SciDAC CCSM Consortium: Software Engineering Activities

Pat Worley (ORNL)

(representing pretty much everyone)

SciDAC CCSM Consortium Group Meeting
March 17, 2005

Overview

- Most software engineering activities reported as part of other task group reports
- Generic activities:
  - Continued development, optimization, and validation on the Cray X1: CAM (EUL and FV), full CCSM.
  - Helping CSEG/SEWG in planning and work.
- Related activities:
  - Porting to new systems: BG/L, Cray XD1, Cray XT3, Cray X1e, IA64 clusters, SGI Altix, …

Outline

- Recent Results:
  - Papers in special issue of IJHPCA
- Physics optimizations
- Issues/Futures
- Talks tomorrow
  - Subgrid Orography
  - BG/L performance

Special Issue of IJHPCA

Preface — A Patrinos

Overview of the Software Design and Parallel Algorithms of the CCSM
G Carr, J Drake, P Jones

Performance Portability in the Physical Parameterizations of the Community Atmospheric Model
P Worley, J Drake

A Scalable Implementation of a Finite-Volume Dynamical Core in the Community Atmospheric Model
A Mirin, W Sawyer

Cross-Platform Performance of a Portable Communications Module in the NASA Finite Volume General Circulation Model — W Putnam, S-J Lin, B-W Shen

High Resolution Mesh Convergence Properties and Parallel Efficiency of a Spectral Element Atmospheric Dynamical Core — J Dennis, A Fournier, W Spotz, A St-Cyr, M Taylor, S Thomas, H Tufo

Vectorizing the Community Land Model (CLM)
F Hoffman, M Vertenstein, H Kitabata, J White III

A Performance Model of the Parallel Ocean Program
D Kerbyson, P Jones

The Model Coupling Toolkit: A New Fortran90 Toolkit for Building Multi-Physics Parallel Coupled Models
J Larson, R Jacob, E Ong, Guo

cpl6: The New Extensible, High-Performance Parallel Coupler for the Community Climate System Model
A Craig, R Jacob, B Kauffman, T Bettge, J Larson, E Ong, C Ding

MxN Communication and Parallel Interpolation in CCSM3 Using the Model Coupling Toolkit
R Jacob, J Larson, E Ong

Coupling Multi-Component Models by MPH on Distributed Memory Computer Architectures
Y He, C Ding

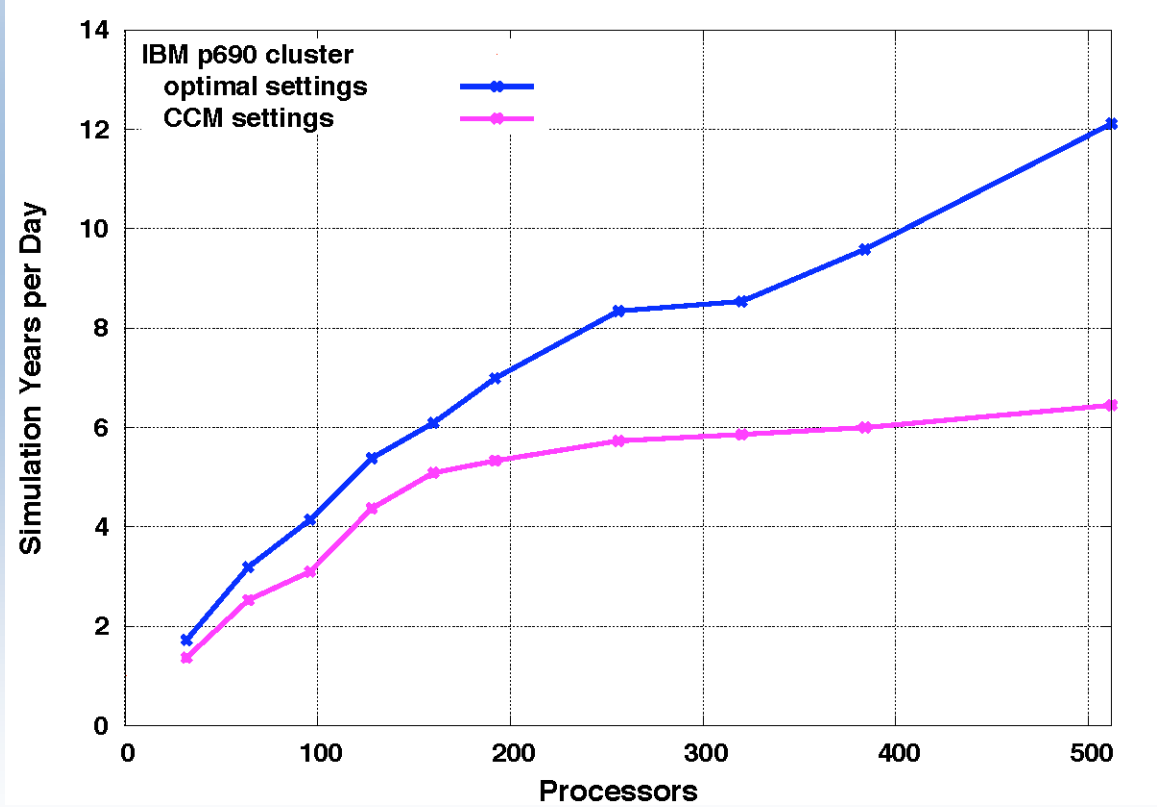
Distributed and Implementation of Earth System Modeling Framework Components
W Collins, G Theurich, C DeLuca, A Trayaonv, P Li, W Yang, C Hill

IBM p690 cluster benchmarks

Comparing performance of cam3.0p1 using original CCM pcols (256-258) and load balancing (-1) and the optimal settings.

