CAM performance on Intel®-based clusters

D. Shkurko
Intel Corp., pr. ak. Lavrentieva 6/1, Novosibirsk, Russia, 630090
dmitry.v.shkurko@intel.com

P. Worley
Oak Ridge National Laboratory
worleyph@ornl.gov

12th Annual CCSM Workshop, 19-21 June 2007

Executive summary

- Performance and scalability of FVCAM have been evaluated for wide spectrum of core numbers
- Intel compilers have been successfully used to port and optimize FVCAM
- Important FVCAM tuning options were applied to improve performance of the D-grid workload on Intel®-based clusters

Notations and abbreviations

- **Conroe processor** – Dual Intel® Core™2 Duo processor 6800 series
- **Woodcrest processor** – Dual Core Intel® Xeon® processor 5100 series
- **Clovertown processor** – Quad Core Intel® Xeon® processor 5300 series
- **ppn** – processes per node (the number of cores used on a node)
- **OpenMPI** – an MPI implementation from Univ.-s Of Tennessee, Indiana, and Stuttgart, and Los Alamos National Laboratory

Workload overview

- Finite Volume dynamical code running on .5x0.625 degree resolution computation grid (also called “D grid”)
  - 361x576x26 (latitude by longitude by vertical) grid
  - Finite volume dycore is the preferred (required among current options) for atmospheric chemistry due to its conservation properties.
- Initial production grid size will be 1.9x2.5 degree resolution (96x144x26), so 15 times smaller than D grid
- More details can be found in the references

System configurations

	CPUs	Memory	OS	Interconnect
32 2-way Clovertown nodes	2.67GHz 4 cores 4MB/2 cores Clovertown	16Gb per node	RHEL4 (U3), OFED-1.2	DDR Infiniband fat-tree network
256 2-way Woodcrest nodes	3.0GHz 2 cores 4MB/2 cores Woodcrest	8Gb per node	RHEL4 (U2), OFED-1.2	
64 1-way Conroe nodes	2.93 Ghz 2 cores 4MB/2 cores Conroe	8Gb per node	RHEL4 (U2), OFED-1.1	

Intel® Fortran and C compilers v10.0 for Linux

Advantages of using Intel® Compilers

Efficiency:

Inherent ability to highly optimize codes for all Intel multi-core processors

Ease of Use:

Automatic optimization features make it easier to obtain highly optimized target code

OpenMP support simplifies threaded application development

Supports other Intel tools:

Code instrumentation for
- Cluster Tools and
- Threading Tools

Enhanced debug information for Intel debugger

Intel® compilers use:

- SSE/SSE2/SSE3/SSE4 instructions
- Interprocedural optimizations
- Auto-vectorization for Intel 64
- OpenMP v2.5

Advanced high-level optimizations

Some useful Intel® Compiler Options

-O2/-O3 Turn optimizations for speed, including high-level optimization

-ip/ipo Enables single-/multi-file interprocedural optimizations

-x*X*/-mtune=… Tuning for particular processor
-xT Enables auto-vectorization with the use of specific Woodcrest instructions

-fno-alias, -safe-cray-ptr, –assume nodummy_aliases These allow for making different aliasing assumptions making more optimizations possible

-openmp Turns on OpenMP directives compilation into threaded code

-tcheck/-tcollect/-openmp_profile Turns on instrumentation for checking or profiling with other Intel tools like Intel Thread Checker, Intel Thread Profiler and Intel Trace Collector

Controlling FP calculations

High Level option: -fp-model <spec>

Accuracy increases →

fast=2 | **fast=1** | **precise/source** | **strict**

← Speed increases

Lower Level options

Speeding up expensive operations

- **-ftz** denormals are treated as zero
- **-no-prec-{div,sqrt}** faster but less accurate division and/or sqrt

-fp-speculation fast|safe|off asks to speculate or not to speculate on FP operations

Lower Level options

-assume protect_parens forces handling of parenthesis according to Fortran standard

-fpe*N* Specifies the floating-point exception handling that possibly impacts performance

