Earth System Modeling at the Extreme Scale

15th ECMWF HPC Workshop
1~5 October 2012

Per Nyberg
Director, Business Development
nyberg@cray.com

CRAY®
THE SUPERCOMPUTER COMPANY

Topics

- Cray Update
- Earth System Modeling at the Extreme Scale
- Climate Knowledge Discovery: A community initiative
  - *Time permitting*

Cray Is Supercomputing

We build solutions to help solve "Grand Challenges" in science, engineering and knowledge discovery

HPC Systems

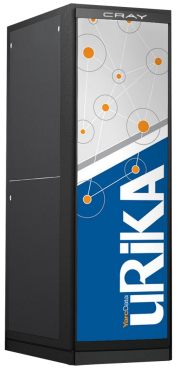
Breakthrough performance and scalability

Storage & Data Management

Scalable, manageable high performance storage systems

Big Data Solutions

Large scale relationship analytics



YarcData
A DIVISION OF CRAY INC.

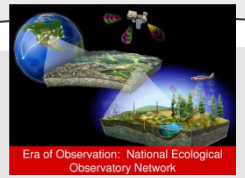
As Science Evolves, Cray is at the Forefront of Technology Development

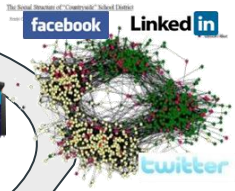
- **"…without the productivity of new disciplines based on data, we cannot solve important problems of the world"**
  - Iwata, S.: Scientific "Agenda" of Data Science. Data Science Journal, 7, 54 (2008)
- **As the nature of science is evolving to include data intensive and data driven approaches, Cray is developing platforms that provide unique advantages:**
  - Massive scalability with distributed memory.
  - Architectures specifically designed for direct analytical queries and knowledge synthesis.
  - Integrated storage technologies.

Evolution of Science and Knowledge Discovery

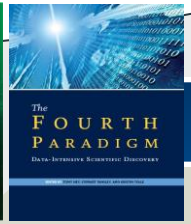
Sensors, devices, simulations, social,…



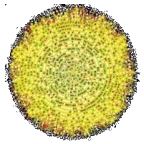



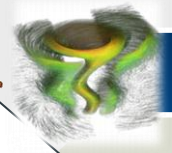


Multi-disciplinary, multi-institutional



Data Intensive / Data Driven Science





Computational Simulation of Complex Phenomena

$$\mathbf{x}_a - \mathbf{x}_b = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}_b)$$

Theoretical Research



Experimental Research

Cray National Leadership Systems Worldwide
Experience in Production Petascale Facilities

Cray Cascade System

- **Next generation follow-on the Cray XE6**
- **Aries Interconnect:**
  - Low Radius, High Bandwidth Network
  - "Dragonfly" topology with scalable global bandwidth
  - Adaptive routing
  - Electro-Optical Signaling
  - Support for globally addressable memory
    - Supports Shmem, UPC, CAF, Global Arrays
- **Move to Intel processor roadmap including Xeon Phi**
- **Continued support for NVIDIA**
- **Customer announcements to date include:**
  - NERSC 2 Pflop/s
  - CSCS 750 Tflop/s
  - CSC Finland
  - CSIRO Pawsey Centre
  - HLRS Germany
  - Kyoto University

Cray's Storage Solution Ecosystem

- Cray Sonexion
- Extreme Scalability
- Tested with CLE
- Fully Supported
- Fast Installations
- Cluster-attached

Modular, Integrated Systems

Strategic Partners and OEMs

- Flexibility
- Cray Integrated
- Tested with CLE
- Fully Supported

NetApp™

xyratex

DataDirect NETWORKS

OpenSFS Open Scalable File Systems, Inc.

SPECTRA

Quantum

Bright Computing
Advanced cluster management made easy

whamcloud

Collaborative Innovation.

Single point of support.

