SciDAC-2 CCSM Consortium: Software Engineering Update (*miscellaneous items not covered in other talks*)

Patrick Worley

Oak Ridge National Laboratory

(On behalf of all the consorts)

Software Engineering Working Group Meeting
13th Annual CCSM Workshop
June 19, 2008
Village at Breckenridge
Breckenridge, CO

Other Presentations with SciDAC content

- Software Engineering Working Group Presentations
  - Dave Bader on high resolution studies at LLNL
  - Mark Taylor on HOMME and integration into CAM
  - Ray Loy on PIO
  - Phil Jones on POP2 infrastructure
  - Mariana Vertenstein on CSEG activities, including some funded by SciDAC
- Poster Presentations
  - Art Mirin and Pat Worley on recent scalability improvements in CAM
  - Pat Worley and Art Mirin on recent performance results for the new CAM benchmarks

Acknowledgements

- The work described in this presentation was sponsored by the Atmospheric and Climate Research Division, Office of Science, U.S. Department of Energy. In particular, Worley was supported under Contract No. DE-AC05-00OR22725 with UT-Battelle, LLC.
- These slides have been authored by a contractor of the U.S. Government under contract No. DE-AC05-00OR22725 with UT-Battelle, LLC. 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.
- This work used resources of the National Center for Computational Sciences at Oak Ridge National Laboratory, which is supported by the Office of Science of the Department of Energy under Contract DE-AC05-00OR22725, and of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

SciDAC-2

Two coordinated SciDAC-2 projects that include CCSM software engineering activities:

- SEESM: A Scalable and Extensible Earth System Model for Climate Change Science (Science Application: Climate Modeling and Simulation)
  - Immediate software (and performance) engineering needs, 5 year duration
- PENG: Performance Engineering for the Next Generation Community Climate System Model (Science Application Partnership: Computer Science)
  - More speculative, longer-term performance engineering activities, 3 year duration with possibility of 2 year extension

Project Goals

- Software
  - Performance scalability
  - Performance portability
  - Software engineering
- Model Development
  - Better algorithms
  - New physical processes (esp. chemistry, biogeochemistry)

Remainder of Talk: Odds and Ends

- MCT update
- Brief summary of material in Mirin and Worley posters
- Update on performance analysis capabilities

Backstory in all SciDAC software engineering activities

- *Porting and optimization on Cray XT, IBM BG/L and BG/P, and LLNL Atlas cluster*
- *Functionality and performance bug identification and elimination, especially for new platforms or system software, new configurations and high(er) resolution*

MCT Developments (since last workshop)

- Two major (2.4.0, 2.5.0) and two minor (2.4.1 and 2.5.1) releases have been made. Most recent was May 22, 2008.
- Much development focused on continuing to squash scaling bugs as they appear. These are either excessive memory growth or poor algorithm choices. MCT-based cpl7 now runs routinely on 1000+ processors.
- Major new feature is the ability to handle GlobalSegmentMaps which are not monotonically increasing. Simplifies code in the new coupler.
- Also: Made MCT's Accumulator usage more intuitive. Made it easier to initialize MCT for complex model/processor configurations.
- MCT-based coupler will be included in the CCSM4 release.

