Accelerated Prediction of the Polar Ice and Global Ocean (APPIGO)

Eric Chassignet
Center for Ocean-Atmosphere Prediction Studies (COAPS)
Florida State University
PO Box 3062840
Tallahassee, FL 32306-2840
phone: (850) 645-7288 fax: (850) 644-4841 email: echassignet@fsu.edu

Award Number: N00014-13-1-0861
https://www.earthsystemcog.org/projects/espc-appigo/

Other investigators:

Phil Jones (co-PI)	Los Alamos National Laboratory (LANL)
Rob Aulwes	Los Alamos National Laboratory
Tim Campbell	Naval Research Lab, Stennis Space Center (NRL-SSC)
Mohamed Iskandarani	University of Miami
Elizabeth Hunke	Los Alamos National Laboratory
Ben Kirtman	University of Miami
Louis Vernon	Los Alamos National Laboratory
Alan Wallcraft	Naval Research Lab, Stennis Space Center

LONG-TERM GOALS

Arctic change and reductions in sea ice are impacting Arctic communities and are leading to increased commercial activity in the Arctic. Improved forecasts will be needed at a variety of timescales to support Arctic operations and infrastructure decisions. Increased resolution and ensemble forecasts will require significant computational capability. At the same time, high performance computing architectures are changing in response to power and cooling limitations, adding more cores per chip and using Graphics Processing Units (GPUs) as computational accelerators. This project will improve Arctic forecast capability by modifying component models to better utilize new computational architectures. Specifically, we will focus on the Los Alamos Sea Ice Model (CICE), the HYbrid Coordinate Ocean Model (HYCOM) and the Wavewatch III models and optimize each model on both GPU-accelerated and MIC-based architectures. These codes form the ocean and sea ice components of the Navy's Arctic Cap Nowcast/Forecast System (ACNFS) and the Navy Global Ocean Forecasting System (GOFS), with the latter scheduled to include a coupled Wavewatch III by 2016. This work will contribute to improved Arctic forecasts and the Arctic ice prediction demonstration project for the Earth System Prediction Capability (ESPC).

OBJECTIVES

The objective of this effort is to create versions of the Los Alamos Sea Ice Model (CICE), the HYbrid Coordinate Ocean Model (HYCOM) and the Wavewatch III models that can perform optimally on both GPU-accelerated and MIC-based computer architectures. These codes form the ocean and sea ice components of the Navy's Arctic Cap Nowcast/Forecast System (ACNFS) and the Navy Global Ocean Forecasting System (GOFS), with the latter scheduled to include a coupled Wavewatch III by 2016. This work will contribute to improved Arctic forecasts and the Arctic ice prediction demonstration project for the Earth System Prediction Capability (ESPC).

APPROACH

We will utilize an incremental acceleration approach to ensure we maintain code fidelity while improving performance. We will begin by improving the performance of selected sections of each code and expanding those regions until we have accelerated the three application codes. Acceleration may start with directive-based mechanisms like OpenACC and OpenMP, but may also include targeted kernels written in CUDA or other lower-level accelerator libraries. This approach provides early successes and opportunities to test the changes as they are made. A second approach will redesign code infrastructure to incorporate a multi-level parallelism by design. The modified codes will be validated both on a single component basis and within the forecast systems.

WORK COMPLETED

As described above, work during the first year was mainly directed at setting up benchmark cases, performing profiling and initial implementation of performance improvements on advanced architectures using directive-based approaches. Work on framework development and configuring a science application test case have also been initiated. Finally, the team has organized or participated in advanced architecture workshops to develop broader expertise in the use of these new systems.