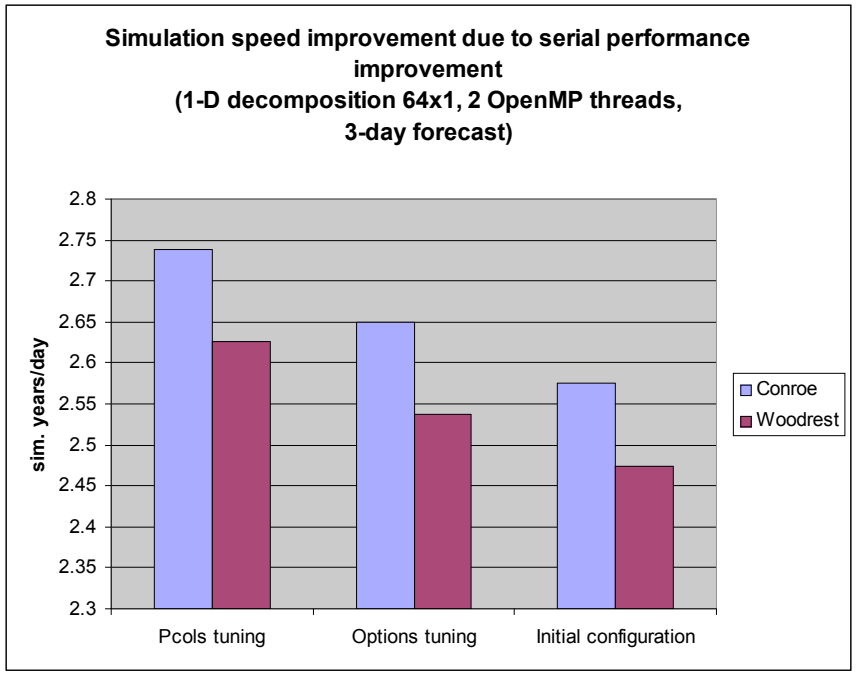
Deprecated options

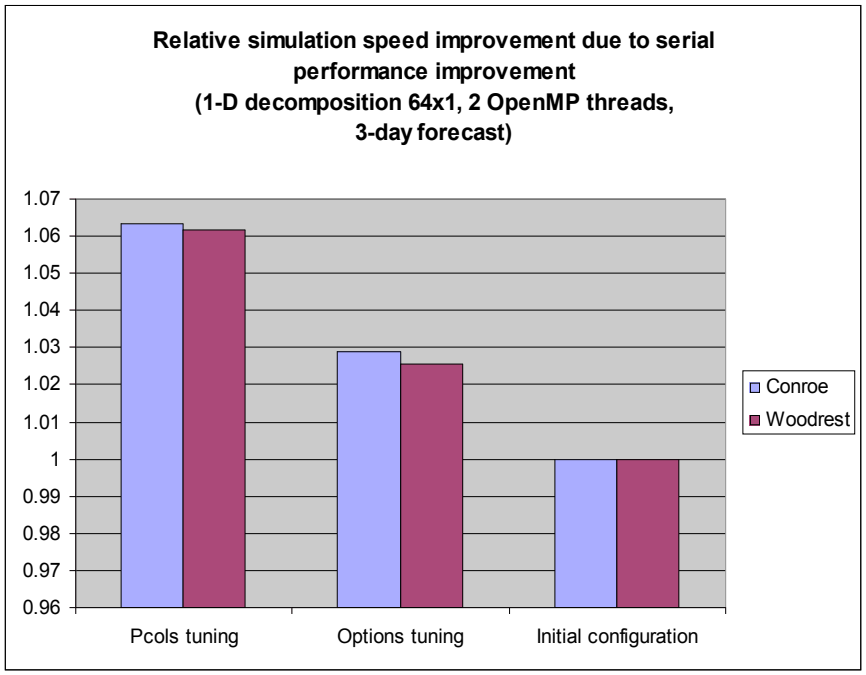
-mp,-mp1,-fltconsistency
Superseded by new options

CAM 3.1p2 build configuration

- OpenMPI 1.2.2
- Compilers:
  - ifort 10.0.023
  - icc 10.0.023
- Compiler options:
  - -O2 –xT –ftz –fp-model precise –g -openmp

