SEMI-ANNUAL PROGRESS REPORT

Collaborative Design and Development of the Community Climate System Model

Department of Energy
Office of Biological and Environmental Research
Scientific Discovery through Advanced Computing

Principal Investigators
Bob Malone (LANL) and John Drake (ORNL)

Co-Principal Investigators
Chris Ding (LBNL), Steve Ghan (PNNL),
Doug Rotman (LLNL), John Taylor (ANL),
Jeff Kiehl (NCAR), Warren Washington (NCAR),
S.-J. Lin (NASA/DAO)

May 2002

1 Overview

The expansion of the project from Coupler and Atmosphere components (as the ACPI Avant Garde) to all the component models and to scientific development of chemistry and biogeochemistry has been accomplished by the formation of new teams and working relationships involving six DOE National Laboratories (Argonne, Lawrence Berkeley, Lawrence Livermore, Los Alamos, Oak Ridge, and Pacific Northwest), the NASA Data Assimilation Office (DAO), and NCAR. Chemistry, Biogeochemistry and the Land model component represent newly formed technical collaborations with regard to scientific model development, software engineering and implementation. The Ocean component design and development is already well established with strong interactions and participation in CCSM Working Groups, but the software coordination with CCSM is new. Both Ocean and Sea Ice component teams have worked to develop software requirements documents as the first steps towards an evolving CCSM model following an "open" design concept.

Progress in code development in accordance with the proposed designs of the atmosphere and coupler components has resulted in performance improvements to the soon to be released version of CCSM2 and will facilitate modular developments and future performance improvements. Tasks with significant progress discussed in this report are the optimization of the 2-D decomposition of the Lin-Rood (finite volume) dynamical core, the introduction of load balanced 2-D parallel decomposition of the physics calculations and "chunking" for cache performance, and the introduction of a new communications module by the NASA DAO into the PILGRIM utility library that utilizes MPI-2 one-sided communication. Development of the CPL6 coupler has proceeded and new functionality and hardening been incorporated with the Model Coupling Toolkit (MCT) utility. Substantial progress has also been made in several aspects of the ocean and sea ice models. Two approaches to improving the performance of POP have been investigated.

The Project refined its administrative procedures with the development of a Management Plan (included as a Milestone/Deliverable) and the drafting of a Memorandum of Understanding (MOU) with the NASA Earth System Modeling Framework (ESMF).

During this reporting period, the following milestones have been attained, resulting in the indicated deliverables. The third column shows the section in this report where the relevant work is discussed in more detail.

Milestone	**Deliverable**	**Section**
2-D decomposition of DAO finite-volume dycore	Verified code in CAM2.0	2.1
Physics chunking and load balancing	Verified code in CAM2.0 Working Note in progress	2.2
Subgrid precipitation parameterization tested	Refereed publication	2.3
Requirements document for POP drafted	Draft document in review	3.1

Cache-friendly chunking in POP2.0	Beta release of POP2.0	3.2
MLP implemented in POP	Alpha version of POP1.4.4	3.3
HYPOP tested in Lagrangian limit	Draft document in review	3.4
MCT now supports PCM-like coupling	Expanded version of MCT	4.1
New MPH (multi-processor handshaking) library	MPH3.0 available	4.2
MPDATA replaced by incremental remapping	New CICE3.0, 30% faster	5.1
Requirements document for land model	Draft document in review	6.1
Chemistry solver interface defined	Proposed interface	7
Management plan completed	Plan document http:/www.scidac.org/CCSM	9.1
MOU between SciDAC and NASA ESMF	Approved MOU document	9.2

2 Community Atmosphere Model (Topic Coordinator: John Drake)

2.1 Lin-Rood Optimization (Art Mirin)

Work centered on the two-dimensional decomposition of the DAO finite-volume (FV) dynamical core. All known aspects of CAM now work with the 2-D decomposition. The latest to be converted was the history module (both 64-bit and 32-bit accumulation buffers and the initial history file).

Our main area of work has been performance optimization of the FV dynamical core, with the order of attack being (1) MPI performance of the 1D (latitudinal) decomposition, (2) MPI performance of the multi-2D decomposition, and (3) OpenMP performance. We have been working with Will Sawyer of NASA/DAO on the 1-D decomposition. This has included implementation of DAO's mod_comm library (which contains lighter-weight versions of selected Pilgrim routines), and addition of ghost points to some of the prognostic arrays (aiming to reduce data buffering). Aside from being lighter-weight, mod_comm supports one-sided MPI-2 communications; this has resulted in significant performance improvement on the SGI but has had no measurable effect on the IBM. We have been comparing dynamical cores in CAM and FVGCM (DAO's version of the code), in an attempt to form a unified core with the best features of each. This work is partially complete.

The motivation for revisiting OpenMP performance is that with the 2-D decomposition, one of the decomposition directions (vertical) is the same as the primary OpenMP direction, thereby reducing the degree of attainable OpenMP parallelism. For the tracer advection, we find that combining the tracer index and vertical level into a single loop is a way to provide increased OpenMP parallelism.

We have been regularly validating the 2-D decomposition as the code has evolved. This has included regular interaction with Pat Worley on the physics chunking, where a main aim has been to improve load balance.

2.2 Physics Chunking and Load Balancing (Pat Worley)

At the end of the Avant Garde project, the physics and dynamics had been split and a prototype implementation of the chunking incorporated. Performance measurements showed that the optimal chunk size was different on each machine and that performance gains were possible by setting the chunk size to less than 64 (the default setting for a T42 latitude slice). For the IBM SP3, the optimal size is 16-32 while for the Compaq it is 8-16. But with smaller chunk sizes, we have an extra degree of freedom for parallelization. Several options were implemented for assigning chunks "on processor" as inherited from the dynamics, for using more processors than used for the dynamics, and for load balancing the chunks. The implementation of load balancing the physics calculations (especially the short wave) was taken as a high priority due to the potential benefit to performance. Pat Worley implemented this and the modifications have been incorporated on the CAM development branch. A Working Note is being written to document the performance improvements.