P	MPI x OMP	pcols	lbal	alltoall	improve-ment
32	32x1	16	2	0	28%
64	64x1	16	3	1	28%
96	96x1	16	3	1	36%
128	128x1	16	3	1	25%
160	40x4	16	3	1	21%
192	48x4	16	3	1	33%
256	128x2	16	3	1	47%
320	40x8	16	3	1	48%
384	48x8	16	3	1	62%
512	128x4	16	3	1	91%

T85L26, Eulerian Spectral Dycore, 30 days (September)

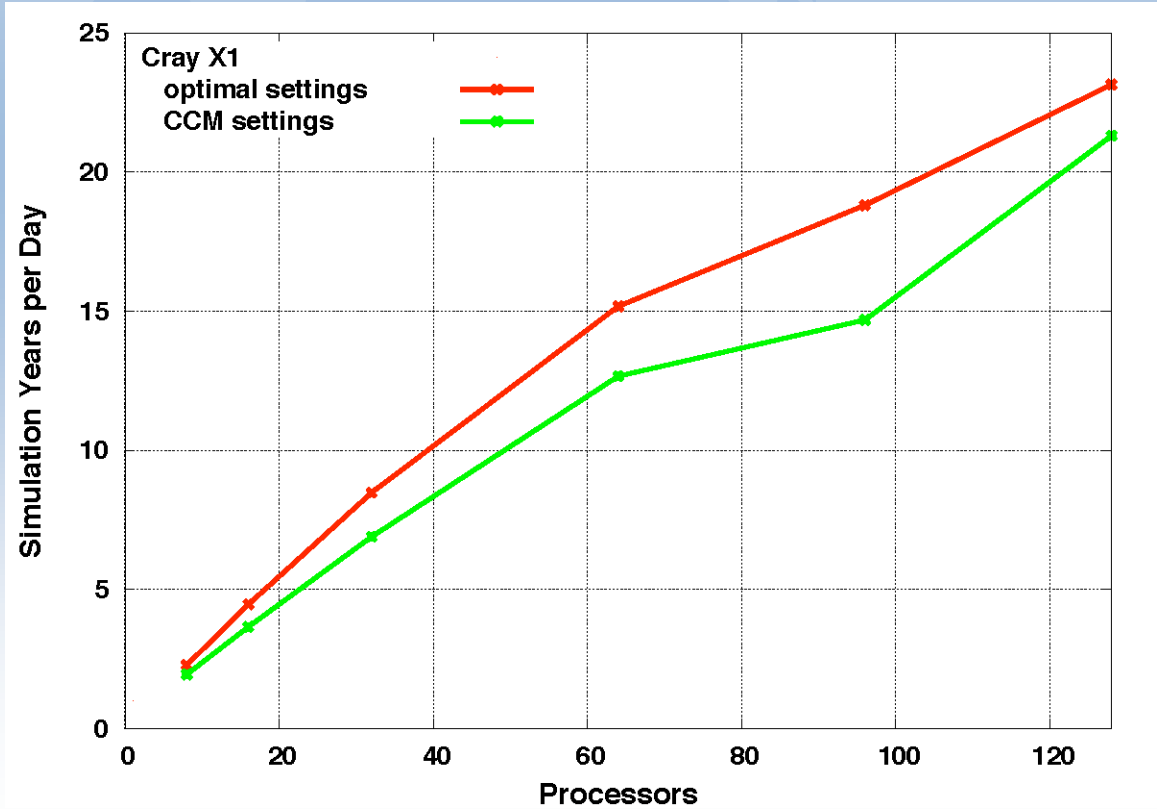


Cray X1 benchmarks

Comparing performance of cam3.0p1 using original CCM pcols (256-258) and load balancing (-1) and the optimal settings.

P	MPI x OMP	pcols	lbal	alltoall	improve-ment
8	8x1	1026	2	0	19%
16	16x1	1026	2	0	23%
32	32x1	1026	2	0	24%
64	64x1	514	2	0	21%
96	96x1	514	2	0	30%
128	128x1	258	2	0	10%

T85L26, Eulerian Spectral Dycore, 30 days (September)



