

OpenPOWER and NVIDIA GPUs

Jeremy Appleyard, Ph.D.; jappleyard@nvidia.com ; NVIDIA, Oxford, UK

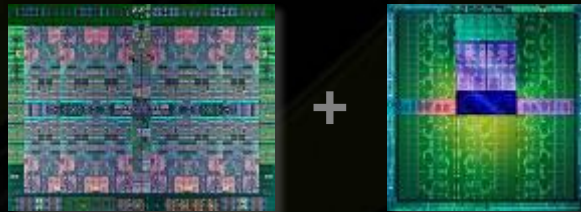
Peter Messmer, Ph.D.; pmessmer@nvidia.com ; NVIDIA, Zurich, CH

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

Computing with IBM Power CPU + NVIDIA GPU

Next-Generation IBM Supercomputers and Enterprise Servers

Long term roadmap integration



POWER CPU

Tesla GPU

OpenPOWER Foundation

Open ecosystem built on Power Architecture



TYAN

& 50+ more...

First GPU-Accelerated POWER-Based Systems Available Oct 2014

IBM Power CPU + Tesla GPU Available 2014

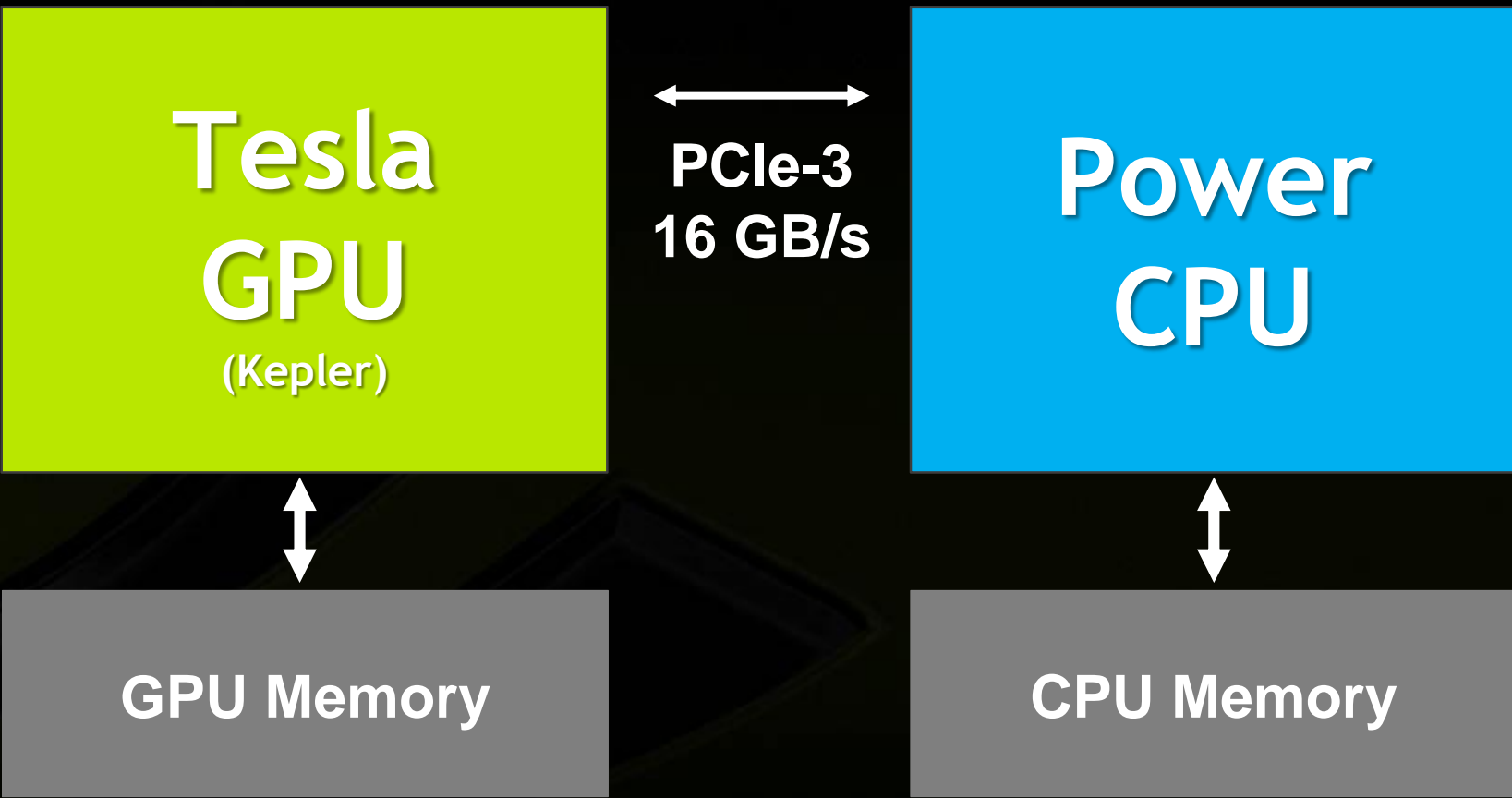


IBM POWER S824L

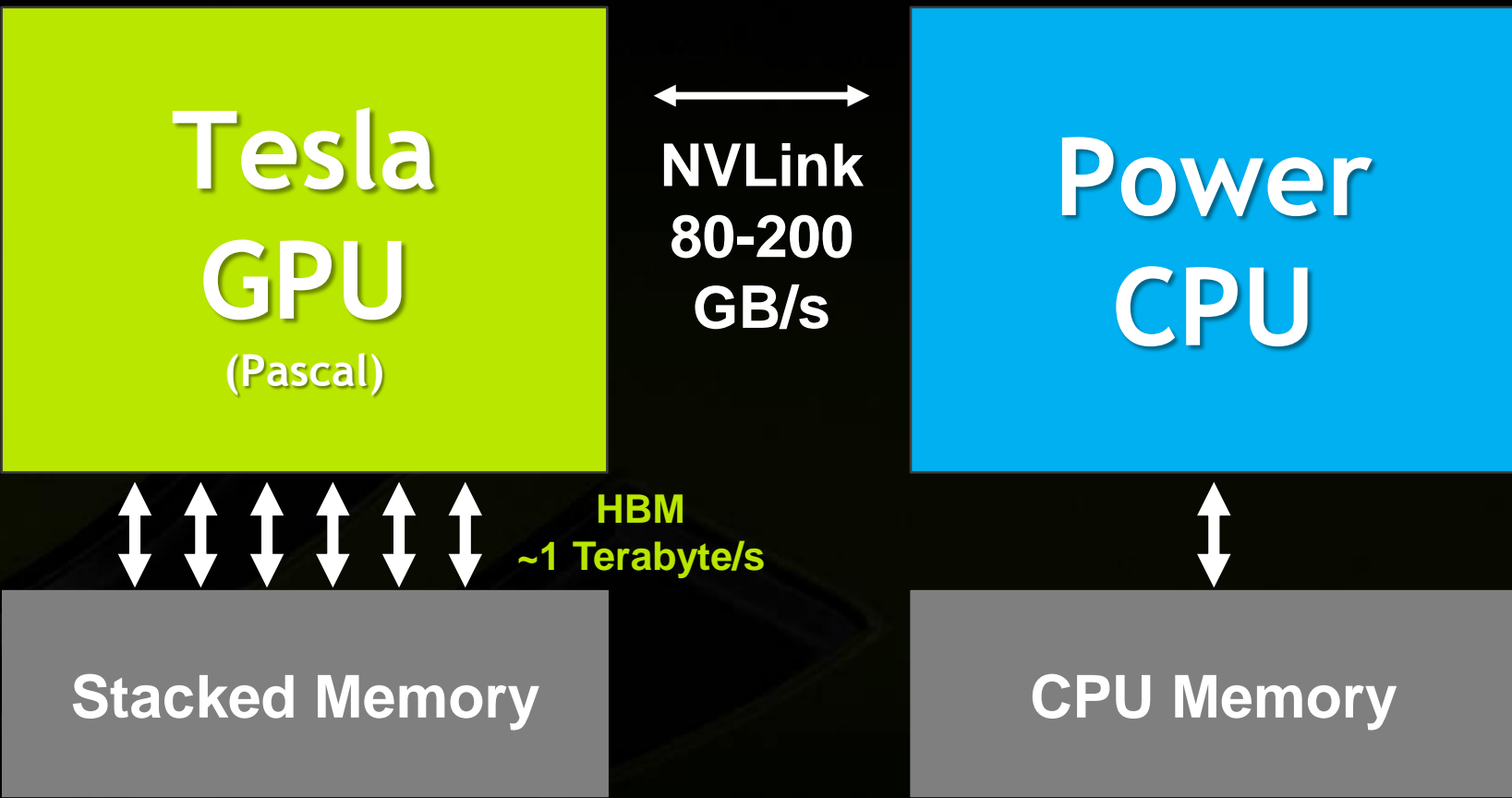


- **2 x POWER8 CPUs**
