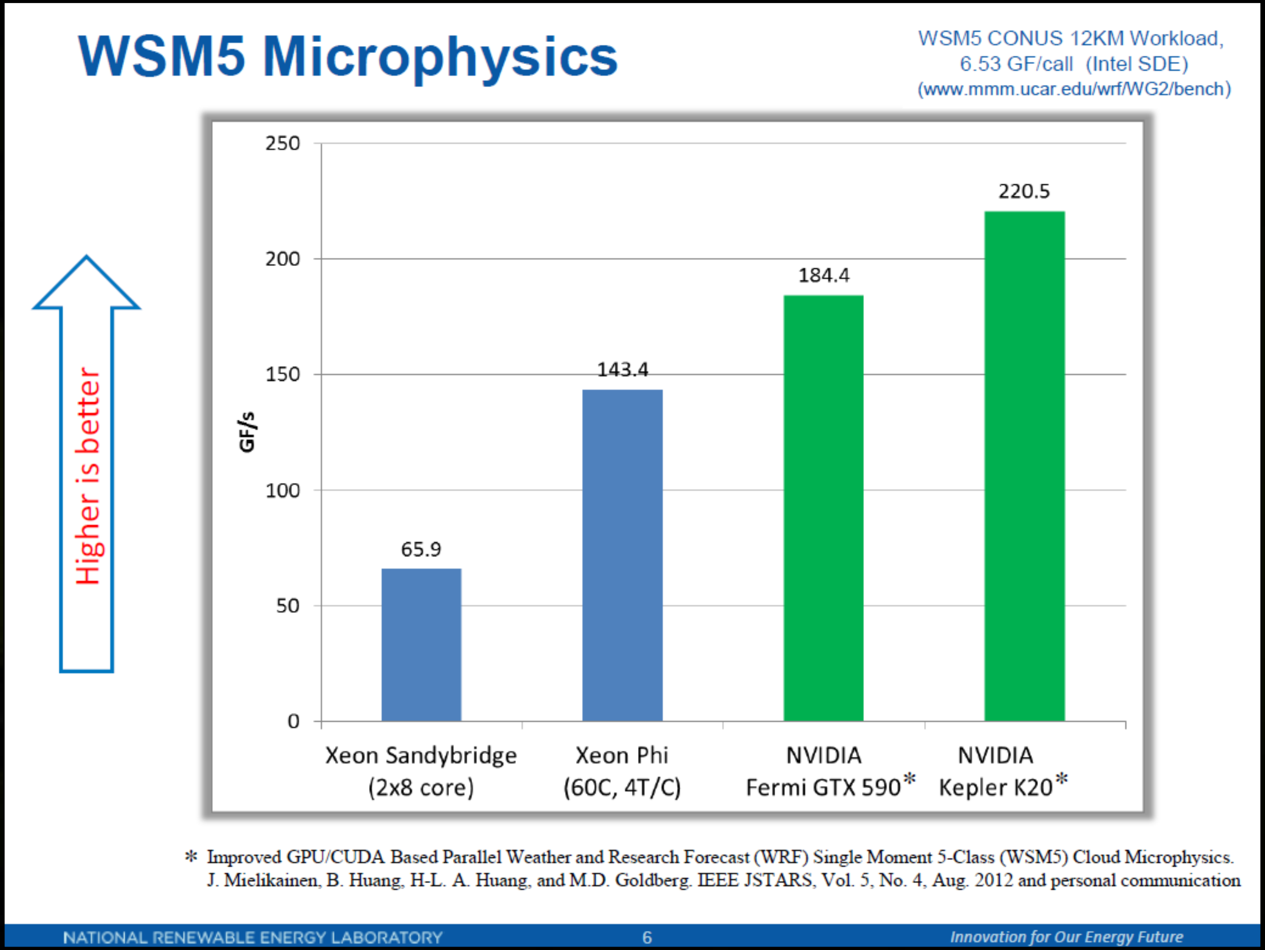
NOTE: Times with no CPU data transfer

NREL Verification of WSM5 from SSEC

High performance computing enhancements to WRFV3.5

- John Michalakes, National Renewable Energy Laboratory, USA

www.mmm.ucar.edu/wrf/users/workshops/WS2013/WorkshopPapers.php



14th Annual WRF Users' Workshop, 24 – 26 Jun 2013, Boulder, CO, USA

- GigaFlop ratings of WSM5 Thompson microphysics scheme (not a full model run)
- Data transfer times with CPU excluded in all results