HYCOM Performance (Alan Wallcraft, NRL-SSC; Louis Vernon, LANL)

During the first year, some initial refactoring of HYCOM at NRL was necessary to prepare for performance optimization on new architectures. HYCOM was updated to use dynamic memory allocation, because its original static memory approach wastes memory when HYCOM is running in parallel with other components in a coupled system. Ocean/land masking has also been revised. HYCOM's original method for avoiding calculations at land points, do-loops over ocean only, is not suitable for most attached processors. So it has been replaced by land/sea masks, which are implemented by MACROs that allow the option to have the mask arrays be replaced by .true. at compile time in order to calculate everything over land. In initial tests, the masks are about 5% slower than the do-loops on existing systems without attached processors. The "calculate everything" approach may be the most efficient on attached processors. It has not yet been fully implemented, but will require about 10% more MPI tasks than a land skipping version for a global domain.

For the purposes of testing single-node performance, a 320 by 384 by 41 layer GLBgx1v6 (x1 grid, average 1-degree resolution) HYCOM-only test case was configured. It can be run using MPI or OpenMP or both OpenMP and MPI, and it will be the new baseline for the addition of OpenACC directives into HYCOM. Since small test cases, such as this x1 grid, do not have the same land/sea distribution, after domain decomposition, as practical (much larger) cases a "bathtub" variant of this test case has been produced with no land except on the south and north edges of the grid. This will simplify profiling and the initial porting to attached processors via OpenACC.

In addition, the standard DoD HPCMP HYCOM 1/25 degree global (9000 by 6595 by 32 layer) benchmark case was tested on the Navy DRSC's new Cray XC30 with 24 Intel Xeon E5-2697v2 cores per node. Figure 1 shows the total core hours per model day on 3 generations of systems at the Navy DSRC. Perfect scalability would produce a horizontal line, and the Cray XC30 is scaling well out to 16 thousand cores (680 nodes). It scales better than earlier generations due to a higher performing network between nodes. Note also that the per core performance is virtually identical between the two year old IBM iDataPlex and the Cray XC30, both of which are using Intel Xeon processors. However, the IBM has 16 cores per node vs 24 cores per node on the XC30. The per node performance is significant because attached processors are provisioned one or two per node. For example, on the Cray XC30 a node can have either two 12-core Xeons or a single 10-core Xeon and an attached processor (Intel Phi or NVIDIA Tesla K40). So we would need to see 30 to 40 node hours per model day from the attached processor nodes on the XC30 to reach parity with the standard 24-core nodes.

While the focus at NRL has been on Intel PHI systems, HYCOM work at LANL has been exploring GPU-accelerated systems. Initially, the serial performance of the src_2.2.18_22 release of HyCOM was profiled on a LANL accelerator testbed using the ATLb2.00 test case. It was found that approximately 30% of the run-time was spent in the momtum and mxkppaij routines. OpenACC directives were used to port several of the existing dense, nested loops in the momtum routine to GPU kernels. In all sections of momtum.f the straightforward translation from OpenMP parallel region to OpenACC kernel, with some consideration of persistent data, resulted in decreased performance due to data movement. Some smaller loops were moved to the GPU with only minor overhead and 40% GPU utilization was achieved. Profiling of the newer src_2.2.94i release with the x1 test case described above was also begun. To date, the code was built and profiled using only MPI with no application of OpenACC. For a case using 24 MPI ranks on the GLBgx1v6 problem, the majority of the time was spent in MPI blocking calls (mpi_waitall_f). This was primarily due to the computationally asymmetric domain decomposition of the x1 problem.

