The ECCO Report Series [1]

Computational requirements for ECCO in support of CLIVAR and GODAE

Detlef Stammer and Carl Wunsch

Scripps Institution of Oceanography, Massachusetts Institute of Technology

Report Nr. **2**

November, 1999.

---

[1] The ECCO Project is funded through a grant from the National Oceanographic Partnership Program (NOPP). Copies of this Report are available at www.ecco-group.org or from Detlef Stammer, Scripps Institution of Oceanography, La Jolla CA 92093-0230, ph.: (858) 822-3376; fax: (858) 534-4464; e-mail: dstammer@ucsd.edu.

Introduction

Over the last decade, substantial progress has been made in ocean modeling and ocean state estimation. Circulation models are now run routinely in realistic configurations on both basin scales and globally. In some cases a spatial resolution of 1/6° has been achieved in global runs, and regional resolutions are even finer (better than 1/10°). Although the models still show deficiencies, e.g., in the degree of their simulated variability, eddy scales, water mass structures, mixing, etc., they have nonetheless achieved a remarkable degree of realism, particularly when compared to the modeling state prior to WOCE.

With this new realism, ocean modeling has become a central element in oceanographic research, and results from numerical ocean simulation are now being used in all aspects of physical oceanography. Related branches of geosciences have started to make use of ocean circulation models for their own purposes: tracer simulations, biology, earth rotation studies, geodesy, etc.

At the beginning of WOCE, the state estimation problem (assimilation) appeared quite daunting. But here too, substantial progress has been made. Demonstrations of global scale estimates of the time-varying ocean circulation and its transport processes have been achieved. It is now technically feasible, on the global scale, to include in the assimilation problem the entire suite of ocean in situ and satellite data in combination with a full GCM. Today, proposed large-scale experiments such as CLIVAR's Basin Wide Extended Climate Studies (BECS) are firmly organized around numerical ocean models and ocean state estimation.

Oceanography is now rapidly migrating from its previous PI-based individual research towards a situation more like that in meteorology. A global observing system is being put in place, and fast data links should assure near-real time ocean observations. Ocean state estimation (data assimilation) is considered the central element in all those activities which will permit the synthesis of the diverse observations into a dynamically consistent and quantitative picture of the time-evolving ocean circulation. But a substantial burden is being put on ocean models, which need to be run with very high resolution to simulate the physics with an accuracy consistent with the observations and the scientific and practical requirements. As a result, the computational resources required are very large.

As part of the effort to bring ocean state estimation forward from its present experimental state to semi-operational application, several Federal agencies recently agreed under the National Ocean Partnership Program (NOPP) to fund two groups with this goal. One group (led by E. Chassignet) has strong support from DOD and is oriented towards real-time application in support of mesoscale forecasting and U.S. Navy needs. The other group, ECCO (led by D. Stammer), originated with the goal of scientific applications of ocean state estimation and an emphasis on understanding climate and climate change, using in particular, a full synthesis of the WOCE data sets. Achieving this goal requires activities analogous to those in the meteorological re-analysis projects. Furthermore, much of ongoing and planned ECCO activities over the next five years will be fundamental for CLIVAR's BECs, as ECCO anticipates developing full ocean state estimation capabilities able to provide a dynamically consistent estimate of the ocean on decadal and longer time scales. One seeks (ultimately) calculations of the time-evolving ocean state over the past 50 years.

Obstacles

A few major obstacles can be identified that hinder more rapid progress towards routine ocean state estimations. These obstacles include the lack of wide-spread experience in the oceanographic community (the analogous meteorological development took place over about 20-30 years), and inadequate information concerning the error structure of both the observations and the models.

But the major problem, with implications for all the others, is the gross inadequacy of the available computational resources available to the oceanographic community. This situation is likely to worsen rapidly if no improvement can be achieved over the next two years. As will be explained below, ongoing NOPP activities are already fundamentally hindered by available computational resources up to the point that further progress is in jeopardy. Despite this fact, new NOPP groups to address high-resolution coastal modeling are in the planning phase, and as noted above, there are wide applications of ocean circulation modeling in biology and chemistry which will escalate the needs.