Algorithmic Scalability Improvements in the CAM

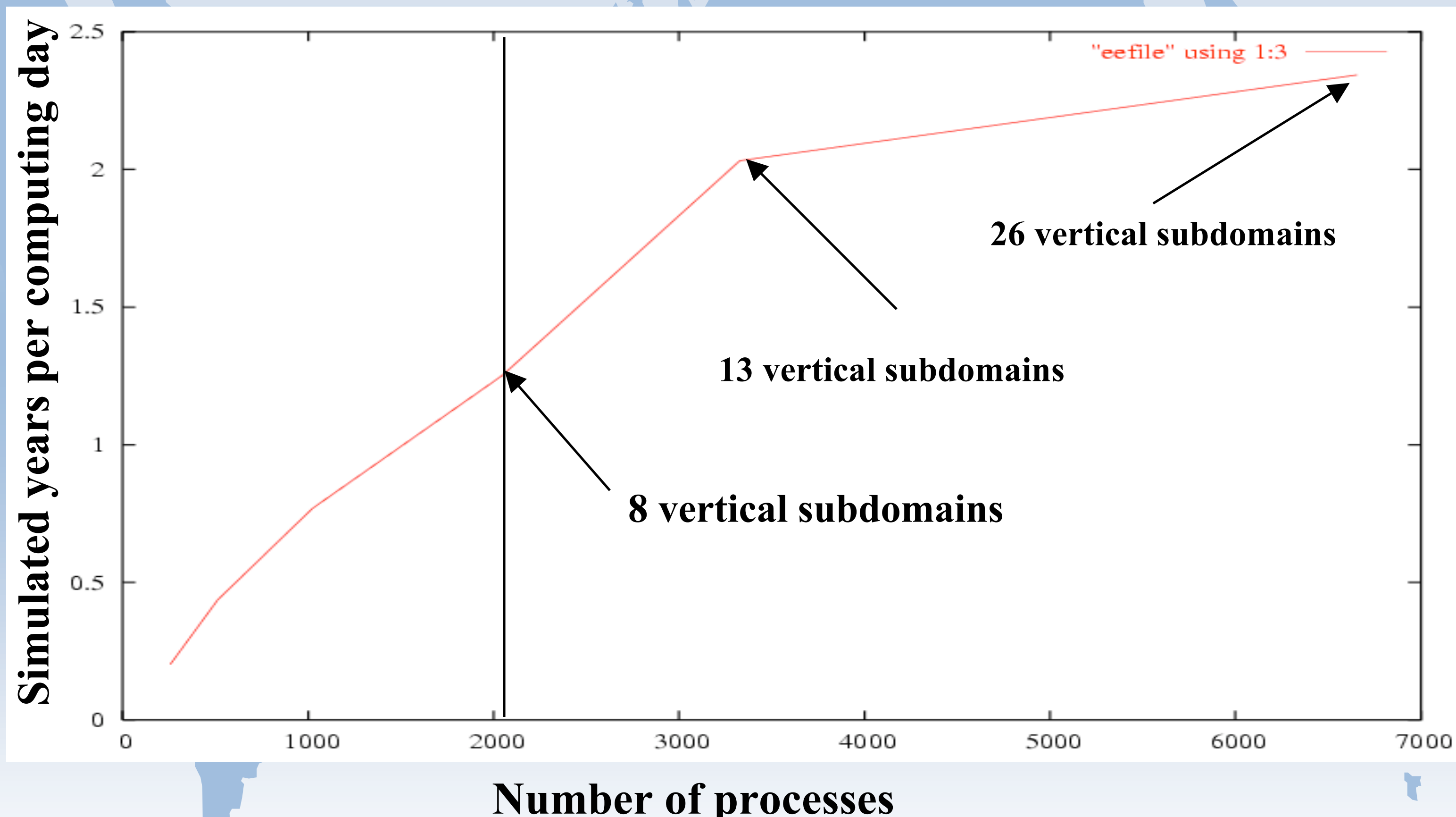
1. Allowing up to 3 times larger vertical decomposition (FV)
   – *useful at all resolutions*
2. Allowing different numbers of active processes for different code sections
   a. inactive processes
   b. larger longitude/latitude than latitude/vertical decomposition (FV)
   c. more physics than dynamics processes, allowing up to 9 times increase in MPI parallelism for FV and PLON times for spectral dycores
   – *currently useful for intense physics, low-to-moderate resolution FV, all resolution spectral. Flexibility will be important for future model developments.*

Algorithmic Scalability Improvements in the CAM

3. Enabling extra MPI processes for tracer advection (FV)
   a. overlapping with main dynamics
   – *useful for moderate tracer count, high resolution*
   b. decomposing over tracer index
   – *useful for high tracer count, all resolutions*
4. Implementing stabilizing terms (Rayleigh friction, additional filtering) to enable tractable operation at 0.25-deg (FV)
   – *not needed at low resolution*