Serial optimizations



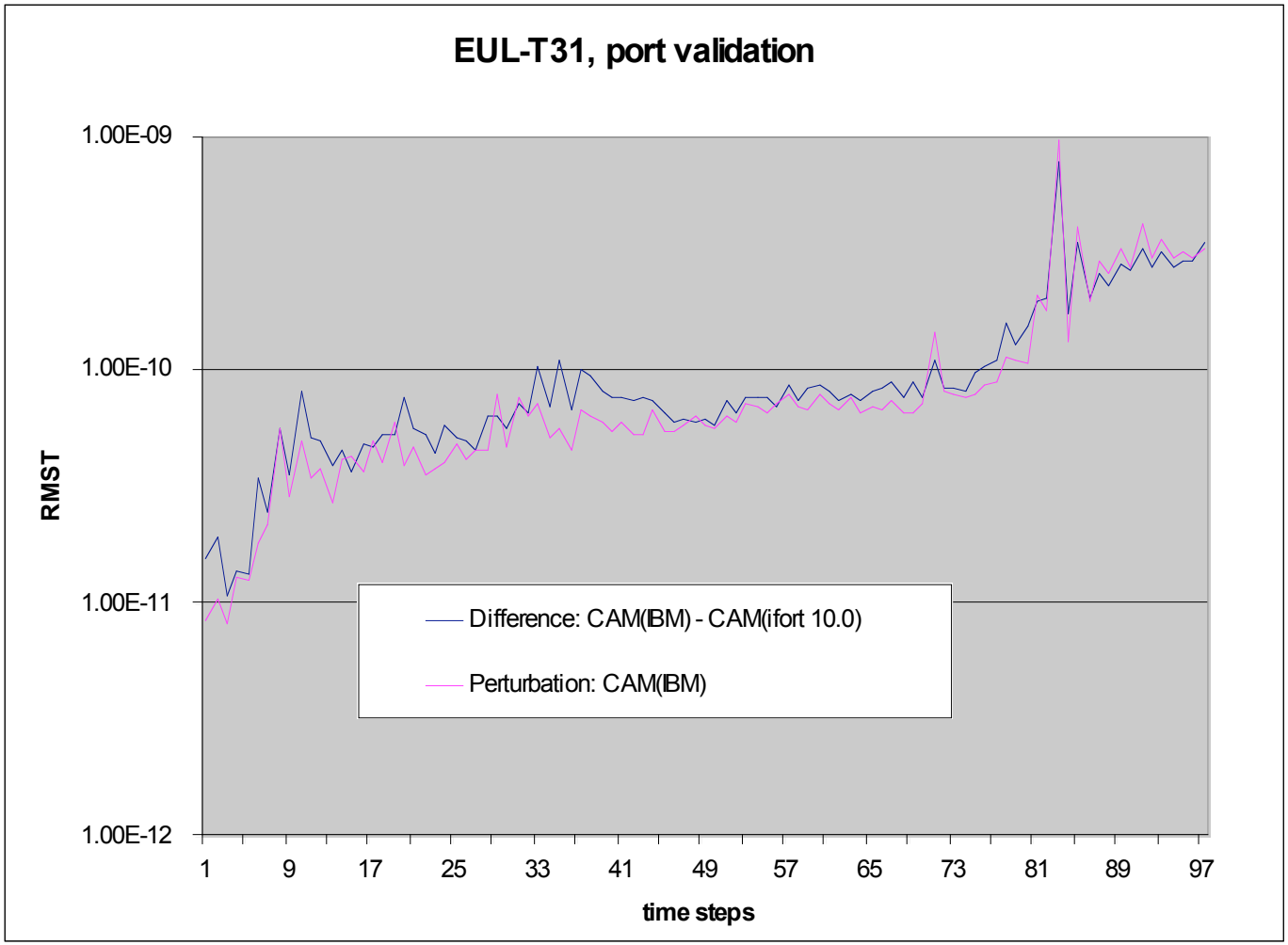


Optimal options: -O3 -xT -g -fp-model precise -fno-alias -ip -no-prec-div -no-prec-sqrt -openmp