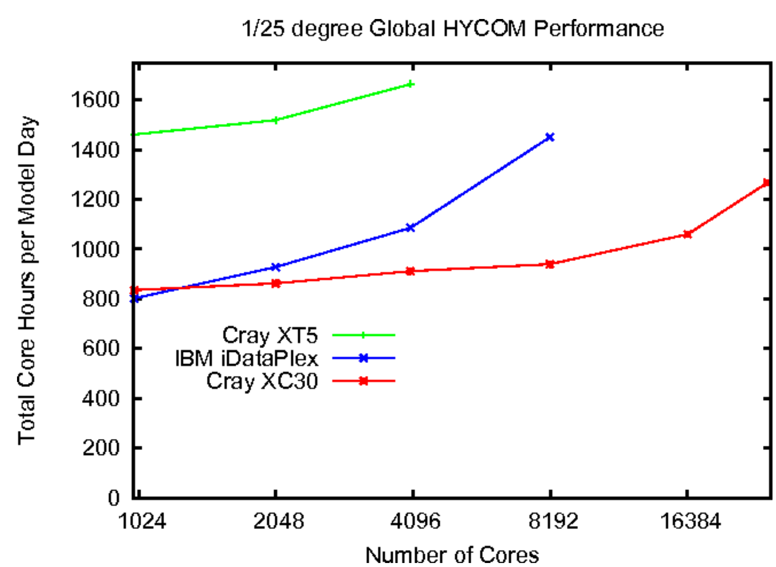

Figure 1. Perfomance (in total core-hours per simulated day) of 1/25-degree HYCOM on Navy DSRC systems.

CICE Performance (Rob Aulwes, Elizabeth Hunke, Phil Jones, LANL)

Performance tuning of CICE began using the profiling tool VTunes and bottlenecks were identified in MPI, due to load-balancing and barrier issues. Some potential solutions are being explored as these bottlenecks will need to be eliminated in order to achieve any speedup using acclerated architectures. In the meantime, initial exploration of GPU acceleration was started using both OpenACC and CUDA.

As in the HYCOM case above, some work was required to prepare CICE for these implementations. In the case of CICE, this was primarily revising the build system to support CUDA Fortran.

One of the computationally expensive routines in CICE is the computation of stress within the sea-ice dynamics formulation. A CUDA kernel of the stress() routine was created and tested. No improvement was realized, likely due to the synchronization required for halos within a subcycling step. The use of CUDA-supported device-to-device communications will likely be required for any further improvement.

The remainder of the year has been focused on accelerating parts of the thermodynamics package in CICE using OpenACC. A major challenge was identifying how to transfer the data arrays efficiently between host and device memory. We explored different strategies to accomplish this, such as using CUDA streams to transfer data asynchronously and overlapping the transfers with CPU and GPU computations. Initial performance profiling showed a slowdown of the x1 test problem by 50% compared to the baseline run. The slowdown is likely due to multiple factors, including large numbers of small data transfers between host and device, insufficient overlap of data transfers with GPU computations, and data dependencies impeding concurrent GPU kernel execution. The next step will use Fortran pointers into a larger allocated memory block in order to perform a single transfer of the block instead of multiple transfers.

Wavewatch III Performance (Tim Campbell, NRL-SSC)

As proposed, work on the Wavewatch will not start until FY15. However, Tim Campbell has performed some work under other projects to improve memory and performance issues to prepare for later work in APPIGO.

Optimized operator frameworks (Mohamed Iskandarani, Miami)

While the majority of the work has been directed at exploring and profiling peformance directly on new architectures, a second approach has been the development of an optimized library designed to simplify operating on variables located on an Arakawa C-Grid. The library will encapsulate the low-level code needed to implement HYCOM using a small number of reusable subroutines. Its primary aim is to shield most of the HYCOM code from the changing hardware environment, while optimizing the code's performance on emerging high performance computing architectures. A rudimentary shallow water code has been developed using this library for the purpose of demonstration and testing. The library is now being retooled to accept arrays laid out according to the HYCOM convention. A MATLAB version of this library has been incorporated in a computational geophysical fluid dynamics class, and has been primarily used for class projects. These included shallow water code that conserve energy and/or potential enstrophy, and a multi-layer shallow water code. As a result of discussions held at the kickoff meeting, additional software developed at LANL for creating communications abstractions is being prepared for release to Miami researchers for use in this new framework. Design and implementation of this code continues, based on interfaces and functionality needed for the HYCOM model.