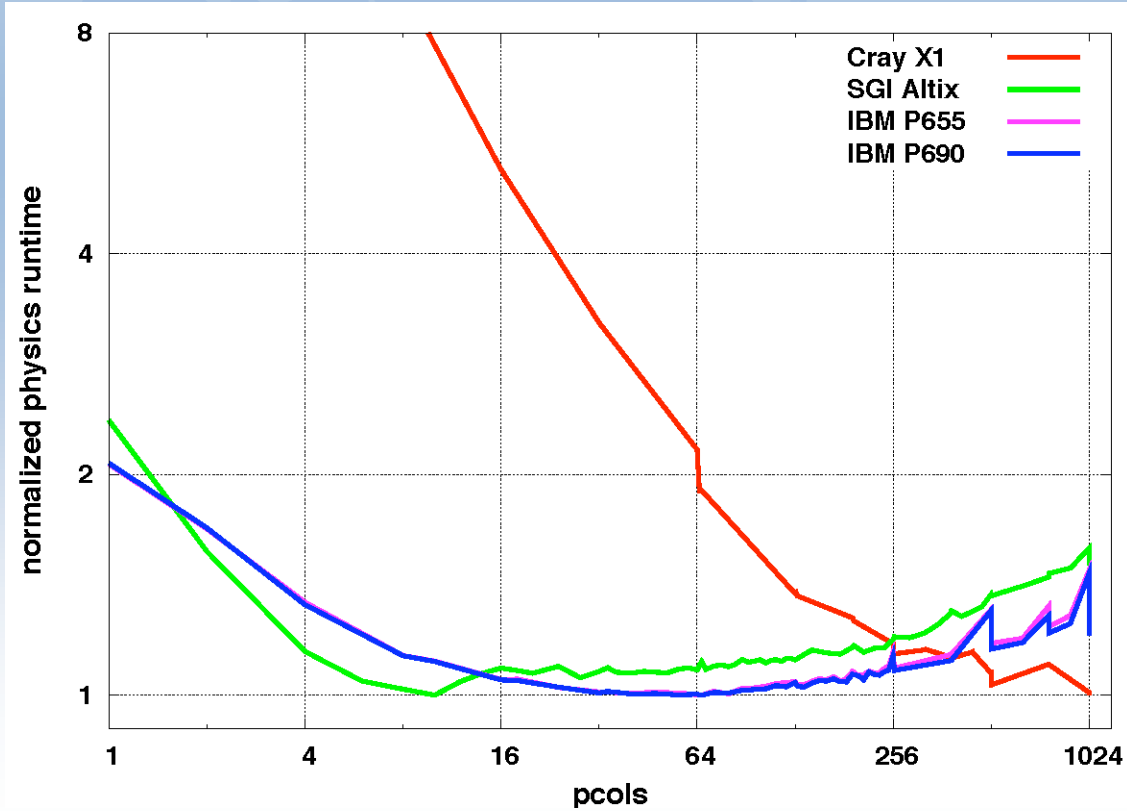
Chunk Size Sensitivity

T21L26, Eulerian Spectral Dycore, 1 day (September 2)

Comparing physics runtime as a function of chunk size when run on 2 processors (1024 columns per processor). Two processors chosen so that they do not share cache. (All other processors in SMP node were idle.) Each curve normalized separately, by minimum runtime for given platform.

Optimal PCOLS

Cray X1:	>= 1026
SGI Altix:	8
IBM P655:	66
IBM P690:	56



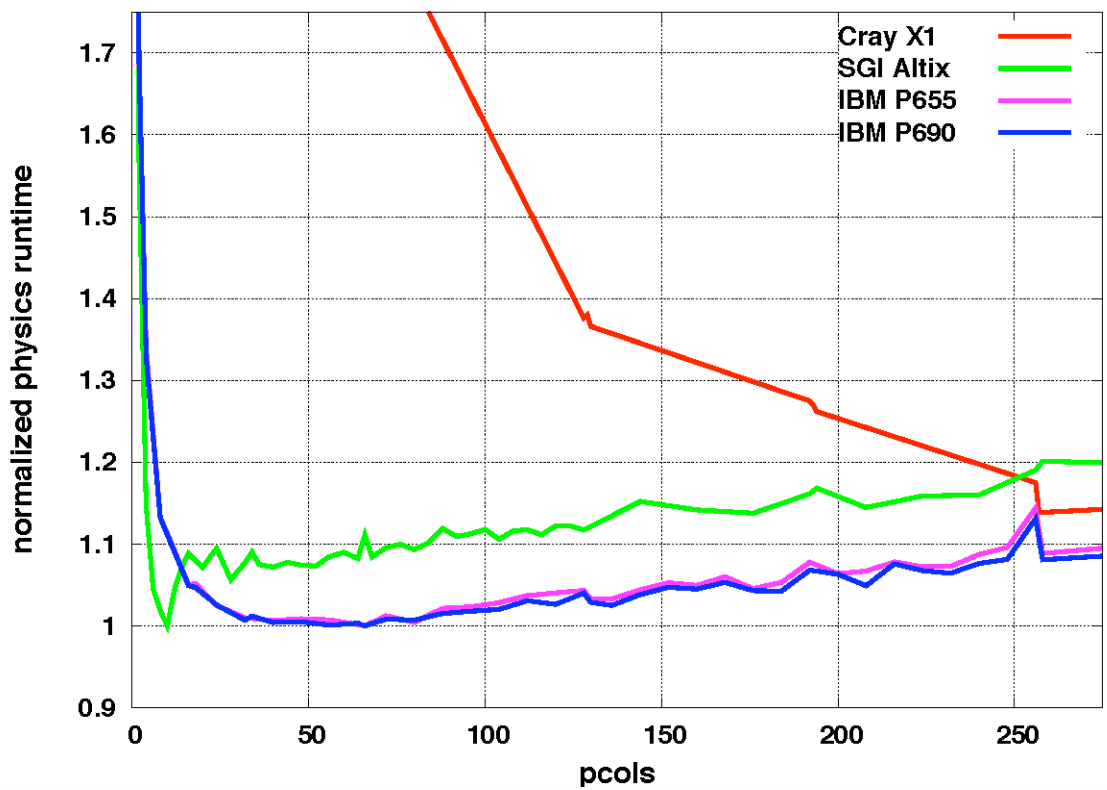
Chunk Size Sensitivity

T21L26, Eulerian Spectral Dycore, 1 day (September 2)

