SciDAC Update

and

High Resolution Runs (at Oak Ridge and at Argonne)

Pat Worley

Oak Ridge National Laboratory

(*including slides provided by J. Drake, P. Jones, M. Taylor, M. Vertenstein, and others*)

Software Engineering Working Group Meeting

CCSM Workshop

Great Divide Lodge

Breckenridge, CO

July 1, 2010

Acknowledgements

- Research sponsored by the Atmospheric and Climate Research Division within the Office of Biological and Environmental Research and by the Office of Advanced Scientific Computing Research, both in the Office of Science, U.S. Department of Energy under Contract No. DE-AC05-00OR22725 with UT-Battelle, LLC.
- This research used resources (Cray XT5) of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725, and resources (IBM BG/P) of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-06CH11357.

SciDAC Update

and

High Resolution Runs (at Oak Ridge and at Argonne)

Low and High

Yes, we still exist! (12 more months)

Ran out of time

Pat Worley
Oak Ridge National Laboratory
(*including slides provided by J. Drake, P. Jones, M. Vertenstein, and others*)

Software Engineering Working Group Meeting
CCSM Workshop
Great Divide Lodge
Breckenridge, CO
July 1, 2010

Who Are Those Guys?

A Scalable and Extensible Earth System Model for Climate Change Science

SEESM – aka SciDAC CCSM

and

Performance Engineering for the Community Climate System Model

PECCSM – aka SciDAC CCSM SAP

Who Are Those Guys?

A Scalable and Extensible Earth System Model
for Climate Change Science
SEESM – aka SciDAC CCSM

We're Baaaack!

Reports of our demise were greatly exaggerated.

and

Son of PENG

We're not dead yet!

Performance Engineering for the
Community Climate System Model
PECCSM – aka SciDAC CCSM SAP

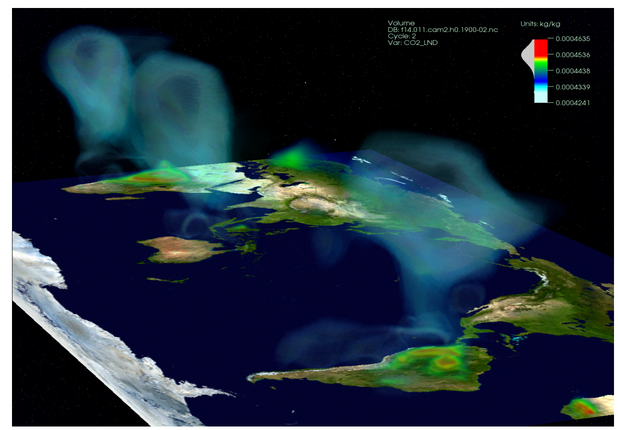
The SciDAC CCSM Consortium:

- ***Former PI:*** John B. Drake (ret), Oak Ridge National Laboratory
- ***Acting and co-PI:*** Phil Jones, Los Alamos National Laboratory
- **Argonne National Laboratory (ANL)** Robert Jacob, Ray Loy, Jay Larson
- **Brookhaven National Laboratory (BNL)** Robert McGraw
- **Lawrence Berkeley National Laboratory (LBNL)** William Collins, Michael Wehner, Inez Fung
- **Lawrence Livermore National Laboratory (LLNL)** Phillip Cameron-Smith, Arthur Mirin
- **Los Alamos National Laboratory (LANL)** Scott Elliot, Philip Jones, William Lipscomb, Mat Maltrud
- **National Center for Atmospheric Research (NCAR)** Peter Gent /Jim Hurrell, Andrew Conley, Tony Craig, Jean-Francois Lamarque, Mariana Vertenstein, **Trey White**, Warren Washington
- **Oak Ridge National Laboratory (ORNL)** Marcia Branstetter, **David Bader**, David Erickson, Kate Evans, James Hack, Forrest Hoffman, Peter Thornton, Patrick Worley
- **Pacific Northwest National Laboratory (PNNL)** Steven Ghan, Xiaohong Liu, Richard Easter
- **Sandia National Laboratories (SNL)** Mark Taylor

- **Scientific Application Partnerships**
  - **Brookhaven National Laboratory Robert McGraw**
  - **Oak Ridge National Laboratory** Patrick Worley
  - **Argonne National Laboratory Kotamarthi Rao**

- **Centers for Enabling Technology Collaborations**

  ESG - Dean Williams **PERI –** Pat Worley
  VIZ – Wes Bethel **TOPS –** David Keyes
  ITAPS – Lori Diachen



Computational Climate End Station for Climate Change Science
- Warren Washington (PI)

Objectives

The primary objective of the two projects is to develop, test, and exploit a first generation of Earth system models based upon the CCSM that will run efficiently on thousands of processors and include significant model enhancements.

Objectives

The primary objective of the two projects is to develop, test, and exploit a first generation of Earth system models based upon the CCSM that will run efficiently on thousands of processors and include significant model enhancements.

Find means for John Drake to participate during his retirement.

Los Alamos NATIONAL LABORATORY EST. 1943 · OAK RIDGE National Laboratory · BERKELEY LAB · Argonne NATIONAL LABORATORY · Pacific Northwest NATIONAL LABORATORY · NCAR