 - 1TB Memory Capacity
 - 384 GB/s Max Mem Bandwidth
- **2 x NVIDIA Tesla K40 GPU Accelerators**
- **Linux Operating System**

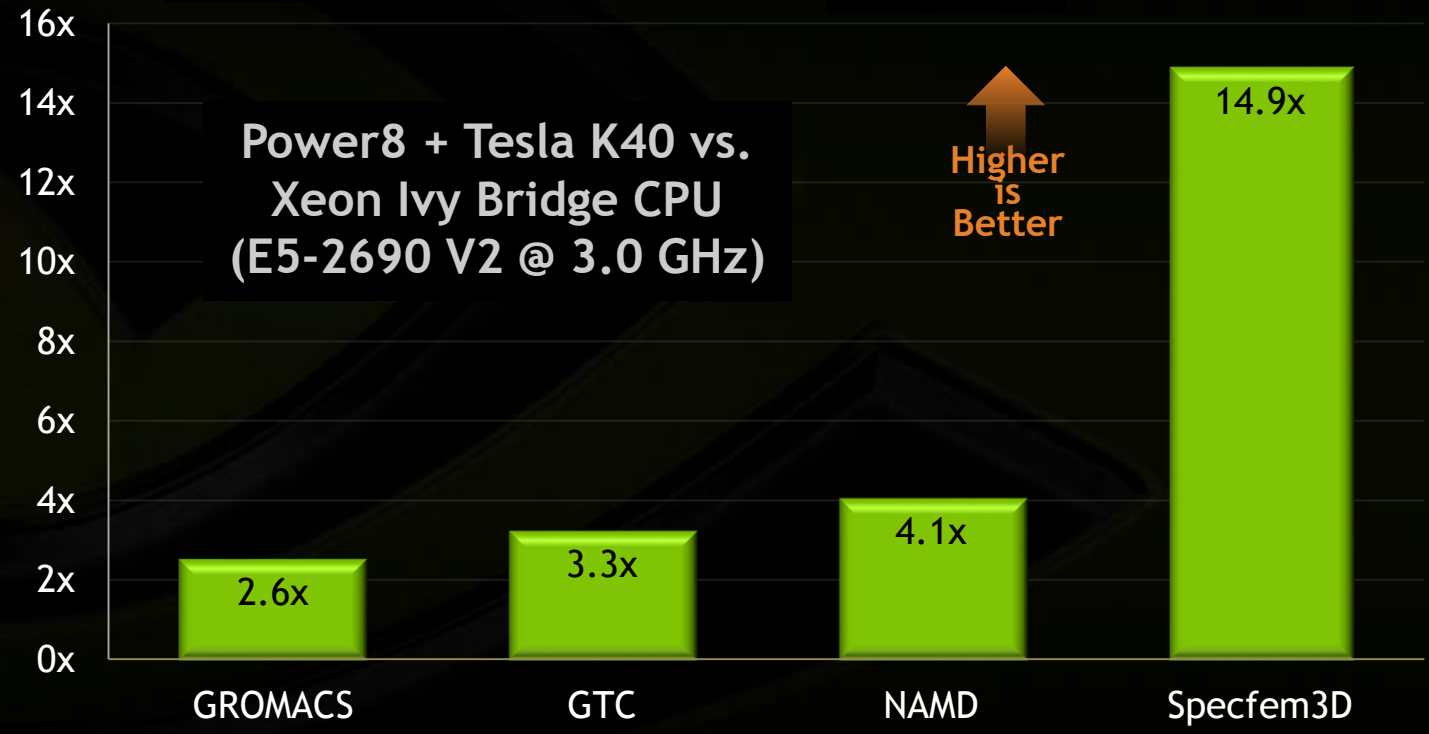
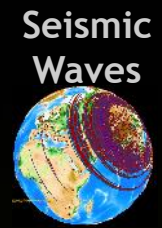
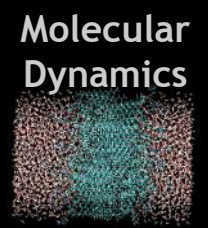
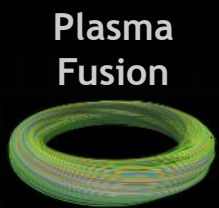
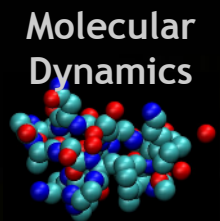
IBM Power CPU + Tesla GPU Available 2014



