High-resolution climate and Community Earth System Model

John Dennis
dennis@ucar.edu

What is CESM?

- Consists of a set of 4->6 geo-components
  - ATM, OCN, LND, CICE, GLC, WAVE
  - Run on potentially different grids
  - Exchange boundary data with each other **via** coupler
    - hub and spoke architecture
- Large code base: >1.3M lines
  - Fortran 90 (mostly)
  - Developed over 20+ years
  - 200-300K lines are critically important
    - Communication, not computational kernels
- CESM is an interdisciplinary collaborative effort
  - DOE, NSF, NOAA, University Community
  - Applied Math, CS, software engineering, climate scientists

NWSC-Yellowstone is now operational

- Compute Nodes
  - Processor: 2.6 GHz Intel Sandy Bridge EP processors
  - Node: dual socket; 32 GB memory; 2 GB/core
  - 4,518 nodes, 72,288 cores total – 1.5 PFLOPs peak
  - 144.6 TB total memory
- High-Performance Interconnect
  - Mellanox FDR InfiniBand full fat-tree
  - 13.6 GB/sec bidirectional bw/node
  - <2.5 usec latency (worst case)
- Central File System
  - 2012: 11 PB
  - 2014: 16.4 PB
  - Bandwidth: 90 GB/sec

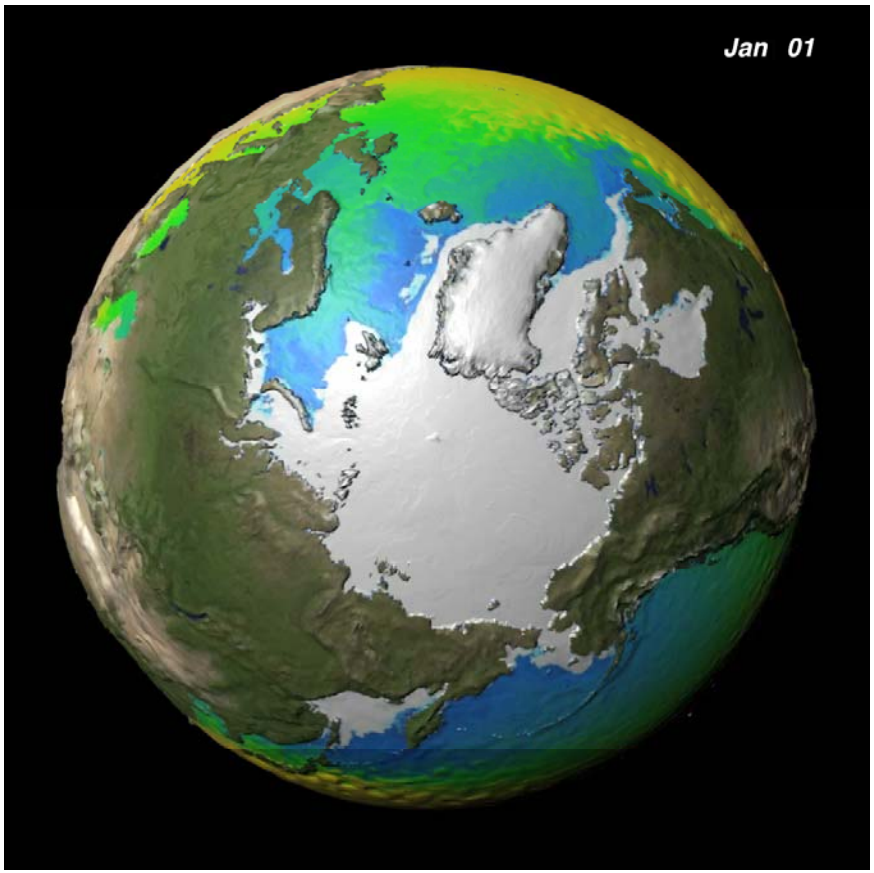
Advanced Scientific Discovery (ASD) project

- Opportunity to use a large piece of newly installed Yellowstone
- “Meso- to planetary-scale processes in a global ultra-high resolution model”
- R. Justin Small, Bailey, Bryan, Danabasogla, Holland, Jochum, Lawrence, Park, Peacock, Tomas, Tribbia, Dennis (NCAR), Saravanan (Texas A&M), Schneider (Hawaii), Kwon (WHOI)
- 47.1 M core hours
  - 25.2 M (2 months) [tuning]
  - 21.9 M (18 months)