SciDAC Scientific Discovery through Advanced Computing

Tasks and Organization

U.S. DEPARTMENT OF ENERGY Office of Science

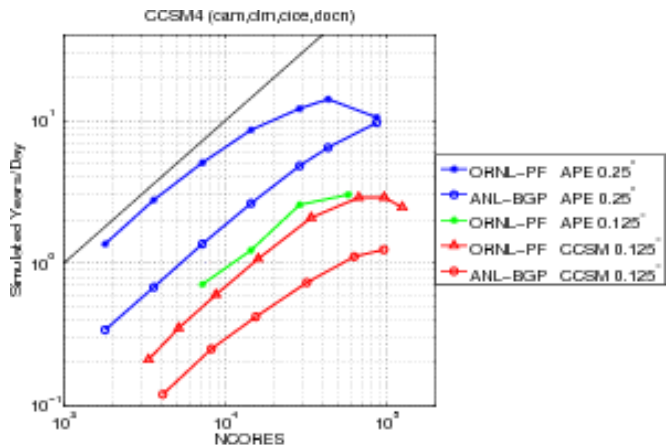
- **Scalable and Extensible Earth System Model (SEESM)**
  - **terrestrial carbon cycle and dynamic vegetation**
  - **atmospheric chemistry and aerosol dynamics**
  - **ocean ecosystems and biogeochemical coupling**
  - **feedbacks between atmospheric composition and biogenic emissions**
- **Model Integration and Evaluation**
  - **Integration and unit testing**
  - **New ice sheet and ocean models**
  - **New atmospheric dynamics: finite volume (cubed sphere), continuous Galerkin, others**
  - **Frameworks for model evaluation**
- **Computational Performance (SEESM and PECCSM)**
  - **scalability to thousands of processors, load balance, (fault recovery)**
  - **performance portability and software engineering**

Los Alamos NATIONAL LABORATORY EST. 1943 · Oak Ridge National Laboratory · Berkeley Lab · Argonne National Laboratory · Pacific Northwest National Laboratory · NCAR

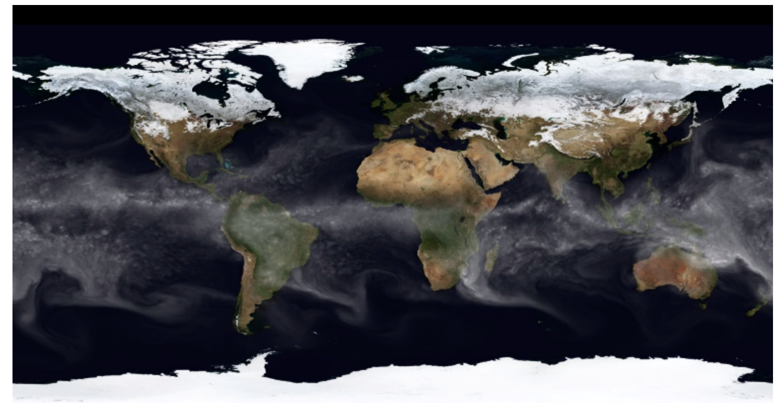
CCSM: Scalable Atmosphere

SciDAC Scientific Discovery through Advanced Computing

U.S. DEPARTMENT OF ENERGY Office of Science



CCSM-HOMME 1/8 degree simulations



- SciDAC led effort to develop CCSM-HOMME
- HOMME: Experimental unstructured grid (finite element based) dynamical core included in CCSM4 release:
  - Excellent conservation (w/o fixers) of tracer mass, moist energy, PV. High-order dynamics (similar to CAM-Eul). Advection: Vertically Lagrangian, quasi-monotone. Supports arbitrary unstructured quadrilateral meshes.
  - Excellent scalability: demonstrated to 86,000 cores at ¼ and 1/8° resolutions in CCSM time-slice configurations. (active atmosphere, land, cice, prescribed SST and ice extent).

- Two configurations included in CCSM4 based on cubed-sphere.
- Moderate resolution for evaluation
  - Atmosphere: 1° cubed-sphere, Land: 2° lat-lon, data ocean and CICE: gx1v6
- High resolution for testing CCSM4 scalability out to O(400,000) cores
  - Atmosphere: 1/8° cubed-sphere, Land: 2° lat-lon, data ocean and CICE: gx1v6

Los Alamos NATIONAL LABORATORY EST.1943 · OAK RIDGE National Laboratory · BERKELEY LAB · Argonne NATIONAL LABORATORY · Pacific Northwest NATIONAL LABORATORY · NCAR

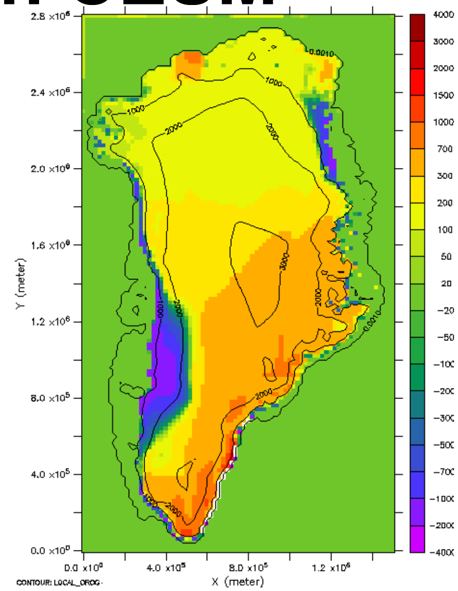
Simulating the Greenland Ice Sheet in CESM

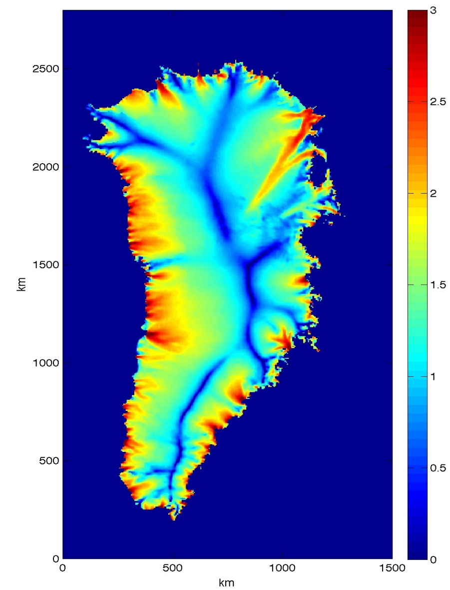
Objective

- Couple a dynamic ice sheet model to the Community Earth System Model, in order to predict ice sheet retreat and sea level rise