Comparing physics runtime as a function of chunk size when run on 2 processors (1024 columns per processor). Two processors chosen so that they do not share cache. (All other processors in SMP node were idle.) Each curve normalized separately, by minimum runtime for given platform.

Optimal PCOLS

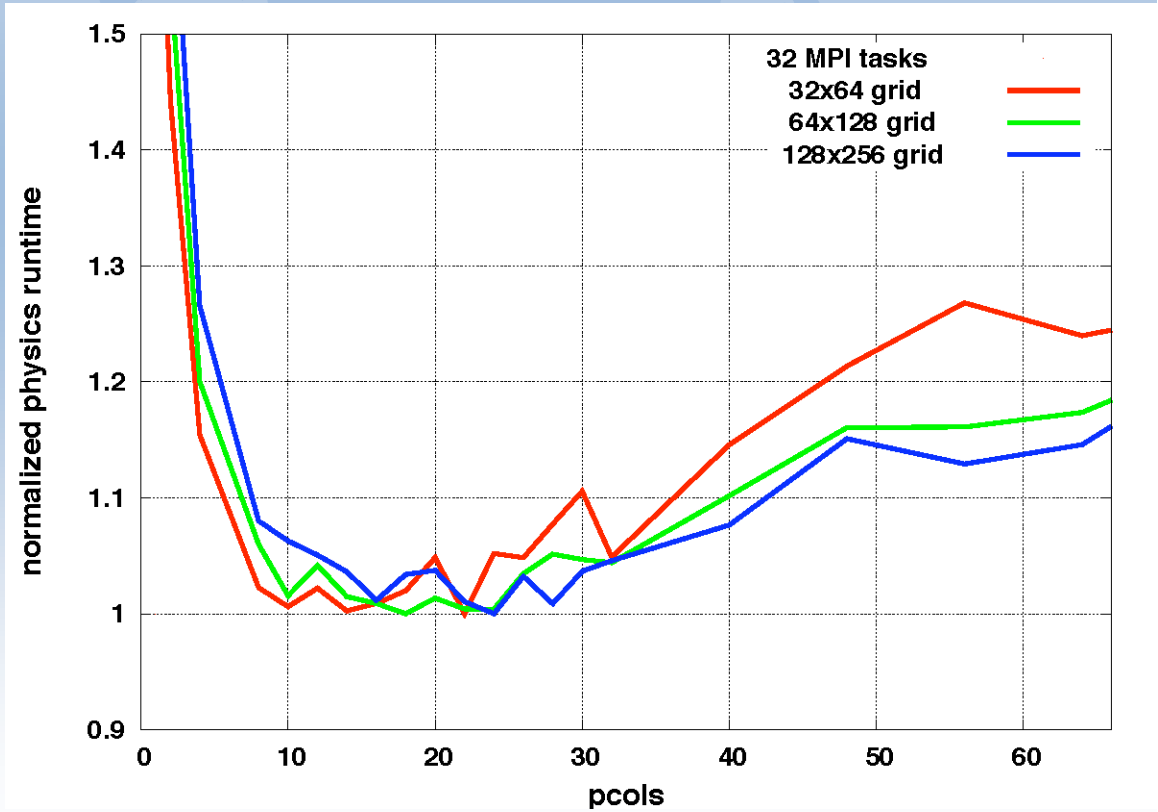
Cray X1:	>= 1026
SGI Altix:	8
IBM P655:	66
IBM P690:	56



Chunk Size Sensitivity for IBM p690

Comparing physics runtime as a function of chunk size when run on 32 processors for T21L26, T42L26, and T85L26. Each curve normalized separately, by minimum runtime for given problem size. Contention decreases optimal chunk size to between 18 and 24, as compared to 56 in previous experiments. The optimal pcols stayed the same in similar experiments on the SGI Altix and the Cray X1.

