Journal of Computational Physics 305 (2016) 877–894

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

# A communication-avoiding implicit–explicit method for a free-surface ocean model

Christopher Newman [a,*], Geoffrey Womeldorff [b], Dana A. Knoll [c], Luis Chacón [d]

[a] Fluid Dynamics and Solid Mechanics Group (T-3), Los Alamos National Laboratory, PO Box 1663 MS B216, Los Alamos, NM 87545, United States
[b] Applied Computer Sciences Group (CCS-7), Los Alamos National Laboratory, Los Alamos, NM 87545, United States
[c] Integrated Design and Assessment Group (XTD-IDA), Los Alamos National Laboratory, Los Alamos, NM 87545, United States
[d] Applied Mathematics and Plasma Physics Group (T-5), Los Alamos National Laboratory, Los Alamos, NM 87545, United States

ARTICLE INFO

*Article history:*
Received 11 May 2015
Received in revised form 4 November 2015
Accepted 5 November 2015
Available online 10 November 2015

*Keywords:*
Ocean modeling
IMEX method
Jacobian-free Newton–Krylov methods
Nonlinear elimination
Communication-avoiding algorithm

ABSTRACT

We examine a nonlinear elimination method for the free-surface ocean equations based on barotropic–baroclinic decomposition. The two dimensional scalar continuity equation is treated implicitly with a preconditioned Jacobian-free Newton–Krylov method (JFNK). The remaining three dimensional equations are subcycled explicitly within the JFNK residual evaluation with a method known as nonlinear elimination. In this approach, the memory footprint of the underlying Krylov vector is greatly reduced over that required by fully coupled implicit methods. The method is second-order accurate and scales algorithmically, with allowed timesteps much larger than fully explicit methods. Moreover, the hierarchical nature of the algorithm lends itself readily to emerging architectures. In particular, we introduce a communication staging strategy for the three dimensional explicit system that greatly reduces the communication costs of the algorithm and provides a key advantage as communication costs continue to dominate relative to floating point costs in emerging architectures.

## 1. Introduction

The long-time behavior of the ocean thermohaline circulation under various forcings is an important scientific question impacting a variety of political decisions. It has been shown that high resolution is necessary for an accurate simulation of ocean circulation [1]. Improved ocean simulations follow from resolution of small-scale features, including mesoscale eddies and topographic features. Eddy-resolving simulations have been carried out in fully coupled Community Climate System Model simulations [2] as well as in ocean-only configurations [1].

Scientists are currently performing simulations with semi-implicit [3] and split-explicit methods [4,5]. Typically, these methods utilize a barotropic–baroclinic splitting that isolates fast (barotropic) and slow (baroclinic) time scales. The faster external gravity waves or barotropic motions are independent of depth, and thus two dimensional, while the slower baroclinic motions are fully three dimensional. For most problems of interest, explicit time discretizations are impractical for these systems, due to short timesteps imposed by the fast waves. Due to numerical stability issues resulting from semi-implicit and split-explicit methods, these tools frequently employ lower-than-desired spatial resolution. Having the ability

---

* Corresponding author. Tel.: +1 505 606 0647.
*E-mail addresses:* cnewman@lanl.gov (C. Newman), womeld@lanl.gov (G. Womeldorff), nol@lanl.gov (D.A. Knoll), chacon@lanl.gov (L. Chacón).

http://dx.doi.org/10.1016/j.jcp.2015.11.008
0021-9991/© 2015 Elsevier Inc. All rights reserved.

to perform such simulations with reduced splitting errors, using high-resolution grids, and using second-order accurate time integration methods will be a significant step forward towards predictive ocean simulations with greater fidelity.

Fully-implicit ocean models have been developed in recent years [6–11]. These approaches have been motivated to remove timestep restrictions, enhance numerical accuracy, and by the need for high spatial resolution. One advantage of these methods is that relatively large timesteps can be taken without sacrificing second-order accuracy, since the timestep is determined by accuracy instead of numerical stability constraints. A disadvantage, however, is that fully-implicit methods require a nonlinear solution for each timestep. In recent years, there has been considerable advancement in both linear and nonlinear solution methods in ocean modeling [12]. In particular, modern methods based on Jacobian-free Newton–Krylov methods (JFNK) [13] have been recently added to the Parallel Ocean Program (POP) [8,14–16] at Los Alamos National Laboratory. The JFNK framework allows tighter coupling of the physics, thus reducing numerical errors and improving stability versus operator splitting techniques. In addition, higher-order implicit time integration schemes can be easily implemented in the JFNK framework.