High Speed Data Interfaces Available 2016



Early Benchmarks: Power8 + Tesla K40 GPUs



Results from NVIDIA PSG Cluster (HQ, USA)



<http://psgcluster.nvidia.com/trac>

- About 32 compute nodes
- Latest IBM and Intel CPUs
- Latest compilers and tools
- This study: Single CPUs and single Tesla K40 GPU

Open for remote login tests, contact NVIDIA for access:

Stan Posey, sposey@nvidia.com

COSMO

Towards GPU-accelerated Operational Weather Forecasting

- Oliver Fuhrer (MeteoSwiss), NVIDIA GTC 2013, **Mar 2013**

Source: <http://on-demand.gputechconf.com/gtc/2013/presentations/S3417-GPU-Accelerated-Operational-Weather-Forecasting.pdf>

Delivering Performance in Scientific Simulations: Present and Future Role of GPUs in Supercomputers

- Thomas Schulthess (CSCS), NVIDIA GTC 2014, **Mar 2014**

Source: <http://on-demand.gputechconf.com/gtc/2014/presentations/S4719-scientific-sims-gpus-supercomputing.pdf>

Implementation of COSMO on Accelerators

- Oliver Fuhrer (MeteoSwiss), ECMWF Scalability Workshop, **Apr 2014**

Source: <http://old.ecmwf.int/newsevents/meetings/workshops/2014/Scalability/>

COSMO Model Operational NWP at Meteo Swiss

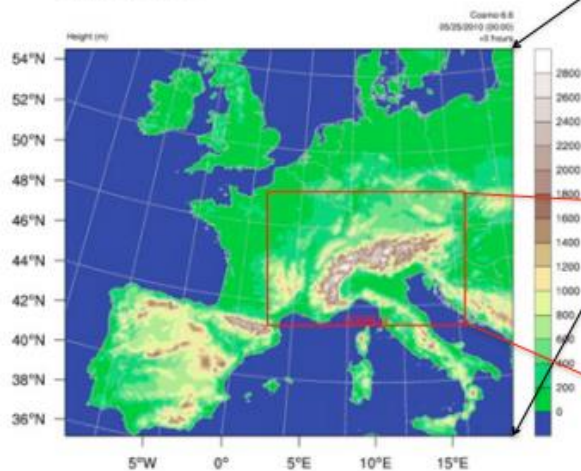


Delivering performance in scientific simulations: present and future role of GPUs in supercomputers