2.3 Subgrid Orography Scheme (Steve Ghan and A. G. Hunt)

Prior to the SciDAC project, the Leung and Ghan subgrid orography scheme had been applied to CCM3.10 and 10-year AMIP simulations had been run with and without the scheme. With SciDAC funding a journal article describing the application and performance of the scheme was prepared, submitted, and successfully revised for publication in *Climate Dynamics*. A design document outlining the application of the scheme to CCSM was prepared, and a CVS branch for applying the scheme to CCSM was created. The scheme has recently been applied to the first (CAM1 atmospheric physics, CLM2 land physics, Eulerian dynamics, no ocean coupling, block domain decomposition) of a variety of CCSM configurations, and multi-month simulations were performed. With this application serving as a template, application to most of the other configurations (semi-Lagrange and finite volume dynamics, ocean coupling) will be relatively straightforward. However, further load balancing will be more difficult. AMIP-style simulations at a variety of horizontal resolutions are planned when testing of the configurations is complete.

2.4 Parallel I/O Library for CAM (Y. Wang and C. Ding)

Developed and tested a parallel I/O method by remapping distributed arrays to/from designated I/O processors for writing/reading disk files efficiently. A core set of I/O functions to perform remapping, netCDF and direct access I/O are written and tested. The existing I/O mode is (i) gathering everything to Proc 0, (ii) switching indices (k and j), and (iii) writing to disk file. In comparison, the new parallel I/O library speeds up by a factor of 10 for 128MB array (http://www.nersc.gov/research/SCG/acpi/IO/index.html contains details). Currently, the library is being integrated into CAM I/O.

A parallel netCDF for IBM SP is being developed by NERSC (Majdi Baddourah). We are actively debugging and evaluating the codes in various domain-decompositions. The measured parallel I/O rates and serial netCDF I/O rates on the IBM SP at NERSC are online: http://www.nersc.gov/research/SCG/acpi/IO/index.html. Currently the reading

rates of parallel netCDF is about 100MB/s, but the writing rates is only about 10MB/s at 64 MB size files and about 70MB/s for large size files. In comparison, sequential netCDF rates are over 120 MB/s for read and 100 MB/s for write. Fortran direct-access writes using GPFS are quite fast, reaching over 500 MB/s. MPI-OI rates follow the GPFS write rates with some overhead penalty. We are also coordinating the parallel netCDF with ANL and Northwestern University.

3 POP ocean model (Topic Coordinator: Phil Jones)

3.1 Software Engineering for POP and CICE (Phil Jones and Elizabeth Hunke)

Many changes are planned during this project for the POP model, including hybrid vertical coordinates, increased emphasis on biogeochemistry and participation in software frameworks like ESMF. It will be important during this process to follow standard software engineering principles to ensure the model satisfies all the new demands. Towards this end, a requirements document for the new ocean model (HYPOP, see section 3.4) has been drafted to include the needs of the new vertical coordinate and especially the requirements for supporting biogeochemical models. The current document is undergoing review by relevant users. A similar document for the CICE model is in progress. Work on architecture documents has also been initiated and design documents will follow the successful review of the requirements documents.

3.2 Ocean Model Performance (Phil Jones)

The current POP model (version 1.4.3) has been re-designed (version 2.0) to achieve better performance on clusters of SMPs, while still maintaining adequate performance for vector processors. The new design involves a decomposition of the computational domain into blocks that can be sized to fit into cache or to permit long vectors. These blocks can be distributed among processors in a manner that balances the computational load, while eliminating blocks that contain only land. Distributing many blocks to each SMP node also permits a hybrid parallelism using shared memory (OpenMP) to communicate between blocks on a node and message passing (MPI) to communicate with blocks on other nodes.

Performance of a global 0.1° simulation with the new decomposition scheme was improved by a factor of two over the currently-released version of POP on SGI Origin 2000 machines. The improvement was a combination of improved cache performance, better load balancing, and land point elimination. Performance on an IBM SP (Winterhawk II) improved by at least 25% for the same problem size, and further experimentation with block sizes and hybrid threading/message-passing is currently in progress to find the appropriate balance on this machine. As mentioned above, block sizes can be increased to achieve vector performance and some preliminary results on Fujitsu vector computers have shown an ability to reach 50% of peak performance.

Performance of the model at climate resolutions (~1°) has thus far not improved significantly. Coarse grids do not offer as much opportunity for land point elimination and already result in small domain sizes on moderate processor counts. At such resolutions, alternative low-latency communications like MLP should provide much better performance improvements.

3.3 MLP in POP (Jim Taft).

The MLP-based effort to improve POP scaling used POP 1.4.3 as the starting point because it is the latest model in wide distribution. The initial scaling effort focused on further optimizations of the MPI version of the code. Reductions in extraneous copies of arrays and some cleanup of boundary condition exchanges achieved approximately a 15% speedup over the original code on 128 Origin-3000 CPUs for the 1° problem.

After a careful analysis of the code, it was determined that the barrier synchronization activity between processes consumes nearly 40% of the run time for the 1° problem when running on 512 Origin-3000 CPUs. This includes the explicit barriers called by the user, as well as those implicit in various other MPI operations. There is a direct approach to solving this problem, which is to reduce the number of active MPI processes and/or reduce the overhead of the barrier-synch process. The cost of the barrier synchronization goes down dramatically with a reduction in the number of heavyweight MPI processes.

As a result of this, conversion of POP 1.4.3 to MLP was undertaken. MLP has significantly less overhead in communication and barrier synchronization than its MPI analogs. A conversion would be expected to significantly increase scaling and overall performance.

From the user's perspective, the barotropic computation is the scaling bottleneck for climate resolutions (1°) and high CPU counts (>128). This calculation can quickly dominate run time on a 512-CPU Origin system for the 1° problem. The reason is due to its frequent exchange of boundary condition information with little intervening work in this 2D computation. MPI latencies, even on a fast shared-memory system like an Origin are too large to allow significant scaling to occur. Thus, the initial work in the MLP conversion was focused on improving the scalability of this computation. The barotropic computation was isolated into a single MLP process which is fine-grained parallelized using OpenMP loop-level parallelism. As a result, run times are now fixed and fast for this simple 2D computation as the CPU count for the remainder of the more scalable work is increased.

The results of the MLP scaling work to date are shown in Figures 1 and 2. Figure 1 shows results for POP1.4.3 and POP1.4.3-MLP for the 1°-resolution problem when executed on a 600-MHz Origin 3000 system. This problem uses a grid of 360x180x40 cells for a total of 2.6 million points. A total of 120 time steps were executed. The chart is a plot of total run time versus CPU count. The figure shows about a 20-40% range in improvements for the MLP code out to 512 CPUs. The barotropic contribution reaches about 25% of the run time for 512-CPU run.