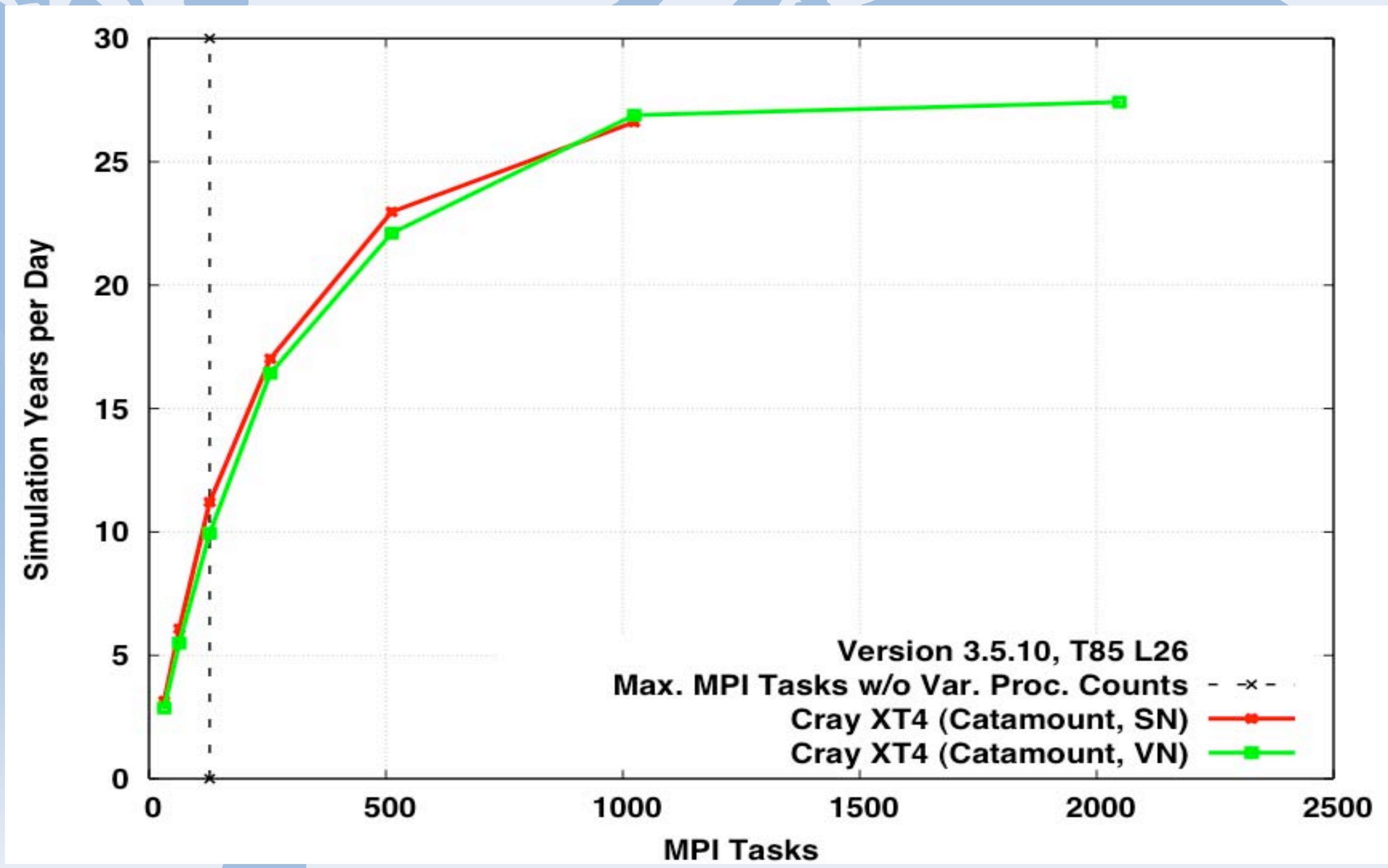
See poster presentation
Scalability Improvements in the Community Atmosphere Model
for details.