High-resolution ASD simulation

- Currently: ~35 years complete
- CAM5-SE (atmosphere model)
  - 28km resolution
  - Scalable spectral-element dynamical core
  - CAM5 physics
  - Fully prognostic aerosol (~50 tracers)
- POP (ocean model)
  - 11km resolution
  - 62 vertical levels
- CICE (sea-ice model)
  - 11km resolution
- CLM (land model)
  - 28km resolution

High-resolution ASD simulation

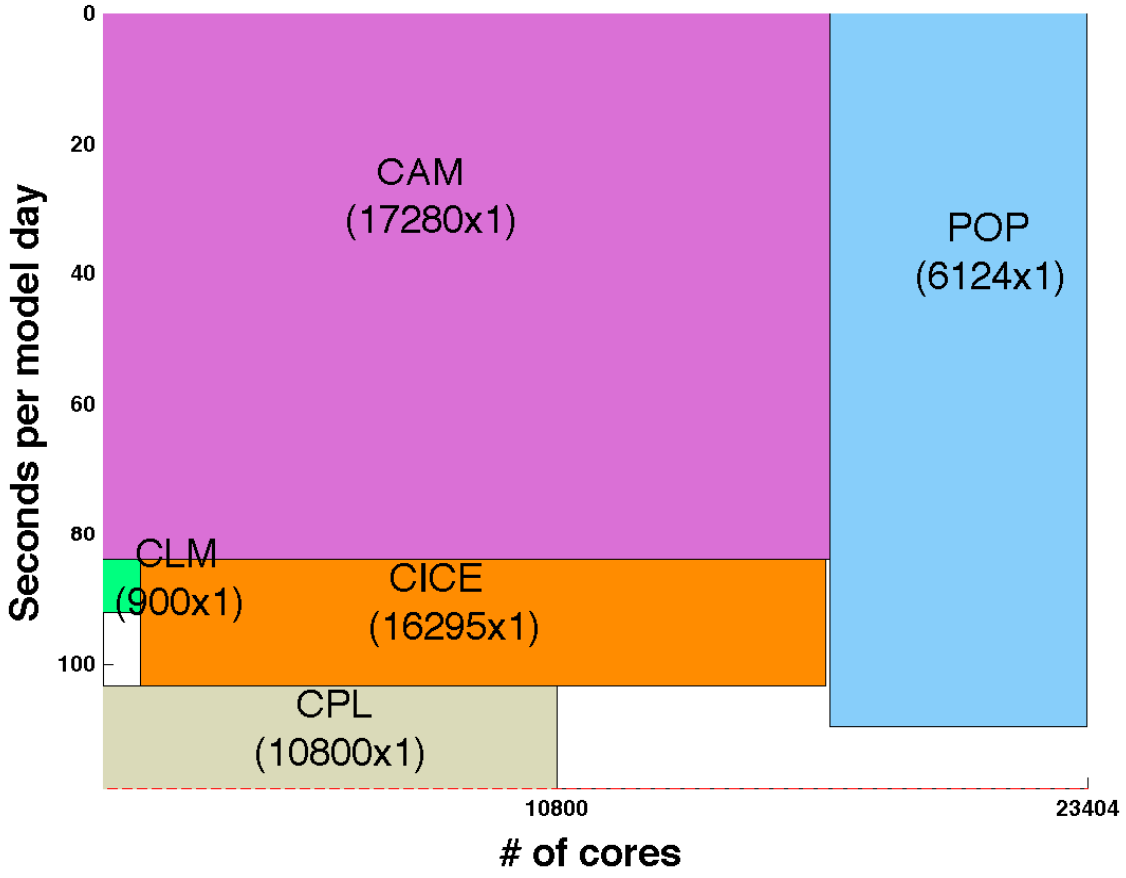


- Improved mean-climate then previous HR runs
- Improved Arctic sea-ice extent
- Ocean temperature bias in the Southern ocean ACC