Approach

- Add the Glimmer ice sheet model to CCSM framework
- Compute the surface mass balance in CLM (subgrid elevation classes) and downscale to Glimmer grid
- Add model improvements as available (higher-order dynamics, ice-ocean coupling)
- Run IPCC AR5 simulations



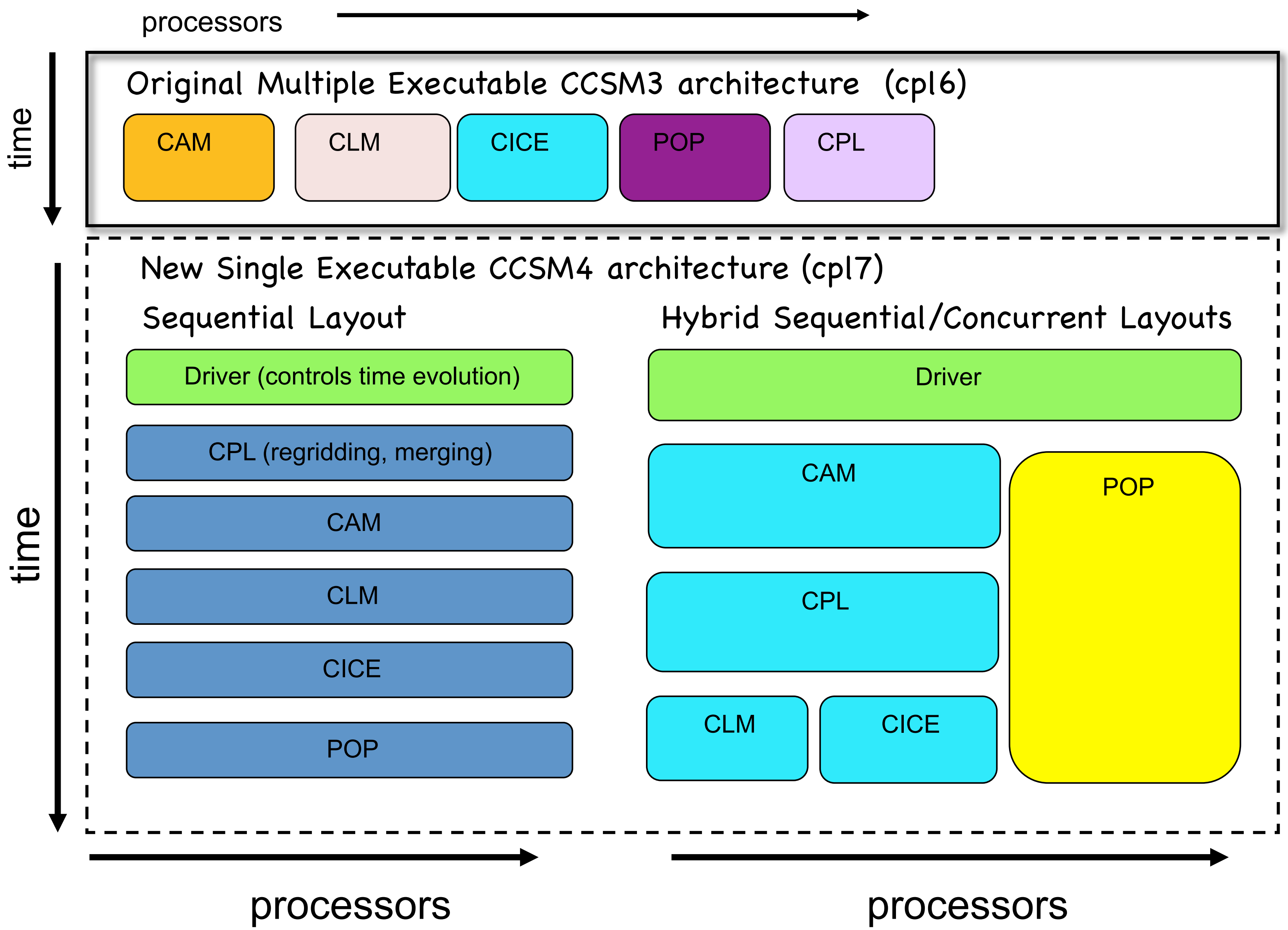


Lipscomb, W., R. Bindschadler, E. Bueler, D. Holland, J. Johnson, S. Price, 2009: A community ice sheet model for sea level prediction, *Eos Trans. AGU*, 90, 23.