Flexible, Open Storage Solutions

- Lustre leaders
- Strong community
- Improved SW testing
- Cray Scale Testing
- Cray Support

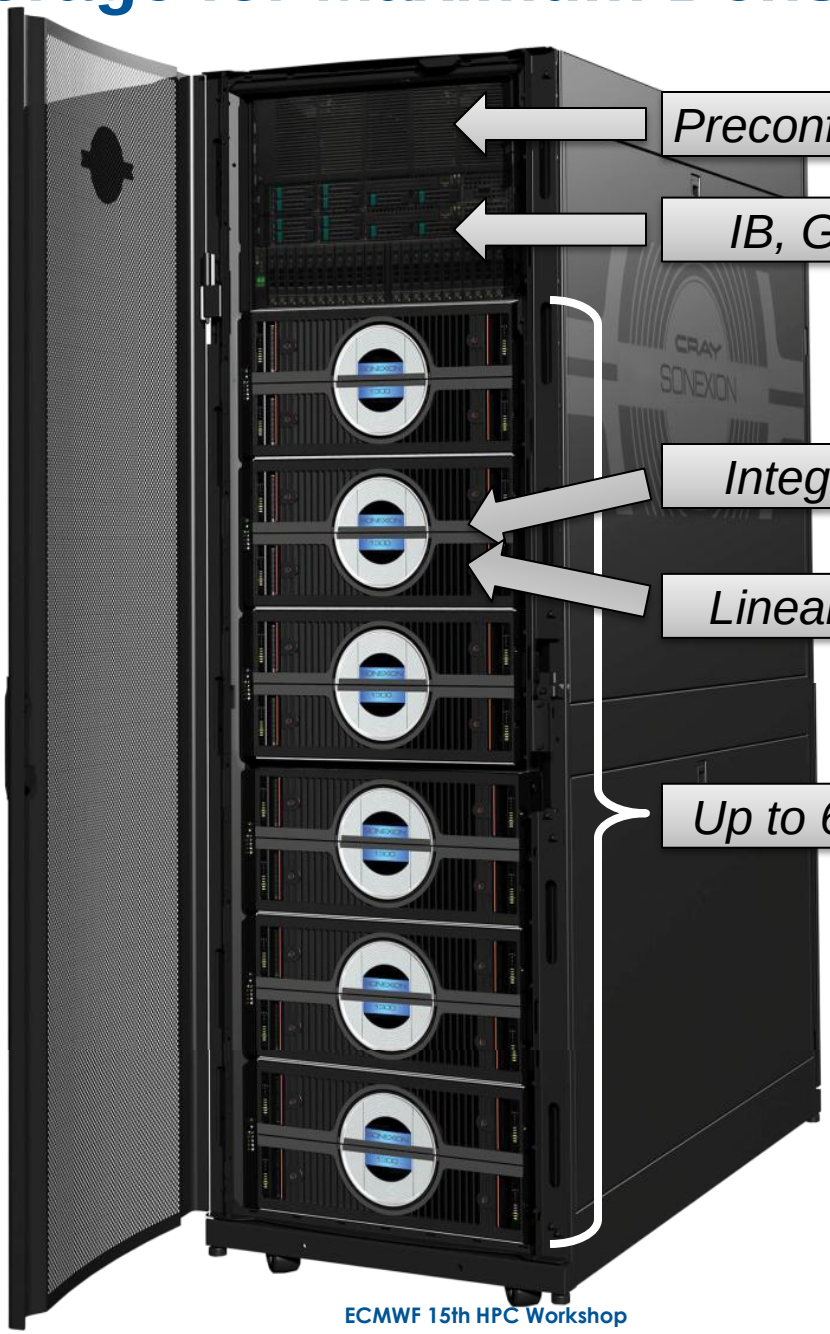
- Integrated
- Tested with CLE
- Fully Supported
- Data movement
- Flexible Logins
- Bright, Dell, Mellanox

Mellanox TECHNOLOGIES

Integrated Storage for Maximum Density

CRAY SONEXION

- **Exclusively designed for Lustre**
- **Fully modular and scalable**
- **Performance scales with capacity**