Performance Results

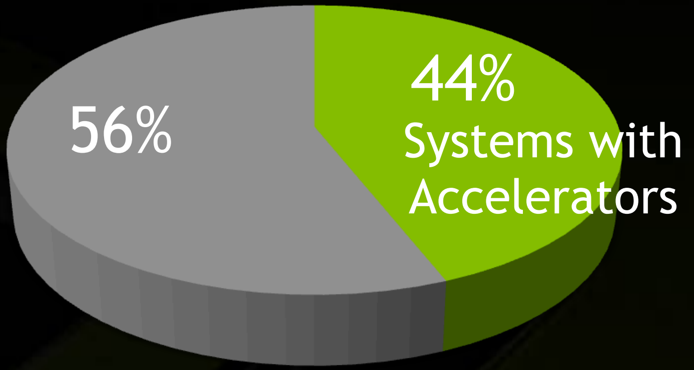
K20X GPU vs. CPU = 3.34x

K20X GPU vs. Phi = 1.54x

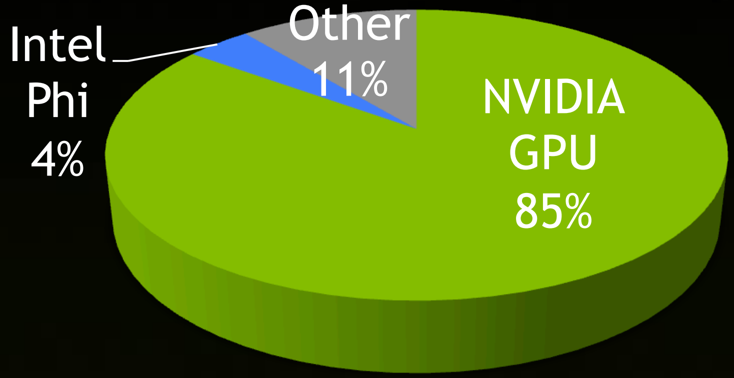
GPU Accelerated Computing Growing Fast

NVIDIA

44% Systems Now Have Accelerators



NVIDIA is Accelerator of Choice



"Intel is not taking share away from NVIDIA but rather both are expanding the use of accelerators."

Intersect360 Research
HPC User Site Census
July, 2013

Accelerator Progress Reports at NCAR Workshop

Programming weather, climate, and earth-system models on heterogeneous multi-core platforms

September 19-20, 2013 at the National Center for Atmospheric Research in Boulder, Colorado

Session:	Focus:
Session 1 - Common Features and Challenges of Using Multi-core Systems	CPU
Session 2 - Optimization for Hybrid GPU systems	GPU
Session 3a - Porting and Optimizing for Intel Xeon and Xeon Phi	Phi
Session 3b - Porting and Optimizing for Intel Xeon and Xeon Phi	Phi
Session 4 - Common Features and Challenges of Using Multi-core Systems	CPU
Session 5 - Porting and Optimizing for Intel Xeon and Xeon Phi	Phi
Session 6a - Optimizing for GPUs using CUDA and OpenACC	GPU
Session 6b - Optimizing for GPUs using CUDA and OpenACC	GPU

Organizers

Ilene Carpenter (NREL), Mark Govett (ESRL), Chris Kerr (GFDL), Rich Loft (NCAR), Bill Putman (GSFC), William Sawyer (CSCS)