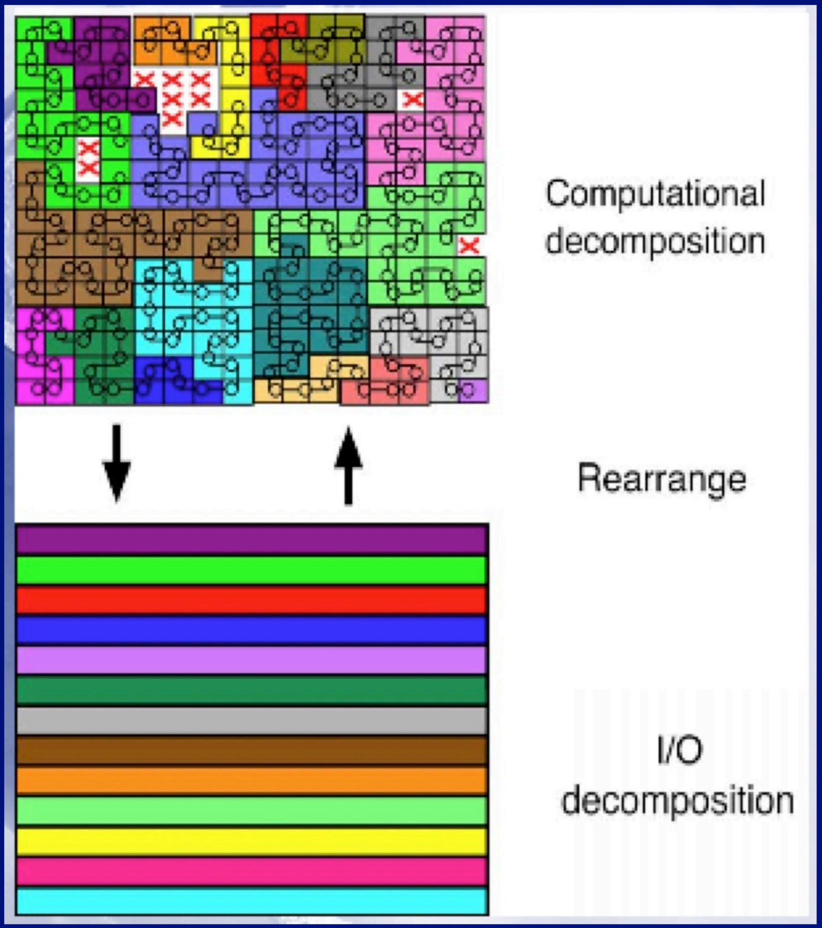
Impact

- Capture runtime feedbacks between ice sheets and other climate components
- Strengthen connections between glaciologists and climate modelers (Land Ice working group)
- Deliver improved sea-level projections for AR5

New CCSM4 Architecture



New Parallel I/O library (PIO)



- Interface between the model and the I/O library. Supports
  - Binary
  - NetCDF3 (serial netcdf)
  - Parallel NetCDF (pnetcdf) (MPI/IO)
  - NetCDF4
- User has enormous flexibility to choose what works best for their needs
  - Can read one format and write another
- Rearranges data from model decomp to I/O friendly decomp (rearranger is framework independent) – model tasks and I/O tasks can be independent

PIO in CCSM

- PIO implemented in CAM, CICE, POP, and CPL
- Usage is critical for high resolution, high processor count simulations
  - Serial I/O is one of the largest sources of global memory in CCSM - *will eventually always run out of memory*
  - Serial I/O results in serious performance penalty at higher processor counts
- Performance benefit noticed even with serial netcdf (model output decomposed on output I/O tasks)

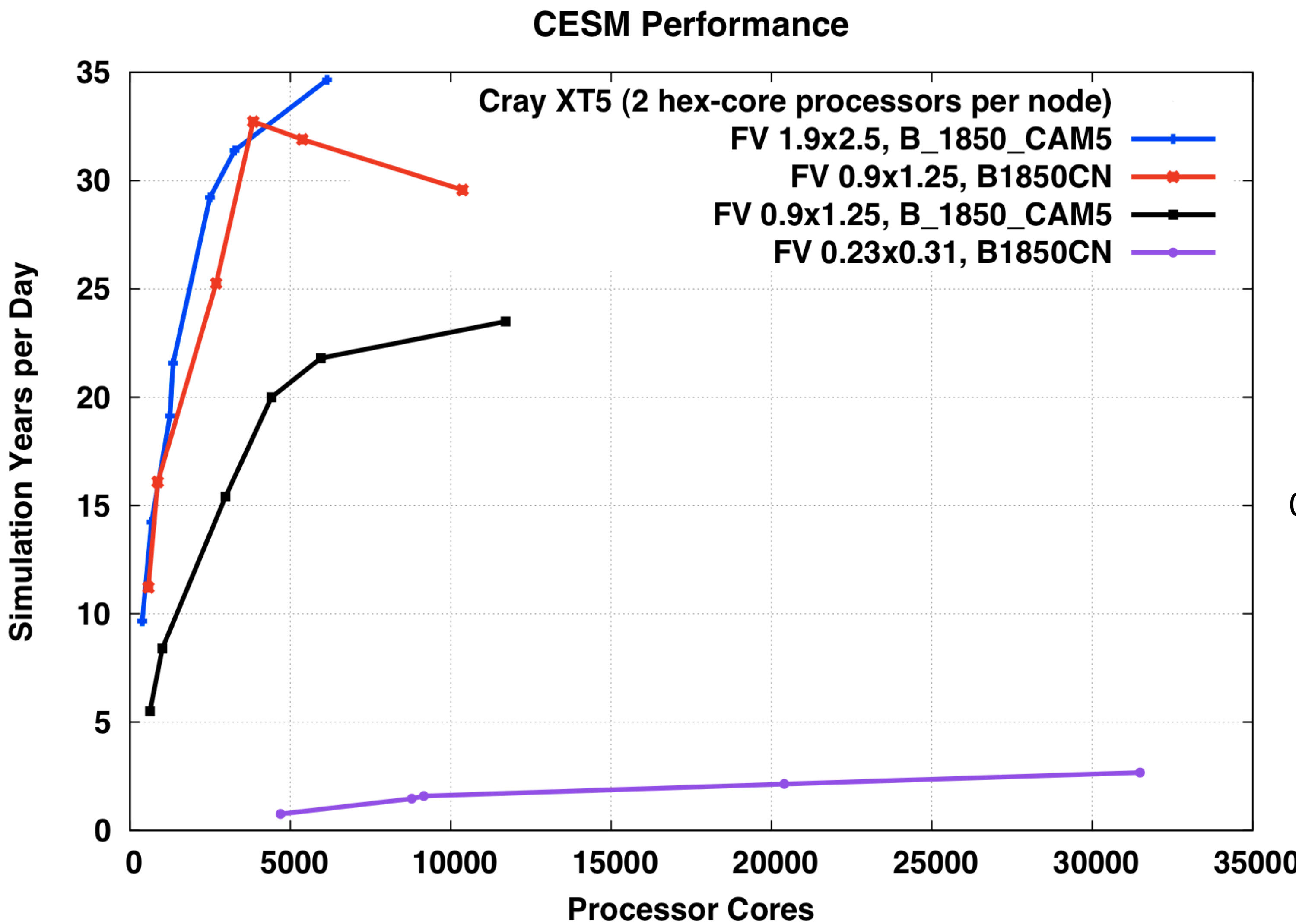
Component Parallelism and Scalability

- MPI/OpenMP
  - CAM, CLM, CICE, POP have MPI/OpenMP hybrid capability
  - Coupler only has MPI capability
  - Data models only have MPI capability
- Parallel I/O (use of PIO library)
  - CAM, CICE, POP, CPL, Data models all have PIO capability
- Robust Communication Algorithms
  - CAM, CICE, CLM, POP, CPL/MCT, PIO all have flow control options in communication algorithms
  - Options are critical for large process count runs