Computational Aspects of ASD simulation

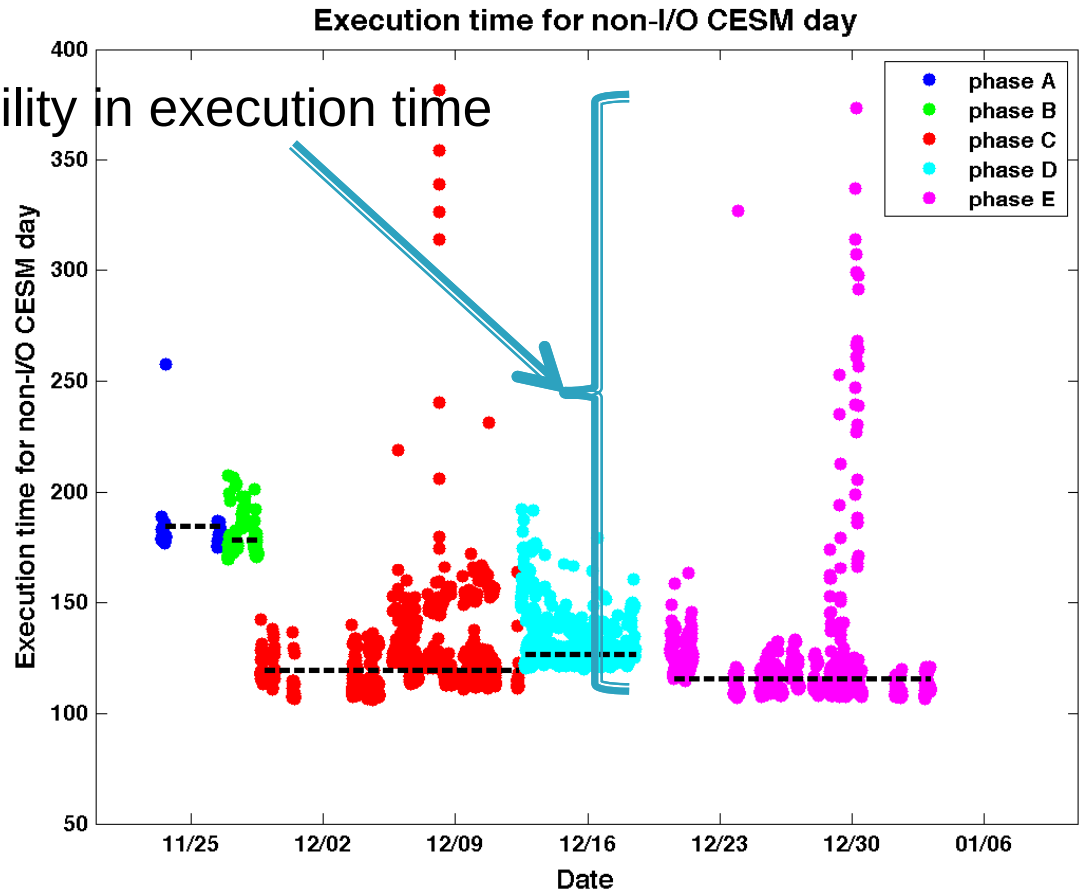
- General statistics:
  - 2.0 simulated years per day (SYPD)
  - 23,404 cores
  - 1 TB of data generated per day
- Component configuration
  - 11km Ocean model (6,124 cores)
  - 11km Sea-ice model (16,295 cores)
  - 28km Atmosphere (17,280 cores)
  - 28km Land (900 cores)
  - Coupler (10,800 cores)