Source: http://data1.gfdl.noaa.gov/multi-core/

NIM

Running the FIM and NIM Weather Models on GPUs

- Mark Govett (NOAA Earth System Research Laboratory)

Source: http://on-demand.gputechconf.com/gtc/2013/presentations/S3429-FIM-NIM-Weather-Models-on-GPUs.pdf

NIM Serial Performance (2013)

- <u>No changes to the source code</u>
- Single Socket Performance
  - 10K horizontal points, 96 vertical levels
- Very efficient CPU performance
  - Measured 29% of peak performance (Intel Westmere)

NIM	Opteron	Westmere	SandyBridge	Fermi	K20x
runtime	143.0	86.8	60.0	25.0	20.7

21%

3x

- Parallel performance
  - Being run on up to 160 GPUs
  - Working on optimizing inter-GPU communications

NIM Parallel Performance

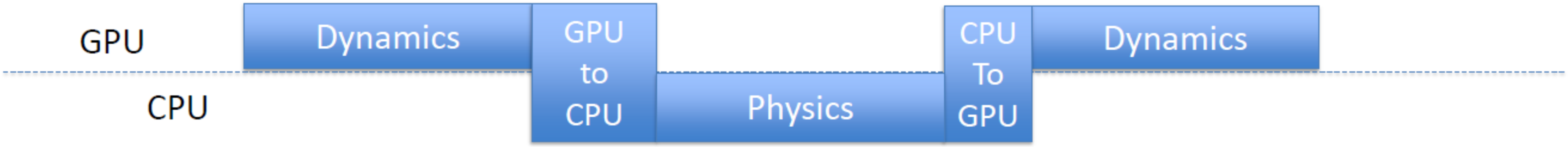
- Weak Scaling with Communications Optimization
  - Moved collective operation to the CPU instead of doing it on the GPU using GPU MappedMemory
    - Too many small writes across the PCIe bus
  - Resulted in a 5-17x speedup for the Pack Operation

GPUs	GPU to GPU Comm Time	Total Time Time (sec)
10	232 (22%)	1034
40	247 (23%)	1054
160	266 (24%)	1076

Operation	Time
Iniitilalization	3 (1%)
Pack Data on GPU	45 (17%)
CPU – GPU Copy	59 (22%)
MPI Comms	82 (31%)
UnPack on GPU	77 (29%)
Total	266

3.5KM NIM on Titan in 2013

- Dynamics on GPU, Physics on CPU + OMP for now



- 10 day forecast, 10,262 horizontal points / GPU

Resolution KM	Vertical Levels	GPUs	Dynamics Time	CPU-GPU Transfer	Physics Time	Total Time In Hours
30	96	80	1700	400	1900	1.1
15	96	320	1700	400	1900	2.2
7.5	96	1280	1700	400	1900	4.4 (1.8%)
3.75	96	5120	1700	400	1900	8.8 (3.6%)

NEMO

Accelerating NEMO with OpenACC

- Maxim Milakov (NVIDIA)

Source: http://on-demand.gputechconf.com/gtc/2013/presentations/S3209-Accelerating-NEMO-with-OpenACC.pdf