-by Dr. Thomas Schulthess, Director CSCS; NVIDIA GTC, March 2014

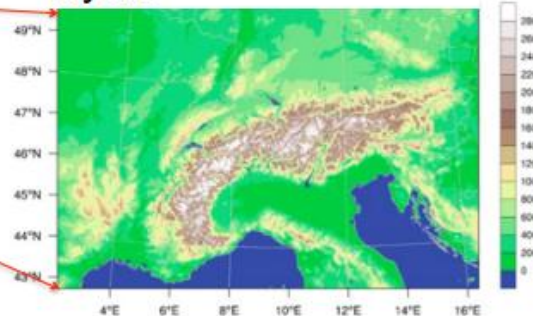
COSMO-7
3x per day 72h forecast
6.6 km lateral grid, 60 layers

Orography of COSMO-7



ECMWF
2x per day
16 km lateral grid, 91 layers

COSMO-2
8x per day 24h forecast
2.2 km lateral grid, 60 layers



Some of the products generated with COSMO-2

- Daily weather forecast
- Forecasting for Swiss air traffic control (Sky Gide)
- Safety management in events of nuclear incidents

Implementation of COSMO on Accelerators

-by Dr. Oliver Fuhrer, MeteoSwiss; ECMWF Scalability Workshop, Apr 2014

Approach

- **Dynamics**
 - 40k lines, 60% runtime
 - Few developers
 - Strongly memory bandwidth bound

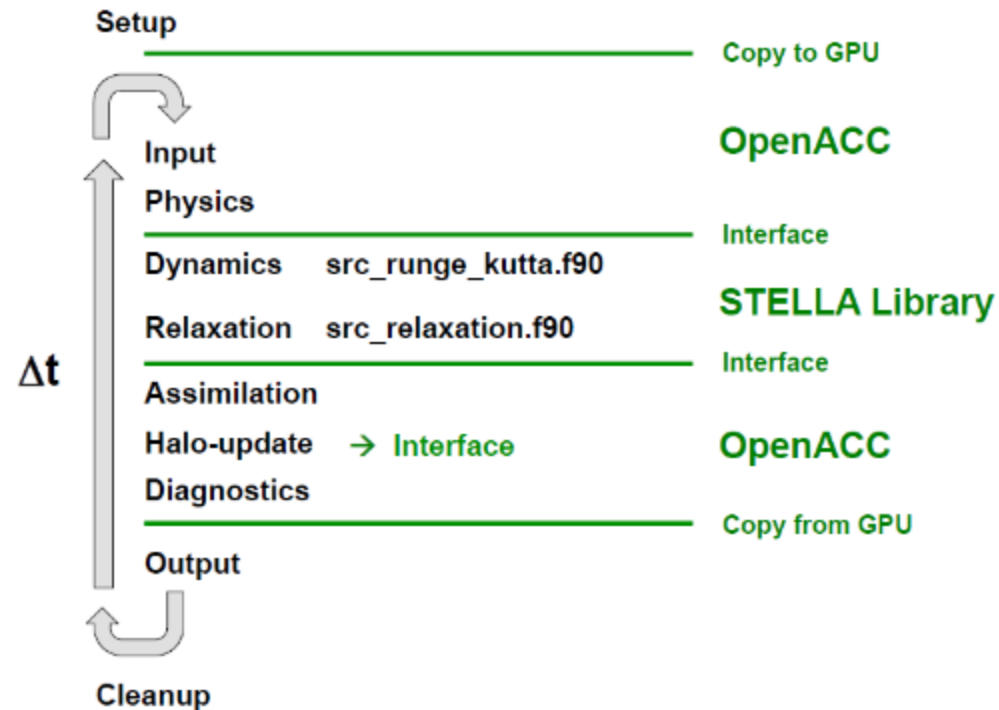
Aggressive rewrite

 - Data structures
 - C++
 - DSEL
- **Physics & Assimilation**
 - 130k lines, 25% runtime
 - Several developers
 - “Plug-in” from other models
 - Less memory bandwidth bound

Port to GPU

 - keep source
 - directives

Implementation



Implementation of COSMO on Accelerators

-by Dr. Oliver Fuhrer, MeteoSwiss; ECMWF Scalability Workshop, Apr 2014

Approach

- **Dynamics**
 - 40k lines, 60% runtime
 - Few developers
 - Strongly memory bandwidth bound