Recent SE Activities

- **Helping** with CCSM4 and CESM1 release
  - Getting OpenMP working in POP
  - Adding flow control communication options (everywhere)
  - Testing, testing, testing; Debugging, debugging, debugging; Evaluating, evaluating, evaluating, especially
    - on Cray XT and IBM BG/P,
    - at high resolutions and on large processor core counts,
    - and for PIO.
- **Helping** with next generation components
  - Spectral element (most recently OpenMP in HOMME and minimizing memory in dynamics/physics interface logic)

CESM Performance



Resolutions

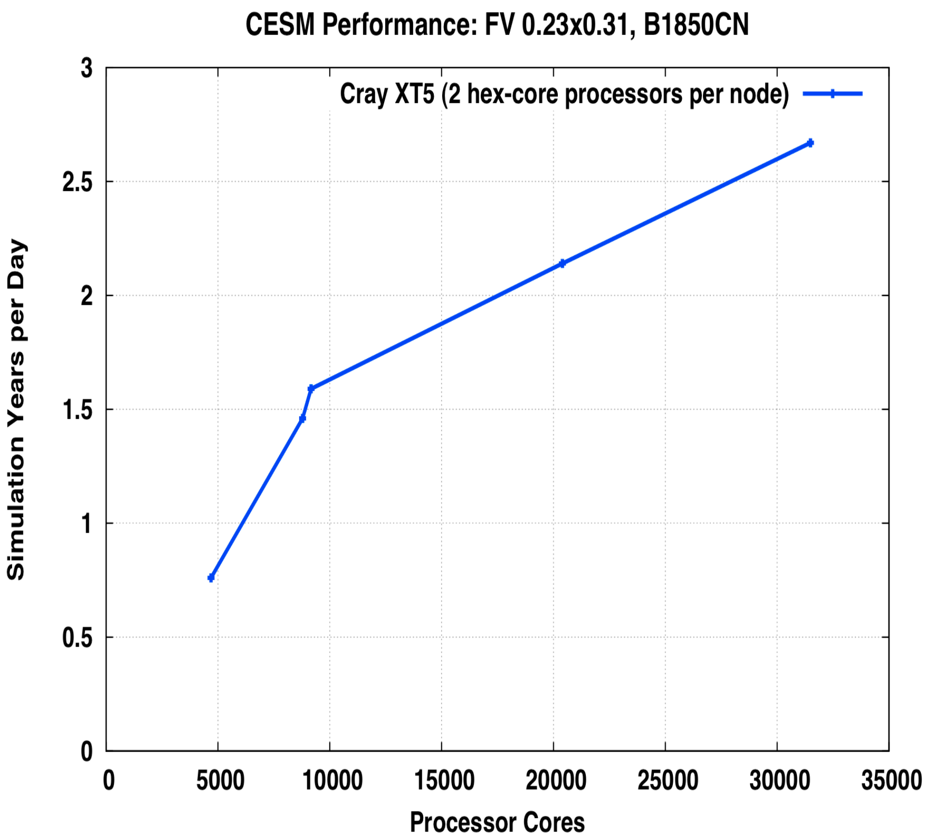
1.9x2.5 ATM/LND
gx1v6 OCN/ICE

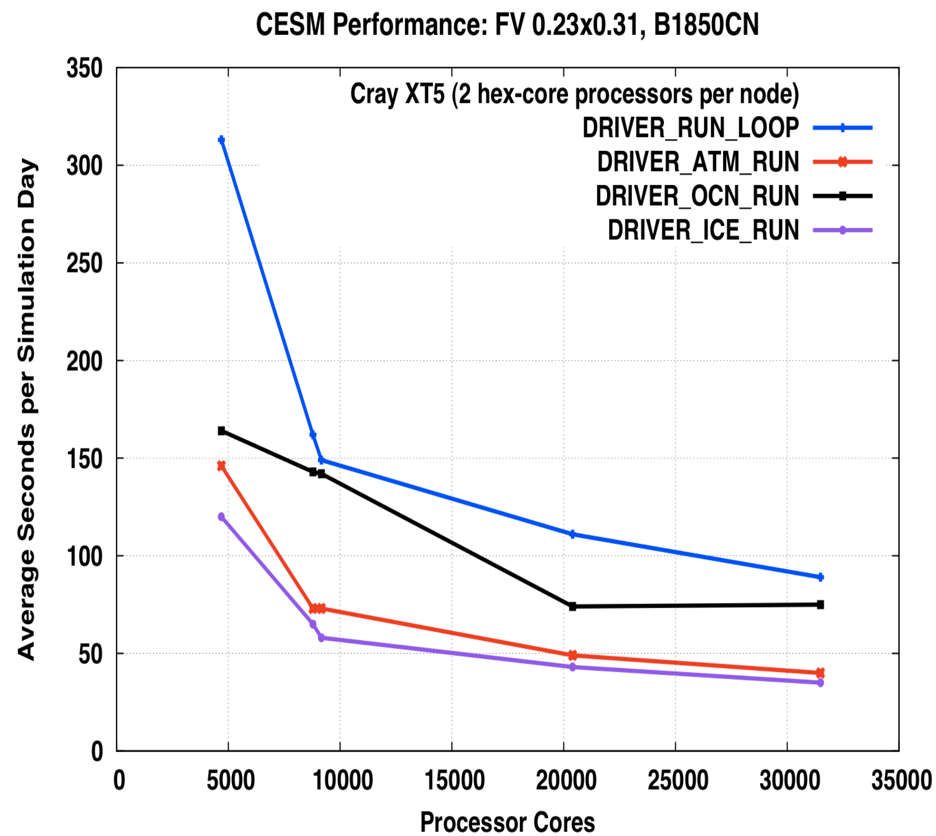
0.9x1.25 ATM/LND
gx1v6 OCN/ICE

0.23x0.31 ATM/LND
tx0.1v2 OCN/ICE

CESM Performance

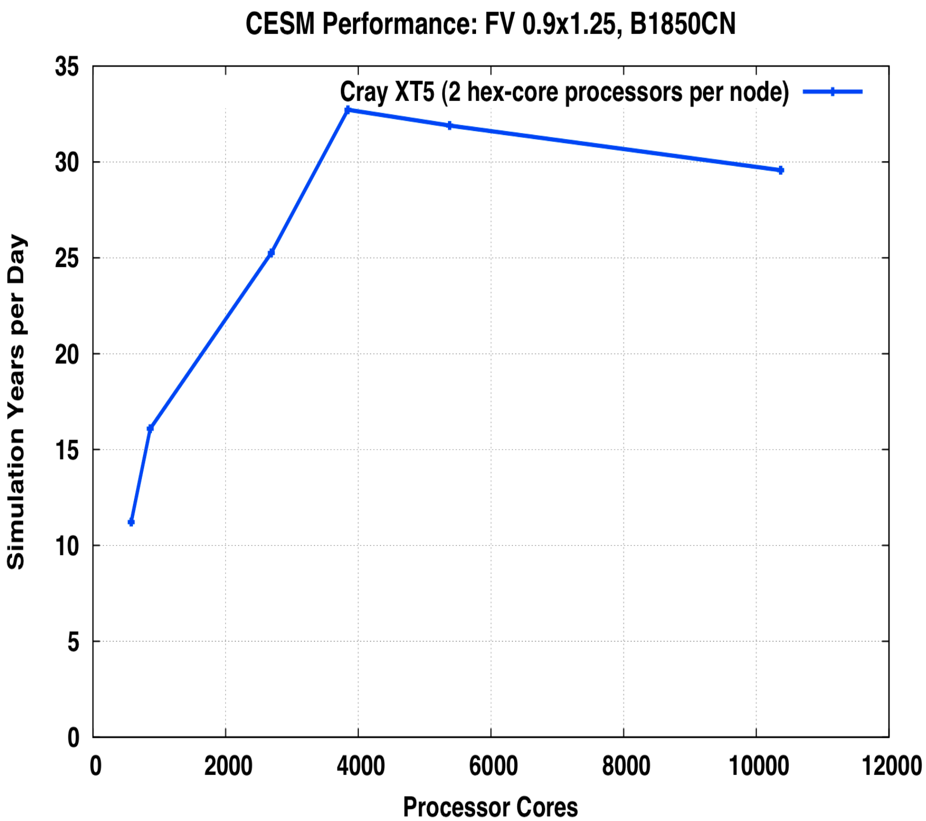
0.23x0.31 ATM/LND
tx0.1v2 OCN/ICE
B1850CN (CAM4)