NEMO Acceleration Using OpenACC

Background

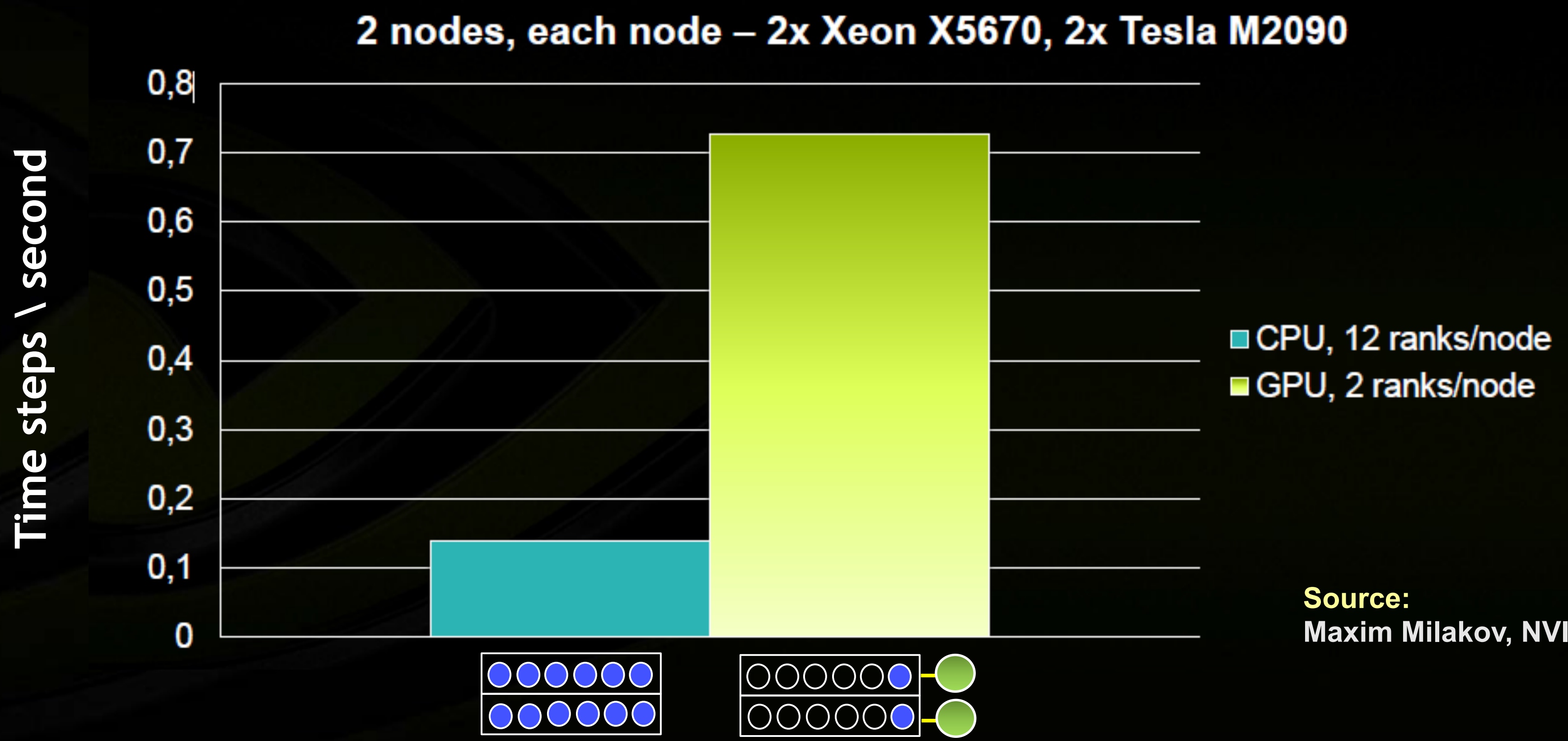
- NEMO ocean modeling framework: http://www.nemo-ocean.eu/
- Used by 240 projects in 27 countries (14 in Europe, 13 elsewhere)

Approach

- Project based on NEMO 3.4, use of PGI Fortran compiler 12.9 preview
- Flat profile, 1st routine is 6%, many routines to accelerate for overall benefit
- OpenACC “present” clause keeps data on the device between subroutine calls
- Directives for 41 routines: rearranged loops in 12, temporary arrays in 13
- Other changes for improved MPI communication, other miscellaneous