Execution time for ASD simulation

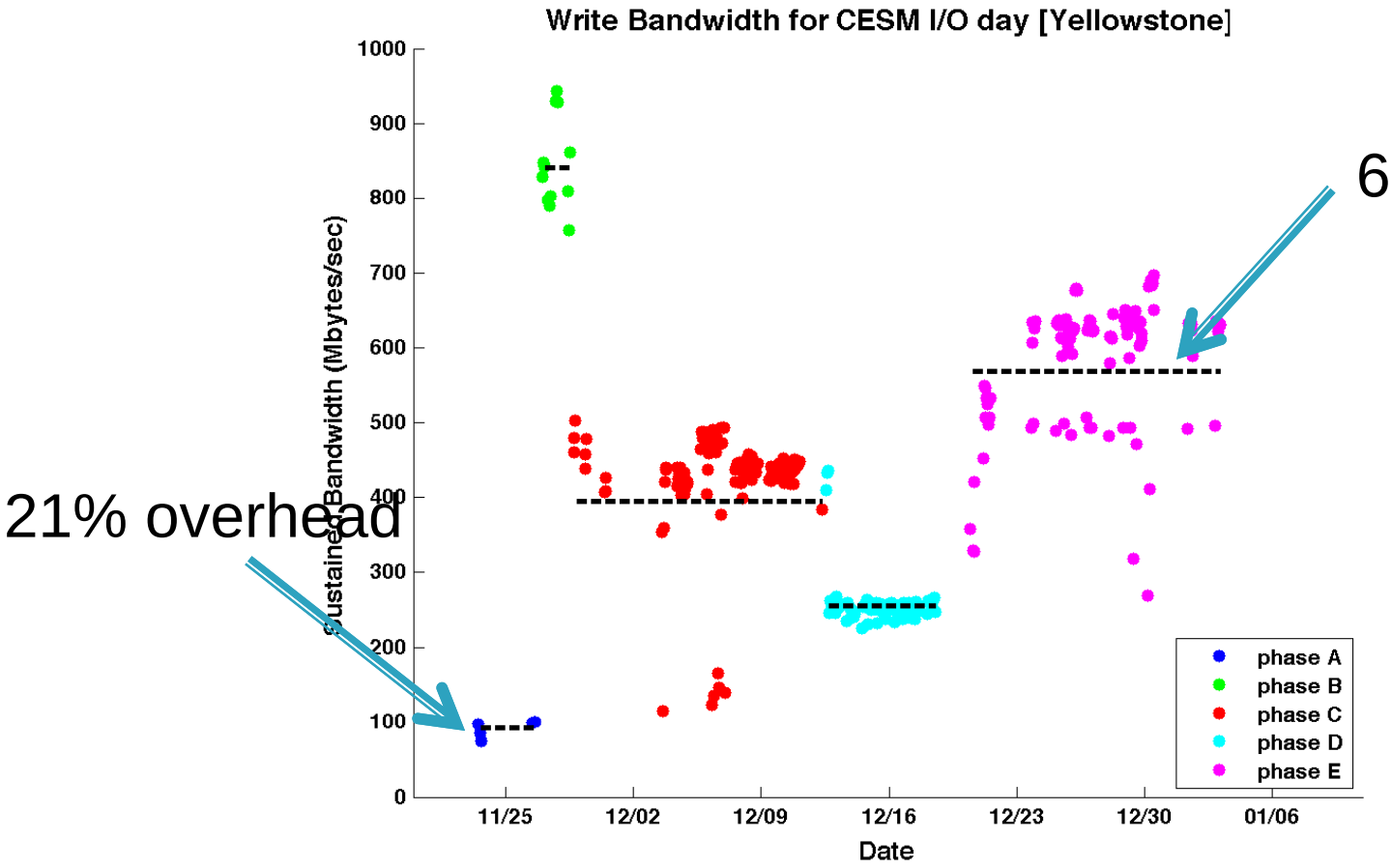


Execution time for ASD on Yellowstone

Large variability in execution time



Write bandwidth for ASD simulation on Yellowstone



Why are we not using more of Yellowstone?

- System stability
  - Significant performance loss (8x) in MPI_Allreduce on occasion (~1%)
  - Collective communication offload failures (33%)
- Queue access for larger then 22k cores
- OS jitter sensitivity [CAM-SE,POP,CICE] (See travel guide)
- Suboptimal CICE partitioning (See travel guide)
- Supoptimal CPL scaling
- Infiniband routing table imbalances
- I/O overhead (6.4%)

Travel Guide

G8-Enabling Climate simulations at Extreme Scale

(See presentation by Marc Snir)

Aachen (Aix-la-Chapelle), Germany
German Research School for Simulation Science (GRS)

- Felix Wolf (f.wolf@fz-juelich.de)
- Scalasca (JSC)
- Research questions:
  - How do you make non-trivial partitioning algorithm development easier? [Monika Luecke]
    - CICE partitioning algorithm: balance communication/computational imbalance
    - Simulated annealing for IFS model
  - Can you identify scaling bottlenecks on large core counts through performance prediction? [Alexandru Calotolu]

Barcelona, Spain
Polytechnic University of Catalonia (UPC)

- Jesus Labata (jesus.labarta@bsc.es)
- paraver, extrae
- Use traces to identify on-node performance problem
- Identified 400 µsec OS jitter affect on Yellowstone
- Identified subtle CPU resource issues -> code restructuring to enable better cache utilization

