High Resolution CCSM, Interactive Ensembles and other Random Stuff

John M. Dennis: dennis@ucar.edu

Nathan Hearn: nhearn@ucar.edu

Mariana Vertenstein: mvertens@ucar.edu

June 17, 2009

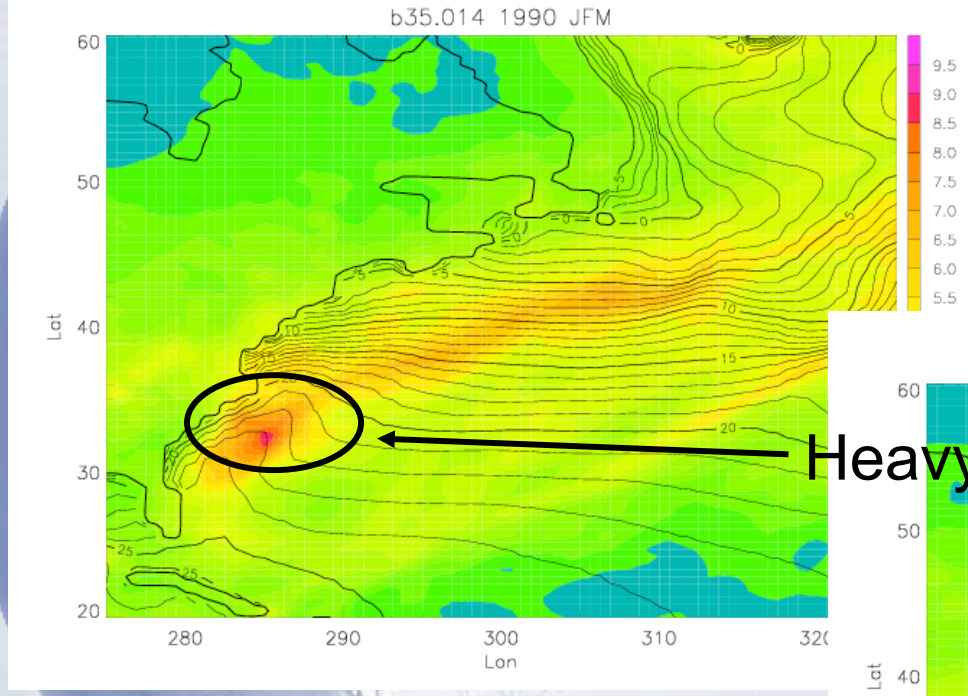
NCAR

Motivation for High-Resolution

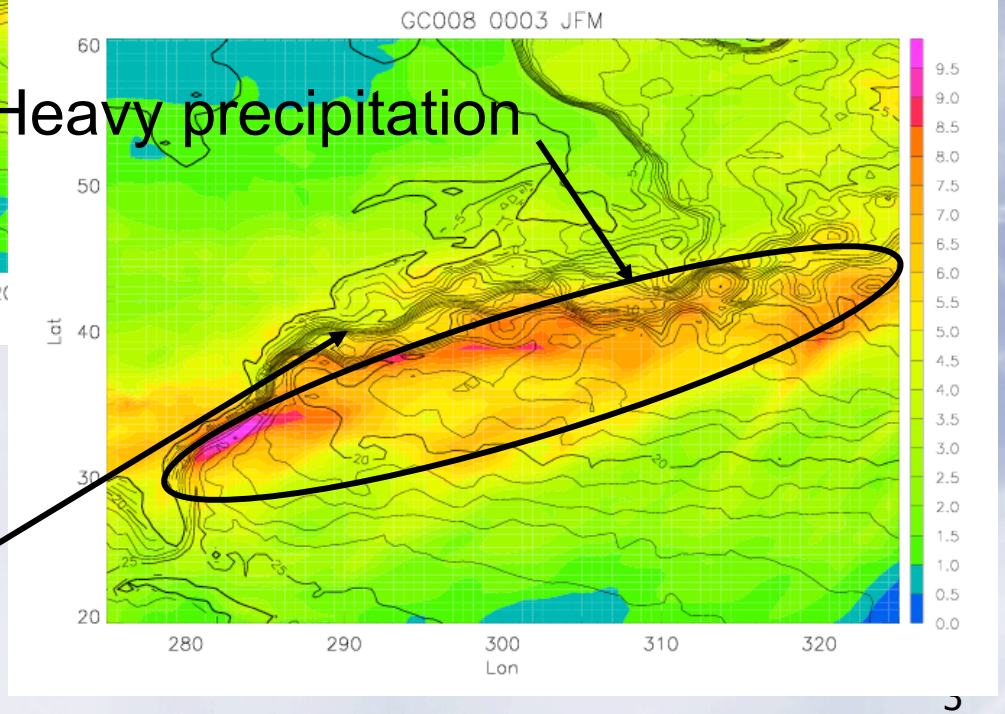
- Stress testing software results in improvements at production resolution
- Reveals some interesting Mesoscale processes
- Makes for pretty pictures! ☺