Eulerian Spectral Dycore, 1 day (September 2)



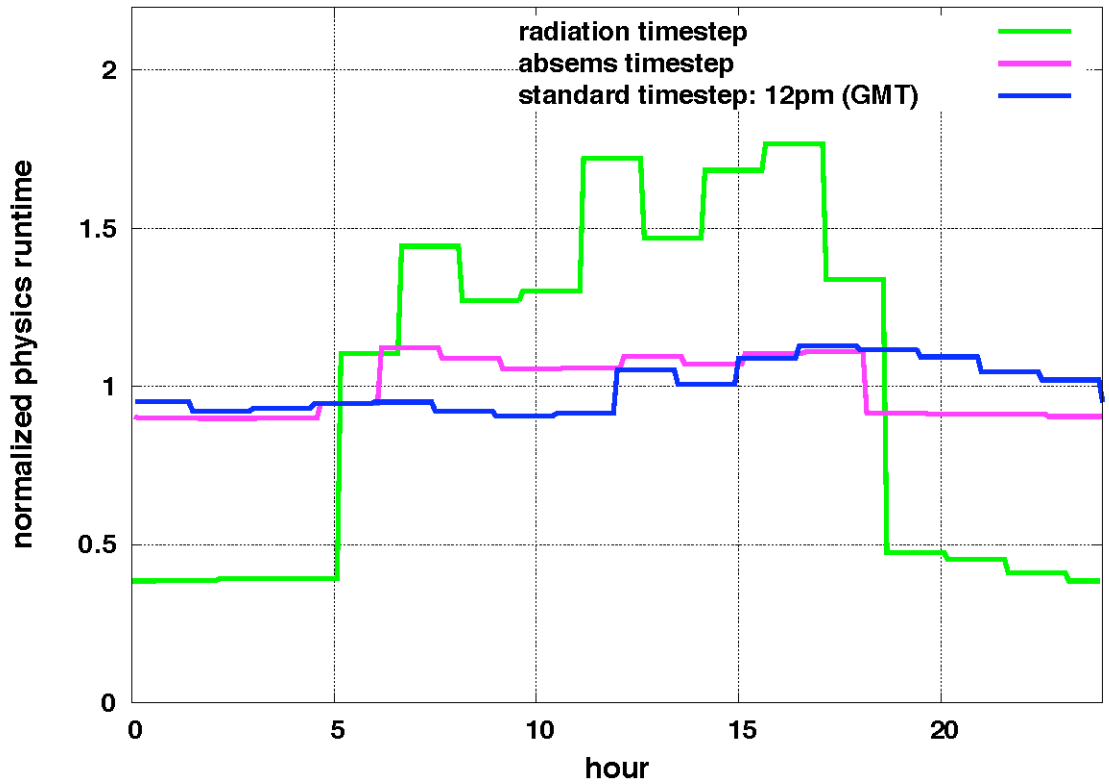
Load Imbalance on IBM Power4

T85L26, Eulerian Spectral Dycore, September 1

Comparing physics runtime for blocks of 16 consecutive columns for different types of timesteps as a function of simulation hour. Data collected from equatorial latitude. Runtime normalized for each type of timestep separately. Summary data for each type of timestep for latitude line:

	steps/day	secs/step	secs/day
standard	120	.078	9.38
radiation	22	.270	5.94
absems	2	1.630	3.26

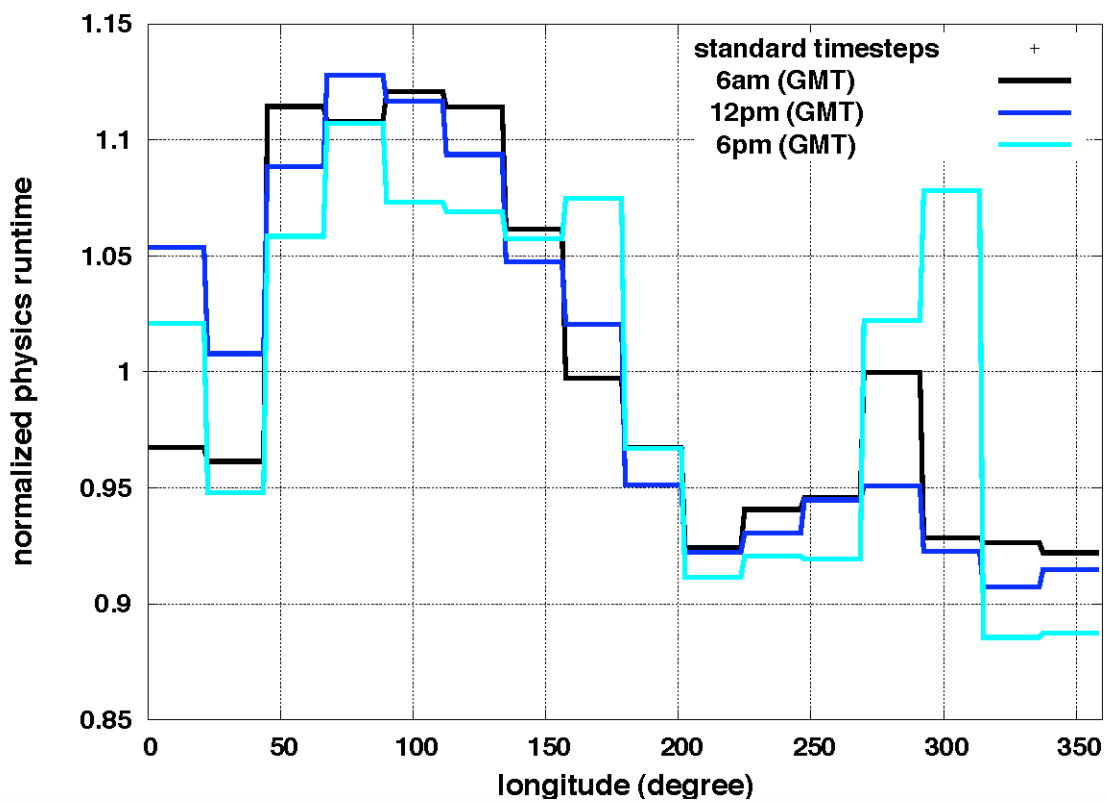
Load imbalances for radiation timestep are large and primarily due to diurnal cycle.