Throughput at 0.25-deg resolution



Additional vertical parallelism immediately exploitable in high resolution studies on Atlas system at LLNL.

Spectral Eulerian Performance



Utility of more MPI processes in physics than in dynamics is most obvious when running with spectral dycores. It is also a useful tuning option for FV even with standard physics, and is a clear win with, for example, superparameterization.

New Round of CAM Benchmarking

- Goals of Benchmarking Activity
  - Quantify change in cost of model due to model evolution and to choice of model configuration, examining new science options and problem resolutions
  - Establish new performance baselines upon which to base future performance optimization activities, and with which to evaluate computing platforms.
- Participants in benchmark specification discussions:
  - Phil Rasch (lead), Philip Cameron-Smith, Brian Eaton, Cecile Hannay, Peter Hess, Jean-Francois Lamarque, Art Mirin, Will Sawyer, Pat Worley
- Initial target platforms: Cray XT4 and IBM BG/P, with IBM POWER6 cluster at NCAR and Atlas system at LLNL to be added soon

New Round of CAM Benchmarking

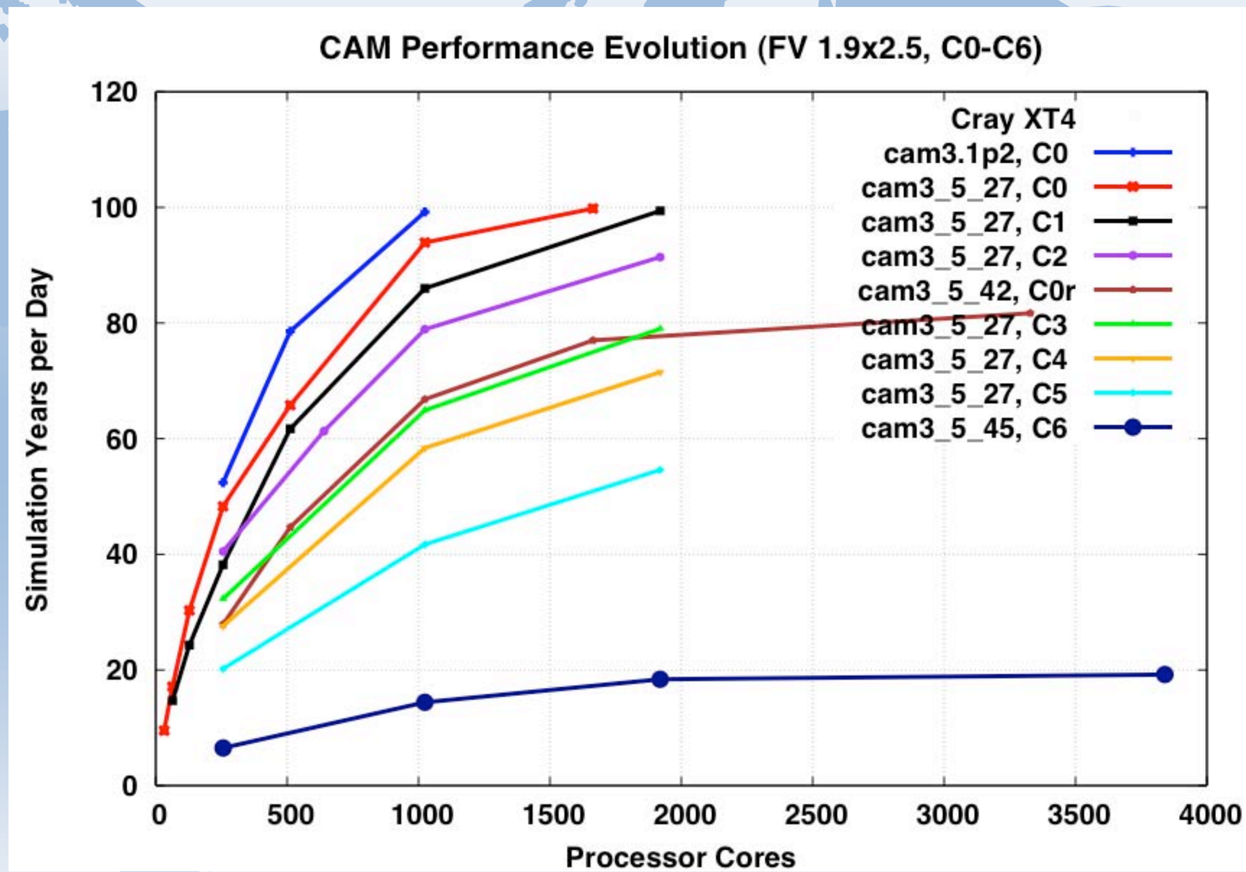
- default configuration (C0): T42L26, T85L26, FV1.9x2.5 L26
  - cam3.0p1, cam3.1p2, cam3_5_27/_45
- C0r: C0 with RRTMG radiation package (FV1.9x2.5 L26 only; cam3_5_42)
- C1: C0 with 30 levels (FV1.9x2.5 only, cam3_5_27)
- C2: C1 with "cam3.5" aerosols
- C3: C2 with UW physics package
- C4: C3 with Morrison Gettelman cloud parameterization
- C5: C4 with predicted aerosol fields
- C6: C4 with full tropospheric chemistry (cam3_5_45)