NCAR

North Atlantic Winter storm track



0.5° atm + 0.1° ocn

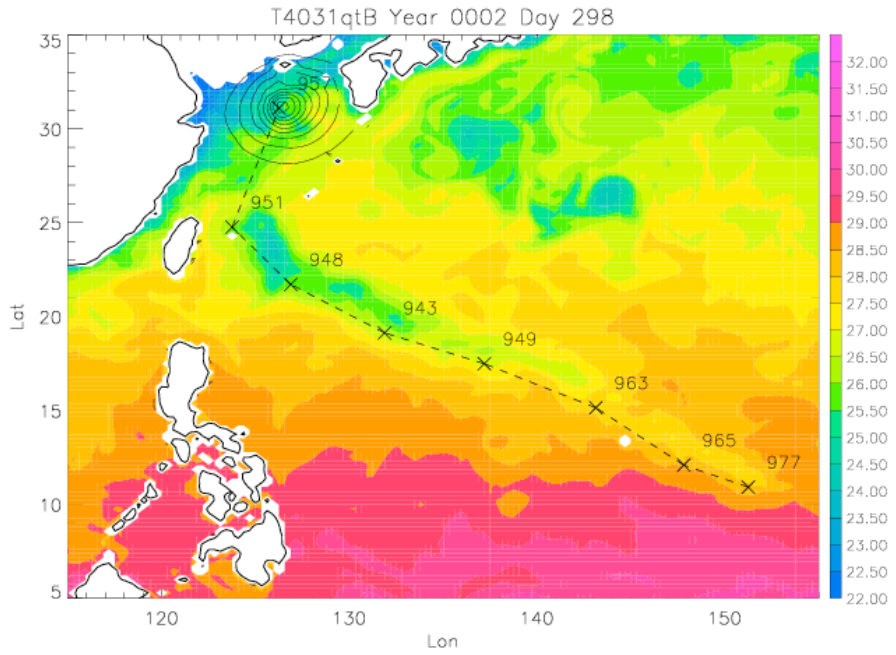


Heavy precipitation

0.5° atm + 1° ocn

Strong SST gradient

Typhoon cold wake (0.25°)



NCAR

Funding Sources

- Department of Energy: CCPP Program Grants
  - DE-FC03-97ER62402 [SciDAC]
  - DE-PS02-07ER07-06 [SciDAC]
- National Science Foundation:
  - OCI-0749206 [PetaApps]
  - OCE-0825754
  - Cooperative Grant NSF01

Some ultra-high-resolution CCSM runs

- 0.25° ATM,LND + 0.1° OCN,ICE
  - 18 years [LLNL]
- 0.50° ATM,LND + 0.1° OCN,ICE
  - 2 years [LLNL]
  - 10+ years [NERSC]
  - 42+ years [NICS]

PetaApps: Interactive Ensembles

- PetaApps project members:
  - J. Kinter, C. Stan (COLA)
  - B. Kirtman (U of Miami)
  - C. Bitz (U of Washington)
  - W. Collins, K. Yelick (U of California)
  - F. Bryan, J. Dennis, R. Loft, M. Vertenstein (NCAR)

NCAR

PetaApps: Interactive Ensembles (con't)

- Interactive ensembles
  - Multiple instances of component models
  - Explore the role of weather noise in climate
    - Test hypothesis that noise is "reddened" and influences low-frequency components of climate system
- 35M CPU hours TeraGrid proposal
  - 6000 core job: 7 months non-stop

NCAR

Progress on PetaApps

- Interactive Ensemble (IE):
  - N. Hearn
  - IE code to CCSM4 code base complete
  - Control run complete
  - 65 yr IE run @ 2 degree complete
  - Validation against CCSM3 underway
- High-Resolution control run on Kraken underway