The key to efficient implementation of JFNK is effective preconditioning. Physics-based preconditioning [8,13,17,18] has been particularly successful. In particular, the work of [8] developed a preconditioner strategy for ocean simulation based on barotropic–baroclinic decomposition. The strategy outlined in [8] effectively preconditions the fully coupled three-dimensional vector system by only inverting a scalar, horizontal, two-dimensional system.

In this study, we propose an implicit/explicit (IMEX) method where the two dimensional scalar continuity equation is treated implicitly with preconditioned JFNK, and the remaining three dimensional equations are driven by the barotropic solution and subcycled explicitly within the JFNK residual evaluation. Thus, the fast barotropic physics are treated implicitly while the slower baroclinic physics are treated explicitly. This approach is known as *nonlinear elimination* since the baroclinic equations are eliminated from the implicit nonlinear residual. After this nonlinear elimination, the resulting nonlinear system consists only of the two dimensional scalar continuity equation [19,20]. An advantage of nonlinear elimination is that the Krylov memory requirements are greatly reduced over the fully implicit method. The work is motivated by successful implementation of an iterated IMEX method for radiation hydrodynamics problems [21,22], related work on sea-ice modeling [23] and kinetic plasma simulation [19,24].

It is important to note that the approach described within this manuscript can be thought of in the context of high order/low order (HOLO) methods. The algorithmic idea of HOLO builds on a well-defined hierarchical description (moments) of widely varying space and time scales. The HOLO approach has been very successful in kinetic plasma simulation [19,24,25], neutral partical transport [26,27], and thermal radiative transfer problems [28,29]. The HOLO approach is a moment-based scale-bridging algorithm, where the coarse scale (LO) problem is obtained via moment integration and is used to accelerate the fine scale (HO) problem. By taking moments, dimensionality is reduced, thus HO and LO refer to higher and lower dimensionality, respectively. The HOLO method should not be confused with other methods where *order* refers to convergence with respect to spatial or temporal discretization. The dimensionality of this LO problem is significantly smaller than the HO problem and, therefore, the LO system is far less expensive to solve. The HOLO approach provides multigrid-like algorithmic acceleration: the LO problem solver relaxes the long wavelength aspects of the solution, while the HO problem solver relaxes only the short wavelength aspects of the solution. Indeed, our approach is characterized by a LO barotropic system obtained by vertical moment of the HO baroclinic continuity and momentum equations.

In both the traditional split-explicit and semi-implicit approaches, the baroclinic or HO system is advanced explicitly with a long timestep ($\Delta t_{\mathrm{HO}} \geq \Delta t_{\mathrm{LO}}$). In the split-explicit approach, the barotropic or LO system is subcycled explicitly within the baroclinic system using timesteps $\Delta t_{\mathrm{LO}}$ necessarily much smaller than the baroclinic timestep, $\Delta t_{\mathrm{HO}}$. The semi-implicit approach treats the barotropic system implicitly within the baroclinic system with $\Delta t_{\mathrm{LO}} = \Delta t_{\mathrm{HO}}$. Our approach differs from the traditional split-explicit and semi-implicit approaches because the barotropic system is advanced implicitly and the baroclinic system is explicitly subcycled within the barotropic system. This allows $\Delta t_{\mathrm{LO}} > \Delta t_{\mathrm{HO}}$ while providing a stable time integration strategy.

Motivated by modern and evolving computer architectures, where communication will be the dominant bottleneck, we have devised and implemented a communication staging strategy designed to minimize communication within our explicit HO solver by reducing the frequency of halo exchanges per timestep. This strategy is a novel combination of ghosting and scheduling, which allows us to partition temporally timesteps contiguously into blocks, in which inter-process communication is performed only at the start of the block. This is to be contrasted against typical implementations, where inter-process communication is performed at each timestep, or each stage in a multistage integration. Our strategy increases the communication halo width from its conventional value of 1, based on the stencil requirements, to a value based heuristically on the amount of subcycling required. For explicit methods, as in the case of our HO solver, the required halo communication per timestep can be determined a priori. This allows the algorithm to perform a single communication call, with larger communication volume, for the HO system per LO timestep. This technique results in a significant reduction of latency, at the expense of performing additional computations in an increased communication halo. We show specifically in Section 7 how this communication staging strategy is particularly advantageous to our IMEX method.

The idea of staged communication is not new. In fact the HYCOM Ocean simulation code [30] utilizes a halo width of 6 which allows communication to be executed once per single Runge–Kutta timestep (recall that Runge–Kutta is a multi-stage integration, where communication is conventionally preformed during every stage). The approach of [31–33] communicates enough halo width to perform multiple explicit timesteps. However, the end goal for [31–33] is optimization of cache fill,