See poster presentation

Performance Results for the new CAM Benchmark Suite

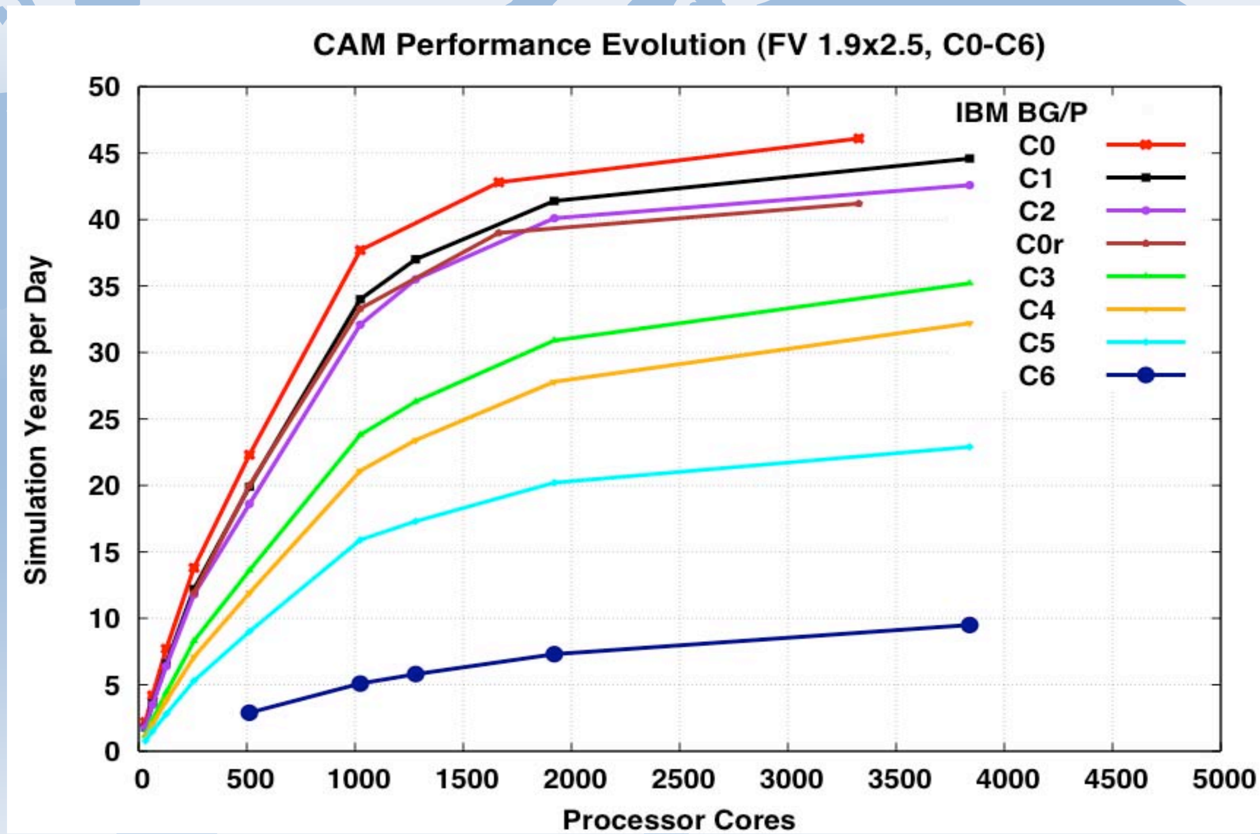
for details.