High Resolution on Kraken

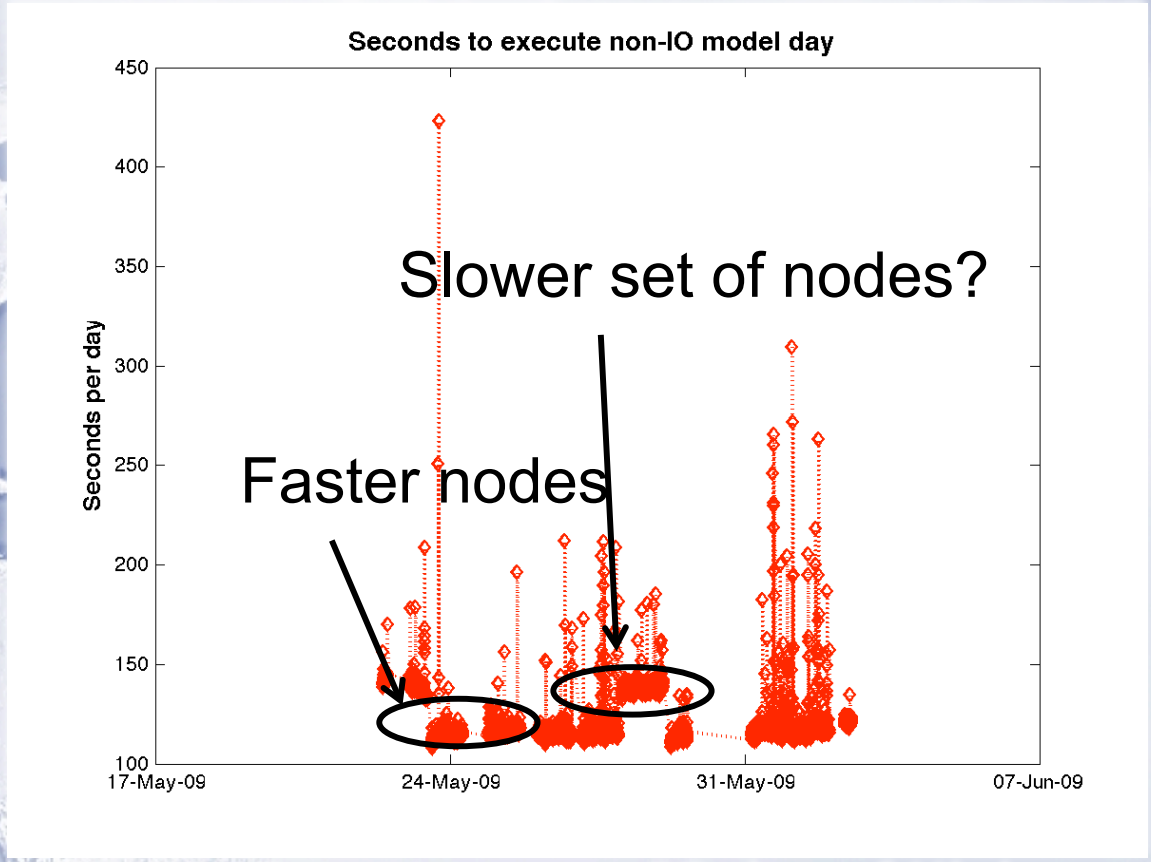
- 42+ years complete
- 5844 processors
- 20-24 hour jobs (~1.6 SYPD)
- 2 GB per core
- File I/O
  - Written to disk: 95MB to 25GB per file (180 files, 1046GB)
  - Archived: (60 files, 414 GB) per job
  - Exported: (60 files, 414 GB) per job
    - gridFTP: ~90MB/sec [78 minutes]