Figure 2 presents results for the high-resolution 0.1° problem (3600x2400x40 cells or 345 million points). The MLP code shows a much greater improvement over MPI in this case (almost a factor of two).

Further significant run time reductions are expected with the adoption of the hybrid OpenMP/MLP approach. When coupled with other single CPU optimizations, another factor of two could be realized in the short term.

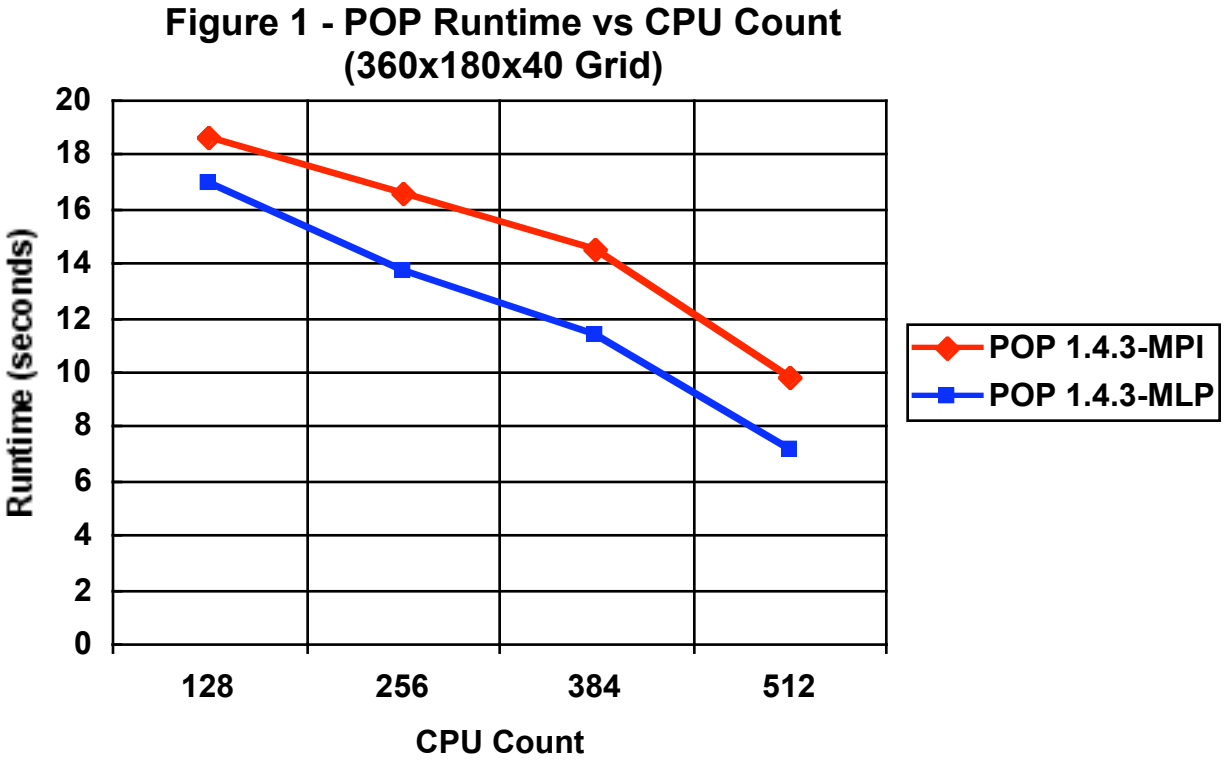
Figure 1 - POP Runtime vs CPU Count
(360x180x40 Grid)
Runtime (seconds)
20
18
16
14
12
10
8
6
4
2
0
128
256
384
512
CPU Count
POP 1.4.3-MPI
POP 1.4.3-MLP

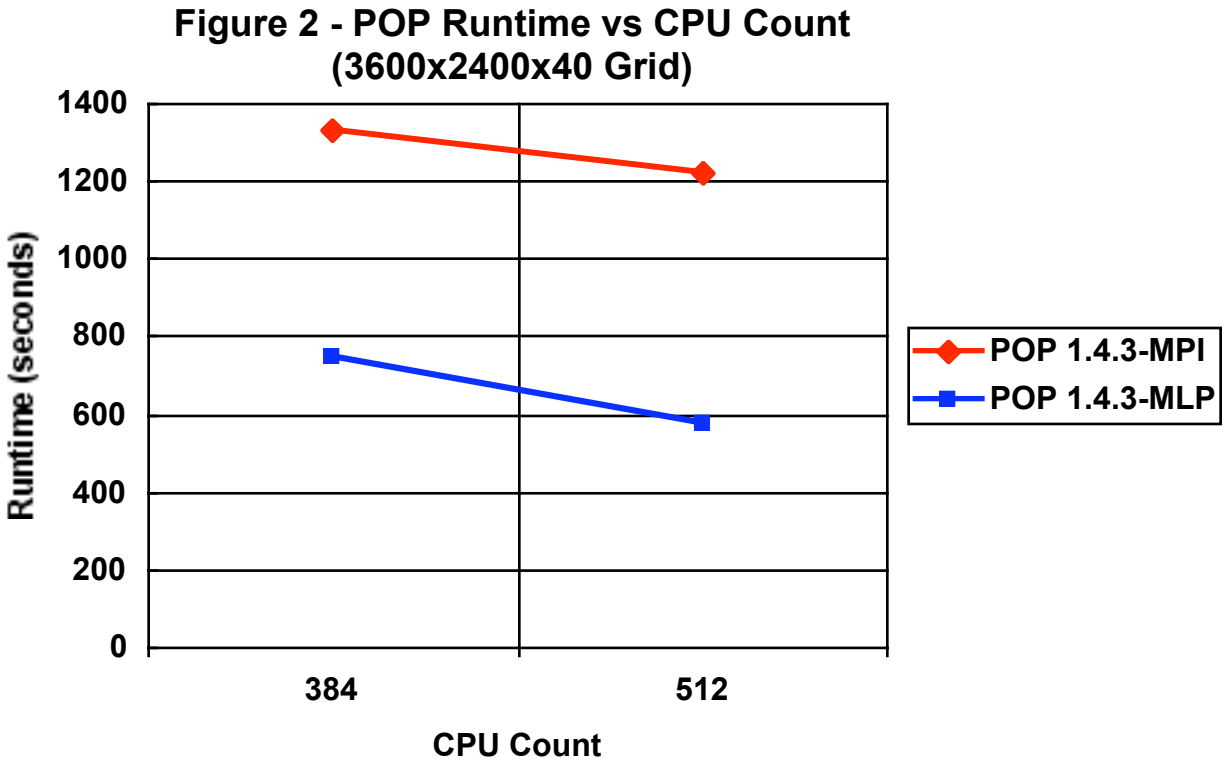
Figure 2 - POP Runtime vs CPU Count
(3600x2400x40 Grid)
Runtime (seconds)
1400
1200
1000
800
600
400
200
0
384
512
CPU Count
POP 1.4.3-MPI
POP 1.4.3-MLP