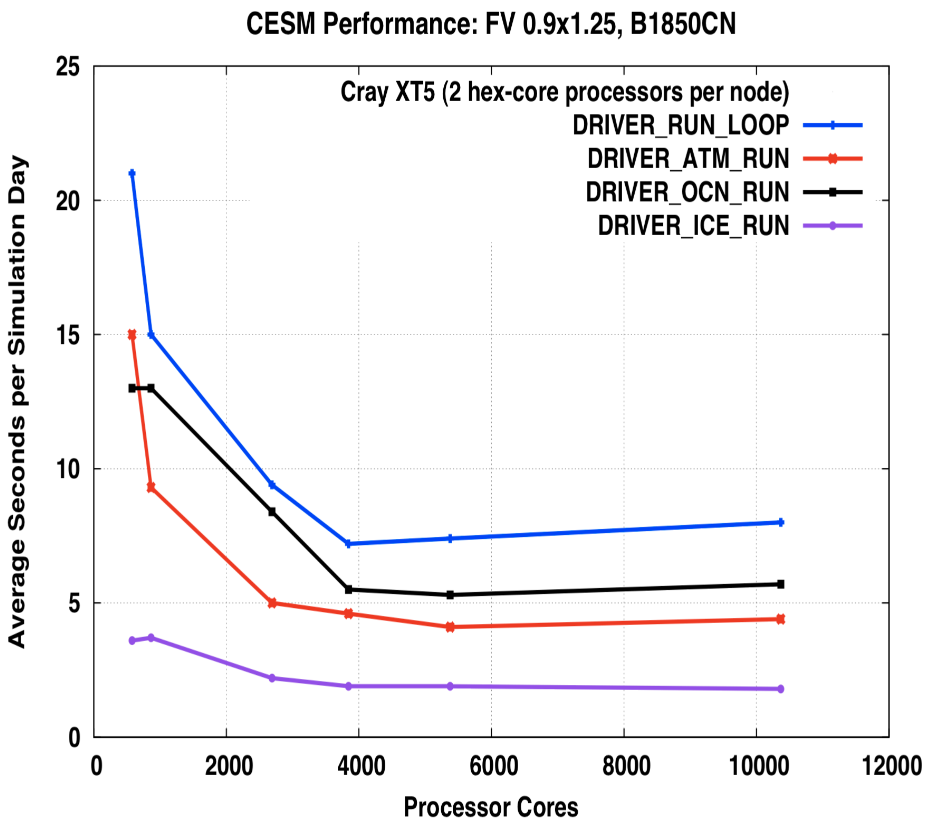




CESM Performance

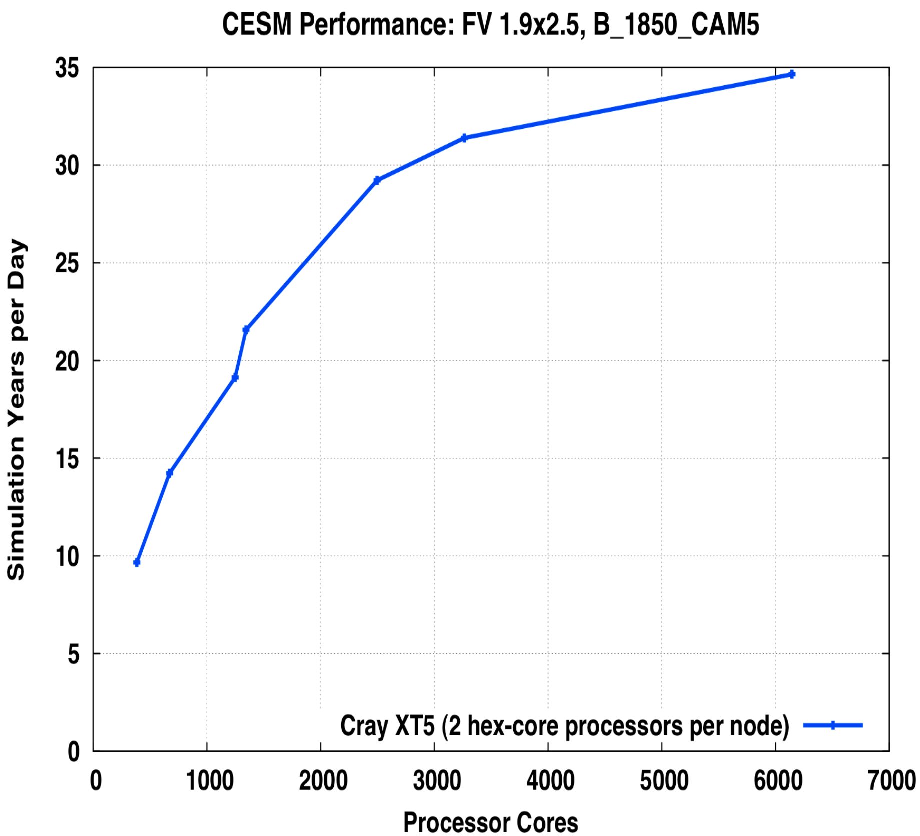
0.9x1.25 ATM/LND
gx1v6 OCN/ICE
B1850CN (CAM4)