For its success, ECCO, in particular, has to rely on insufficient NPACI and UCAR resources. It must be clearly understood that these externally-provided computations are not guaranteed: there are no allocations reserved for ECCO, nor for any future CLIVAR or GODAE activities. The oceanographic community is completely dependent upon outside organizations for a central resource, with no guarantees that any will actually be provided. Problems of computational resources were discussed extensively by Powell (1998).

Current Computational Load.

The now-ongoing ECCO work aims at a global optimization with 1° spatial resolution and extending over the 6 year interval 1992 - 1997. To save computer time, our strategy is to start at a coarse spatial resolution and then move to higher resolution. We have evidence that through this procedure, the savings on computer resources is substantial.

In practice we have been working for more than 15 months on the global optimization with 2° spatial resolution. This problem is already a huge, complex and nonlinear optimization problem with $10^8 - 10^9$ control parameters. These unknowns are determined iteratively through the use of the adjoint of the MIT ocean circulation model in what is sometimes known as the "adjoint-method." With 2° spatial resolution, each iteration takes about 52 CPU hours on the NPACI CRAY T90. One complete optimization requires about 50 such iterations resulting in a total of about 110 CPU days days on the T90. Model performance is about 550 Mflops on the T90 and using 8 CPUs we obtain a speed-up factor of about 3 compared to one CPU. That means that in principle we could run one complete optimization on 2° spatial resolution in 1 month wall clock time. By being extremely patient, we have been able to complete only 42 iterations over a period exceeding one year—because our access to the machine has been only fractional.

Continuing with 1° spatial resolution, we need about 20 more iterations until a 1° model has probably converged. These further computations could require another 175 days CPU time on the

CRAY T90. With multitasking, that requires about 2 more months of CPU time, and in theory a full 1° optimization could be reached in 3-months elapsed time—given a well managed T90 computer with a reasonable computational load.

In practice, the NPACI T90 is completely over-loaded, and is not capable of coping with the 1° problem. Even for the 2° problem, our situation looks grim: we have to spend 50-75 % of our time waiting in queues and in conflicts with other big competing applications from all fields of science and industrial users. We are wasting weeks and months of time. With the computers that have been available to us thus far, the ability of ECCO to carry out its required computations is doubtful. It should also be noted that current policy CSL does not allow to perform our runs on one of the NCAR/CSL computers because their duration is substantially longer than the longest jobes allowed.

What is required to support CLIVAR and GODAE?

Over the next 5 years, the ECCO goal is to obtain a global optimization with up to 1/4° spatial resolution and covering a 15+ years time frame. Relative to the 2° experiment, a factor of about 1000 more capacity is required to reach our goal. But in practice, many more experiments are needed to obtain a good solution because we are still in the experimental stage where runs fail and where parameters have to be tested. Moreover, regional solutions with even higher resolution are necessary to support ongoing and anticipated community science efforts. Therefore, the above numbers are a severe lower limit on what is actually required. Projecting all this out to a practical system which can support CLIVAR, ocean state estimation will need at least a factor of 10,000 more computer time than we have now. (Part of the immediate ECCO plan is to shift to a new version of all of the system codes, a version capable of running on a broad variety of parallel computer architectures. This shift, while making much easier the evolution of code to follow changing machine architectures, does not fundamentally change the computational problem.) Without a substantial increase and a longer term strategy in computational support, we cannot make any serious progress.

Short-term Solution:

It is widely accepted that the problem we are trying to solve is incompatible with other user needs. For the same reason, meteorological centers long ago concluded that they were unable to share their central computing machines. Progress towards better support will come therefore, only by migrating to computers dedicated to ECCO, CLIVAR and related work. Various scenarios can be envisioned, but the most attractive at the present time would be some form of dedicated computing hardware, preferably located and maintained by, e.g., one of the existing national computer centers. Such a solution would take advantage of the infrastructure existing in buildings, personnel, and experience at the national centers, while permitting oceanographers to focus primarily on the scientific issues.