3.4 HYPOP Model Development (John Dukowicz, John Baumgardner, Bill Lipscomb, and Kirk Bryan)

The most fundamental aspect of the HYPOP model is its ability to treat in a robust manner the purely Lagrangian dynamics of constant-mass layers as they inflate and deflate in regions of intersecting bottom topography. Most of the effort over this report period has been in identifying and evaluating a set of algorithms that successfully achieve this crucial goal. We believe we have succeeded in this objective. Following is a summary of the approach we have adopted.

First of all, we perform a splitting of the non-Boussinesq primitive equations into a barotropic part that corresponds to the fast external-gravity wave dynamics and a baroclinic part that corresponds to the remaining slower dynamics. The defining principle in making this split is that the baroclinic mode should have no dynamics arising from the pressure when the density is uniform. To accomplish this, the pressure gradient is split into a 'baroclinic' part that vanishes and a 'barotropic' part that does not vanish when the density is uniform. Because ocean stratification is relatively weak, the baroclinic part of the pressure gradient is relatively small and can drive only slow dynamics. The resulting barotropic momentum equation is simple and involves only a Coriolis term and a term involving the gradient of the free surface height.

Our horizontal spatial discretization uses a C-grid, with velocities on cell faces and masses, tracers, and layer heights located at cell centers. Pressures are defined on layer interfaces but aligned vertically with the cell centers. Our time discretization employs an adaptive third-order Runge-Kutta method to advance the fast barotropic equations in a subcycled manner over a baroclinic interval. This Runge-Kutta scheme dynamically finds the appropriate barotropic step size, is self-starting, and satisfies a simple form of energy conservation. The time stepping of the baroclinic system utilizes a second-order-accurate staggered forward-in-time method, in which the baroclinic velocities are staggered by a half step relative to the barotropic cell-mass values.

Although both the barotropic continuity equation and the baroclinic layer continuity equations utilize perfectly conservative transport schemes, ensuring consistency in the column masses between the barotropic and baroclinic solutions is crucial, but nontrivial. Our approach applies the vertically-averaged baroclinic velocities at the middle of the barotropic interval in the solution of the barotropic continuity equation and the time-averaged barotropic velocities at the middle of the baroclinic interval in the solution of the baroclinic layer continuity equations. To achieve a high degree of consistency in column masses, the time-integrated barotropic face mass fluxes are compared with the vertically-integrated baroclinic face mass fluxes after updating the baroclinic continuity equations. Vertically uniform face correction velocities derived from the face flux differences are used to advect small amounts of cell mass to yield column mass consistency. The only regions where such consistency is not achieved to a high degree of accuracy is where the topmost layer thins to zero against topography. But these small discrepancies are all but eliminated by initializing the subsequent barotropic interval with the surface height field obtained by summing the baroclinic layer thicknesses. (This relatively clean and simple strategy yields a tight coupling and high degree of consistency between barotropic and baroclinic modes.) This strategy redistributes the mass

discrepancy within layers, rather than across layers, thus preserving conservation within layers and avoiding spurious diapycnal mixing across layers.

Another aspect of our approach relates to highly accurate mass and tracer transport. We have established that the 2-D incremental remapping method developed originally for discretizations with velocities at cell corners (i.e., B-grid discretizations) can be generalized to yield the same highly accurate performance when velocities are specified at face centers. Incremental remapping is a method with some outstanding properties for computing the conservative transport of mass and tracers. Not only is it conservative, positivity-preserving, and monotonicity-preserving, but also it naturally preserves the compatibility property of tracers and is very efficient for large numbers of tracers because the additional cost per tracer is low. Therefore, our overall approach for HYPOP includes use of this 2-D incremental remapping method for solving the layer continuity and tracer transport equations but a considerably less expensive split 1-D remapping scheme for the barotropic continuity equation. Because the barotropic time steps are small relative to the baroclinic step size and the vertically averaged velocities tend to be modest, the splitting errors in the inexpensive split scheme tend to be small and acceptable. Hence we believe we have an optimal strategy relative to accuracy and cost for the advection aspects of the HYPOP model.

Our focus thus far has been to ensure we have a robust, accurate, and efficient formulation that works in the Lagrangian limit. Our next focus will be to test this set of algorithms in the Eulerian limit where we remap in the vertical direction to an original Eulerian grid every few baroclinic time steps. After we verify HYPOP is this regime, we will have a relatively high level of confidence that our Arbitrary Lagrangian-Eulerian (ALE) approach is indeed ready to handle the regime of practical interest in which deep ocean is treated in a nearly Lagrangian manner and the surface mixed layer is treated in a nearly Eulerian way. For this a practical algorithm will be needed for specifying the vertical grid that distributes the available layers such that the grid is essentially isopycnic (follows "orthobaric" density surfaces) below the mixed layer and is uniformly distributed within the mixed layer. We plan to use the KPP mixed-layer parameterization to specify the location of the mixed layer. Kirk Bryan has been developing and testing a version that explicitly constructs the grid from the top and the bottom and blends the two. This summer we plan to try an alternative that uses a variational algorithm for redistributing the layers within the mixed layer and isopycnal region based on the variation of the KPP diffusivity.

4 Coupler Development (Topic Coordinator: Jay Larson)

4.1 Model Coupling Toolkit (Jay Larson and Rob Jacob)

MCT work in the first half of FY02 focused on software quality, increased functionality and performance improvements. In the first category, we have found workarounds or otherwise fixed Fortran90 incompatibilities regarding pointer status and improved the handling of MPI versus other runtime errors.