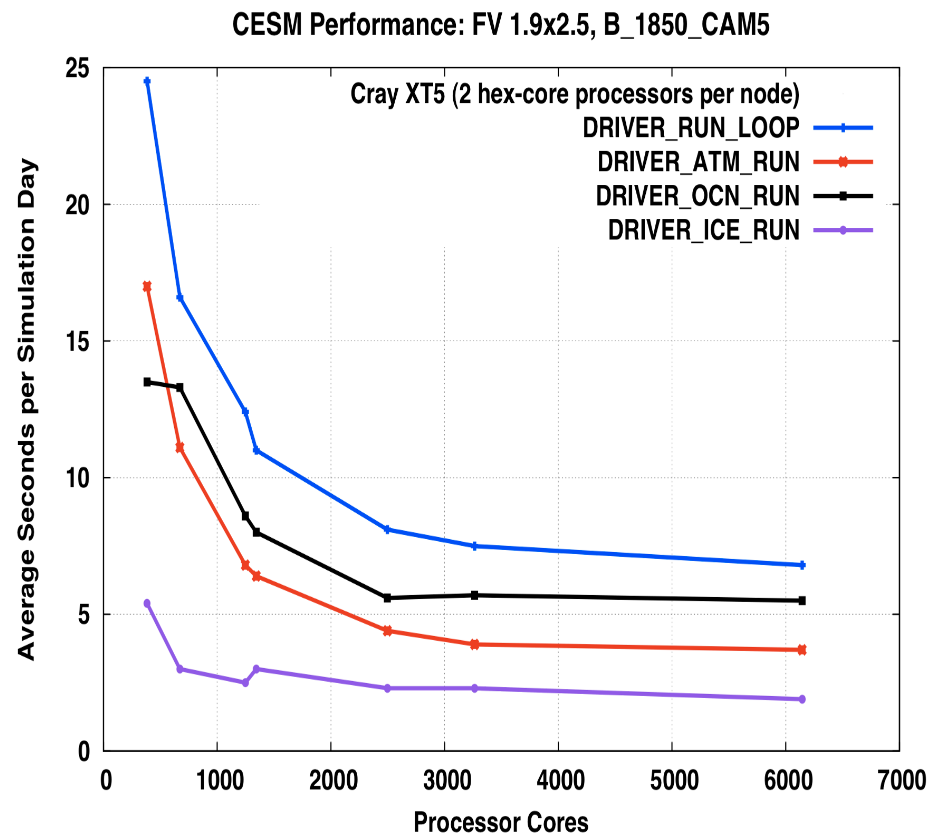




CESM Performance

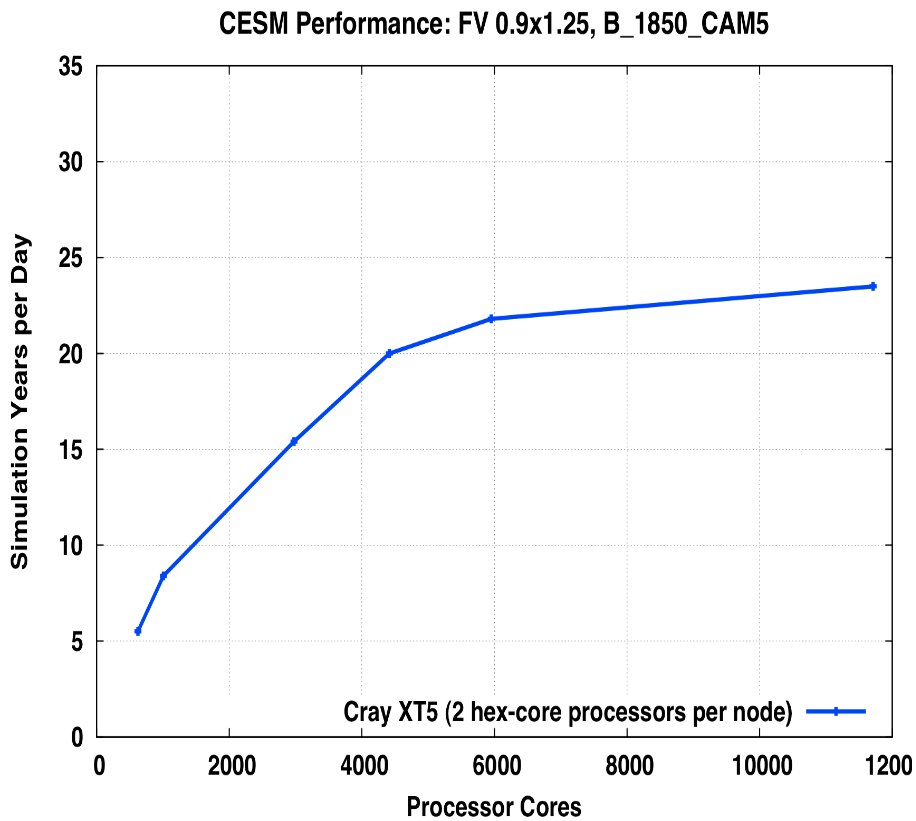
1.9x2.5 ATM/LND
gx1v6 OCN/ICE
B_1850_CAM5 (CAM5)

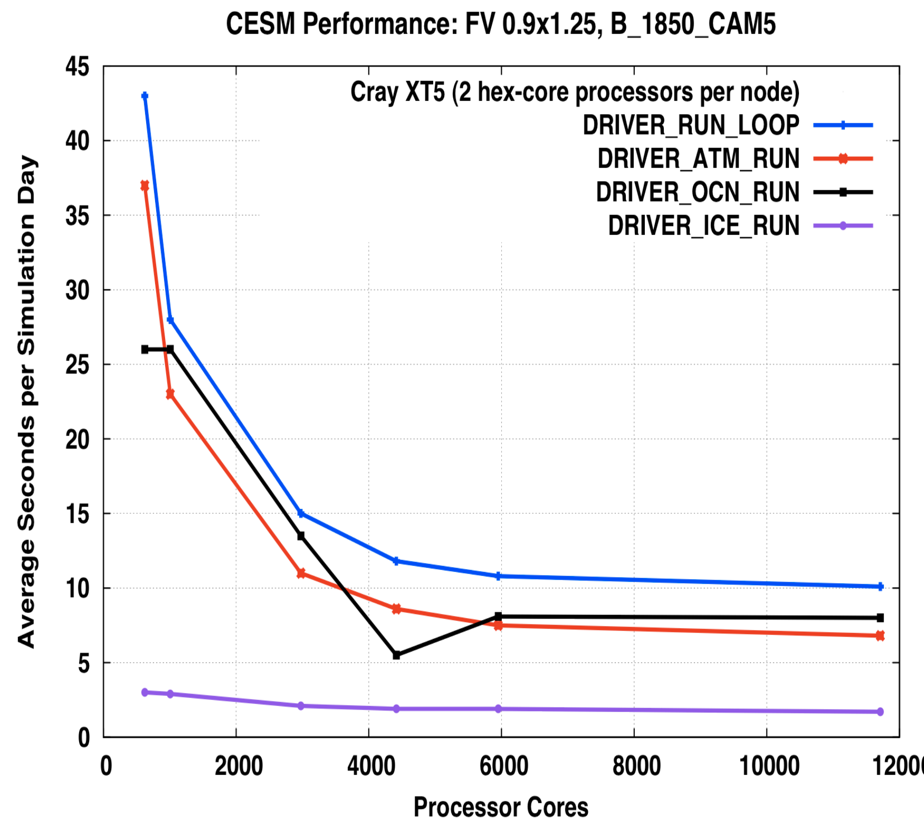




CESM Performance

0.9x1.25 ATM/LND
gx1v6 OCN/ICE
B_1850_CAM5 (CAM5)





CESM Performance

Comments:

- CAM5 (30 levels, MAM3 chemistry, …) not that much more expensive than CAM4 (26 levels, no chemistry) at scale
- At both low and high resolutions, ATM or ATM + CICE are the bottlenecks
- OpenMP really helps in OCN (which is why OCN is not the performance limiter at the moment)
  - Without OpenMP in OCN, best performance for B1850CN at 0.9x1.25 ATM/LND and gx1v6 OCN/ICE drops from 33 SYPD to 18 SYPD.
- Flow control in communication algorithms is critical on the Cray XT systems
  - Without flow control enabled, can't run 30,000 processor 0.23 ATM/LND, tx0.1v2 OCN/ICE B1850CN experiment on jaguarpf.

CESM Performance

- Just Getting Started: resolutions and compsets
  - 0.47x0.63 FV with 0.1 degree OCN/ICE, both CAM4 and CAM5
  - Trop_mozart, WACCM, …
- Just Getting Started: platforms
  - IBM BG/P (Intrepid), IBM Power6 (BlueFire), …
- Just Getting Started: ATM dycores
  - Spectral Eulerian (EUL), Spectral Element (HOMME), …
- Just Getting Started: performance optimization
  - Component load balancing
  - High process count decompositions for OCN and ICE
  - Detailed component performance analysis and further optimization
  - I/O