Load Imbalance on IBM Power4

Comparing physics runtime for blocks of 16 consecutive columns for standard timestep as a function of longitude.
Data collected from equatorial latitude at three different times (same day). Runtime normalized for each timestep separately.
Load imbalances are small, and primarily spatial.

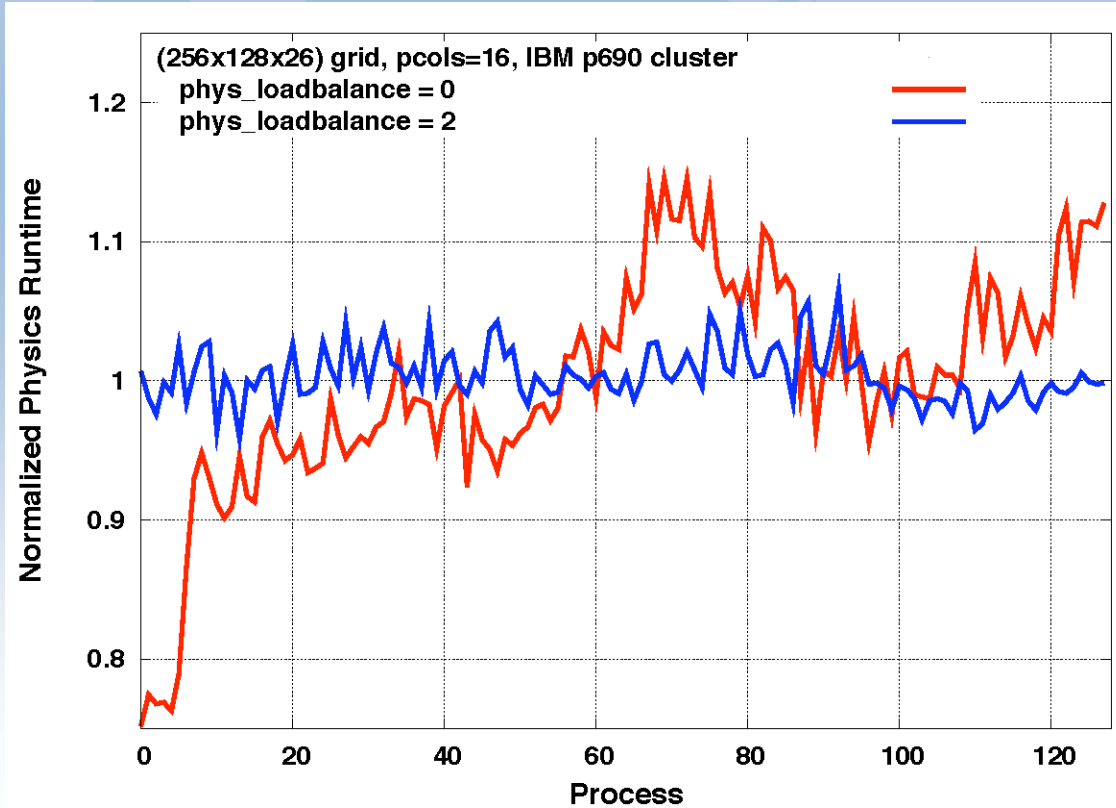
T85L26, Eulerian Spectral Dycore, September 1



Load Imbalance on IBM p690 cluster

Comparing physics runtime with and without load balancing when each processor is assigned a single latitude line. Communication cost of load balancing is not included. Both curves normalized by the mean runtime for the no load balancing case.