Increased functionality represents the largest category of work and includes 12 new functions to retrieve subsets of data from MCT Attribute Vector and General Grid data types and additional communication functions for the GlobalSegMap. With a new

module, Rearranger, MCT can now support two forms of general data transfer between parallel components. The first is between components that share no processors and the second is for transfer of data between components that share all or a portion of their processor space. The Rearranger module and some changes to MCTWorld initialization provide complete support for PCM-style coupling.

Finally we have increased the performance of the MCT Router initialization (with assistance from Jace Mogill of Cray Research) as well as the cache performance of MCT's matrix-vector multiply routines and inter-component data transfer (MCT_Send and MCT_Recv).

We continued our collaboration on the development of CPL6 with NCAR. We maintain a separate copy of MCT software within the CPL6 CVS repository at NCAR and have performed conflict resolution and any necessary debugging as each new release has been checked in. We have also participated in the continuing design of CPL6 architecture and datatypes.

An active collaboration between the Model Coupling Toolkit developer Jay Larson and the Common Component Architecture developer Wael Elwasif has resulted in a plan to prototype the MCT Attribute Vector module as a CCA component. Since the MCT is a F90 implementation and the CCA framework is C++, the interfaces must be translated through an intermediary function. The interface definitions have been written and issues with inter-language operability have been clarified.

Jay Larson went to SNL on 26 Feb to give a talk on the Model Coupling Toolkit and discussed collaborations involving CCA and possible uses for the MCT outside of the climate community. In particular, there was some interest from the CRFRS SciDAC application center.

In other SciDAC related activities, we have begun collaborating with the Center for Component Technology for Tera-scale Simulation Software ISIC. Collaborative efforts include attending two Common Component Architecture workshops and exploring the recasting of MCT into CCA components. We have also participated in the requirements process of the Earth System Modeling Framework through attending meetings and teleconferences.

4.2 MPH (multi-processor handshaking) library for coupling component models (Y. He and C. Ding)

MPH3, the hierarchical integration of multi-component executables, was completed. All codes have been modified according to NCAR guidelines, ProTex was added, and a complete manual has been written. MPH2 is currently used in CCSM2. Dave Randall's team at Colorado State is currently using MPH for their coupled model as well. We've had a few email exchanges with Don Dazlich of Randall's group regarding MPH. See http://www.nersc.gov/research/SCG/acpi/MPH/ for information.

4.3 CPL6 Development (Tom Bettge, Tony Craig, and Brian Kauffman)

CPL6 is the high-level application which uses MCT as the foundation for distributed, parallel coupling functionality.

During the first half of FY02, development of CPL6 progressed along several fronts. High-level CPL6 datatypes were identified and defined. These so-called bundles are

built upon the basic datatype (attribute vector) defined in MCT. The interface to the component models was finalized. A general "dead" component was built for prototyping. Several user level modules, which replicate the functionality of CPL5, were identified, created, and tested – including messaging, history writing, and mapping modules.

Initial timing of the message passing and mapping functions of CPL6 has been performed. As a result, additions, modifications, and enhancements to MCT have been made in close collaboration with the MCT team.

5 Sea Ice Model (Topic Coordinator: Bill Lipscomb)

5.1 Incremental Remapping for Sea Ice and Ocean Transport (Bill Lipscomb, Elizabeth Hunke, and John Dukowicz)

Incremental remapping is an efficient, second-order accurate, monotonicity-preserving scheme for horizontal fluid transport. It was originally developed at LANL by John Dukowicz and John Baumgardner (2000). During the past several months we have modified and enhanced the incremental remapping scheme for use in the Los Alamos sea ice model, CICE, and the new ocean model, HYPOP.

CICE previously used the MPDATA scheme for transport of conserved quantities: area, volume, and internal energy of ice and snow. When the model was upgraded to allow for multiple ice thickness categories and a better-resolved vertical profile, MPDATA proved inefficient, and transport became the most expensive model component. We therefore implemented an incremental remapping scheme that proved to be three times faster than MPDATA in the standard configuration, giving a total model speedup of about 30 percent. Several innovations were needed to use this scheme for sea ice. For example, we developed methods for constructing and integrating cubic polynomials (the original scheme was limited to quadratics) and for transforming velocities between different coordinate systems on a non-rectangular grid. Incremental remapping is now the standard transport scheme in CICE Version 3.0 and is being proposed for the NCAR Community Sea Ice Model. A paper (Lipscomb and Hunke, 2002) is being submitted to Monthly Weather Review that describes the scheme and demonstrates its advantages relative to the MPDATA and donor-cell schemes.

5.2 Sensitivity analysis and parameter tuning test of the CICE code (Jong Kim)

We have conducted sensitivity studies of the Los Alamos sea ice model, CICE, based on Automatic Differentiation (AD)-generated derivative code. The numerical experiments were performed for the simulation of 1997 at 2-hour intervals. The comparison of the AD-computed derivatives with the results calculated by the finite-difference (FD) method indicates that AD provides accurate sensitivities for the multivariate CICE model. The major modeling parameters to control the sea ice thickness computation were the ice-albedo constants, densities and emissivities of ice and snow, and salinity constant. The AD-processed CICE code was further used to provide the gradient information in the parameter-tuning experiment. The preliminary results, using a bound-constrained minimization method, L-BFGS-B, with simulated observational data show that satisfactory convergence with the tolerance of 1.0E-15 was obtained about within the 28 minimization iterations, which required about 39 functional computations.

6 Land Surface Model and River Transport Model (Topic Coordinator: Forrest Hoffman)

6.1 Land Model Development Activity (Forrest Hoffman and Marcia Branstetter)

A software requirements document was initiated. At the same time the CLM2 was undergoing a major rewrite to accommodate new process modules. SciDAC software engineering is focused on the interface to the atmospheric model, through the physics module or through the coupler. The elimination or reduction of parallel gather-scatter operations and the introduction of load-balanced parallel distributions are the initial target. These improvements should provide performance gains while not disrupting the rapid development of the CLM2.