YarcData uRiKA: Big Data Appliance for Graph Analytics

- **Cray Hardware Engine**
  - Originally designed for deep analysis of large datasets
  - Very large *scalable* shared memory
    - Architecture can support 512TB shared memory
  - Unique highly multithreaded architecture
    - 128 hardware threads per processor
    - Extreme parallelism, hides memory latency
  - Highly Scalable I/O
    - Up to 350 TB/hr
- **Multithreaded Graph Database**
  - Highly parallel in-memory RDF quad store
  - High performance inference engine
- **Industry Standard Front End**
  - Based on Jena open source semantic DB
  - All standard Linux infrastructure and languages
  - Lustre parallel file system

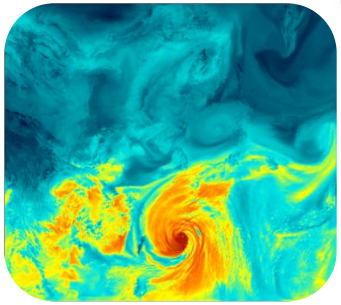
Gartner

"YarcData's uRiKA Shows Big Data Is More Than Hadoop and Data Warehouses." (Sept 2012)

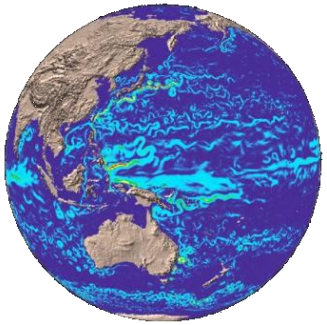
- *"YarcData has designed uRiKA with three technologies to minimize or eliminate the costs of the irregular, unpredictable leaps in graph processing.* ***A unique approach to processor design****, YarcData's Threadstorm chip, shows no slowdown under the characteristic zigs and zags of graph oriented processing…* ***the data is held in-memory in very large system memory*** *configurations, slashing the rate of file accesses. Finally, a* ***global shared memory architecture provides*** *every server in the uRiKA system* ***access to all data****."*

Cray's Presence and Experience in the Earth System Modeling Community

- **Earth System Modeling is a key area on Cray Systems worldwide:**
  - Dedicated operational NWP and research centers.
  - Multi-disciplinary research centers.
  - From Teraflops to Petaflops.
- **Cray Petascale systems have been key:**
  - In enabling transformational science
  - As development platforms for preparing earth system models for extreme scale capabilities.
  - Either you have run on 100,000 cores or you have not.
- **NCSA Blue Waters and ORNL Titan will further push the boundary of Petascale computing.**



KMA Simulation of Typhoon Man-Yi



NERSC High Resolution 1/10deg POP ocean model currents

MeteoSwiss 0.55km Resolution Topography over Alps

Some Recent Successes

MeteoSwiss

- Operational NWP
- Upgrade to 48.3 Tflop/s Cray XE6

TerraMe-Galileu Laboratory

- Located at the Brazilian National Institute for Space Research (INPE)
- 22.6 Tflop/s Cray XE6m

Mongolian National Agency of Meteorology and Environmental Monitoring (NAMEM)

- Operational NWP
- 6.46 Tflop/s Cray XE6

Arctic Region Supercomputing Center (ARSC)

- Research in arctic systems, including weather and climate, oceans and ice and permafrost.
- 41.75 Tflop/s Cray XK6m with 48 NVIDIA Tesla GPU processors

Naval Research Laboratory (NRL)

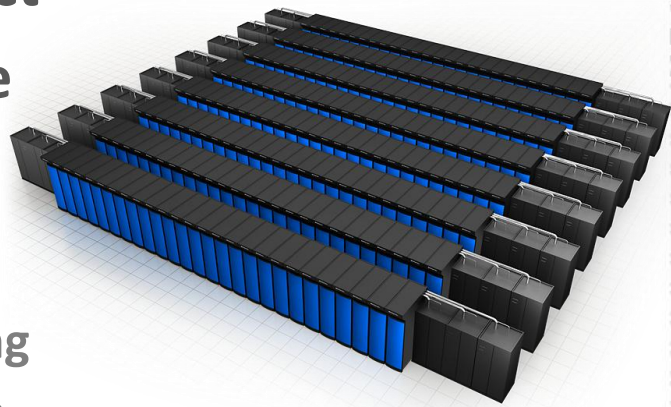
- Advanced environmental research capabilities in support of US Navy
- 54 Tflop/s Cray XE6m

Consortium for the Advanced Research of Transport of Hydrocarbons in the Environment CARTHE)

- Cray XE6m located at the University of Miami's Rosentiel School of Marine & Atmospheric Science
- Will be used studying the surface ocean currents that transport pollutants in real time.