XT4 FV: C0 vs. C1 vs. … vs. C6



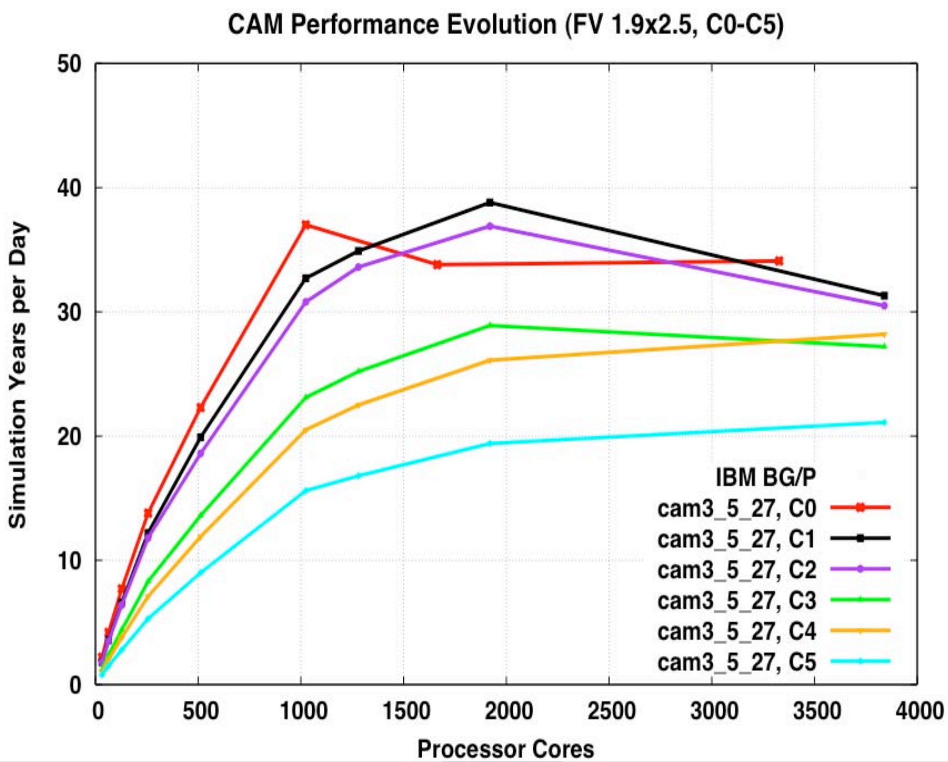
For 1024 processor cores, the normalized cost progression (C0 through C6) is .95 => 1.00 => 1.09 => 1.19 => 1.45 => 1.61 => 2.25 => 6.52. C0r is 1.41 times as expensive as C0. (Note: using enhanced algorithm scalability.)