A draft Community Land Model (CLM) Requirements Document (http://clm.ornl.gov/) was developed by Hoffman and Mariana Vertenstein (NCAR) with significant input from Branstetter, Gordon Bonan, and Peter Thornton. The draft was presented to the Land Model Working Group (LMWG) at its spring meeting in Boulder for consideration and review.

Hoffman and Taylor will serve on a code review committee for the new version of CLM, tentatively called CLM2.1. The committee is headed by Peter Thornton and consists of a total of 10 members from both the BGCWG and the LMWG. The committee will review and accept or reject the new architectural changes to the data structures.

The River Transport Model (RTM) is also being reviewed and redesigned. In the current CCSM2, the RTM is a single- processor implementation called from the land model. A new data structure with a direct interface to the MCT attribute vector methods will allow use of the parallel sparse matrix utility for a distributed parallel implementation of the transport of chemical species as well as water along the river network. In addition, the encapsulation of the RTM will allow alternative coupling strategies. Marcia Branstetter (ORNL) is working on these options.

7 Atmospheric Chemistry (Topic Coordinator: Doug Rotman)

Atmospheric Chemistry is being implemented into the CAM atmospheric model through a collaboration with LLNL (Rotman, Cameron-Smith, Connell, Tannahill) and NCAR/ACD (McKenna, Walters). In addition, John Taylor at ANL and David Erickson are involved in module development and analysis. This work is being done as a complement to the ongoing WACCM effort at NCAR (a combined CGD and ACD effort). Chemistry work builds upon the off-line chemistry models at LLNL (IMPACT model) and NCAR (MOZART model).

FY01/02 Milestone: Identify optimal location and implement initial software interface for chemistry solution in the CCSM

The chemistry solver interface was defined and the location for calling the solution package was also identified. The solver's interface is as follows,

Call Chemistry (species_distribution, pressure, temperature, cloud_water, cloud_fraction)

The call to this solver occurs in the column physics package of the CCSM, specifically from the routine, TPHYSAC.

FY02 Milestone: Identify and implement an initial chemistry solver and (reduced) chemical mechanism into the CCSM

An initial chemistry solver has been identified and is in the process of implementation into the CCSM. The solver is that used in the NCAR MOZART model and implementation is being done by the ACD Division at NCAR. At LLNL we have carried out a steady state simulation using a comprehensive tropospheric/stratospheric chemical mechanism in an off-line chemistry model (the LLNL IMPACT model). Distributions are now being evaluated. Importantly, we have completed an assessment of process budgets. This analysis has identified the tropospheric ozone burden and how the burden is influenced by advection, diffusion, convection, scavenging, and chemical reactions. Concurrent with this simulation, we have identified an initial ~20 species chemical mechanism (see details in the Appendix) focused on tropospheric ozone. This 20 species reaction set does not include all those species and reactions that are known to be important to tropospheric ozone. However, with the existing complete simulation, and the analysis of budgets, we will refine this mechanism to produce proper distributions and radiative forcings for quality, yet efficient climate-relevant chemistry. Over the next months, a steady-state simulation will be carried out using the simple mechanism for comparison to the comprehensive simulation to assess its strengths and weaknesses.

John Taylor (ANL) is working with LLNL on the analysis of global atmospheric methane and carbon monoxide, including an analysis of their respective distributions and fluxes. We are initially performing the analysis of the off-line model simulations performed using the LLNL IMPACT model, performed with more complex chemistry than can be efficiently included in CCSM. This analysis will provide a base case for future analyses as we simplify the chemical mechanism. We will be performing continuing analyses of the distributions and fluxes of methane and carbon monoxide, and their role in influencing the distribution of ozone in the troposphere and stratosphere, as the chemistry solver and chemical mechanism are developed and incorporated into CCSM.

The Marine Aerosol-Gas Phase Interactions (MAGPI) model has been extended to include reactive bromine chemistry and has entered the diagnostic and validation phase. Taylor (ANL) has begun working with David Erickson, Marcia Branstetter and Bill Keene of University of Virginia on the development of the marine aerosol and gas phase interactions MAGPI code. Taylor has completed an initial code redesign which included making the code FORTRAN 90 compliant, breaking the code into subroutines representing the functionality of the code, adding Protex headers and reformatting the code consistent with the NCAR documentation requirements, creating a CVS repository and adding the new code to the repository, and building an automatic installation and test script for the MAGPI code.

8 Biogeochemistry (Topic Coordinator: David Erickson)

Ocean biogeochemistry is the newest topic to be addressed by all of the participating institutions, each of which is struggling to define its own capabilities. This will make the job of the Topic Coordinator more difficult. Good progress is being made in this regard.

8.1 Fine-resolution POP simulations (Scott Elliot, Shaoping Chu, and Mat Maltrud)

Fine-resolution POP biogeochemistry simulations (global one-fifth degree lat/long) have recently been completed, driven by single phytoplankton-bin ecodynamics and embracing complete iron, nitrogen and sulfur cycles. Modeled chlorophyll distributions closely matched the available satellite data (both CZCS and SeaWiFS). Dimethyl sulfide concentration patterns were accurately portrayed at the scale of the biogeographical province, but difficulties arose at the basin scale (e.g. gyre values could be underpredicted). We are currently testing a new POP biogeochemistry package which addresses key deficiencies in earlier results. A carbon cycle has been added to those of the nutrients and sulfur. The diatoms have been segregated as a separate tracer and plant taxon, such that a silicon cycle is included as well. DMS distributions are improved at the larger scales by specifying species-dependent, cell-internal dimethyl sulfonium propionate concentrations and by introducing first-order bacterial consumption. Improved geocycling and ecodynamics will soon be run on the fine grid. PCO2 and dimethyl sulfide results will be compared with the available climatologies. The Los Alamos biogeochemistry group has in coarse resolution OGCM frameworks simulated the surface dynamics of other trace gases which cross the sea-air interface to enter the lower atmospheric photochemical system. Carbonyl sulfide, the methyl halides, and simple nonmethane hydrocarbons have all been represented successfully as they are produced by phytoplankton or the photolysis of dissolved organics. Precursor and product species are included as well. We are currently embarking upon the literature search which will be required to add surface ocean cycling for the molecules carbon monoxide, isoprene, methane and nitrous oxide. This summer the plan is to test all the modelable trace gases in a coarse mode POP mixed layer, then move the routines into the CCSM framework. These first few runs will be conducted with concentration boundary conditions prescribed throughout the lowest layer of the atmosphere. During the next year of SciDAC activity, Los Alamos hopes to relax the boundary constraints by coupling with simplified atmospheric photochemistry under development at LLNL and NCAR.