ORNL "Titan"

- Upgrading Jaguar (XT5) to Titan (XK6)
- Peak performance 10~20 petaflops
- Upgrade will be completed by the end of 2012
- Efficiency & scalability of unprecedented scale
- Hybrid of GPUs/CPUs & Gemini interconnect
- ORNL step towards pre-exascale architecture
- Focus on programming environment:
  - To exploit all levels of parallelism.
  - Compiler directives hold the promise of allowing developers to efficiently exploit the GPUs while preserving portability to other platforms.
  - Partnership with Cray.



OpenACC

A common, open directives-based programming standard for parallel computing and accelerators

Offers portability between operating systems, host CPU's, accelerators and compliers with a single code base

OpenACC Partners

CRAY®

CAPS

NVIDIA®

PGI®

Blue Waters System 11.5 PF
Sustained Petascale Performance

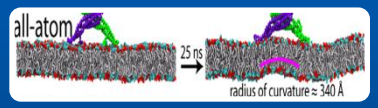
Cray System & Storage cabinets:	• >300
Compute nodes:	• >25,000
Usable Storage Bandwidth:	• >1 TB/s
System Memory:	• >1.5 Petabytes
Memory per core:	• 4 GB
Gemini Interconnect Topology:	• 3D Torus
Usable Storage:	• >25 Petabytes
Peak performance:	• >11.5 Petaflops
Number of AMD processors:	• >49,000
Number of AMD x86 core:	• >380,000
Number of NVIDIA GPUs:	• >3,000

NSF NCSA ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Cray Sonexion Storage

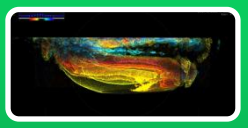
Blue Waters Science

More than 25 science teams
25 distinct research fields
selected to run on the new Blue Waters

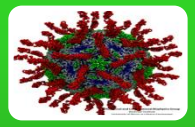


Nanotechnology

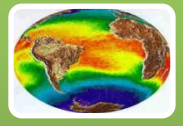
Astronomy



Earthquakes and the damage they cause



Viruses entering cells



Severe storms

Climate change

Blue Waters Climate/Weather Petascale Computing Resource Allocations (PRACs)

- **Understanding tornadoes and their parent supercells through ultra-high resolution simulation/analysis**
  - Principal Investigator: Robert Wilhelmson, University of Illinois at Urbana-Champaign
- **Using petascale computing capabilities to address climate change uncertainties**
  - Principal Investigators: Donald Wuebbles and Xin-Zhong Liang, University of Illinois at Urbana-Champaign
- **Testing hypotheses about climate prediction at unprecedented resolutions on the NSF Blue Waters system**
  - Principal Investigators: Benjamin Kirtman, University of Miami; William Large, University Corporation For Atmospheric Research; David Randall, Colorado State University; Cristiana Stan, Institute of Global Environment and Society