Science application and test case (Eric Chassignet, Alexandra Bozec, FSU)

In order to validate the code changes being explored above and to demonstrate the improvement of the model, a scientifically useful case has been configured. HYCOM has been implemented as an ESPC ocean component in a standalone configuration. HYCOM was then configured on a Parallel Ocean Program (POP) glbx1v6 dipolar grid (320x384) grid and bathymetry. This configuration is a 1-degree climate case from the Community Earth System Model (CESM) also used by CICE above to provide a direct comparison to simulations performed by the CESM model. Several routines of the HYCOM source code (v2.2.86) have been modified to include the proper reading of the CORE-II forcing and a new passive ice component has been added, based on the CESM DICE. This new option allows HYCOM to evolve with a prescribed ice cover derived from SSMR/SSMI NSIDC climatology (Cavalieri et al., 1997). In addition to those changes, options to use a spatially varying sea surface salinity and/or temperature relaxation as well as a correction of the precipitation based on the global salinity have been introduced to comply with the POP simulation parameters used for comparison. Several 30-year runs of HYCOM have been performed with the CORE-II (Large and Yeager, 2009) normal year atmospheric fields to assess the sensitivity of the model to several parameters (reference pressure, thermobaricity, and isopycnal smoothing).

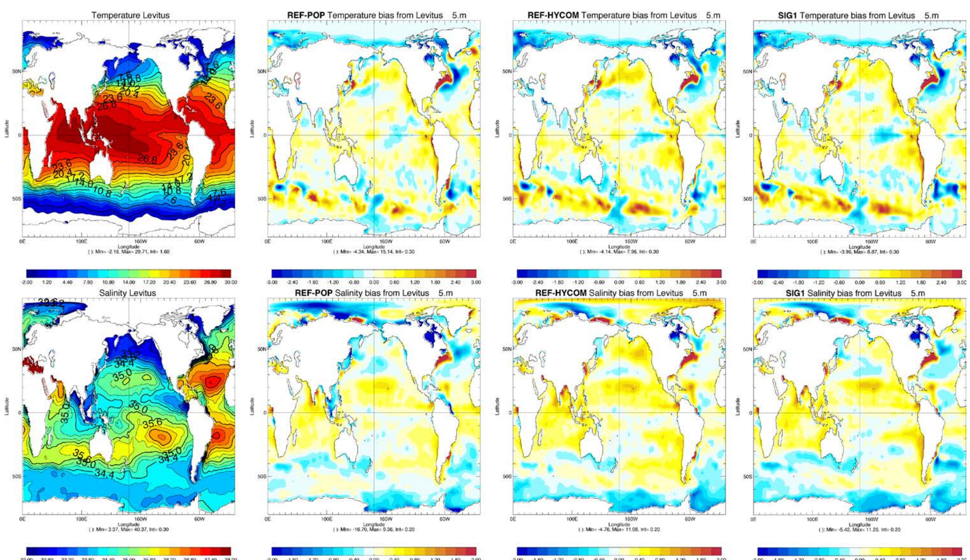

Figure 2. SST and SSS bias from Levitus PHC2 for POP (2nd column) and HYCOM-SIGMA2 (3rd column) and HYCOM-SIGMA1 (right column).