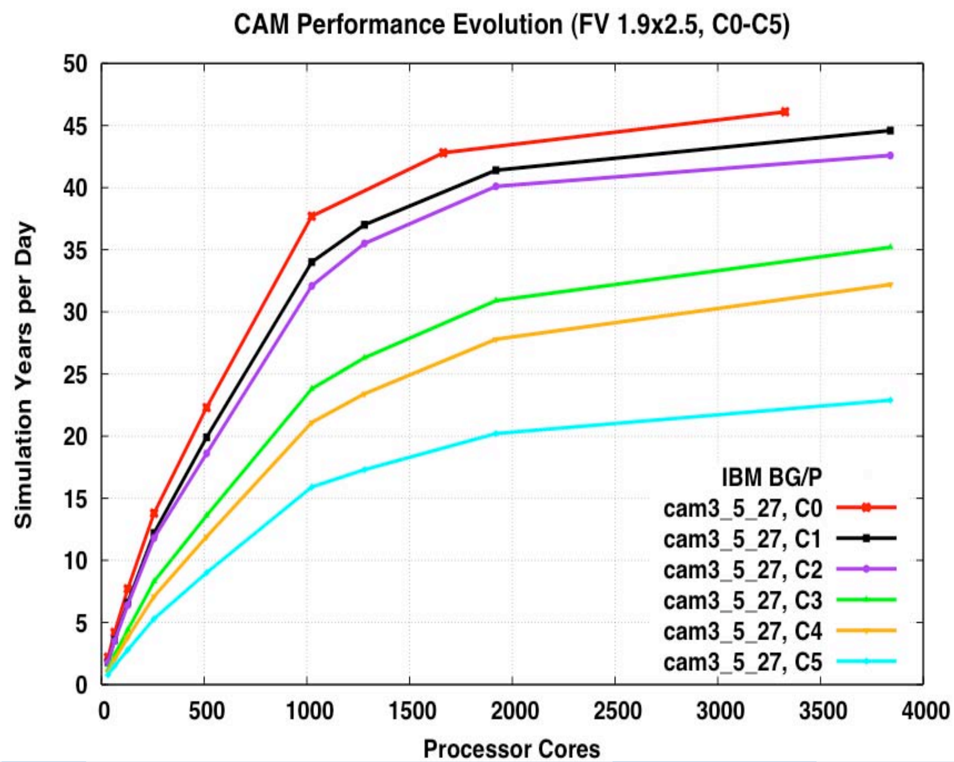
BGP FV: C0 vs. C1 vs. … vs. C6



For 1024 processor cores, the normalized cost progression (C0 through C6) is 1.00 => 1.13 => 1.20 => 1.60 => 1.80 => 2.37 => 7.39, and C0r is 1.13 times as expensive as C0.