9 Coordination and Management (Bob Malone and John Drake)

9.1 Management Plan

In a collaboration such as this, which involves eight major research institutions, it is imperative that roles and responsibilities be clearly defined. Toward this end, a Management Plan (MP) has been produced, edited, and reviewed by at least the lead investigator at each laboratory. The MP defines key topical areas in which this SciDAC project intends to contribute to CCSM. Each DOE laboratory has decided to which areas it is qualified to contribute and has selected a "Topic Leader" in each such area to oversee that laboratory's activities and serve as a contact point in coordinating with other institutions in the Consortium. The PIs have appointed one such Topic Leader to be the Topic Coordinator; this person has the responsibility to coordinate activities among the six DOE laboratories and with the NCAR counterpart in a given topic area.

The MP also indicates the mechanisms by which this SciDAC Consortium project will coordinate its activities with those of the other programs contributing to the development and validation of the CCSM. First, there is the management hierarchy of the CCSM

project itself, funded by NSF: the Working Groups, the Scientific Steering Committee, and the CCSM Advisory Board. Second, the DOE Climate Change Prediction Program continues to make important contributions to CCSM. More recently, the NASA-sponsored Earth System Modeling Framework (ESMF) project has come into existence. Finally, many other aspects of the SciDAC program are relevant to CCSM. The MP addresses coordination with all these.

9.2 Memorandum of Understanding between SciDAC and ESMF

A particularly difficult issue arose between SciDAC and ESMF, regarding perceived competition between ESMF and the DOE-sponsored Model Coupling Toolkit (MCT) developed at ANL. This was dealt with by developing a Memorandum of Understanding between ESMF and the SciDAC Consortium project. The MOU has been approved by appropriate levels of management on both sides.

9.3 Coordination with SciDAC Integrated Software Infrastructure Centers and Collaboratories

The dialog with the Integrated Software Infrastructure Centers (ISICs) has continued over the first half year of the project. The Earth System Grid collaboratory and the Performance Evaluation Research Center (PERC) are two of the most relevant SciDAC projects related to the CCSM model development. In addition to these two, we have worked closely with the Common Component Architecture (CCA) ISIC and made progress defining the component structures for the Model Coupling Toolkit. The following reports on these and other tasks that are being performed in collaboration with the ISICs.

Earth System Grid (ESG) is collaborating on the rapid transfer of model output between computing and analysis centers. We can report that a number of prototype tools have been tested including gridftp, with varying degrees of success. The availability of production quality tools appears to be hampered by firewalls, rapidly changing security procedures and operational differences among the centers.

The Performance Evaluation Research Center (PERC) has provided results that now include the PCM code on a variety of machines. Benchmarking, performance instrumentation and visualization of PCM results using the SvPablo tools was performed by Celso Mendes (see http://bugle.cs.uiuc.edu/~cmendes/pctm/). Discussions have taken place to define with PERC the objective measures of performance that can be derived from hardware performance monitors. The PAPI tools and others like SvPablo and Vampir, when applied to the CCSM, will present a new level of information on memory utilization important for optimization of the code. Work with the CAM will begin when CCSM2 is released. The older climate codes (e.g. CCM/MP-2D) continue to be useful in early evaluation and scaling experiments. PERC is involved in an ongoing diagnosis of unexpected performance degradation and scalability on the IBM Power systems. Further information is available on performance of climate kernels, component models and fully coupled models from Pat Worley (see http://www.csm.ornl.gov/evaluation).

An active collaboration between the Model Coupling Toolkit developer Jay Larson and the Common Component Architecture developer Wael Elwasif has resulted in a plan to prototype the MCT *attribute vector* module as a CCA component. Since the MCT is a F90 implementation and the CCA framework is C++, the interfaces must be translated

through an intermediary function. The interface definitions have been written and issues with inter-language operability have been clarified. Jay Larson visited both ORNL and SNL to talk on the Model Coupling Toolkit and discussed collaborations involving CCA and possible uses for the MCT outside of the climate community. In particular, there was some interest from the CRFRS SciDAC application center.

The Scientific Data Management (SDM) Center has also become actively involved with climate data concerns. Discussions on a broad range of issues including analysis methods for ongoing simulations, factor analysis, metadata and hierarchical data managers are ongoing. Two particular topics have resulted in focused collaborations. First, the SDM group looking at NetCDF parallel implementations on top of MPI-IO with Bill Gropp, has begun coordination with the SciDAC CCSM group at LBL working with Chris Ding on parallel implementations of NetCDF and I/O issues for the IBM system. NetCDF is a self-describing data format that has become the *de facto* standard in the climate community and in particular, is the data format implemented in the CCSM code. Second, the some climate analysis tools have been modified to allow direct access into HPSS archives. This eliminates one step in a climate scientist's data management practice. This ability has been prototyped in *ncview* and plans are to also modify the GRADS tool and the PCMDI tools.

Discussions continue on model formulation and numerical requirements with the TOPS, TSTT and APDEC centers.

10 Current tasks and levels of effort

The following tables give an accounting of how the SciDAC resources at each Laboratory have been allocated during the time interval covered by this progress report. "FTE" is the level of effort in units of fraction of the person's time. The general task area(s) are indicated, followed by the number of the section in this report in which the details of progress are given.