    - scp: ~5MB/sec [23 hours]

Not possible without GridFTP

NCAR

Performance variability on Kraken

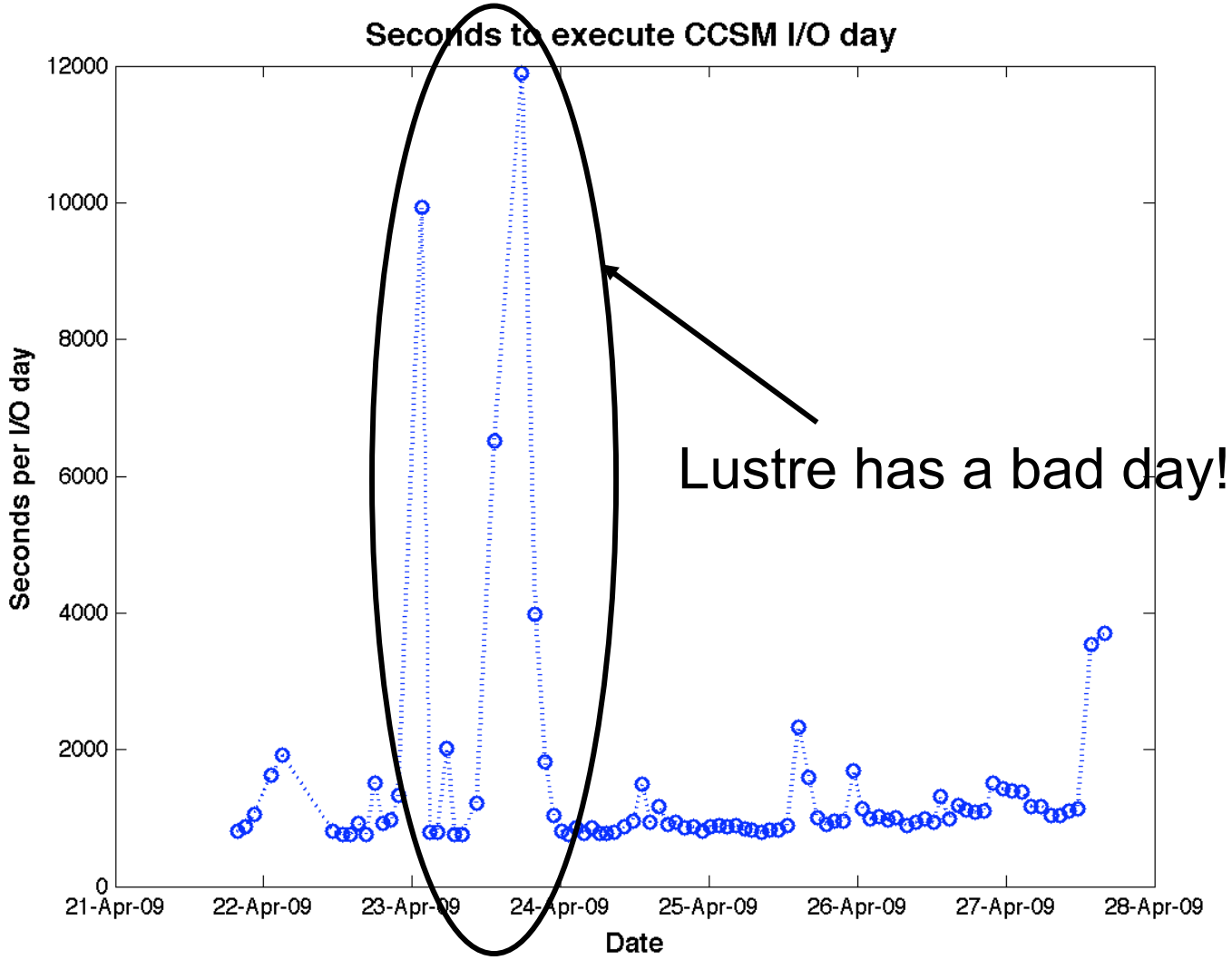
- Large variability in performance of CCSM during production run
- Running virtually nonstop for 1+ week blocks
- Look at time for {IO,non IO} day
- Solutions:
  - Short term: {little-endian, Rosinski mods}
  - Long term: PIO



