NVIDIA HPC Directions for Earth System Models

Stan Posey, HPC Program Manager, ESM Domain, NVIDIA (HQ), Santa Clara, CA, USA

TOPICS OF DISCUSSION

- NVIDIA HPC DIRECTIONS
- ESM GPU PROGRESS
- PGI UPDATE – D. NORTON

NVIDIA HPC Investments for the ESM Community

- Investment in ESM domain as a strategic HPC focus with worldwide support
- Host CPU flexibility through 'virtual' offload – x86; OpenPower; ARM64
- GPU performance focus on "cost-practical" high memory bandwidth speeds
- GPU scalability design of fewer heavy nodes (vs. many thin CPU nodes)
- Programming models (OpenACC, OpenMP), task-parallelism, libraries, tools, etc.
- Technical collaborations and alliances with system OEMs on ESM projects:

IBM – CORAL: ORNL, LLNL; Deep Thunder; NCAR-Yellowstone

Cray – CSCS/MeteoSwiss; NOAA; NCSA/BL; ORNL/TITAN; ROKAF

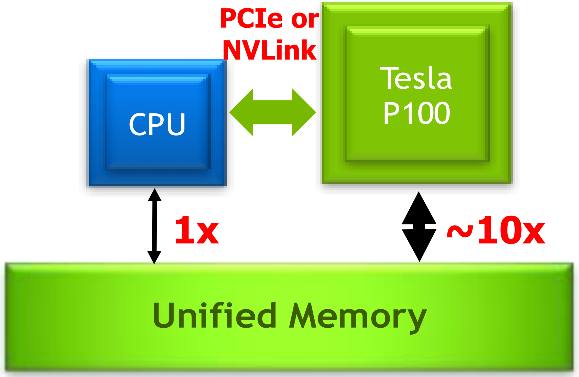
SGI – NASA GSFC/ARC, US DoD AFRL, NIES (JP), NCAR-Cheyenne (Q1 2017)

NVIDIA Support for Earth System Modeling Domain

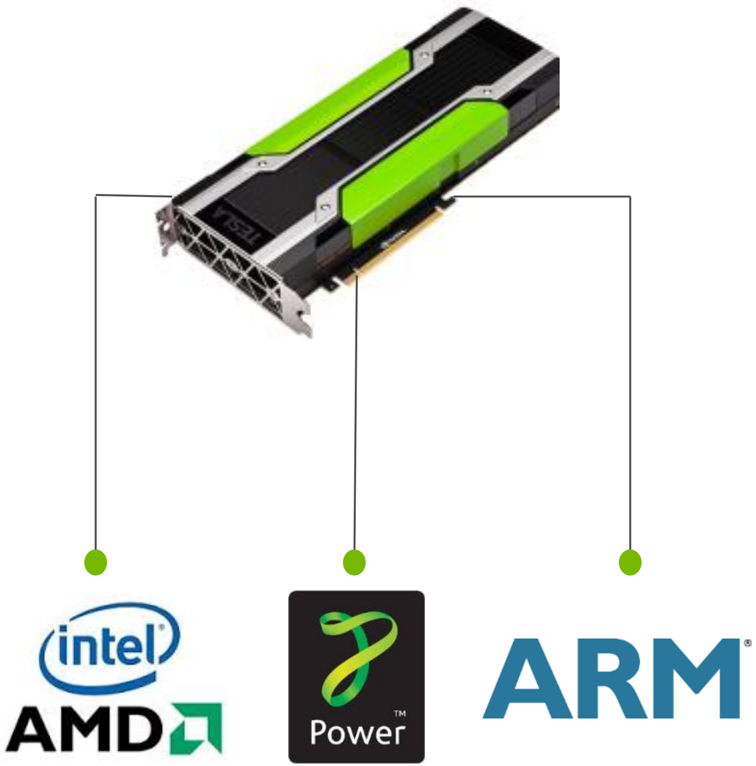
ESM Program Manager: [No People Management]	Stan Posey - Santa Clara, CA (HQ)	sposey@nvidia.com
Developer Technology:	Carl Ponder, PhD - Austin, TX, US Jeff Larkin - Oak Ridge, TN, US Jeremy Appleyard, PhD - Oxford, UK Peter Messmer, PhD - Zurich, CH Akira Naruse, PhD - Tokyo, JP . . .	WRF, MPAS-A, FV3, GEOS-5, COAMPS, GEM CAM-SE, all ACME component models IFS, NEMO, UM/GungHo, GOcean IFS/*ESCAPE*, COSMO, ICON JMA-GSM, ASUCA, NICAM
PGI Applications Eng:	Dave Norton, Lake Tahoe, CA	*All Models that use PGI compilers*
Business Alliances:	Steve Rohm - Charlotte, NC, US Greg Branch - Boulder, CO, US Jeremy Purches - Bristol, UK Stefan Kraemer - Würselen, DE Frederic Pariente - Toulouse, FR . . .	US East: NOAA, EC, DOE, DoD, NASA US West: NOAA, NCAR, DOE, NASA ECMWF, UKMO, STFC, No. Europe DWD, DKRZ/MPI-M, MCH, Central Europe MF, IPSL, CNRS, CERFACS, So. Europe
Solution Architects:	*Several Worldwide*	Contact sposey@nvidia.com