Conclusions

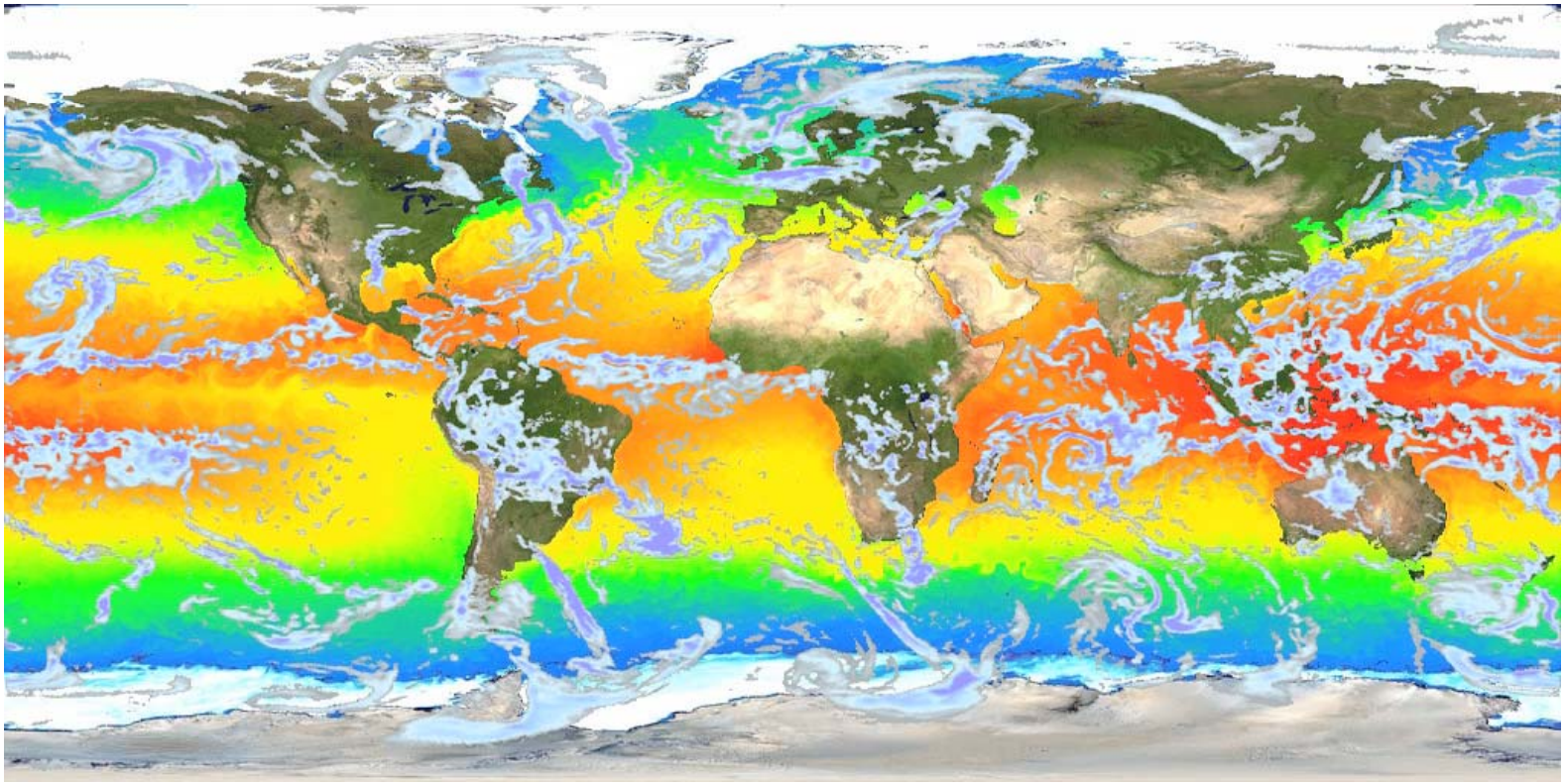
- CESM has High-resolution capability in release code
- “Yellowstone” has enabled very large scale simulation capability at NCAR
- Challenging issues still remain for large scale simulations
- Engaging CS community through G8 initiative

Acknowledgements

- NCAR:
  - D. Bailey
  - F. Bryan
  - T. Craig
  - B. Eaton
  - J. Edwards [IBM]
  - N. Hearn
  - K. Lindsay
  - N. Norton
  - M. Vertenstein
- COLA:
  - J. Kinter
  - C. Stan
- U. Miami
  - B. Kirtman
- U.C. Berkeley
  - W. Collins
  - K. Yelick (NERSC)
- U. Washington
  - C. Bitz
- NICS:
  - M. Fahey
  - P. Kovatch
- ANL:
  - R. Jacob
  - R. Loy
- LANL:
  - E. Hunke
  - P. Jones
  - M. Maltrud
- LLNL
  - D. Bader
  - D. Ivanova
  - J. McClean (Scripps)
  - A. Mirin
- ORNL:
  - P. Worley

and many more…

- Grant Support:
  - DOE
    - DE-FC03-97ER62402 [SciDAC]
    - DE-PS02-07ER07-06 [SciDAC]
  - NSF
    - Cooperative Grant NSF01
    - OCI-0749206 [PetaApps]
    - CNS-0421498
    - CNS-0420873
    - CNS-0420985
- Computer Allocations:
  - TeraGrid TRAC @ NICS
  - DOE INCITE @ NERSC
  - LLNL Grand Challenge
- Thanks for Assistance:
  - Cray, NICS, and NERSC



Questions?

dennis@ucar.edu