Time to Execute a non-IO day of High-resolution CCSM on Kraken

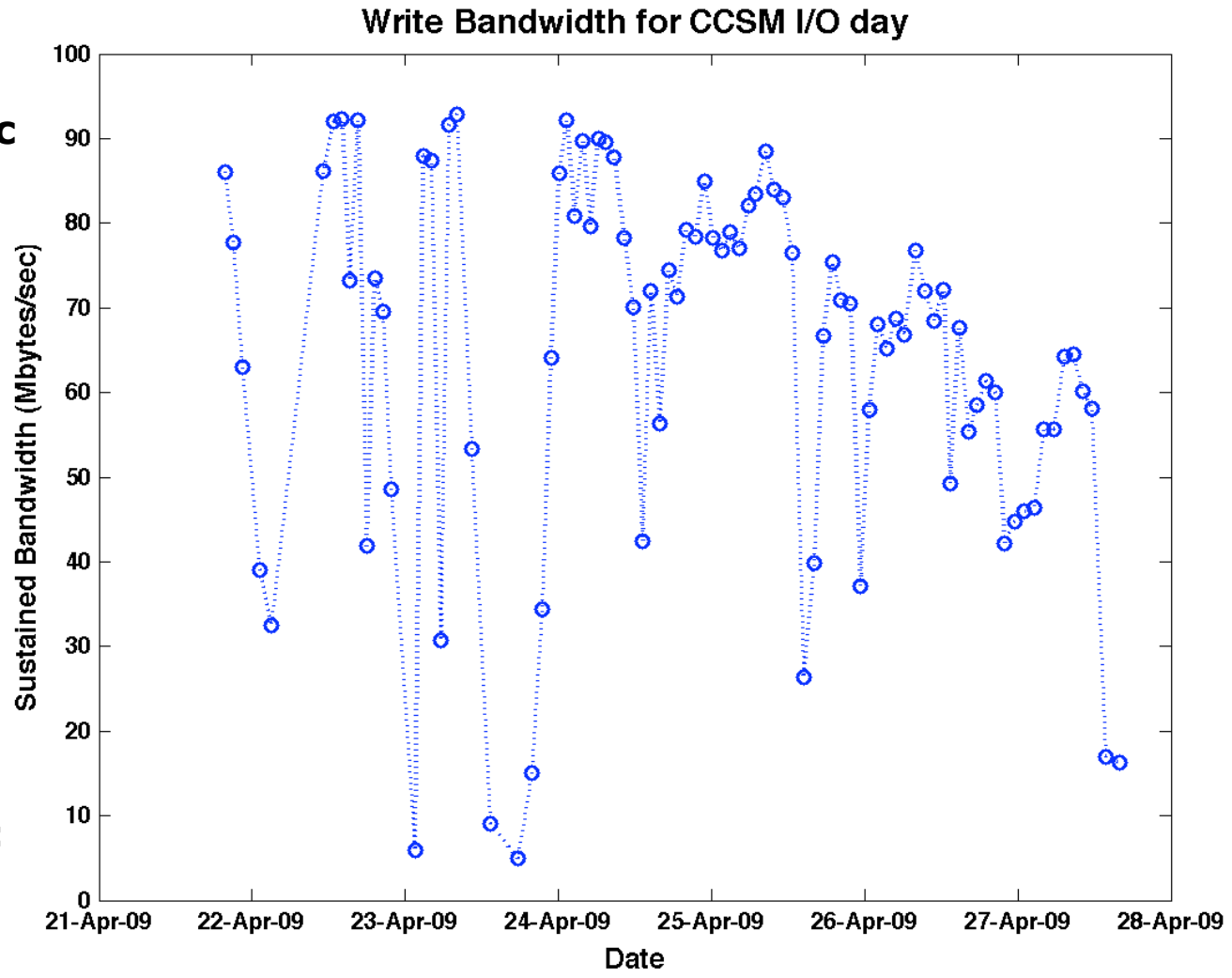
IO days

Variability of CCSM File Write Times on Kraken



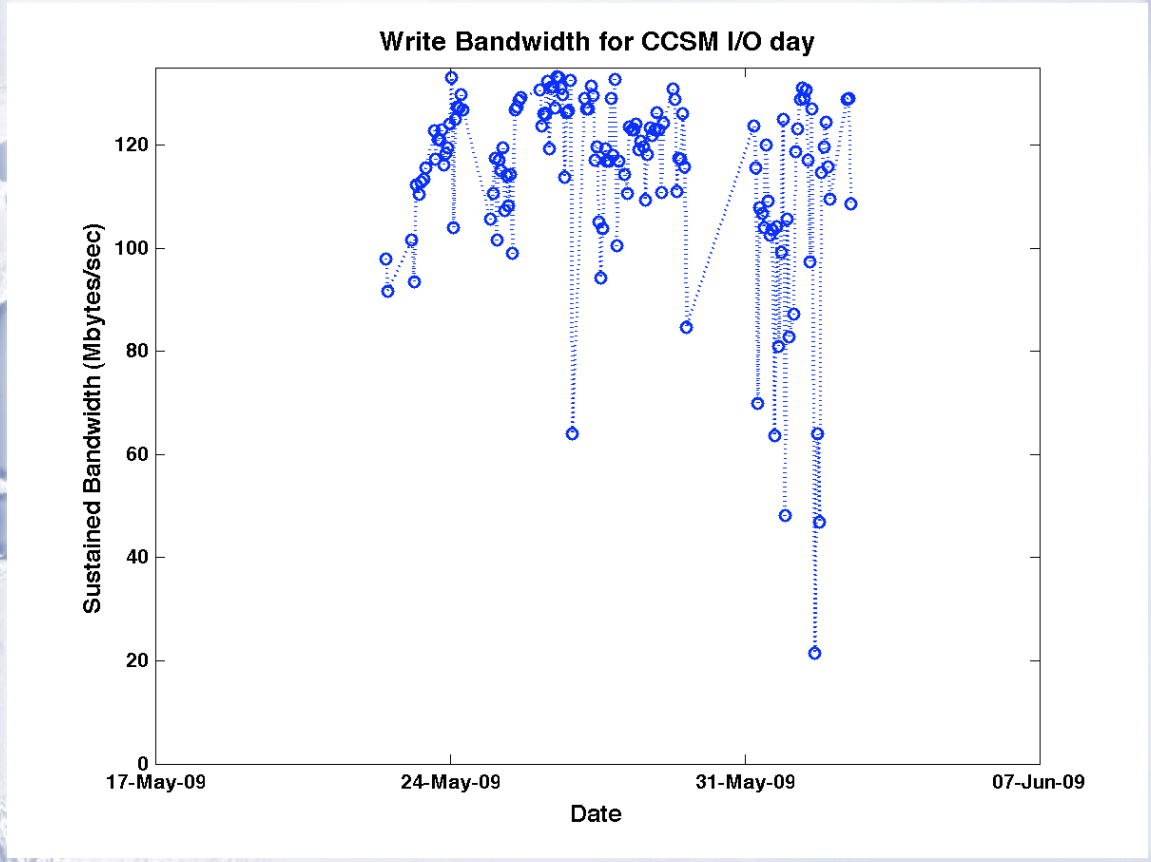
CCSM Sustained Output Bandwidth on Kraken [using big-endian]