NVIDIA GPU: Status and Hardware Features

GPU Introduction

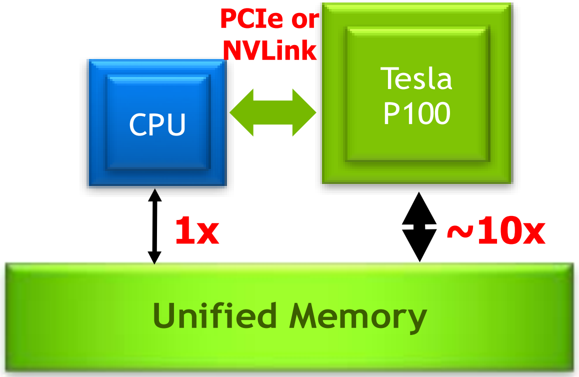


- Co-processor to the CPU
- Threaded Parallel (SIMT)
- CPUs: x86 | Power | ARM
- HPC Motivation:
  - Performance
  - Efficiency
  - Cost Savings



NVIDIA GPU: Status and Hardware Features

GPU Introduction



- Co-processor to the CPU
- Threaded Parallel (SIMT)
- CPUs: x86 | Power | ARM
- HPC Motivation:
  - Performance
  - Efficiency
  - Cost Savings

Schematic of 8 GPU Server by Facebook

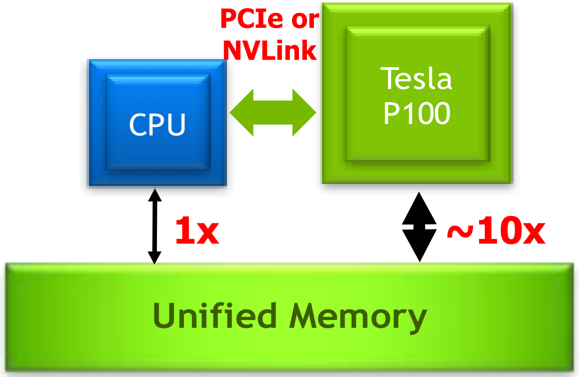
ORNL Titan
#3 Top500.org
18,688 GPUs

Image Credit: Facebook

IMAGE: Facebook's new Big Sur GPU server
http://venturebeat.com/2016/08/29/facebook-gives-away-22-more-gpu-servers-for-a-i-research/

NVIDIA GPU: Status and Hardware Features

GPU Introduction



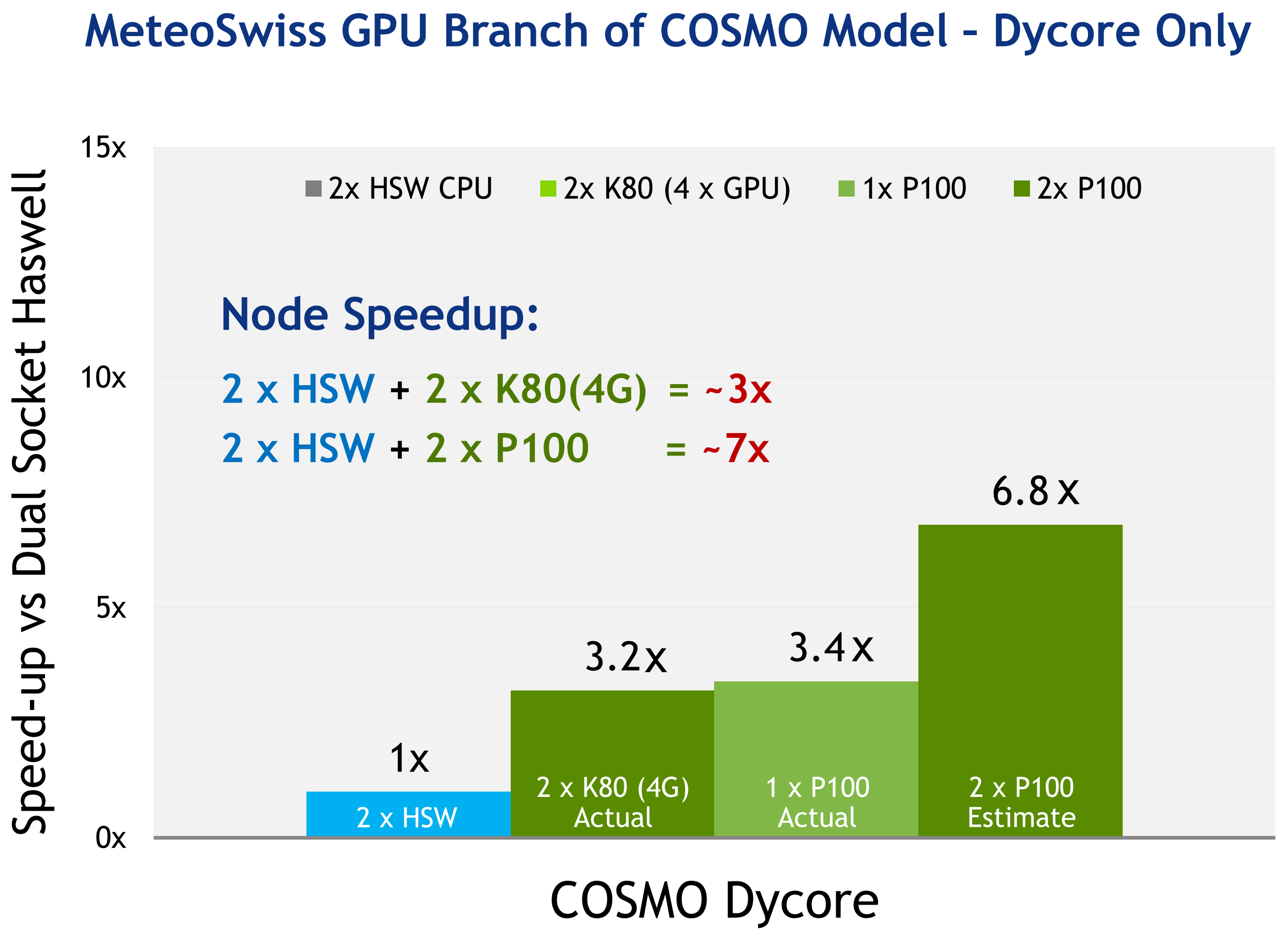
- Co-processor to the CPU
- Threaded Parallel (SIMT)
- CPUs: x86 | Power | ARM
- HPC Motivation:
  - Performance
  - Efficiency
  - Cost Savings