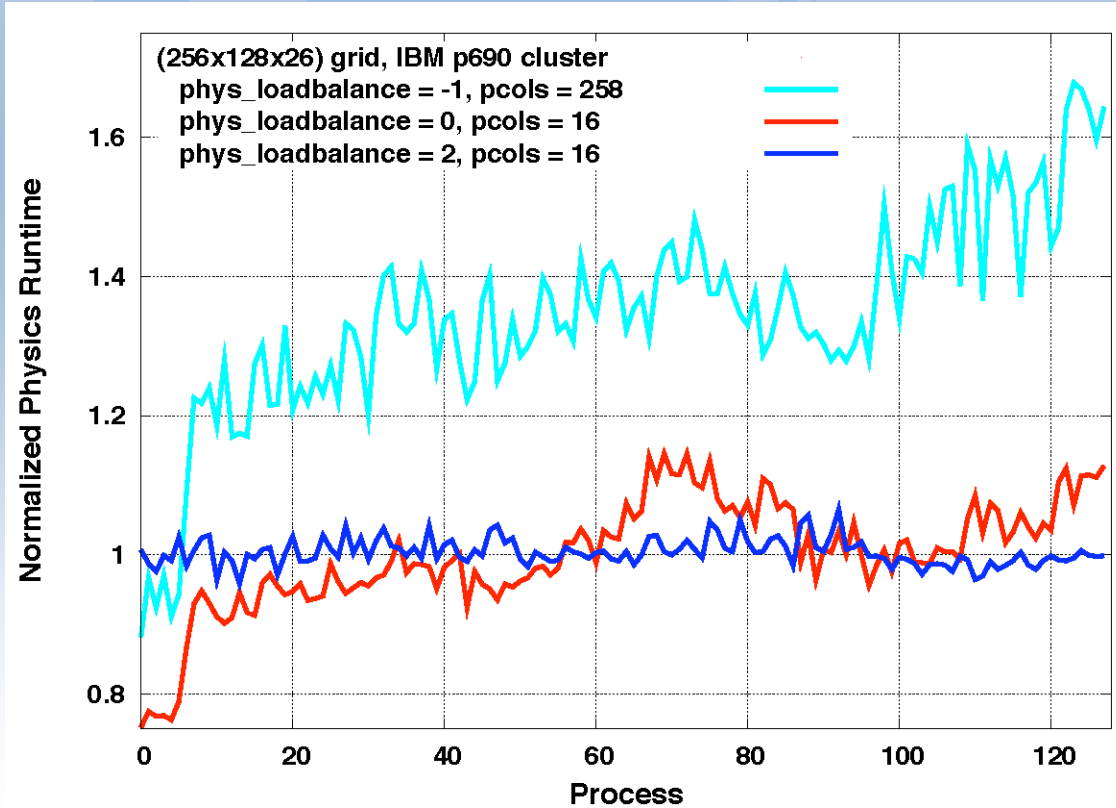
T85L26, Eulerian Spectral Dycore, 1 day (September 2)



Load Imbalance on IBM p690 cluster

T85L26, Eulerian Spectral Dycore, 1 day (September 2)

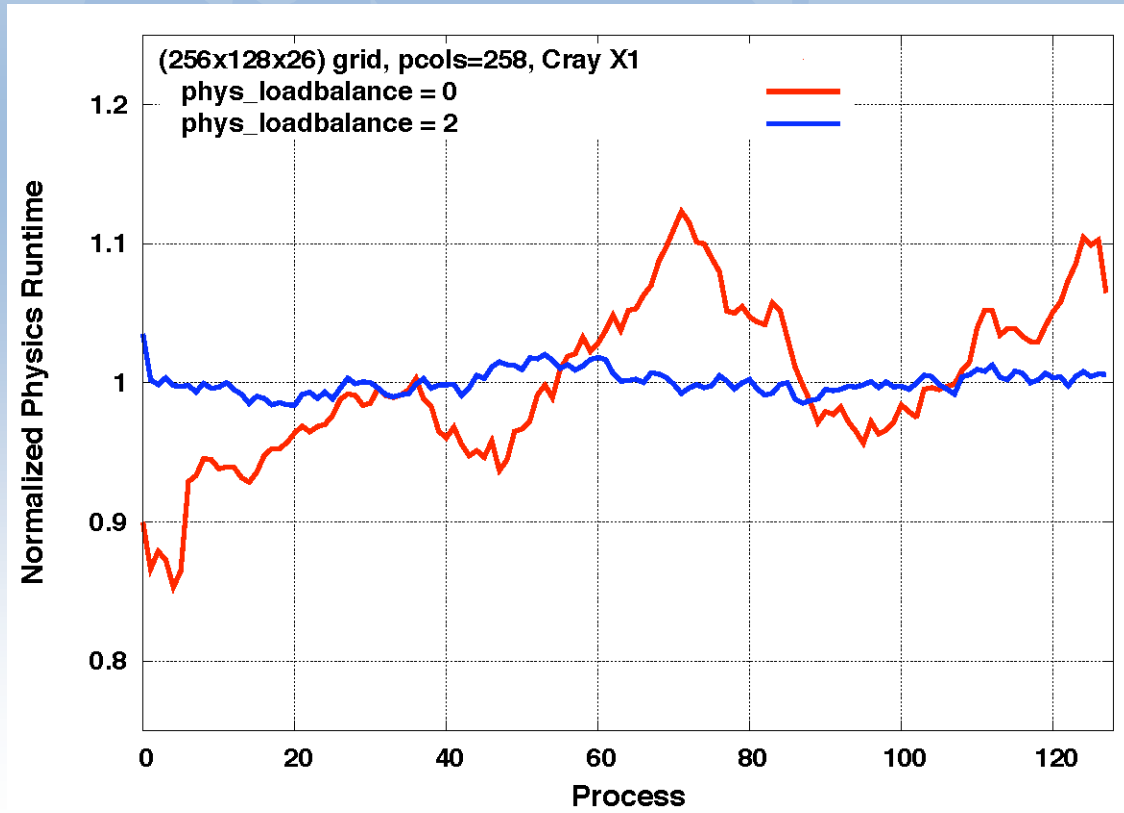
Comparing physics runtime with load balancing and optimal pcols, without load balancing but with optimal pcols, and without load balancing and with CCM pcols. Each processor is assigned a single latitude line. Communication cost of load balancing is not included. All curves normalized by the mean runtime when using optimal pcols but no load balancing.