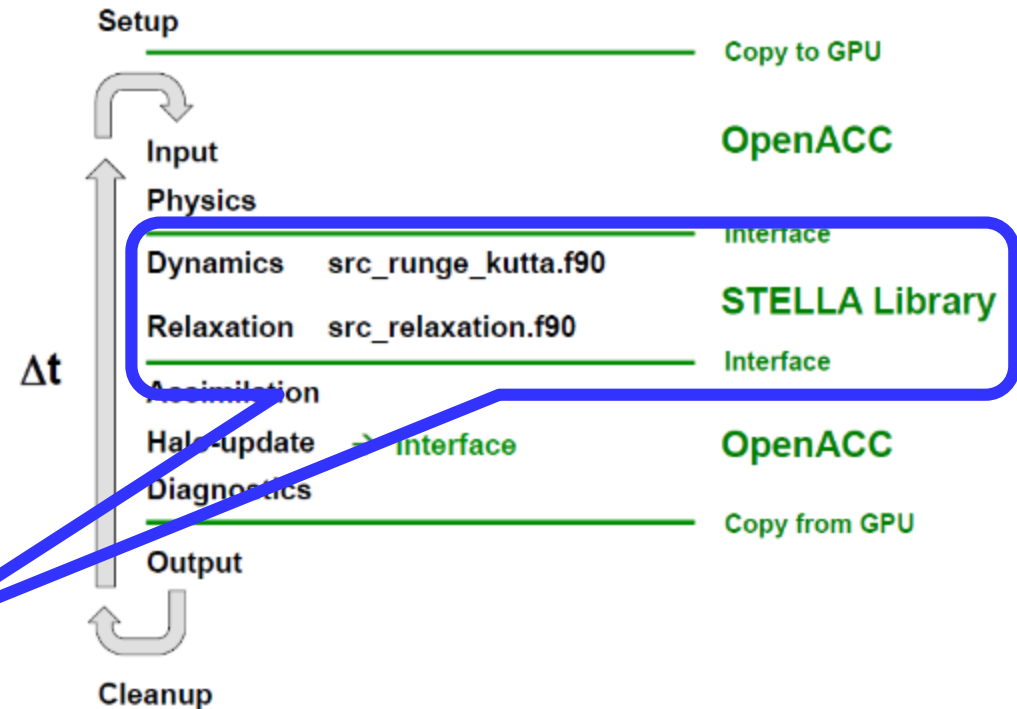
Aggressive rewrite

 - Data structures
 - C++
 - DSEL
- **Physics & Assimilation**
 - 130k lines, 25% runtime
 - Several developers
 - “Plug-in” from other models
 - Less memory bandwidth bound