Computer hardware is developing with such speed that it is hard to project a definitive system even for the next few years. But based on available and anticipated hardware, we can at least sketch out a system here that would serve our immediate needs, so as to provide some sense of scale and budget requirements.

Our estimate is based on a COMPAQ Alpha cluster. There are interesting developments going on within Sun, IBM and HP. But those systems are not fully mature now and will not be substantially different in their price range from the one drawn out here.

Consider a 64 CPU COMPAQ Alpha server with 128 GByte RAM and 500GBytre scratch disc. Such a system will lead to a sustained performance of about 10GFlop/s and will enable us to run one iteration with 1/4° resolution in 15 - 20 amounts of days (see attached benchmark). (The 64 CPU configuration serves here as an example. In practice a 128 CPU server with similar amount of RAM would substantially speed up the computations which can be achieved with only a relatively small extra investment.)

This server can be constructed by clustering together 16 individual ES40 servers to a 64 CPU machine. Each unit has 4 CPUs and 8 Gbyte RAM and cost today about $ 160,000 (see enclosed quote). This gives a total of about $ 2.6 Mil for the harware acquisition.

To run clustered server configurations, substantial know-how is required, which in practice means that hardware maintenance and system support is not the only cost. In addition there is support required for at least one high-level scientific programmer, at about $150K per year. Hardware maintenance cost amounts to about $100K per year and run-time cost (such as electricity, cooling, licenses, etc.) amounts to $30K per year. This includes the cost of 500 Gbyte scatch disk space. Running at a larger computer center, such as SDSC, permits use of their existing mass storage system as a backup device.

With an anticipated lifetime of 3 years, this totals to about $ 2.9 million, or somewhat less than $ 1 million per year. These are not large amounts of money compared to the existing national investment in observation systems (ship and spaceborne), nor compared to the on-going investment in computing facilities.

Longer-term Strategy:

The ocean is a global, turbulent time-evolving fluid. Thus the movement of the scientific community towards observing and understanding it through global observations and dynamical models is unlikely ever to be reversed. To the extent that the ocean will remain an important element of climate change, fisheries, pollutant movement, sealevel rise etc., we will need the computational resources adequate to address these problems. This evolution of oceanography is a very recent development (from the last decade); consequently we lack the organizations and infrastructure to address these new problems. Some kind of national initiative appears to be inevitable—if major scientific progress is to occur.

```
Performance numbers for ES40 running MITgcm and for memory-channel
cluster interconnect performance with MITgcm.
====================================================================
(prepared by Chris Hill at MIT (cnh@ocean.mit.edu; ph.: 617 253-6430).

Basic parameters
 1/4 degree, 35 level global ocean simulation
 Timestep ~15mins.

 Computational parameters
 Grid in X - 1440
 Grid in Y -  720
 Grid in Z -   35
 Memory per model field - 276MB
 Minimal storage for forward model ~ 8GB ( fields + grid + forcing,
                                          prognositc fields 64-bit FP,
                                          others 32-bit FP)

 Distribute onto 16 x 4-way cluster as 8 x 8 XY grid.

 Per proc. grid:              X - 180, Y - 90, Z - 35
 Per proc. / per field memory  4.3MB

 With four proc. nodes
 Per node. grid               x - 360, Y - 180, Z - 35
 Per node. / per field memory  17.2MB

 Minimal memory per node       512MB
 for forward runs

 Adjoint assimilation estimate * 16 memory => 8GB per node, 128GB total.
 On-node disk storage * 16 => 128GB per node
 Central storage           => 1000GB

Tests
 Single proc. 180x90x35 run     - 200MFlop/s per proc.
 Four-way     4x180x90x35 run   - 150MFlop/s per proc.
                                  ( aggregate == 600MFlop/s per node )
```

```
Memory channel MITgcm low-level interface (IMC based) performance to
8 end-points.
global sum                      -   10 usec per sum
exch_xy    (for 180x90 grid)   -  108 usec per exch_xy
           (==> 4KB nearest-neighbor transfers)

exch_xyz                        -  3.7 msec per exch_xyz

4x16 cluster ==>
Total comm. time per time step - 0.25 sec
Total comp. time per time step - 1.5  sec
```