Argonne National Laboratory

Person	FTE	Current tasks (section in progress report)
J. Taylor	0.81	Coordination and management (1) Biogeochemistry(8) Atmospheric Chemistry(7) Land Surface Model(6.1)
J. Larson	0.75	Coupler Development(4) Land Surface Model(6.1)
R. Jacob	0.73	Coupler Development(4)
E. Ong	0.75	Coupler Development(4)
J. Kim	0.125	Sea Ice Model (5)
Total FTEs	3.165	
Annual budget	$700K	@ average annual cost per FTE = $220K

Lawrence Berkeley National Laboratory

Person	FTE	Current tasks (section in progress report)
C. Ding	0.40	I/O, Coupler, and Coordination (2.4)
Y. He	0.60	Coupler (MPH), Pilgrim, OpenMP (4.2)
W. Yang	1.00	I/O library (2.4)
Total FTEs	2.00	
Annual budget	$400K	@ average annual cost per FTE = $200K

Lawrence Livermore National Laboratory

Person	FTE	Current tasks (section in progress report)
D. Rotman	0.20	Coordination, Atmospheric Chemistry (9)
P. Connell	0.10	Atmospheric Chemistry, mechanism definition (7)
J. Dignon	0.20	Atmospheric chemistry emissions (7)
A. Mirin	0.40	Software Engineering, optimization of Lin-Rood (2.1)
P. Cameron-Smith	0.60	Atmospheric Chemistry, radiation balance (7)
J. Tannahill	0.35	Software Engineering, Chemistry (7)
P. Duffy	0.20	High Resolution climate modeling (8)
K. Caldeira	0.17	Ocean Biogeochemistry (8)
B. Govindasamy	0.17	High Resolution climate modeling (8)
S. Mickels	0.07	Clerical support
Total FTEs	2.46	
Annual budget	$700K	@ average annual cost per FTE = $285K

Los Alamos National Laboratory

Person	FTE	Current tasks (section in progress report)
R. Malone	0.50	Coordination and management (9)
J. Dukowicz	0.40	HYPOP development (3.4)
B. Lipscomb	0.40	HYPOP (3.4), CICE performance (5.1)
J. Baumgardner	0.50	HYPOP development (3.4)
P. Jones	0.50	POP SWE (3.1) and performance (3.2)
J. Taft	0.40	POP performance via MLP (3.3)
S. Chu	0.50	Biogeochemistry (8.1)
Total FTEs	3.20	
Annual budget	$800K	@ average annual cost per FTE = $250K

National Center for Atmospheric Research

Person	FTE	Current tasks (section in progress report)
W. Yu	1.0	Software Engineering, Testing, Performance Assessment and Optimization
D. Miner	1.0	Software Engineering, Infrastructure Support and Development, Repository Management
Total FTEs	2.0	
Annual budget	$400K	@ average annual cost per FTE = $200K

Oak Ridge National Laboratory

Person	FTE	Current tasks (section in progress report)
J. Drake	0.90	Coordination and management (1) Atmospheric Model (2) Atmospheric Chemistry (7) Land Surface Model (6.1)
P. Worley	0.50	Software Engineering Atmospheric Model (2)
D. Erickson	0.75	Biogeochemistry (8) Atmospheric Chemistry (7)
F. Hoffman	0.50	Land Surface Model (6)
M. Branstetter	1.00	Land Surface and River Routing (6)
Total FTEs	3.65	
Annual budget	$850K	@ average annual cost per FTE = $230K

Pacific Northwest National Laboratory

Person	FTE	Current tasks (section in progress report)
S. Ghan	0.33	Subgrid orography scheme (2.3)
A. Hunt	0.30	Subgrid orography scheme (2.3)
T. Shippert	0.03	Subgrid orography scheme (2.3)
Total FTEs	0.66	
Annual budget	$150K	@ average annual cost per FTE = $230K

11 Publications during 10/1/01-4/30/02

Ding. C., and Y. He, 2001. A ghost cell expansion method for reducing communications in solving PDE problems. Proceedings of SC 2001.

Ghan, S.J., X. Bian, A. G. Hunt, and A. Coleman, 2002. The thermodynamic influence of subgrid orography in a global climate model. *Climate Dynamics*, accepted for publication.

He, Y., and C. Ding, 2002. MPI and OpenMP paradigms on cluster of SMP architecture: the Vacancy Tracking Algorithm for multi-dimensional array transpose. Submitted to SC'02.

Kim, J. G., P. D. Hovland, and J. A. Taylor, 2001. A Parameter Tuning Scheme of Sea ice Model Based on Automatic Differentiation Technique. Eos. Trans. AGU, 82(20), Spring Meet. Suppl., Abstract U42A-04.

Lipscomb, W. H., and E. C. Hunke, 2002. Modeling sea ice transport using incremental remapping. Submitted to Monthly Weather Review.

Ong, E.T., J.W. Larson, and R.L. Jacob, 2002. A Real Application of the Model Coupling Toolkit. To appear in Proceedings of the 2002 International Conference on Computational Science (ICCS), to be published by Springer-Verlag.

12 References

Dukowicz, J., and J. Baumgardner, 2000. Incremental remapping as a transport/advection algorithm. *J. Comp. Phys.*, **160**, 318-335.

Appendix: Names, affiliations, and primary interests

The following table contains the full names, affiliations, and primary interests of individuals referred to in this document. Individuals whose names are italicized are funded by their own SciDAC projects, which are closely related to the present project.

Last name	First name	Laboratory	Primary Interest
Baumgardner	John	LANL	Ocean model
Branstetter	Marcia	ORNL	River runoff model
Bettge	Tom	NCAR	Coupled model
Bryan	*Kirk*	Princeton	Ocean model
Chu	Shaoping	LANL	Ocean biogeochemistry
Craig	Tony	NCAR	SW Engineering
Ding	Chris	LBNL	SW Engineering
Drake	John	ORNL	Atmospheric model
Duffy	Phil	LLNL	Hi-res atmosphere
Dukowicz	John	LANL	Ocean and sea ice models
Elliott	Scott	LANL	Ocean biogeochemistry
Erickson	David	ORNL	Biogeochemistry
Ghan	Steve	PNNL	Atmospheric model
Hoffman	Forrest	ORNL	Land-surface model
Hunke	Elizabeth	LANL	Sea-ice model
Jacob	Rob	ANL	Coupler, SWE
Jones	Phil	LANL	Ocean model
Larson	Jay	ANL	Coupler, SWE
Lin	S. J.	NASA/DAO	Lin-Rood dycore
Lipscomb	Bill	LANL	Sea ice model
Malone	Robert	LANL	Project coordination
Maltrud	Mat	LANL	Ocean analysis
McKenna	Daniel	NCAR	Atmospheric chemistry
Mirin	Art	LLNL	Atmospheric model
Randall	*David*	CSU	Geodesic grid models
Rotman	Doug	LLNL	Atmospheric chemistry
Sawyer	Will	NASA/DAO	SW Engineering
Smith	Rick	LANL	Ocean model
Taylor	John	ANL	Biogeochemistry
Worley	Pat	ORNL	Model performance