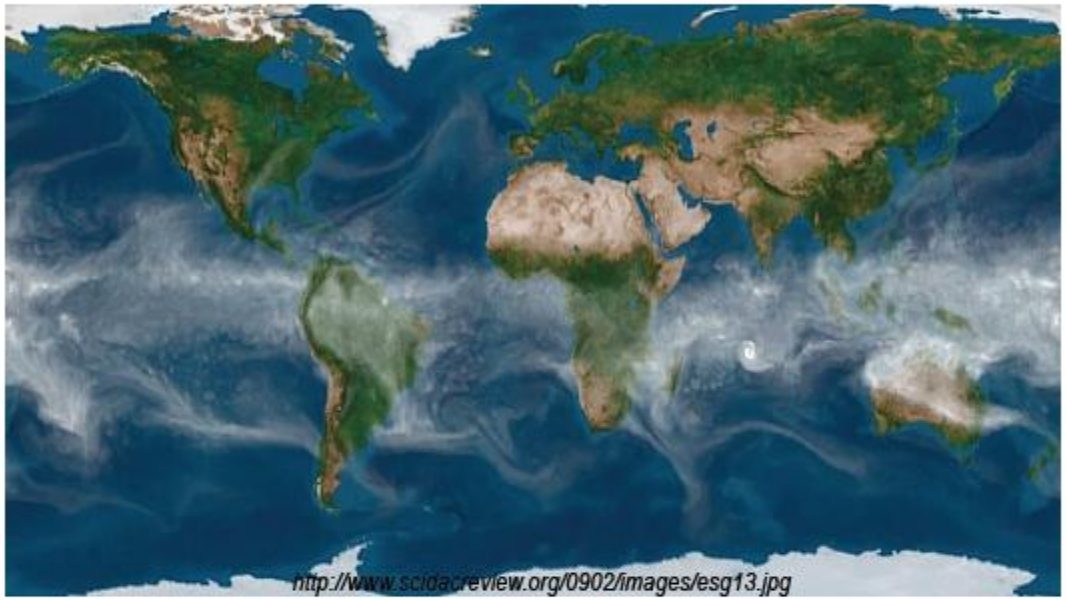
Earth System Modeling at the Extreme Scale

- **To achieve 1000x extreme scale computing will require the successful solution to a collection of interrelated science and technology challenges.**
  - Incremental change is not enough.
  - Need for standards-based disruptive technologies (!?)
  - Tension between performance, portability and programmability.
- **Infrastructures and workflows for exascale facilities will be impacted by growth in:**
  - Data set sizes
  - Real-time nature of data
  - Multi-disciplinary nature of science
- **Today HPC applications assume that the data is static and co-located (ie: on attached filesystems)**
- **In an exascale world:**
  - Data is unlikely to be at a single location.
  - Dynamic streaming data models may need to be considered.
  - Greater impact on time-to-solution by pre/post-processing.

Applications R&D for the Extreme Scale

- **Cray has been engaging in a number applications focused petascale and exascale activities involving the earth system modeling community, including:**
  - SciDAC CESM Performance Optimization
    - Science on current Petascale systems through improving scalability of CESM
  - Swiss Platform for High-Performance and High-Productivity Computing - HP2C
    - Implementations of key applications on hybrid architectures
  - UK Met Office / NCAS Project UPSCALE
    - Science on current Petascale systems through improving scalability of UM
  - ICOsahedral-grid Models for EXascale Earth system simulations (ICOMEX)
    - Exascale implementations of next generation models
  - Titan Early Science Applications
    - Pre-exascale implementations of key applications using GPUs
  - Blue Waters Petascale Computing Resource Allocations (PRAC) Teams
    - Sustained Petaflop applications
  - CRESTA Project – EU funded exascale co-design
    - Co-design approach to exascale application implementation

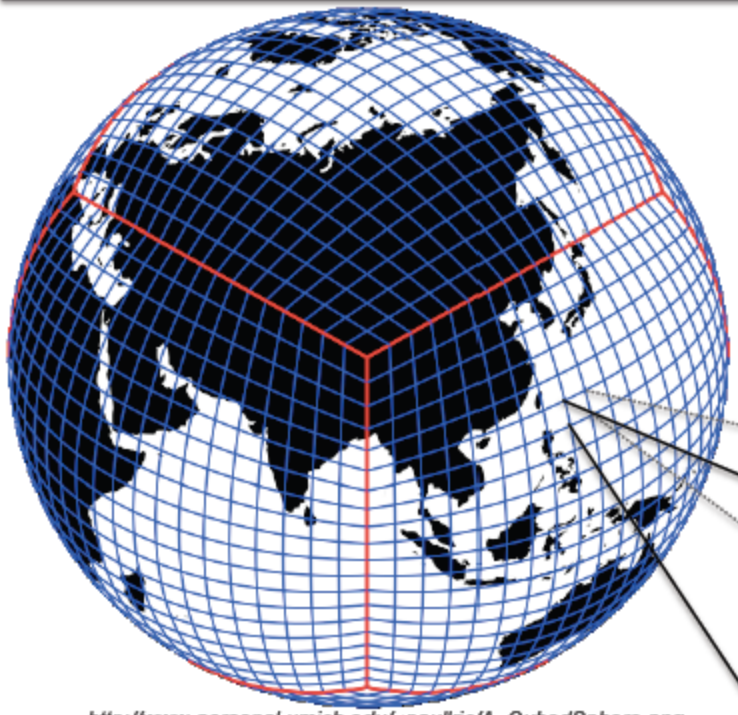
Porting The Spectral Element Community Atmosphere Model (CAM-SE) To Hybrid GPU Platforms