NEMO Acceleration Using OpenACC

GYRE_50 Configuration, I/O disabled, OpenACC 1.0: Speedup ~5x

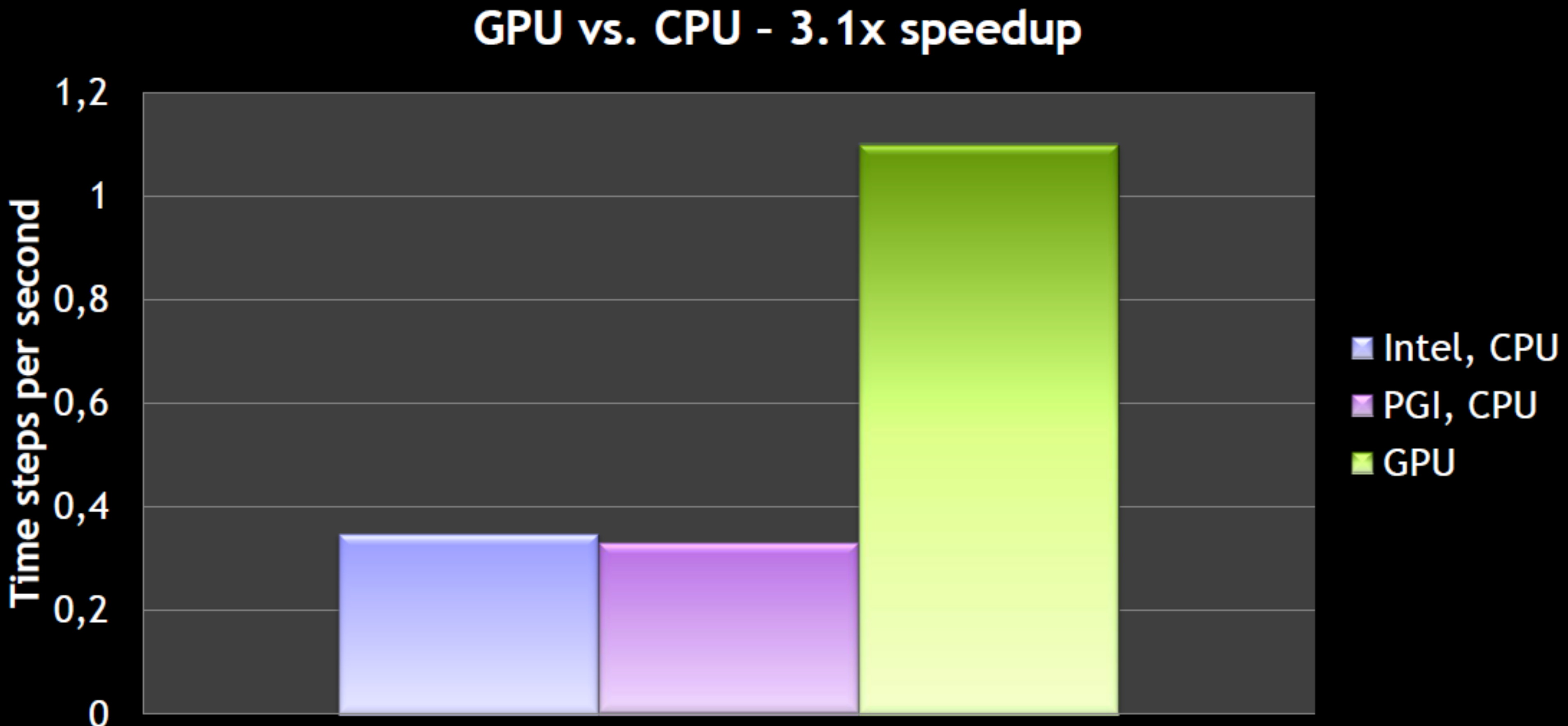


Source:
Maxim Milakov, NVIDIA

Benchmarking - hardware

- “Sandy Bridge + Kepler” nodes, each having:
  - CPU: 2 sockets * Xeon E5-2670 (Sandybridge), 2.6GHz (3.3GHz Turbo Boost), 8 cores, 64 GB RAM
  - GPU: 2x Tesla K20X, ECC off, 6GB RAM each
  - 4x FDR Infiniband (56 Gb/s)
- Running configuration is GYRE_50 (1/4 degree), requires about 23GB of total RAM, fits 4 K20X
- The code is running on 2 nodes
- The performance is measured by running 1000 time steps, startup and shutdown overheads are not included in figures