	Next GPU (Q4 2016)	Current GPUs Since 2014		
GPU Feature	**Tesla P100**	**Tesla K80**	**Tesla K40**	
Stream Processors	3584	2 x 2496	2880	
Core Clock	1328MHz	562MHz	745MHz	
Boost Clock(s)	1480MHz	875MHz	810MHz, 875MHz	
Memory Clock	1.4Gbps HBM2	5Gbps GDDR5	6Gbps GDDR5	
Memory Bus Width	4096-bit	2 x 384-bit	384-bit	
Memory Bandwidth	720 GB/sec	2 x 240 GB/sec	288 GB/sec	2.5x
VRAM	16GB	2 x 12GB	12GB	
Half Precision	21.2 TFLOPS			
Single Precision	10.6 TFLOPS	8.74 TFLOPS	4.29 TFLOPS	
Double Precision	5.3 TFLOPS	2.91 TFLOPS	1.43 TFLOPS	3.7x
GPU	GP100	GK210	GK110B	
Transistor Count	15.3B	2 x 7.1B	7.1B	
Power Rating	300W	300W	235W	
Cooling	Passive	Passive	Active/Passive	
Manufacturing Process	TSMC 16nm FinFET	TSMC 28nm	TSMC 28nm	
Architecture	Pascal	Kepler	Kepler	

NOTE: P100 nodes available for ESM community remote access on NVIDIA PSG cluster

COSMO Dycore Speedup on P100 GPU

http://www.cosmo-model.org/



Results from NVIDIA Internal Cluster (US)

(Preliminary – Mar 2016)

- COSMO MCH branch
  (Based on COSMO 4.19)
- 128x128, x80 Vertical
- Time steps 10
- CPU: x86 Xeon Haswell
  - 10 Cores @ 2.8 GHz
- GPU: Tesla P100
- Use of single node
- CUDA 8

Node Speedup:

2 x HSW + 2 x P100 = ~7x

[P. Messmer – NVIDIA]

ESM Opportunities with Pascal Architecture

ESM Characteristic	Pascal Feature	ESM Opportunity
Requirement of memory bandwidth vs. FLOPS	HBM (High B/W Memory)	Increase of ~3x memory bandwidth over Kepler
Flat profiles - requires GPU port of entire code to avoid host data exchange (PCIe)	NVLink	NVLink hides data xfer; incremental port now practical vs. full model
Large legacy Fortran code, some projects waiting for programming improvements	UM (Unified Memory)	UM manages host data exchange, simplifies programming effort

Performance

Programming

WRF Progress with OpenACC: Incremental Approach

- Project to implement OpenACC routines into full model WRF 3.6.1
  - Several dynamics routines including all of advection
  - Several physics schemes (10):
    - Microphysics (4) – Kessler, Morrison, Thompson, WSM6
    - Radiation (2) – RRTM (lw), Dudhia (sw)
    - Planetary boundary layer (2) – YSU, GWDO
    - Cumulus (1) – Kain-Fritsch
    - Surface layer (1) – Noah

- Routines completed:
  - Dynamics (11)

```
dyn_em/module_advect_em.OpenACC.F
dyn_em/module_bc_em.OpenACC.F
dyn_em/module_big_step_utilities_em.OpenACC.F
dyn_em/module_diffusion_em.OpenACC.F
dyn_em/module_em.OpenACC.F
dyn_em/module_first_rk_step_part1.OpenACC.F
dyn_em/module_first_rk_step_part2.OpenACC.F
dyn_em/module_small_step_em.OpenACC.F
dyn_em/module_stoch.OpenACC.F
dyn_em/solve_em.OpenACC.F
dyn_em/start_em.OpenACC.F
```