High = 92 MB/sec



Low = 5 MB/sec

Lustre variability doubled cost of disk I/O from 11% to 23% of total



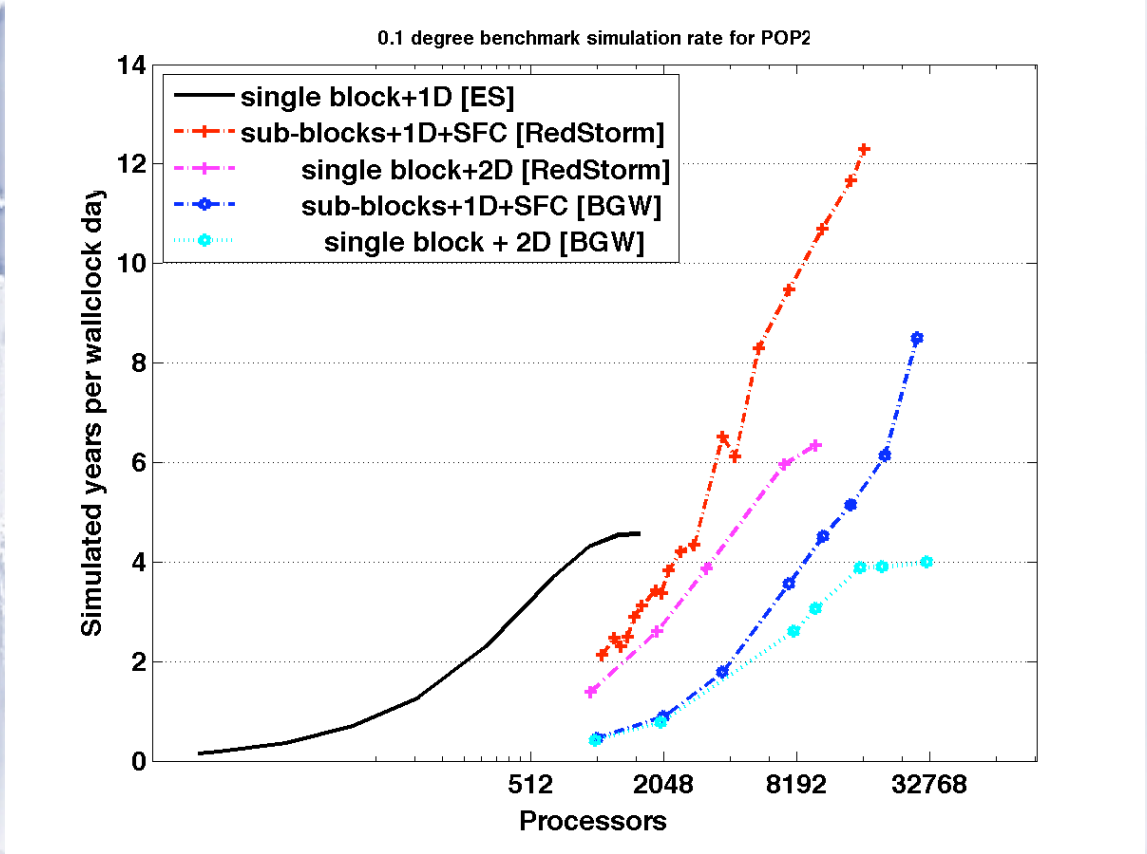
Write bandwidth for I/O day for high resolution CCSM on Kraken [little-endian]

Eliminated page size [4 Kb] I/O ops for POP restart

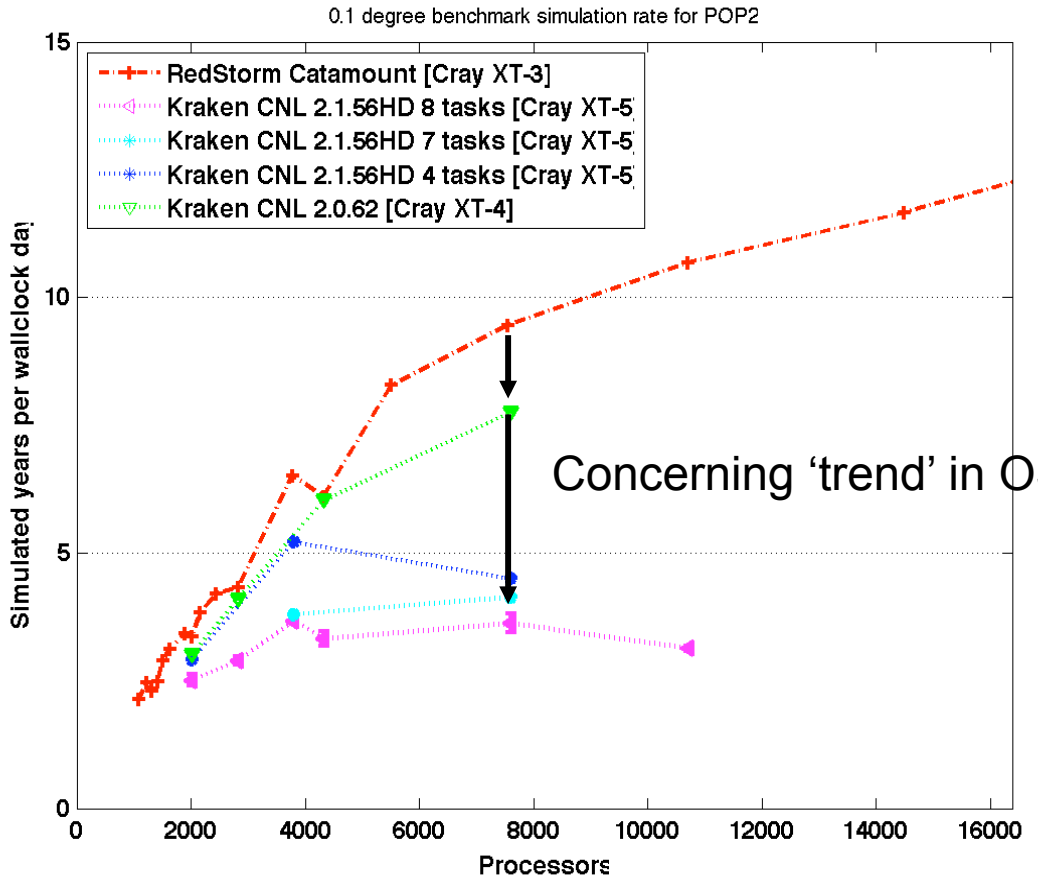
OS jitter and POP scalability

The good old days, back when the OS was scalable!