Port to GPU

 - keep source
 - directives

Implementation

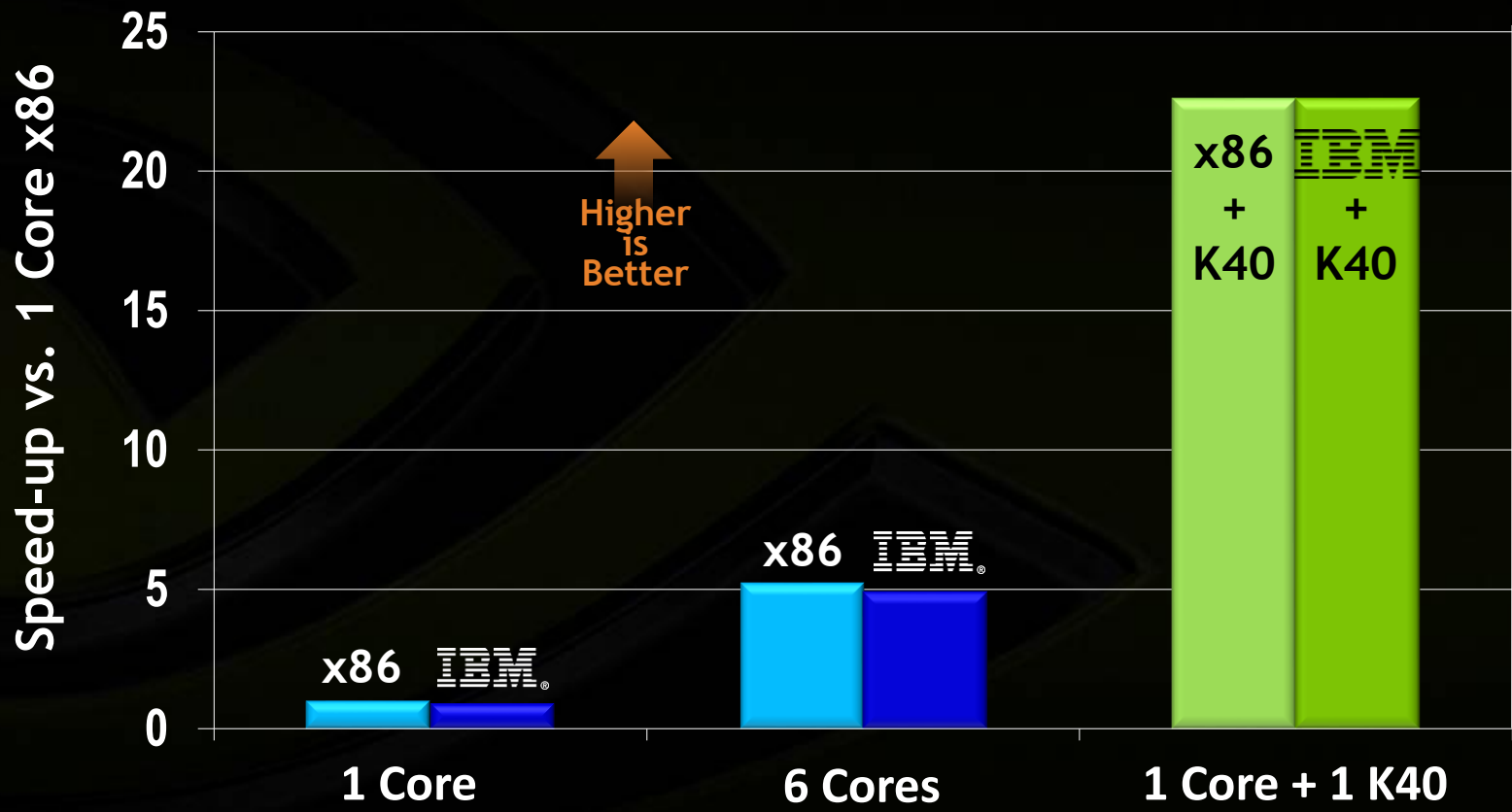


**Dynamics only test
for Power8 + K40**

COSMO Dynamics Results for Power8 + K40



Numerical Results Validation Confirmed by MeteoSwiss



Results from NVIDIA PSG Cluster (HQ, USA)



<http://psgcluster.nvidia.com/trac>

- COSMO SVN revision 6307
- PerformanceUnittest.Dycore case, 122 x 122 (x60)
- All results single socket
- IBM Power 8:
 - 12 Cores @3.9 GHz
- Power 8 Compiler: gcc 4.9.1
- x86 Xeon E5-2690 v2:
 - 8 Cores @ 3.0 GHz
- X86 Compiler: gcc 4.8.2
- NVIDIA Tesla K40 GPU
- CUDA 5.5

Summary: OpenPOWER and NVIDIA GPUs



- **OpenPOWER offers an alternative to x86-based GPU accelerated HPC**
 - First GPU accelerated, Power-based systems available during 2014
- **IBM and NVIDIA have developed a tightly integrated HPC roadmap**
 - New technologies in 2016: Power roadmap, Pascal GPU, NVLink, more
- **The COSMO model was successfully demonstrated on Power8 + K40**
 - Dynamics test achieved numerical validation and performance expectations

Thank You and Q&A

Jeremy Appleyard, Ph.D.; jappleyard@nvidia.com ; NVIDIA, Oxford, UK

Peter Messmer, Ph.D.; pmessmer@nvidia.com ; NVIDIA, Zurich, CH

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