- Physics (18)

```
phys/module_bl_gwdo.OpenACC.F
phys/module_bl_ysu.OpenACC.F
phys/module_cu_kfeta.OpenACC.F
phys/module_cumulus_driver.OpenACC.F
phys/module_microphysics_driver.OpenACC.F
phys/module_microphysics_zero_out.OpenACC.F
phys/module_mp_kessler.OpenACC.F
phys/module_mp_morr_two_moment.OpenACC.F
phys/module_mp_thompson.OpenACC.F
phys/module_mp_wsm6.OpenACC.F
phys/module_pbl_driver.OpenACC.F
phys/module_physics_addtendc.OpenACC.F
phys/module_physics_init.OpenACC.F
phys/module_ra_rrtm.OpenACC.F
phys/module_ra_sw.OpenACC.F
phys/module_sf_noahlsm.OpenACC.F
phys/module_sf_sfclayrev.OpenACC.F
phys/module_surface_driver.OpenACC.F
```

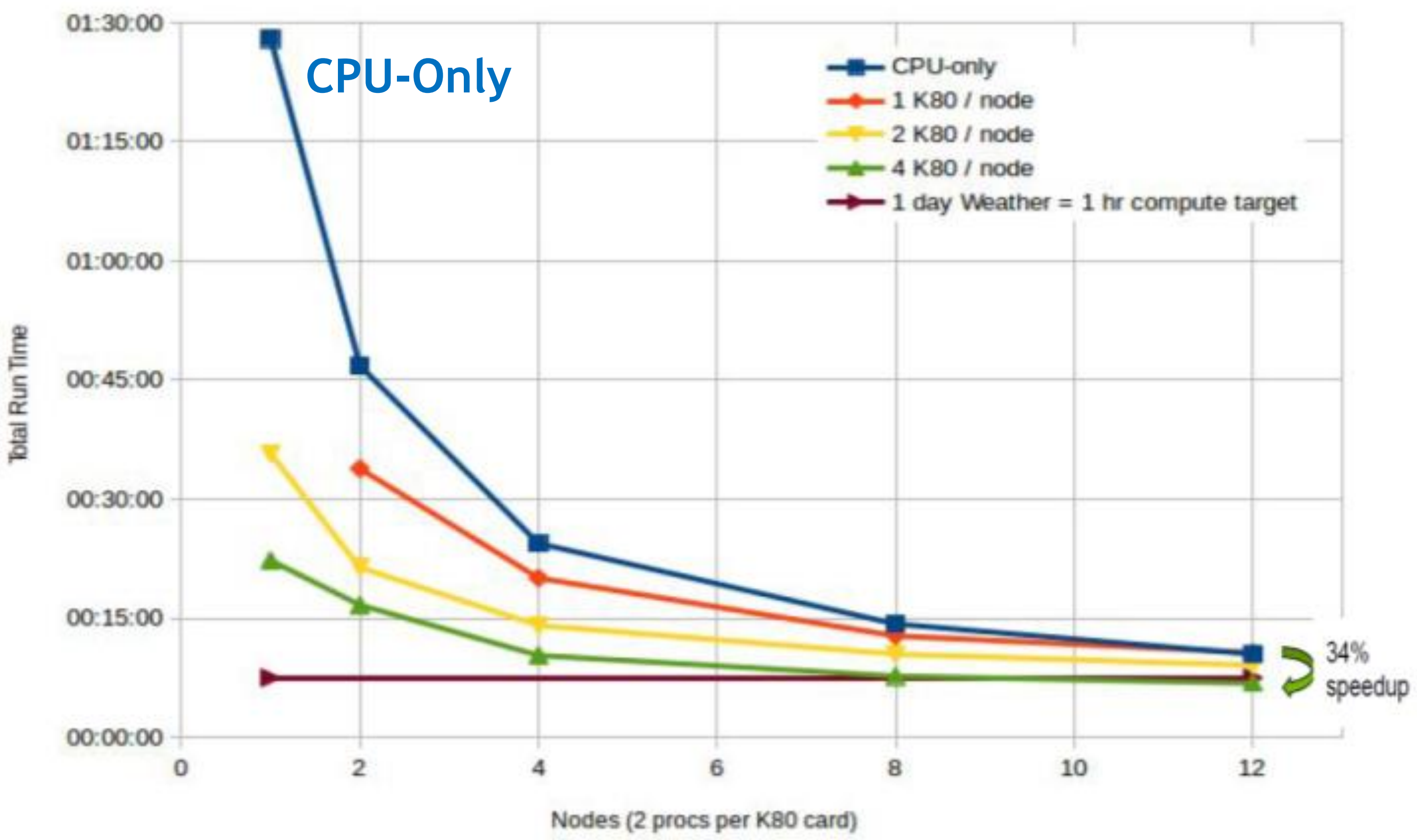
- Other (8)

```
frame/module_dm.OpenACC.F
frame/module_domain_extra.OpenACC.F
frame/module_domain.OpenACC.F
frame/module_domain_type.OpenACC.F
frame/module_integrate.OpenACC.F

share/mediation_integrate.OpenACC.F
share/module_bc.OpenACC.F
share/wrf_bdyin.OpenACC.F
```

WRF Hybrid CPU + GPU Performance (Mar 2016)

OpenACC Scalability Results on 10 Nodes with K80 GPUs



Results from NVIDIA PSG Cluster (USA)