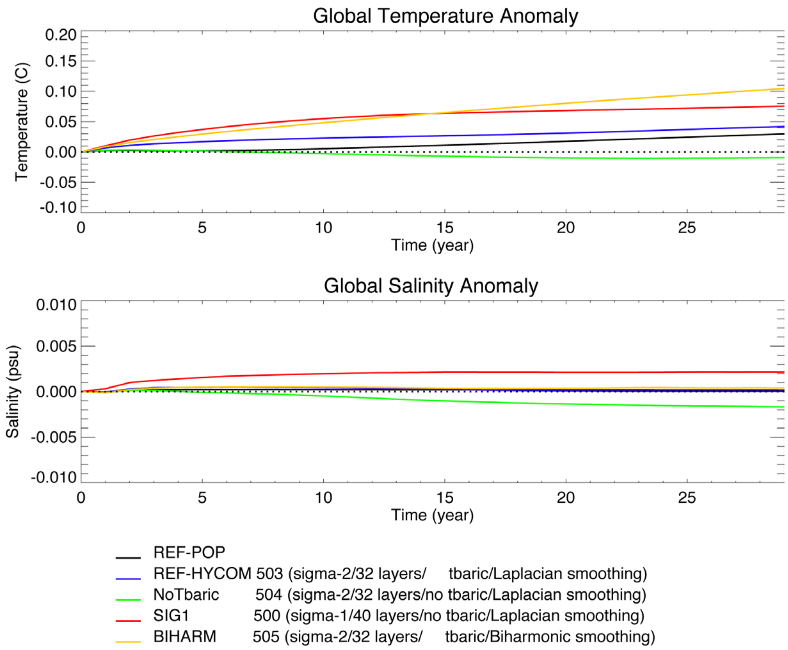

Figure 3. Evolution of the global temperature and salinity anomaly (from initial state) for POP (black), HYCOM-SIGMA2 (blue), HYCOM-SIGMA1 (red), HYCOM-SIGMA2 no thermobaricity (green), and HYCOM-SIGMA2 biharmonic instead of Laplacian (yellow).

Figure 2 shows the SST and SSS biases of POP and HYCOM with a reference pressure at 2000 and 1000 meters, respectively. The three simulations exhibit similar features and intensity over the most part of the ocean except over the North Atlantic subpolar gyre region where HYCOM shows a cold bias and POP, a warm bias. The evolution of the global temperature and salinity is shown in Figure 3 for the 30 years of the simulations. The global temperature increases in all experiments, except for the simulation with no thermobaricity that has a small decrease. As expected since a global correction is applied at every time step, the global salinity remains almost constant for the duration of all experiments, except for the simulation without thermobaricity.

Meetings and workshops:

An initial kick-off meeting for the entire program was held on November 20-21, 2013, with all projects presenting their proposed work. Discussion both within the project and across related projects were helpful in further defining tasks. In addition, project web space was provided and project web sites and mailing lists were set up.

In June, 2014, Rob Aulwes organized an OpenACC workshop with Nvidia engineers to disseminate best practices for GPU-accelerated architectures.

Finally, this project was chosen to participate in a special "Hackathon" at the Oak Ridge Leadership Computing Facility (OLCF). During this workshop, both OLCF and vendor personnel will be assigned as mentors to each of the project codes with a focused effort to produce accelerated code. Six project members will attend this workshop.

RESULTS

For the first year of this project, progress has been made on initial profiling and implementation of HYCOM and CICE on advanced architectures. Experience in the broader computational performance community has been that these initial ports often result in slower code as data transfer costs to attached accelerators dominate. This project has demonstrated similar results, with as high as 50% degradation in peformance. However, the experience gained through these initial prototypes should lead to more effective implementations and future improvements in computational performance as the project continues.

IMPACT/APPLICATIONS

Model performance improvements under this project will result in high-performance codes to enable improved future Arctic prediction, through improved resolution, increased realism or an ability to run ensembles.

RELATED PROJECTS

This project builds on the core model development activities taking place at the partner sites, including:

The Climate, Ocean and Sea Ice Modeling (COSIM) project that includes the primary development of the Los Alamos Sea Ice Model (CICE), funded by the US Department of Energy's Office of Science.

The ongoing development of the Arctic Cap Nowcast-Forecast System (ACNFS) and Global Ocean Forecast System (GOFS) at the Naval Research Lab – Stennis, funded by the US Navy.

Continued development of the Hybrid Coordinate Ocean Model (HYCOM) at Florida State University, funded by the National Science Foundation, Department of Energy and US Navy.

REFERENCES

Large, W.G. and S.G. Yeager, 2009. The global climatology of an interannually varying air-sea flux data set. *Climate Dynamics,* **33**, 341-364, doi:10.1007/s00382-008-0441-3.

Cavalieri, D. J., C. I. Parkinson, P. Gloersen, and H. J. Zwally. 1997. Arctic and Antarctic Sea Ice Concentrations from Multichannel Passive-Microwave Satellite Data Sets: October 1978 to December 1996, User's Guide. NASA Technical Memorandum 104647. 17 pages.