Optimal pcol parameter: 8

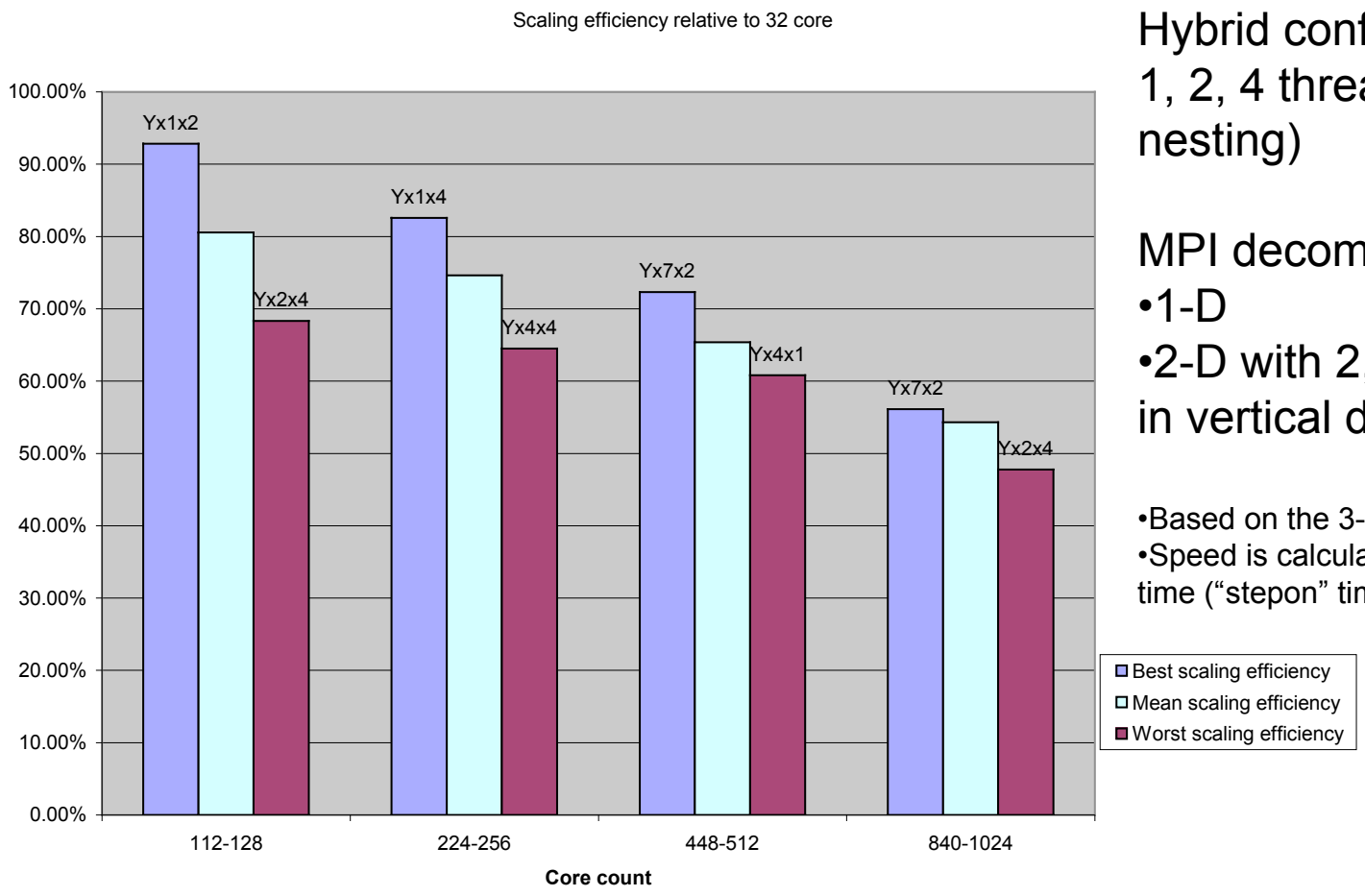
- Based on the 3-day forecasts
- Speed is calculated from integration time (“stepon” timer)