NVIDIA PSG Cluster

http://psgcluster.nvidia.com/trac

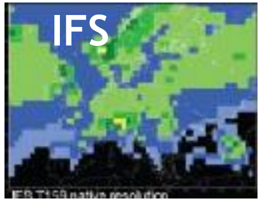
- WRF revision 3.6.1
- 3km, 50 vertical layers
- 1080 ts, 3 simulated hrs
- GPU accelerated WRF code:
  - MP = Thompson
  - Rad = RRTM+Dudhia
  - Most dynamics
- CPU: x86 Xeon Haswell
  - 10 Cores @ 2.8 GHz
- GPU: NVIDIA Tesla K80
- 10 nodes (80 GPUs total)
- CUDA 7.0.27

TOPICS OF DISCUSSION

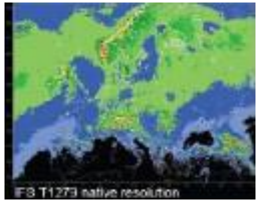
- NVIDIA HPC DIRECTIONS
- **ESM GPU PROGRESS**
- PGI UPDATE – D. NORTON

GPUs Motivated by Earth System Model Directions

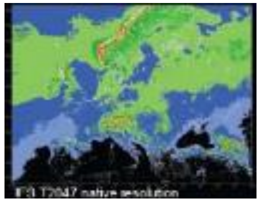
- **Higher grid resolution with manageable compute and energy costs**
  - Global atmosphere models from 10 km today to cloud-resolving scales of 3 km



128 km



16 km



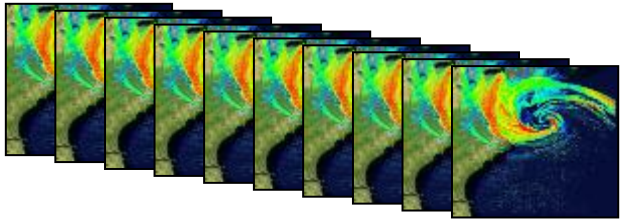
10 km

?

3 km

http://www.wxmaps.org/athena/home/

- **Increase of ensemble use and ensemble members to manage uncertainty**



Number of Jobs 10x - 50x

- **Fewer model approximations (NH), more features (physics, chemistry, etc.)**

Accelerator technology identified as a cost-effective and practical approach to future HPC requirements

Select NVIDIA ESM Highlights During 2016

- **WW ESM growth in GPU funded-development:** NOAA, NCAR, ECMWF, DOE, DoD
- **New ESM-driven GPU systems (K80/P100):** NOAA, ECMWF, CSCS, NIES
- **First ever GPU-based operational NWP:** MeteoSwiss with COSMO (Mar 2016)
  - ~4x speedup with ~5x less energy vs. conventional CPU-only in heavy GPU server node configuration
- **DOE climate model ACME-Atm v1 production on TITAN using PGI OpenACC**
- **New NCAR + KISTI collaboration agreement; and GPU Hands-on Workshops**
  - https://www2.cisl.ucar.edu/news/summertime-hackathon (focus on GPU development of MPAS-A)
- **ECWMF selected NVIDIA as partner for ESCAPE Exascale weather project**
  - https://software.ecmwf.int/wiki/display/OPTR/NVIDIA+Basic+GPU+Training+with+emphasis+on+Fortran+and+OpenACC
- **NEMO Systems Team selected NVIDIA as member of HPC working group**
  - Following successful NVIDIA OpenACC scalability of NEMO for ORCA025 configuration (NEMO UGM 2014)
- **ESM team's participation and positive outcomes in DOE-led GPU Hackathons**
  - DOE/ACME, NRL/COAMPS, MPI-M/ECHAM6, ODU/FVCOM, NRL/HYCOM, NOAA GFDL radiation models
- **New GPU opportunities developing in new ESM solution focus areas**
  - DL in climate and weather; BI for Ag and Actuary; Air quality monitoring (CN, KR); Commercial WRF start-up TQI

GPU Funded-Development Increasing for ESM

HPC Programs with Funding Specifically Targeted for GPU Development of Various ESMs

- SENA – NOAA funding for accelerator development of WRF, NGGPS (FV3), GFDL climate, NMMB
- ESCAPE – ECMWF-led EUC Horizon 2020 program for IFS; NVIDIA 1 of 11 funded partners
- ACME – US DOE accelerated climate model: CAM-SE, MPAS-O, CICE, CLM, SAM, PISCEES, others
- AIMES – Govt's from DE, FR, and JP for HPC (and GPU) developments of ICON, DYNAMICO, NICAM
- SIParCS – NCAR academia funding for HPC (and GPU) developments of CESM, DART, Fields, others
- AOLI – US DoD accelerator development of operational models HYCOM, NUMA, CICE, RRTMG
- GridTools – Swiss Gov't fund of MCH/CSCS/ETH for accelerator-based DSL in COSMO, ICON, others

NOTE: Follow each program LINK for details; Programs listed from top-down in rough order of newest to oldest start date

GPUs Deployed for ESM and NWP Modeling (Sep 16)