Matthew Norman	ORNL
Jeffrey Larkin	Cray
Richard Archibald	ORNL
Valentine Anantharaj	ORNL
Ilene Carpenter	NREL
Paulius Micikevicius	Nvidia
Katherine Evans	ORNL

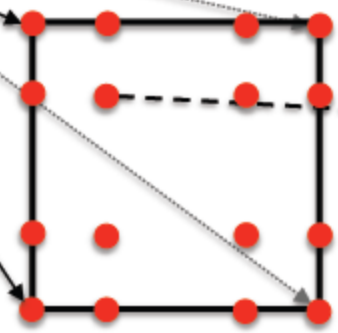
2012 Programming weather, climate, and earth-system models on heterogeneous multi-core platforms

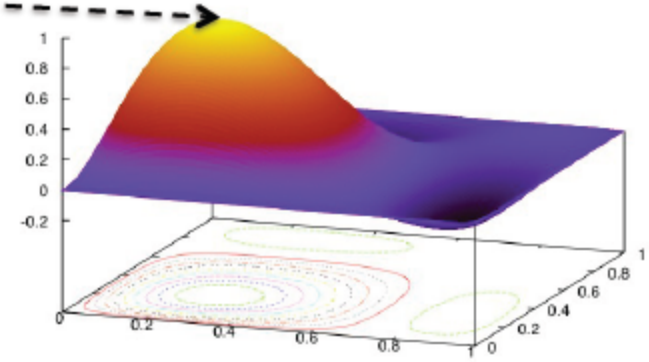
Gridding, Numerics, & Target Run



http://www-personal.umich.edu/~paullric/A_CubedSphere.png

- Cubed-Sphere + Spectral Element
- Each cube panel divided into elements
- Elements spanned by basis functions
- Basis coefficients describe the fluid