Compiler options validation



Compiler options and “pcols” parameter were tuned for D-grid workload

CAM scaling on Woodcrest cluster

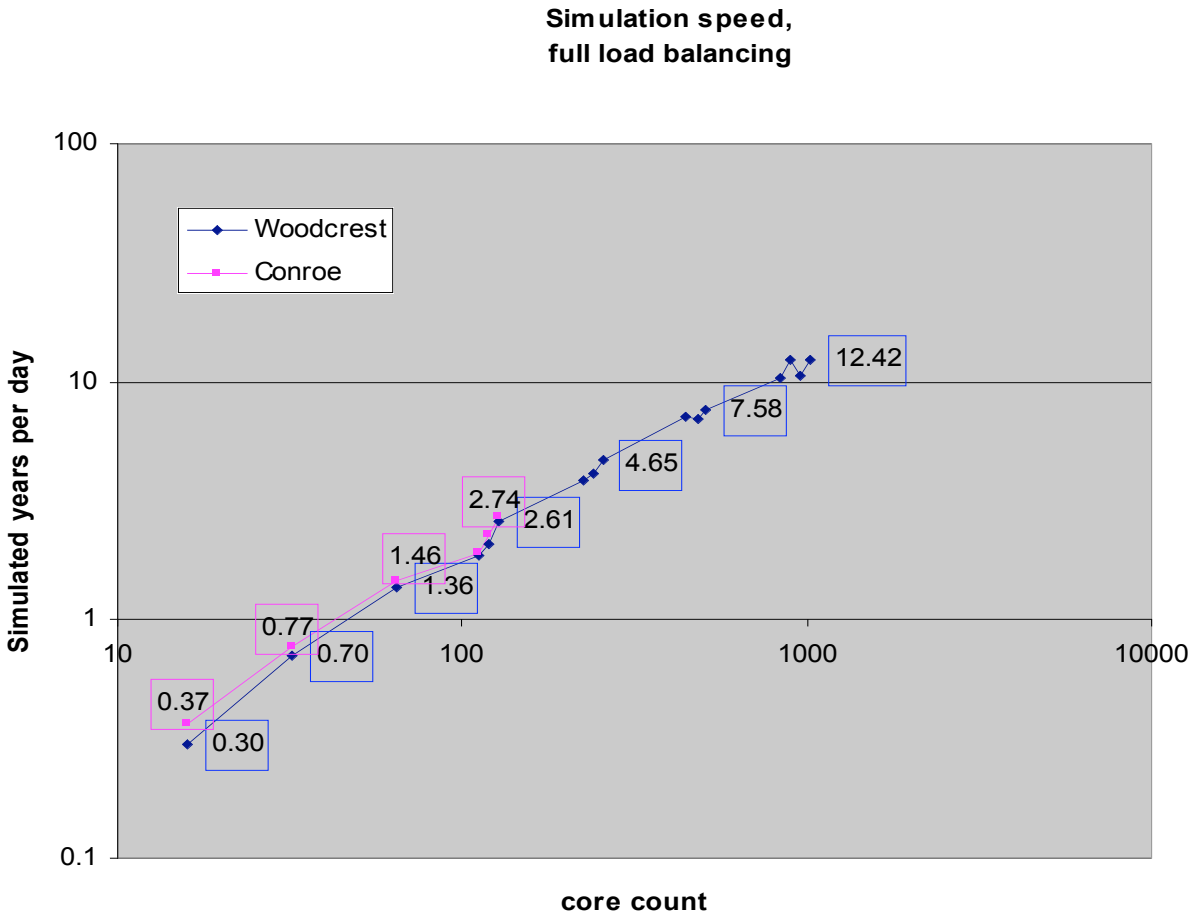


Hybrid configurations:
1, 2, 4 threads (no nesting)

MPI decompositions:
- 1-D
- 2-D with 2, 4, 7 processes in vertical direction

- Based on the 3- or 6-day forecasts
- Speed is calculated from integration time ("stepon" timer)