Benchmarking - results

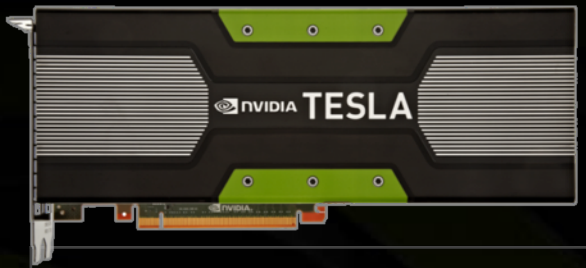


Agenda: GPU Progress and Directions for Earth System Modeling

- Introduction of GPUs in HPC
- NVIDIA Application Strategy
- GPU Progress in ES Modeling
- **NVIDIA Technology Directions**

Tesla Kepler Family

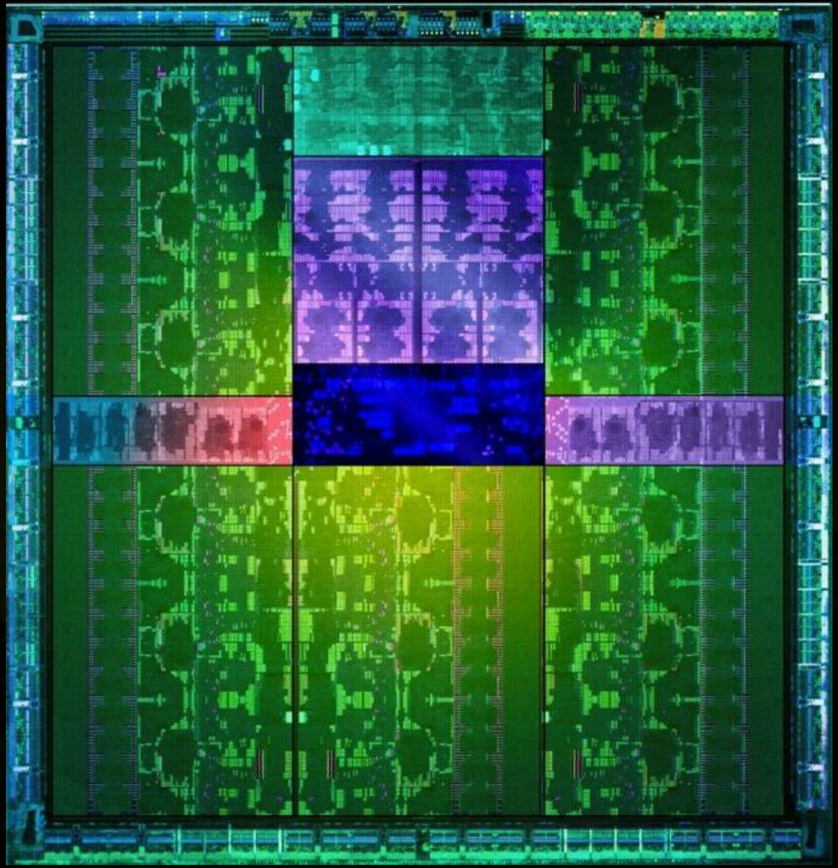
World's Fastest and Most Efficient HPC Accelerators



	GPUs	Single Precision Peak *(SGEMM)*	Double Precision Peak *(DGEMM)*	Memory Size	Memory Bandwidth (ECC off)	System Solution
Weather & Climate, Physics, BioChemistry, CAE, Material Science	**K20X**	3.95 TF *(2.90 TF)*	1.32 TF *(1.22 TF)*	6 GB	250 GB/s	Server only
	K20	3.52 TF *(2.61 TF)*	1.17 TF *(1.10 TF)*	5 GB	208 GB/s	Server + Workstation
Image, Signal, Video, Seismic	**K10**	4.58 TF	0.19 TF	8 GB (4 GB ea.)	320 GB/s	Server only