	Organization	Location	GPUs	System
NextGen	DOE ORNL	Oak Ridge, TN, US	Volta – 200 PF	IBM Power9 – *Summit*
	DOE LLNL	Livermore, CA, US	Volta – 150 PF	IBM Power9 – *Sierra*
	CSCS	Lugano, CH	P100 – 4,500	Cray XC-40
	NOAA	Fairmont, WV, US	P100 – 760	Cray CS-Storm
	NIES	Tsukuba, JP	P100 – 266	SGI ICE
K80/K40	MCH/CSCS	Lugano, CH	K80 – 192	Cray CS-Storm – *Piz Kesch*
	Unnamed	Seoul, KR	K80 – 130	Cray CS-400
	ECMWF	Reading, UK	K80 – 68	Dell
	DoD AFRL	Dayton, OH, US	K40 – 356	SGI ICE X – *Thunder*
	NASA GSFC	Greenbelt, MD, US	K40 – 36	SGI, IBM
	NASA ARC	Mtn View, CA, US	K40 – 64	SGI – *Pleiades*
K20X	DOE ORNL	Oak Ridge, TN, US	K20X – 18,688	Cray – *TITAN*
	CSCS	Lugano, CH	K20X – 5,272	Cray – *Piz Daint, Piz Dora*
	TiTech	Tokyo, JP	K20X – 4,224	NEC/HP – *TSUBAME 2.0*
	NCSA	Urb-Ch, IL, US	K20X – 3,072	Cray – *Blue Waters*
	NCAR	Cheyenne, WY, US	K20X – 30	IBM – *Yellowstone*

NVIDIA and HPC System for Deep Learning

FORTUNE

WHY DEEP LEARNING IS SUDDENLY CHANGING YOUR LIFE

http://fortune.com/ai-artificial-intelligence-deep-machine-learning/

A GLOSSARY OF ARTIFICIAL-INTELLIGENCE TERMS

- ARTIFICIAL INTELLIGENCE
AI is the broadest term, applying to any technique that enables computers to mimic human intelligence, using logic, if-then rules, decision trees, and machine learning (including deep learning).

- MACHINE LEARNING
The subset of AI that includes abstruse statistical techniques that enable machines to improve at tasks with experience. The category includes deep learning.

- DEEP LEARNING
The subset of machine learning composed of algorithms that permit software to train itself to perform tasks, like speech and image recognition, by exposing multilayered neural networks to vast amounts of data.

. . . says Jen-Hsun Huang, CEO of graphics processing leader NVIDIA, which began placing a massive bet on deep learning about five years ago.