POP scalability [3 years ago]



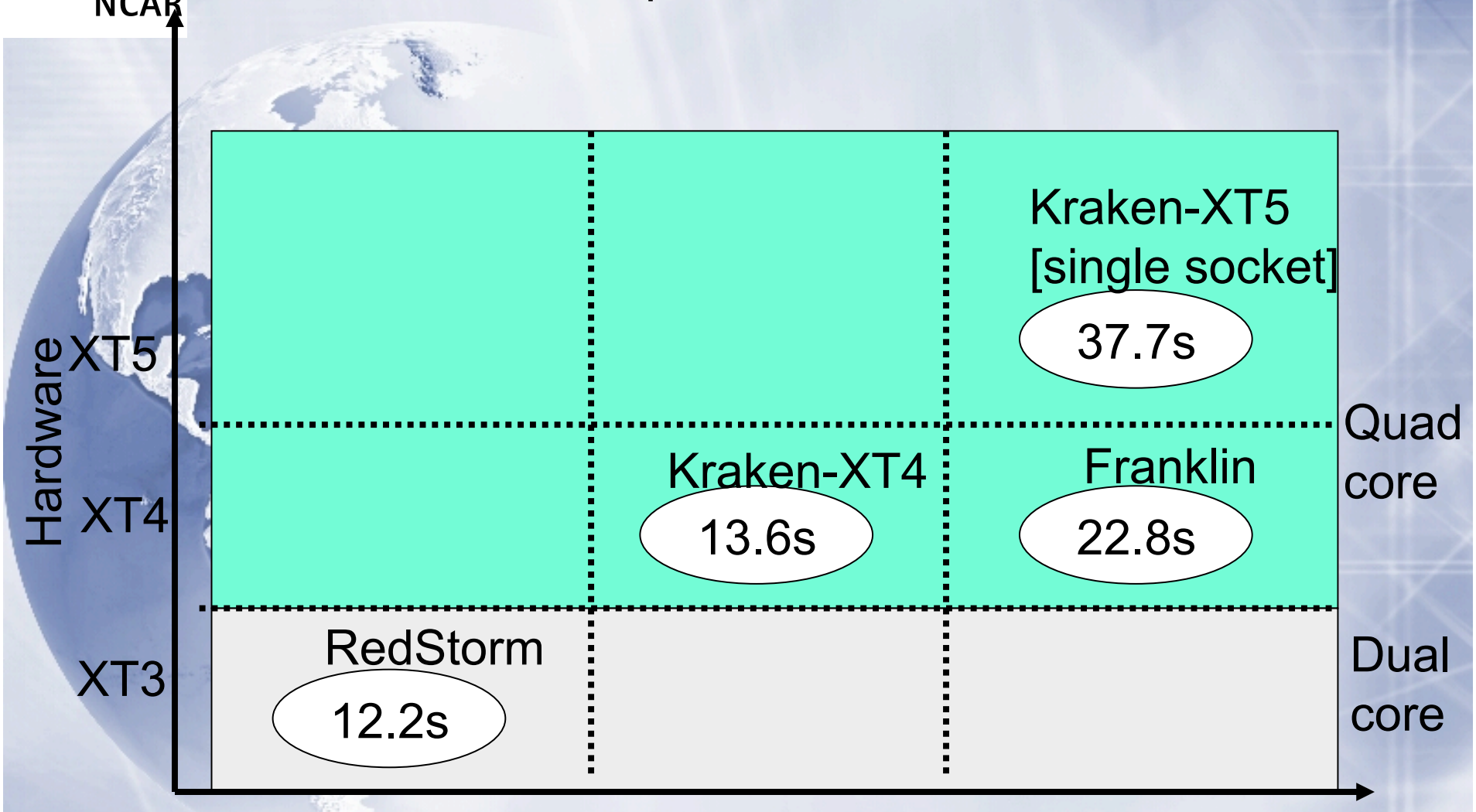
POP scalability [Today]