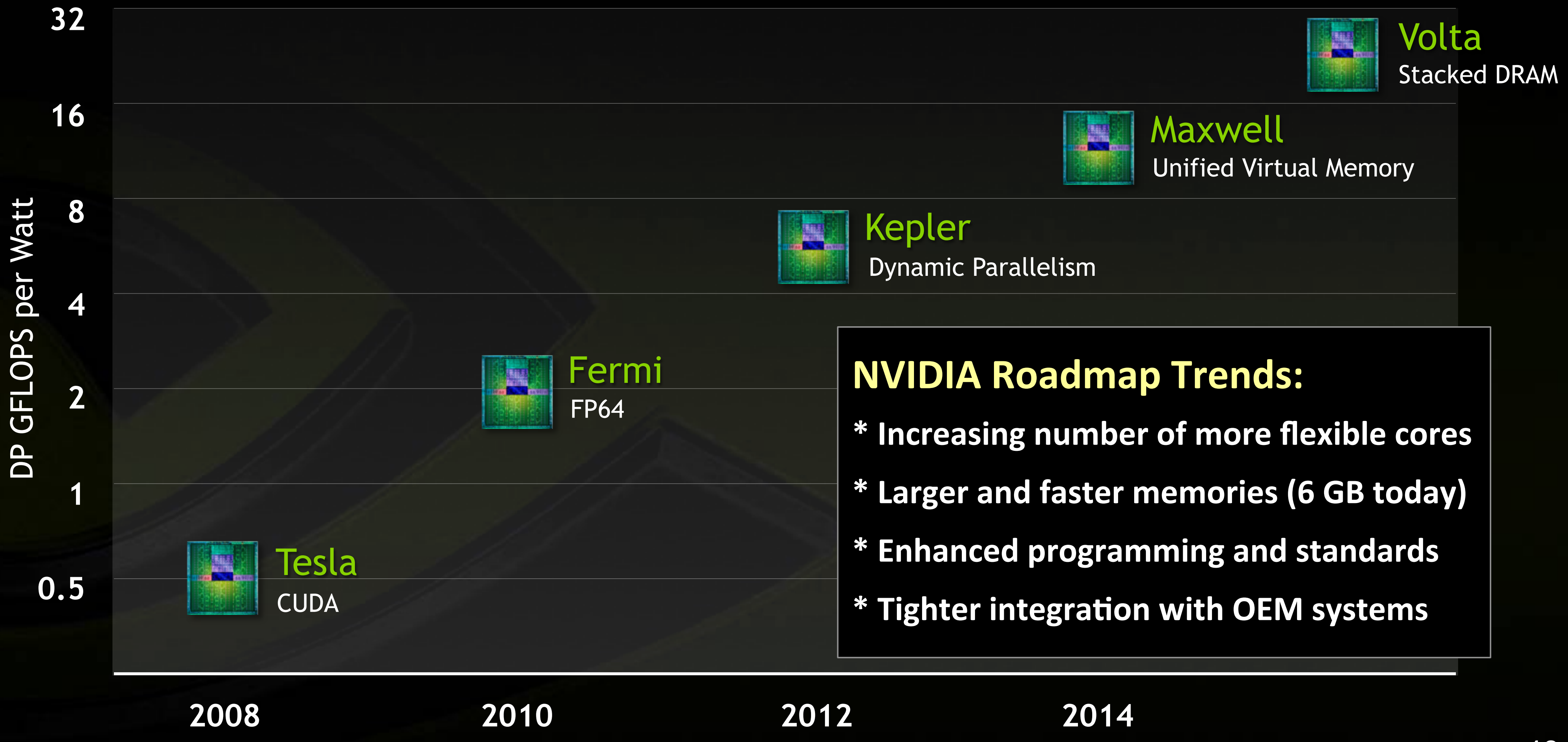
Kepler

Fastest, Most Efficient HPC Architecture Ever



SMX	3x Performance per Watt
Hyper-Q	Easy Speed-up for Legacy MPI Apps
Dynamic Parallelism	Parallel Programming Made Easier than Ever