MeteoSwiss Weather Prediction Based on GPUs

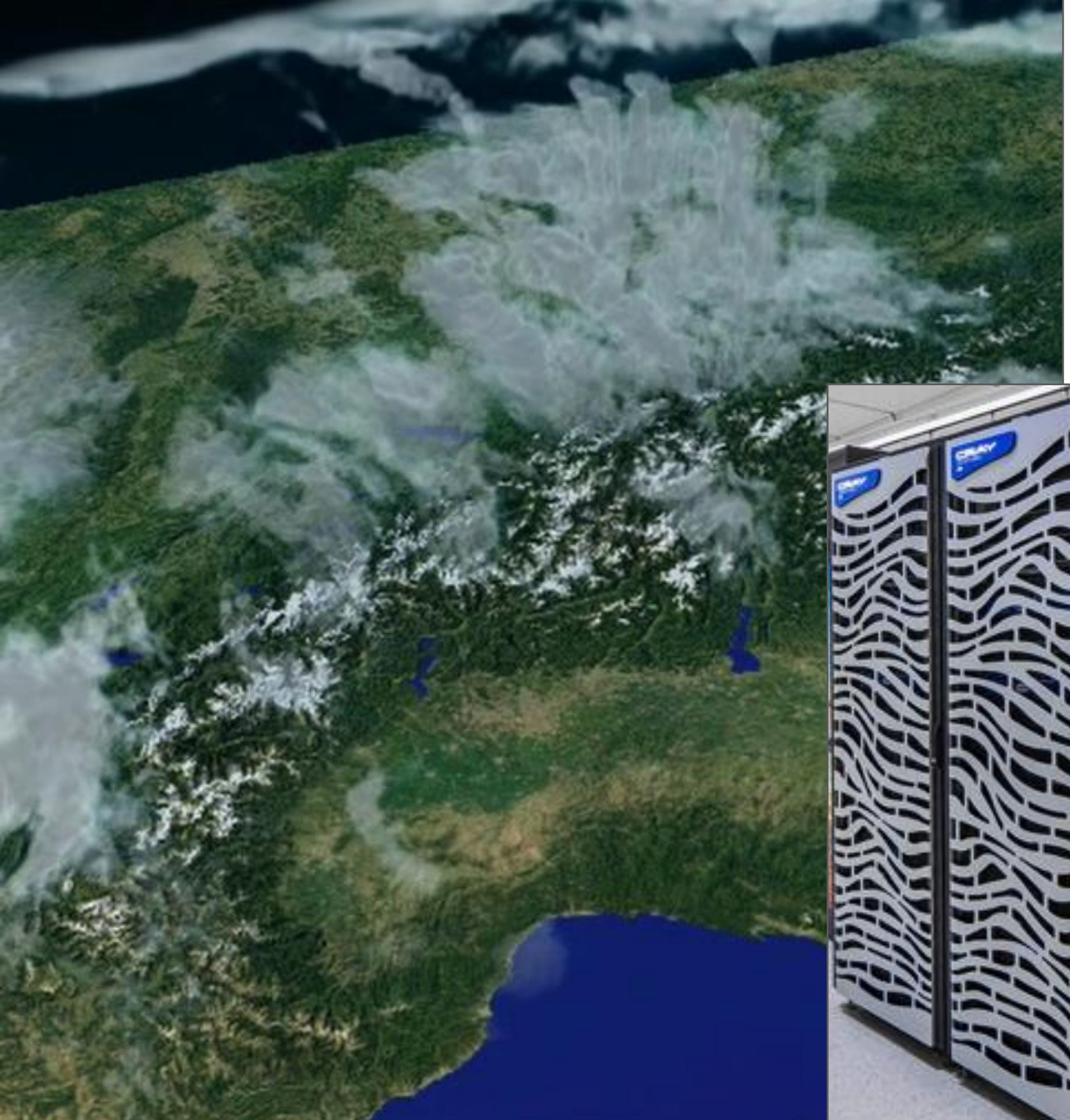


Image by NVIDIA/MeteoSwiss

World’s First GPU-Accelerated NWP

Piz Kesch (Cray CS Storm)

Installed at CSCS July 2015

2x Racks with 48 Total CPUs

192 Tesla K80 Total GPUs

High GPU Density Nodes:

2 x CPU + 8 x GPU

> 90% of FLOPS from GPUs

Operational NWP Mar 16

ACME First GPU Accelerated Coupled Climate Model

- **ACME: Accelerated Climate Modeling for Energy**
  - First fully accelerated climate model (GPU and MIC)
  - Consolidation of DOE ESM projects from 7 into 1
  - DOE Labs: Argonne, LANL, LBL, LLNL, ORNL, PNNL, Sandia
  - Towards NH global Atm 12 km, Ocn 15 km, 80 year

- **ACME component models and GPU progress**
  - Atm – ACME-Atmosphere (NCAR CAM-SE fork)
    - Dycore now in trunk, CAM physics started with OpenACC
  - Ocn – MPAS-O (LANL)
    - LANL team at ORNL OpenACC Hackathon during 2015
  - Others – published OpenACC progress
    - Sea-Ice – ACME-CICE (LANL)
    - Land – CLM (ORNL, NCAR)
    - Cloud Superparameterization – SAM (SBU, CSU)
    - Land-Ice – PISCEES (Multi-lab – LLNL, Sandia)

ACME Accelerated Climate Modeling for Energy

Update on DOE Pre-Exascale CORAL Systems

- **US DOE CORAL Systems**
  - **ORNL Summit at 200 PF Early 2018**
  - **LLNL Sierra at 150 PF Mid-2018**
  - **Nodes of POWER 9 + Tesla Volta GPUs**
  - **NVLink Interconnect for CPUs + GPUs**
- **ORNL Summit System**

  Based on original 150 PF plan:
  - **Approximately 3,400 total nodes**
  - **Each node 40+ TF peak performance**
  - **About 1/5 of total #2 Titan nodes (18K+)**
  - **Same energy used as #2 Titan (27 PF)**