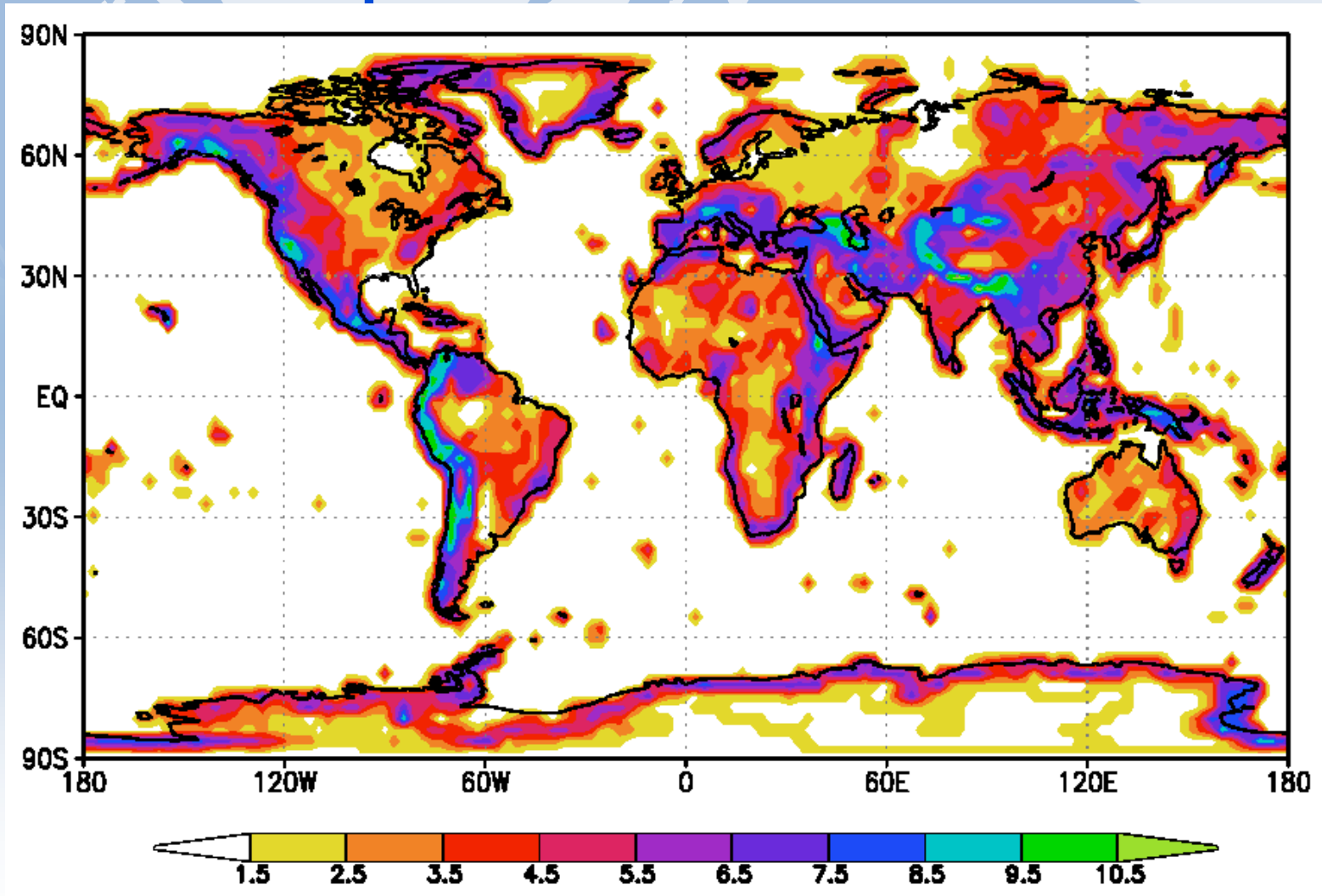
Load Imbalance on Cray X1

Comparing physics runtime with and without load balancing when each processor is assigned a single latitude line. Communication cost of load balancing is not included. Both curves normalized by the mean runtime for the no load balancing case.

T85L26, Eulerian Spectral Dycore, 1 day (September 2)



Computational Burden



Issues

- Helping CSEG achieve what we want/need them to achieve
  - What in particular?
    - Single executable?
    - CAM dynamics interface?
    - FV dynamics validation and optimization?
    - updating ocean model(s)?
    - Biogeochemistry branch?
    - CCSM validation on new platforms?
    - CCSM optimization on new platforms (including performance intrumentation and evaluation technology)
    - ESMF evaluation?
    - LMWG research directions?

Issues

- Things we want to do that CSEG won't do
  - Interfacing new dycores?
  - Performance optimization for *our* science runs?
    - High resolution
    - Atmospheric chemistry
  - Porting to / optimizing on new platforms?
    - What new platforms are reasonable for production (not just benchmarking)?

Futures

- High Resolution and/or "Petaflops" computing?
- More atmospheric chemistry optimizations?
  - Load balancing
  - Efficient transport
- New components?
  - Integrating other dycores (HOMME, Cubed Sphere FV, …)
  - Developing new dycores
  - ???