NVIDIA GPU Roadmap (Details Require NDA)



NVIDIA Roadmap Trends:

* Increasing number of more flexible cores

* Larger and faster memories (6 GB today)

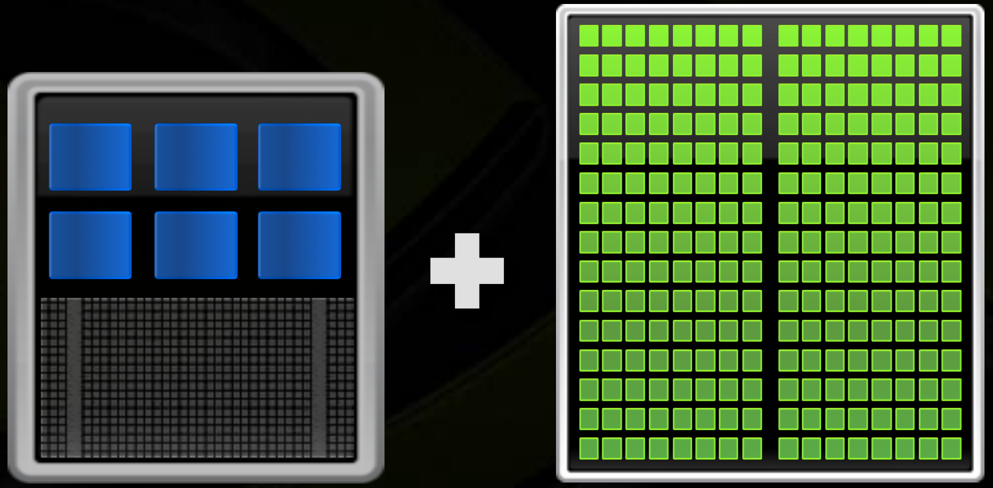
* Enhanced programming and standards

* Tighter integration with OEM systems

ARM Support Now Available Since CUDA 5.5

ARM or x86 CPU
Optimized for Few Serial Tasks

GPU Accelerator
Optimized for Many Parallel Tasks



CUDA 5.5 Highlights

Full Support for ARM Platforms

- Native compilation on ARM

Optimized for MPI

- Faster Hyper-Q for all Linux distros
- MPI workload prioritization

Guided Performance Analysis

- Step-by-step optimization

Available Now
http://developer.nvidia.com/cuda-toolkit

NVIDIA Quadro K6000: Kepler-Based 12GB GPU

Announced July 2013

Available Q3 2013
http://www.nvidia.com/object/quadro-desktop-gpus.html

QUADRO K6000 QUICK SPECS	
CUDA Parallel-Processing Cores	2880
Frame Buffer Memory	12 GB GDDR5
Max Power Consumption	225 W
Graphics Bus	PCI Express 3.0 x16
Display Connectors	DVI-I (1), DVI-D (1) DP 1.2 (2), Optional Stereo (1)
Form Factor	4.376" H x 10.5" L Dual Slot

Summary For GPUs and ES Modeling

- Opportunities exist for GPUs to provide significant performance acceleration for ES Models
  - Improved simulation quality from higher resolutions
  - Faster time to predictions for operational forecasting
  - Cut down energy consumption in IT procedures
- Simulations recently considered intractable are now possible
  - Global models are cloud resolving scale
  - Parameter physics at higher resolutions and more frequent time steps
  - Expanded and more common use of ensembles

Thank you
&
Questions

Stan Posey, sposey@nvidia.com
NVIDIA, Santa Clara, CA, USA