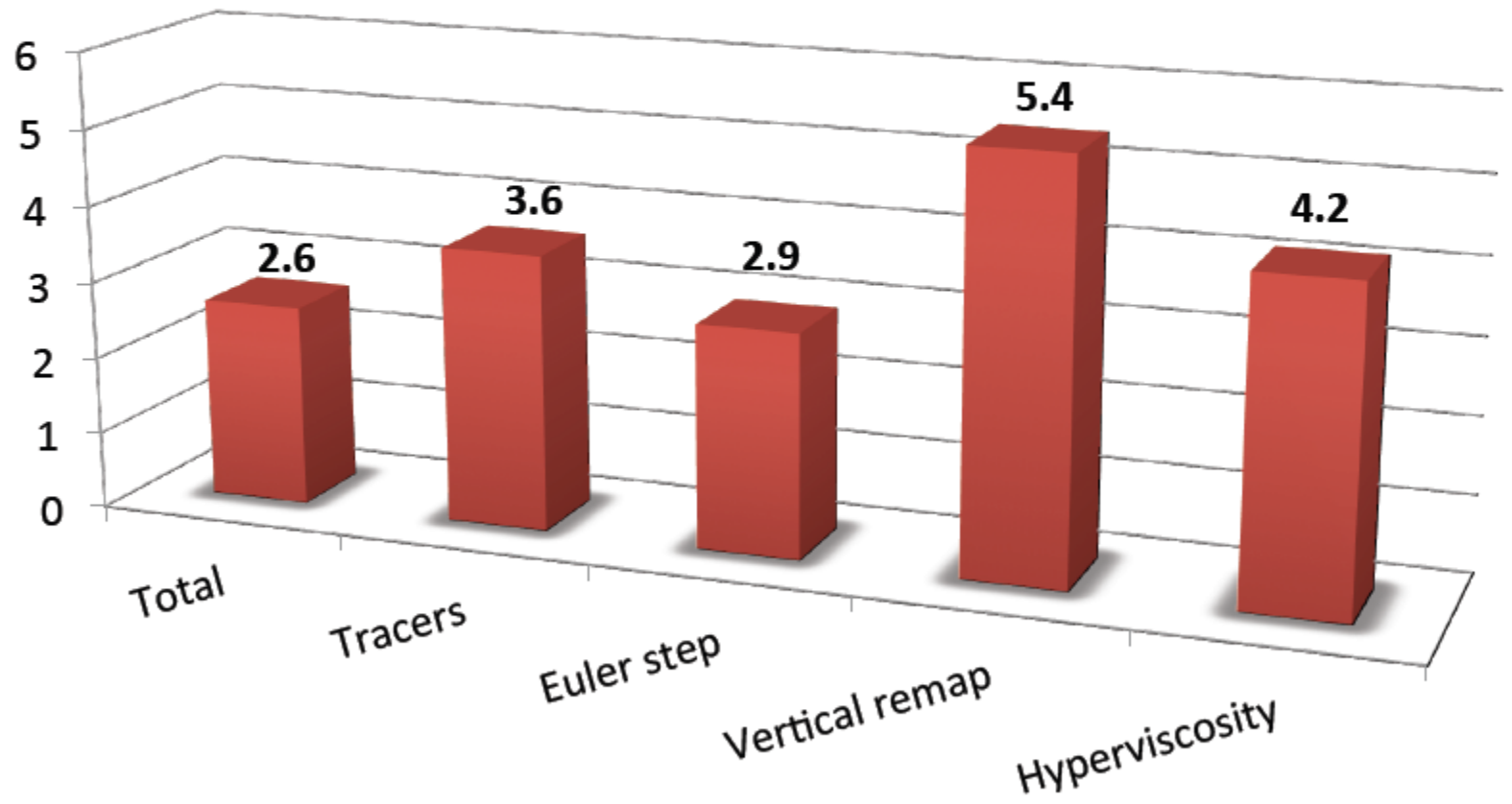


Used CUDA FORTRAN from PGI

OACC Directives: Better software engineering option moving forward

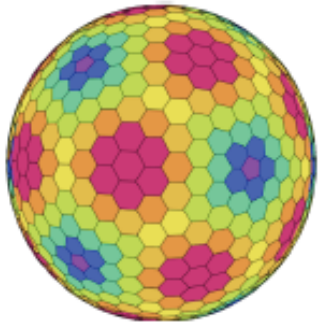
Speed-Up: Fermi GPU vs 1 Interlagos / Node

- Benchmarks performed on XK6 using end-to-end wall timers
- All PCI-e and MPI communication included



Credit: "**G8 Exascale Projects Summary",** William Tang
Princeton University and Princeton Plasma Physics Laboratory

G8 ICOMEX: ICOsahedral-grid Models for Exascale Earth System Simulations [Japan, UK, France, Germany, Russia]



→ Lead PI: **Gunther Zangl**, Univ. of Munich, Germany

Objective: Prepare 4 advanced Earth system models (ESMs) based on icosahedral grids – *with differences in terms of numerics and grid structures*

- NICAM, Special care of moist energy budget, Structured A grid
- ICON, Unstructured triangular C-grid
- MPAS, Special care of energy and vorticity budgets, Unstructured hexagonal C-grid
- DYNAMICO, Spatial discretization very similar to MPAS, except transport, Structured hexagonal C-grid

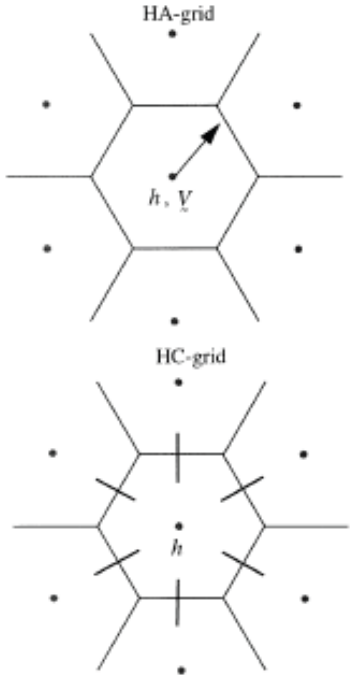
- Exascale computing required

→ *global convection-resolving climate simulations*

- Icosahedral Grids:

→ high scalability up to 10^6 cores with relative uniform mesh size

- → allows combining via explicit time-stepping schemes
- → avoids global communication
- → enables straightforward implementation of variable resolution grids to deal with *multi-scale interactions of atmospheric and/or ocean dynamics*



Summary

- **As the nature of science is evolving to include data intensive and data driven approaches, Cray is developing platforms that provide unique advantages:**
  - Massive scalability with distributed memory.
  - Architectures specifically designed for direct analytical queries and knowledge synthesis.
  - Integrated storage technologies.
- **Cray's technologies are playing a key role in supporting the weather and climate communities:**
  - Enabling unprecedented simulations.
  - Supporting the development of next generation modeling capabilities.
- **Extreme scale computing will require the successful solution to a collection of interrelated science and technology challenges.**
  - Need for standards-based disruptive technologies (!?)
  - Tension between performance, portability and programmability.

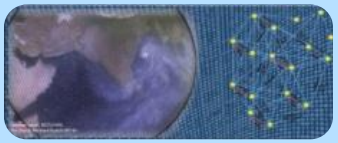
Climate Knowledge Discovery Community Initiative

- **Activity to jumpstart the investigation of a CKD technology stack.**
  - *Support discovery and understanding of unknown complex patterns from data that span across multiple dimensions.*
  - **Data driven** approach - network science, data mining, machine learning…
  - The breakthroughs needed to address CKD challenges will come from collaborative efforts involving several disciplines:
    - End-user scientists, computer and computational scientists, etc…

- **Organizers/Contributors over the past year:**
  - Reinhard Budich, Max-Planck-Institut für Meteorologie
  - Per Nyberg, Cray
  - Karsten Steinhaeuser, University of Minnesota
  - Tobias Weigel, DKRZ
  - John Feo, PNNL
  - Peter Fox, Rensselaer Polytechnic Institute
  - Reik Donner, PIK
  - Jim Kinter, Center for Ocean-Land-Atmosphere Studies
  - …

Climate Knowledge Discovery Initiative Workshops

1st CKD Workshop



- First workshop convened by Jochem Marotzke (MPI-M), Per Nyberg (Cray), Thomas Ludwig (DKRZ).
- Hosted by DKRZ in Hamburg, March 30 to April 1, 2011.
- Brought together 50 experts from various domains, centers and countries.
- Workshop report published.
- Article “Network-Based Approaches to Climate Knowledge Discovery” published in AGU EOS on 22 November.

2nd CKD Workshop

- Part of SC11 programme – Seattle November 2011
- Proceedings will be published in ACM
- Organizers: Reinhard Budich (MPI-M), John Feo (PNNL), Per Nyberg (Cray), Tobias Weigel (DKRZ)

Climate Knowledge Discovery Initiative Workshops

3rd CKD Workshop

- Held as a session at the European Geophysical Union Annual Meeting: "NP2.6 Complex networks: Theory and methods applied to geophysical systems"
- Convener: R. V. Donner (PIK)
- Co-Conveners: J. Kurths (PIK), T. Nocke (PIK), R. Budich (MPI-M), P. Nyberg (Cray), K. Steinhaeuser (U. Minnesota)
- Vienna, 22-27 April 2012

4th CKD Workshop

- Part of SC12 programme – Salt Lake City November 12
- Proceedings will be published in ACM
- Organizers: Reinhard Budich (MPI-M), John Feo (PNNL), Per Nyberg (Cray), Karsten Steinhaeuser (U. Minnesota), Tobias Weigel (DKRZ)

Image Credit: Thomas Nocke, Potsdam Institute for Climate Impact Research (PIK)

Thank you for your attention. Questions ?

More information on CKD activities is available at:

https://redmine.dkrz.de/collaboration/projects/ckd-workshop