Early Impact of Benchmarking



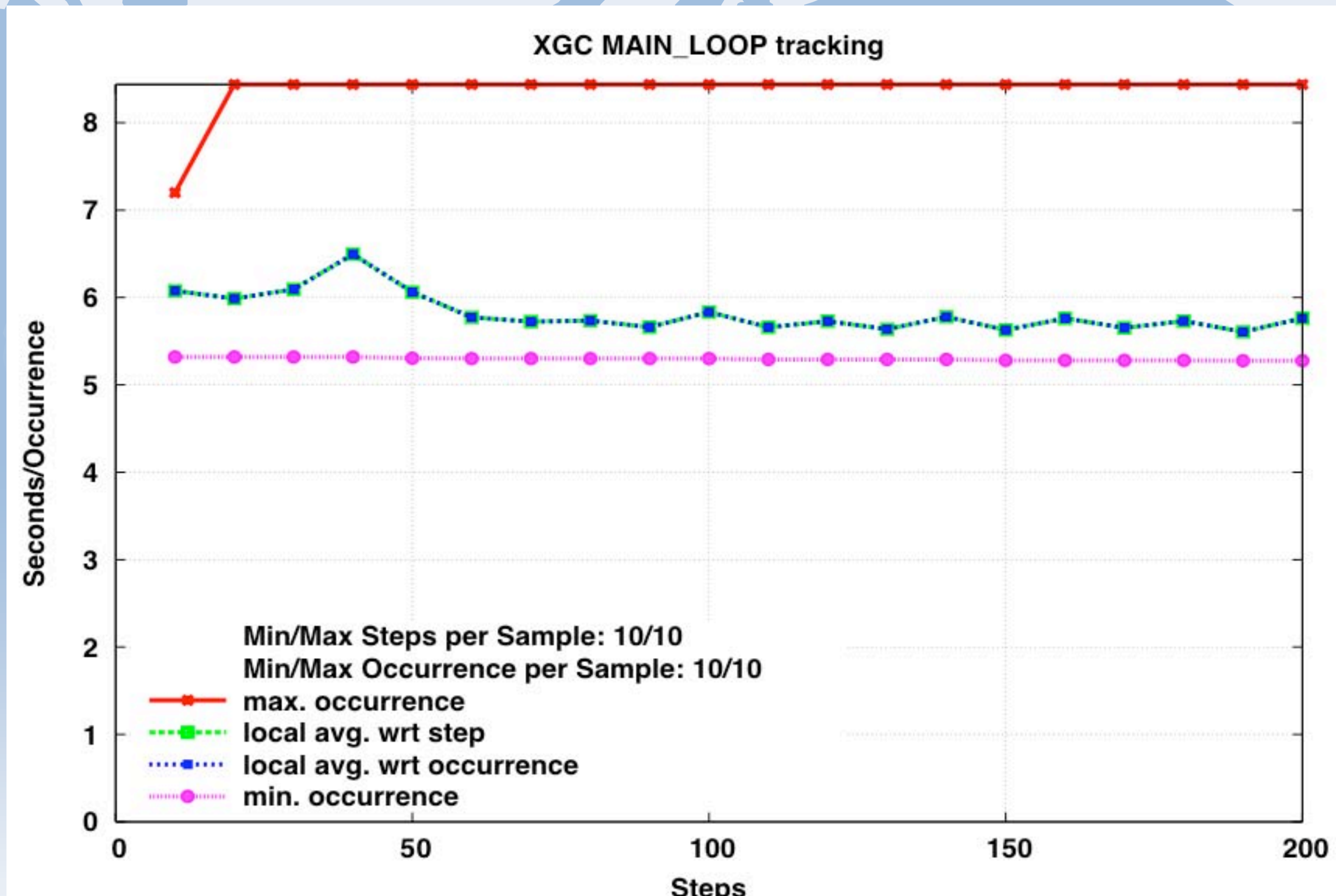


Example performance improvement from the identification and elimination of scalability bottlenecks identified as part of this exercise. Others found when looking at large tracer counts (C6) and high resolution studies (0.47x0.63).

Timing Library Updates

1. Updated to latest version of GPTL library (from Jim Rosinski), adding support for output of global data (min/max across processes) in addition to per process data.
2. Added support to enable and control output of PAPI counters.
3. Added additional support to control amount of data collected and saved (e.g., which processes output data).
4. Added support for writing out timing data periodically during a run, to monitor "performance health".

Timing Library Updates



Example from XGC-1 fusion code (into which CAM timers have been ported) to look at performance variability as a function of simulation time. Same sort of analysis can be generated for CAM timers.

Comments and Caveats

- Not all software engineering-oriented SciDAC activities discussed (in this or in the other talks in this session), including Glimmer, cubed sphere FV integration and evaluation, RRTMG performance optimization, …
- All activities are collaborative or coordinated, e.g.,
  - Parallel I/O (with John Dennis, Jim Edwards, CSEG)
  - Porting to BG/L and BG/P (with John Dennis, Jim Edwards, CSEG)
  - Sequential CCSM (with CSEG)
  - HOMME/CAM integration (with Jim Edwards, CSEG)
  - Benchmarking cubed sphere FV on Cray XT4 (with Chris Kerr, S-J Lin, Bill Putman, Christiane Jablonowski, Peter Lauritzen, CSEG, …)

Questions? Comments?