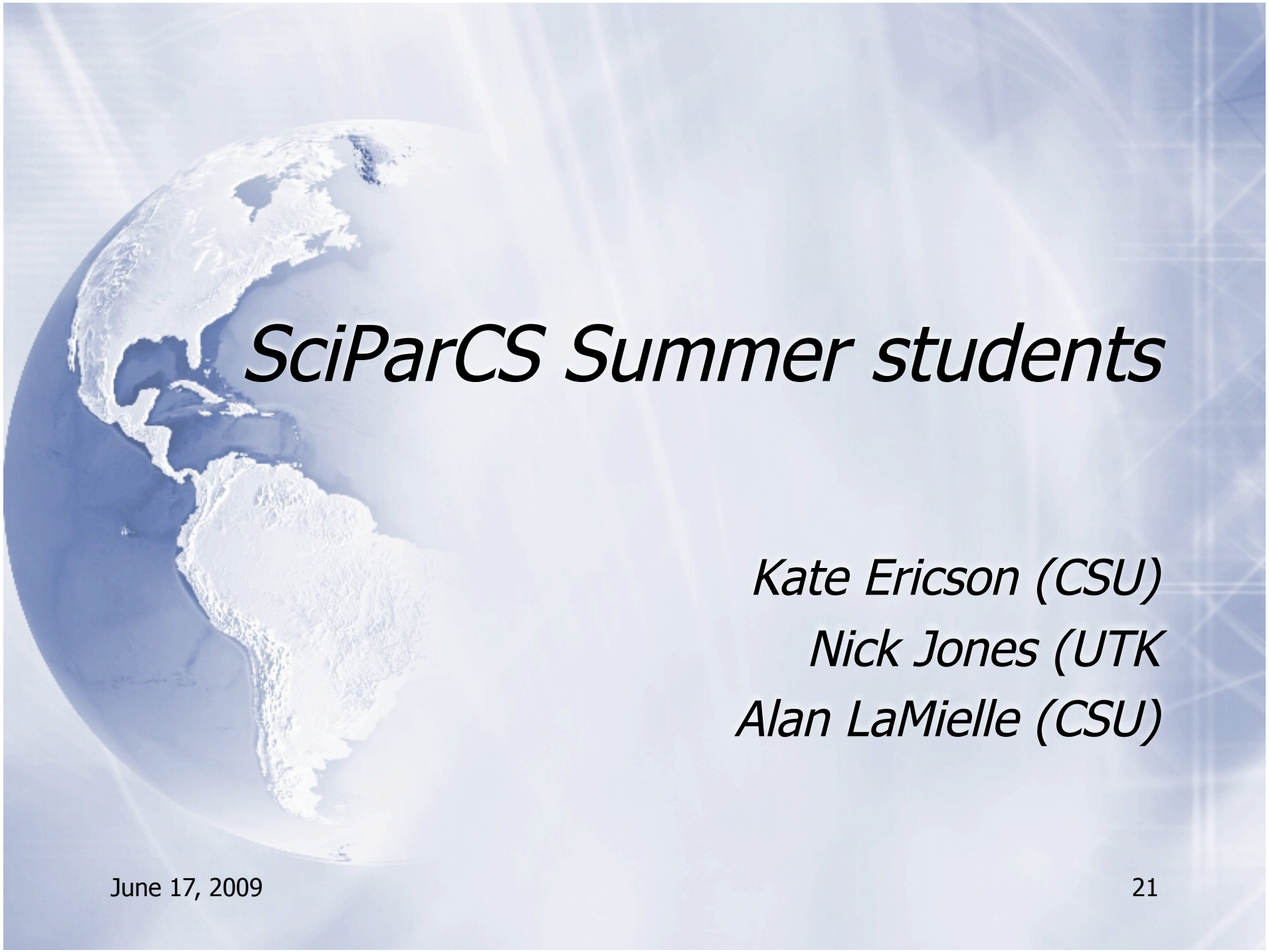


Concerning ‘trend’ in OS/HW evolution?

Execution time of Barotropic section of POP on 7600 cores

NCAR





SciParCS Summer students

Kate Ericson (CSU)
Nick Jones (UTK
Alan LaMielle (CSU)

NCAR

Jobs placement and network congestion in XT torus

- N. Jones [UTK]
- How does job placement impact performance
- Is performance variability due to congestion in torus network?

NCAR

PIO optimization

- K. Ericson [CSU]
- PIO has many tuning options (R. Jacob talk)
- Best configuration for Parallel I/O (PIO) library
  - Maximize bandwidth
  - Minimize memory usage
- Use Active Harmony [U of Maryland]

Parallel Lookup Table

- A. LaMielle (CSU)
- Goal: Reduce required memory footprint
- Decomposed across processors?
  - Sparse or dense access patterns?
  - E.G. All chemical reactions in Polar region?
- Preliminary Results
  - RSF: max of 5% accessed/task, 15% overall

NCAR

Conclusions

- Preliminary ultra-high-resolution science runs
  - Mesoscale processes: Hurricane cold wake
- 5844 cores in production (TeraGrid)
- Data indigestion:
  - Analysis tools are not ready
  - PetaApps project generated 5 TB of data last week!
- Future work:
  - Improve disk I/O performance [10 - 25% of time]
  - Improve memory footprint scalability
  - OS jitter investigation

Acknowledgements and Questions ?

- NCAR:
  - D. Bailey
  - F. Bryan
  - B. Eaton
  - 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
- Colorado State U
  - A. Lamielle
  - K. Ericson

- 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]
    - OCE-0825754
    - CNS-0421498
    - CNS-0420873
    - CNS-0420985
- Computer Allocations:
  - TeraGrid TRAC @ NICS
  - DOE INCITE @ NERSC
  - LLNL Grand Challenge
- Thanks for Assistance:
  - Cray, NICS, and NERSC