CORAL Summit System

5-10x Faster than Titan

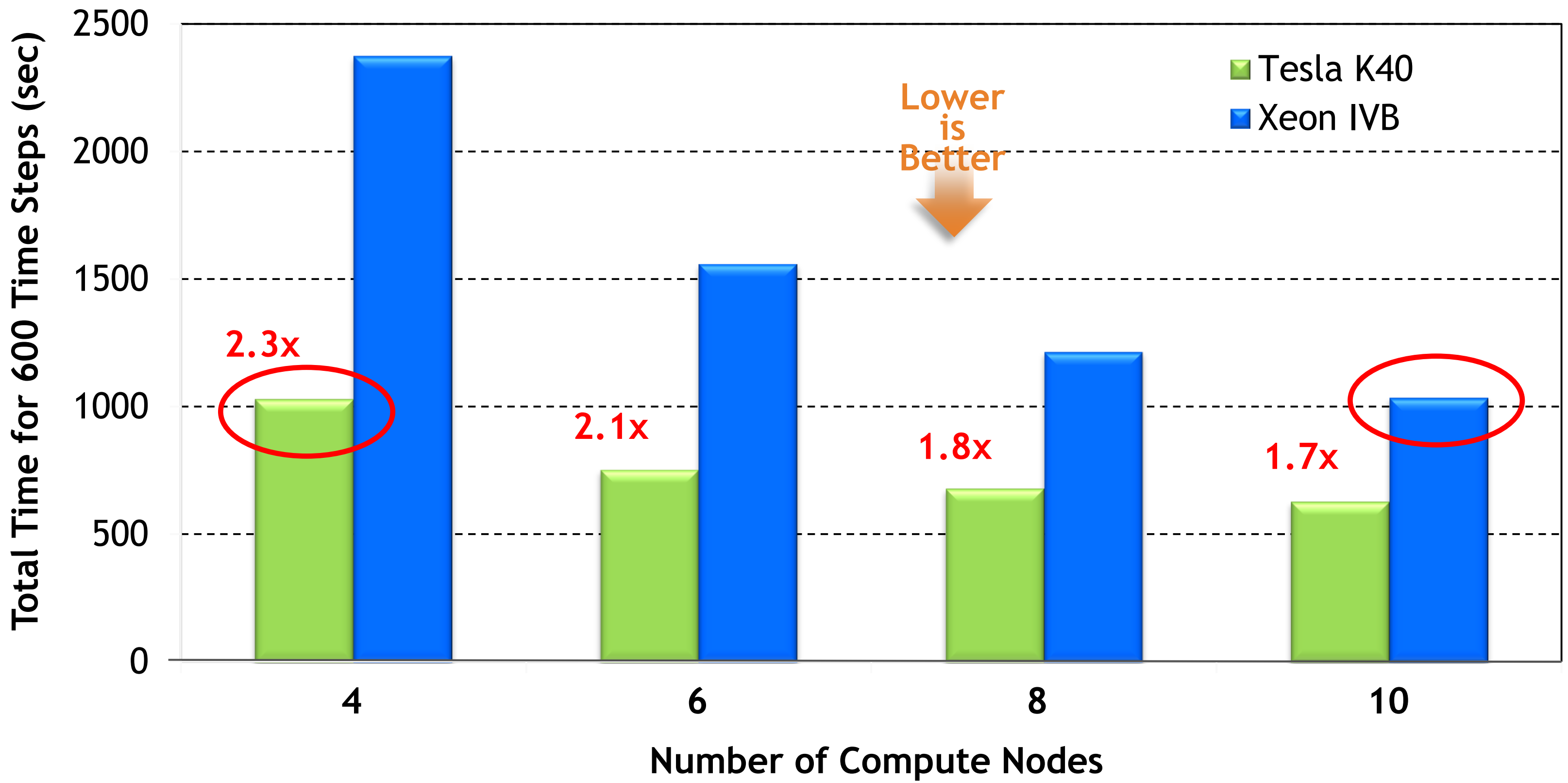
1/5th the Nodes,

Same Energy Use as Titan

(Based on original 150 PF)

NEMO 3.5 Performance with OpenACC and GPUs

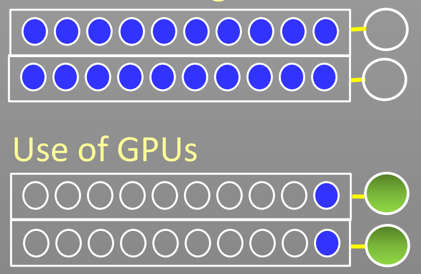
Comparisons for 2 x CPU vs. 2 x GPU for ORCA025 Configuration



Node utilization:
2 x IVB + 2 x K40

Without using GPUs

Use of GPUs



ORCA025 settings:

- Output every 5 days
- Total run: 10 days
- Time steps: 600

NOAA GFDL Radiation Scheme GPU Developments

NCAR UCAR CISL Computational & Information Systems Lab

2016 MultiCore 6 Workshop

September 13 and 14, 2016, at the National Center for Atmospheric Research in Boulder, Colorado

OpenACC Evaluation of Legacy Radiation Scheme

Raymond Menzel, NOAA/GFDL: *Challenges of offloading radiative heating calculations using OpenACC and OpenMP accelerator device directives in the CM2.5 FLOR Earth System Model*

CUDA Development of New Radiation Scheme

Venkatramani Balaji, Princeton University: *On the Application of GPU Hardware for Line by Line Radiation Computations*

NOAA GFDL ENGILITY Your Mission. Our Commitment.

https://www2.cisl.ucar.edu/events/workshops/multicore-workshop/2016/2016-agenda

NVIDIA Air Quality Prediction Project with CAS IAP

GPU Development of ODE's for solution of chemical reactions in NAQPMS

Comparisons of NAQPMS for Single-Core CPU vs. GPUs

	Nodes Num.	Run Time (Sec)	Speedup(X)
CPU[1] Intel(R) Xeon E5-2690 @ **3.0** GHz	473600	24295.7	-
K40	473600	375.9	**64.6**
K80[2]	473600	376.5	**64.5**

1: In the test, only one core is used. We did not parallelize the CPU code.
2: K80 has two GPU chips, and only one chip is used in this test.

From NVIDIA GTC 2016
Apr 2016, San Jose, CA, USA

MBE - A GPU-based fast, robust and precise solver for chemical ODEs

-- by Dr. Fan Feng, CAS IAP

Achieved >64x speedup for modified backward Euler (MBE) solver on K40 and K80 GPUs vs. single-core Xeon CPU

GPUs at Convergence of Data and HPC in ESM

Fusion of Observations from Machine Learning with the Model

- Yandex developments of “ML + Model” for Hyperlocal NWP with WRF:
  Yandex Introduces Hyperlocal Weather Forecasting Service Based on Machine Learning Technology
- Deep Learning primary topic at NCAR workshop Climate Informatics 2015
- IBM acquisition of The Weather Company – applied NWP data analytics

Data Assimilation – Next Phase Following Model Development

- 4DVAR GPU development success with MeteoSwiss and others
- RIKEN study of 10,240 member ensemble with NICAM (Miyoshi, et al.)
  Largest ensemble simulation of global weather using real-world data

Programming Strategies for GPU Acceleration

Applications

GPU Libraries	OpenACC Directives	Programming in CUDA
Provides Fast "Drop-In" Acceleration	GPU-acceleration in Standard Language (Fortran, C, C++)	Maximum Flexibility with GPU Architecture and Software Features

Increasing Development Effort →

NOTE: Many application developments include a combination of these strategies

TOPICS OF DISCUSSION

- NVIDIA HPC DIRECTIONS
- ESM GPU PROGRESS
- PGI UPDATE – D. NORTON