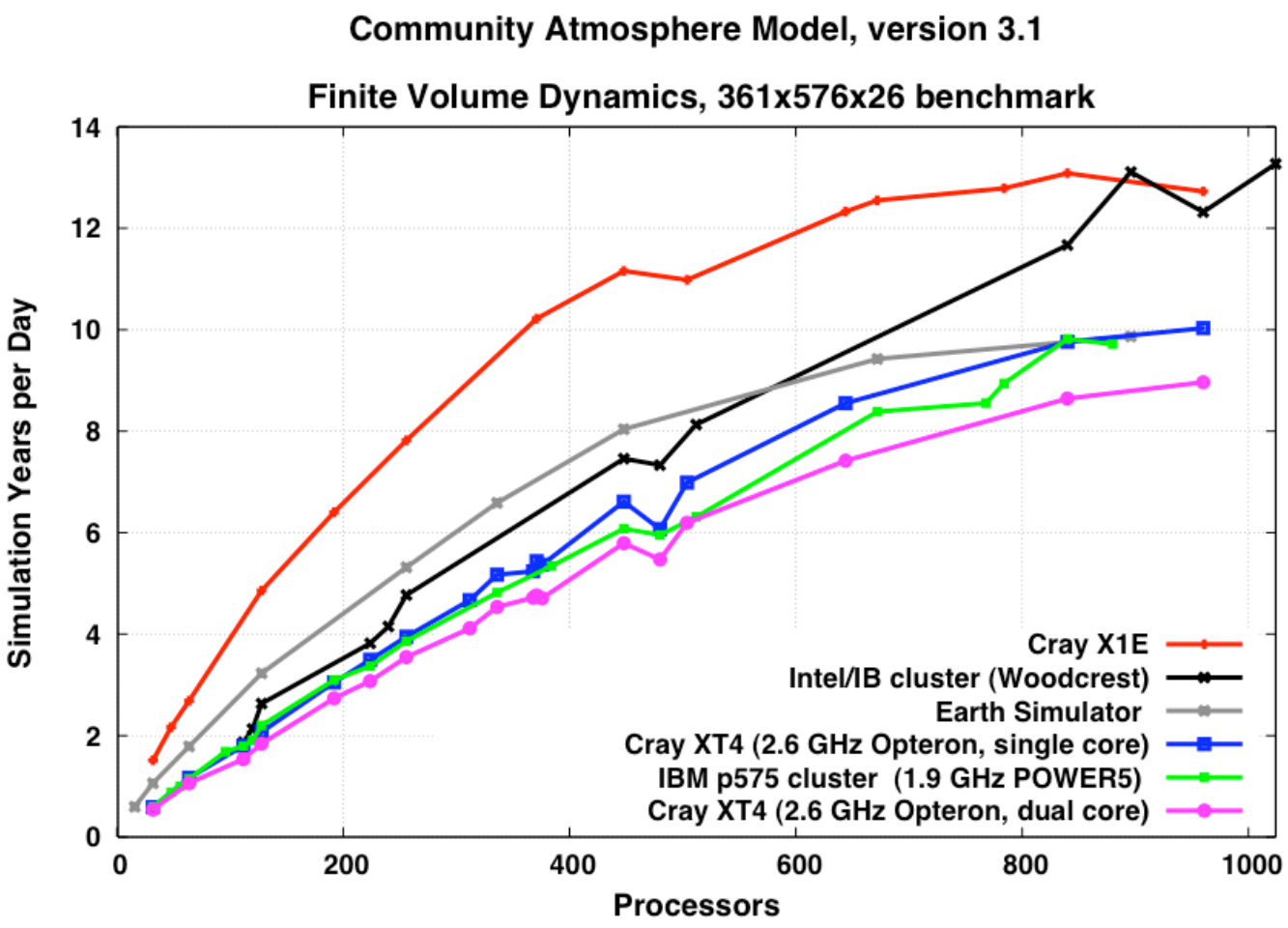
CAM performance



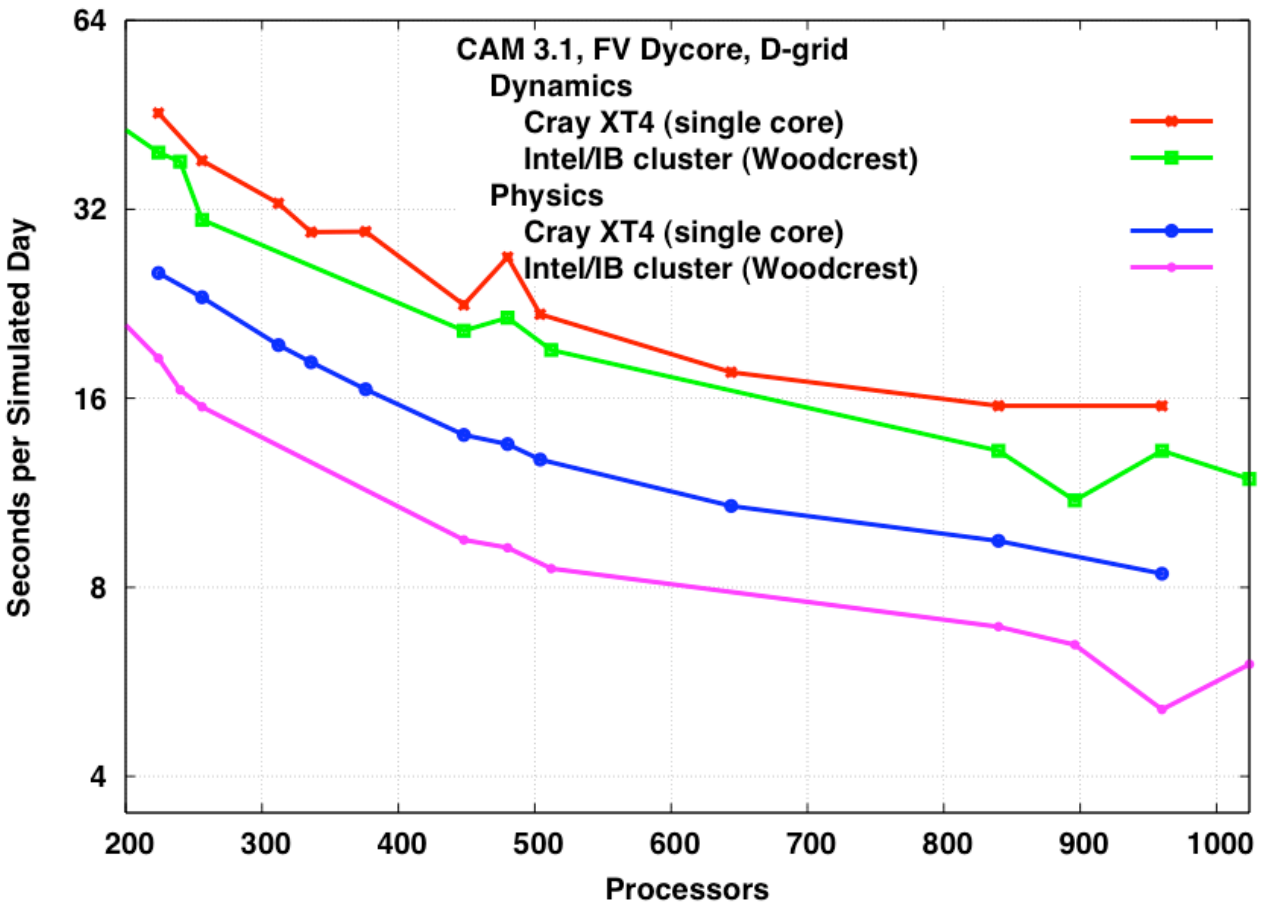
- Drops in the curve are explained by suboptimal decomposition in certain points
- Full load balancing is the best and gives up to 6.5% increase in speed compared to pair wise balancing

- Based on the 3- or 6-day forecasts
- Speed is calculated from integration time ("stepon" timer)

Comparison with other platforms



Performance diagnosis



- Intel cluster outperforming XT4 in both communication-sensitive dynamics and computation-bound physics.
- Intel cluster effectively exploiting OpenMP parallelism, which is not yet an option on the XT4.

Acknowledgements

- The work by Worley was performed under the auspices of the Climate Change Research Division of the Office of Biological and Environmental Research and of the Office of Mathematical, Information, and Computational Sciences, both in the Office of Science, U.S. Department of Energy, by UT-Battelle Oak Ridge National Laboratory under contract No. DE-AC05-00OR22725.
- These slides have been co-authored by a contractor of the U.S. Government under contract No. DE-AC05-00OR22725. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

References:

- A. Mirin, W. Sawyer, L. Oliker, M. Wehner, D. Parks, P. Worley, "*Performance Intercomparison of Community Atmosphere Model on High-End Computing Platforms,"* SIAM Conference on Parallel Processing for Scientific Computing Sir Francis Drake Hotel, San Francisco, CA, February 23, 2006
- P. H. Worley, *"Comparison of Cray XT3 and XT4 Scalability,"* in Proceedings of the 49th Cray User Group Conference, Seattle, WA, May 7-10, 2007