2009

High-order semi-implicit time-integrators for a triangular discontinuous Galerkin oceanic shallow water model

Giraldo. Francis X.; Restelli, M.

INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS
Int. J. Numer. Meth. Fluids (2009)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/fld.2118

High-order semi-implicit time-integrators for a triangular discontinuous Galerkin oceanic shallow water model[‡]

F. X. Giraldo[1,*,†] and M. Restelli[2]

[1]*Department of Applied Mathematics, Naval Postgraduate School, Monterey, CA 93943, U.S.A.*
[2]*Ozean im Erdsystem, Max-Planck-Institute für Meteorologie, Bundesstraße 53, 20146 Hamburg, Germany*

SUMMARY

We extend the explicit in time high-order triangular discontinuous Galerkin (DG) method to semi-implicit (SI) and then apply the algorithm to the two-dimensional oceanic shallow water equations; we implement high-order SI time-integrators using the backward difference formulas from orders one to six. The reason for changing the time-integration method from explicit to SI is that explicit methods require a very small time step in order to maintain stability, especially for high-order DG methods. Changing the time-integration method to SI allows one to circumvent the stability criterion due to the gravity waves, which for most shallow water applications are the fastest waves in the system (the exception being supercritical flow where the Froude number is greater than one). The challenge of constructing a SI method for a DG model is that the DG machinery requires not only the standard finite element-type area integrals, but also the finite volume-type boundary integrals as well. These boundary integrals pose the biggest challenge in a SI discretization because they require the construction of a Riemann solver that is the true linear representation of the nonlinear Riemann problem; if this condition is not satisfied then the resulting numerical method will not be consistent with the continuous equations. In this paper we couple the SI time-integrators with the DG method while maintaining most of the usual attributes associated with DG methods such as: high-order accuracy (in both space and time), parallel efficiency, excellent stability, and conservation. The only property lost is that of a compact communication stencil typical of time-explicit DG methods; implicit methods will always require a much larger communication stencil. We apply the new high-order SI DG method to the shallow water equations and show results for many standard test cases of oceanic interest such as: standing, Kelvin and Rossby soliton waves, and the Stommel problem. The results show that the new high-order SI DG model, that has already been shown to yield exponentially convergent solutions in space for smooth problems, results in a more efficient model than its explicit counterpart. Furthermore, for those problems where the spatial resolution is sufficiently high compared with the length scales of the flow, the capacity to use high-order (HO) time-integrators is a necessary complement to the employment of HO space discretizations, since the total numerical error would be otherwise dominated by the time discretization error. In fact, in the limit of increasing spatial resolution, it makes little sense to use HO spatial discretizations coupled with low-order time discretizations. Published in 2009 by John Wiley & Sons, Ltd.

Received 29 September 2008; Revised 14 May 2009; Accepted 18 May 2009

*Correspondence to: F. X. Giraldo, Department of Applied Mathematics, Naval Postgraduate School, Monterey, CA 93943, U.S.A.
†E-mail: fxgirald@nps.edu

Published in 2009 by John Wiley & Sons, Ltd.

KEY WORDS: discontinuous Galerkin; explicit; finite element; finite volume; implicit; semi-implicit; shallow water; triangle

1. INTRODUCTION

The discontinuous Galerkin (DG) method has come into prominence in the last decade in all areas of numerical modeling; however, it has only been in the last few years that this method has received attention in geophysical fluid dynamics. The high-order (HO) accuracy, geometric flexibility to use unstructured grids, conservation, and monotonicity properties of the DG method makes it a prime candidate for the construction of future ocean and shallow water models. The advantages offered by the DG method will benefit all areas of ocean modeling but, specifically, it will improve coastal ocean models where proper coastline representation, and the ability to handle steep gradients should translate into better modeling of tsunamis [1], storm surges, and hurricanes. Let us now review the literature concerning the application of the DG method to the oceanic shallow water equations.

Schwaneberg and Köngeter [2] first used the DG method for the planar shallow water equations, followed by the work of Li and Liu [3], and Aizinger and Dawson [4]. Dupont and Lin [5], Eskilsson and Sherwin [6], Remacle *et al.* [7], and Kubatko *et al.* [8] constructed shallow water models on triangles using a collapsed local coordinate DG method. Giraldo and Warburton [9] developed a HO DG oceanic shallow water model on unstructured adaptive triangular grids. In that paper, we showed that the model yields exponentially convergent solutions (for smooth problems). However, we used explicit time-integration methods that, while easy to implement, require small time steps in order to maintain stability. To ameliorate this deficiency found in all DG shallow water models, we extend the explicit time-integrators to semi-implicit (SI). To date, there has been no work on the development of SI time-integrators for shallow water DG models; all of the DG shallow water models found in the literature use explicit time stepping, including those discussed above. Furthermore, the only work on DG and SI methods found in the literature are the papers by Dolejsi *et al.* on the compressible Navier–Stokes equations (see [10–14]). Their SI DG formulation is based on low-order polynomial spaces (third order or less) and their approach is fundamentally different from ours in that they rely on a linearization of the nonlinear operators in conjunction with a special flux function that facilitates this linearization. Our approach [15, 16] relies on extracting the linear operators containing the fastest wave speeds in the system and then discretizing them implicitly in time. While both approaches are very effective, our approach is more similar to the classical SI method first proposed by Robert *et al.* [17]. The advantage of this approach is that, once the numerical machinery is developed, it can be applied to any equation set with minimal modifications. Moreover, the SI DG approach is easily extendable to generalized families of linear multi-step time-integration methods as we show here. In the future work, we plan to show that multi-stage methods (such as Runge–Kutta (RK) methods) can also be included into this framework.

The remainder of the paper is organized as follows. Section 2 describes the governing equations of motion used to test our numerical method. In Section 3 we describe the spatial discretization of the governing equations and in Section 4 the time-integrators used. Finally, in Section 5 we present comparisons between the explicit and SI versions of the model. This then leads to a summary on the direction of future work.

Published in 2009 by John Wiley & Sons, Ltd.

DOI: 10.1002/fld

2. CONTINUOUS EQUATIONS

The oceanic shallow water equations are a system of nonlinear partial differential equations that govern the motion of a viscous incompressible fluid in a shallow depth. The predominant feature of this type of flow is that its characteristic horizontal scale length is far greater than its depth. Since this scale ratio characterizes, as a first approximation, the motion of water in the oceans, the shallow water equations are typically used as a first step toward the construction of ocean models. In addition, this simple set of equations is what is currently used in many operational-like tsunami models (e.g. [18–22]).

The shallow water equations in conservation form are

$$\frac{\partial \mathbf{q}}{\partial t} + \nabla \cdot \mathbf{F}(\mathbf{q}) = S(\mathbf{q}) \tag{1}$$

where $\mathbf{q} = (\phi, \mathbf{U}^{\mathscr{T}})^{\mathscr{T}}$ are the conservation variables,

$$\mathbf{F}(\mathbf{q}) = \begin{pmatrix} \mathbf{U} \\ \dfrac{\mathbf{U}\otimes\mathbf{U}}{\phi} + \dfrac{1}{2}(\phi^2 - \phi_B^2)\mathscr{I}_2 \end{pmatrix} \tag{2}$$

is the flux tensor and

$$S(\mathbf{q}) = -\begin{pmatrix} 0 \\ f(\mathbf{k}\times\mathbf{U}) - \phi_S \nabla \phi_B - \dfrac{\boldsymbol{\tau}}{\rho H} + \gamma \mathbf{U} \end{pmatrix} \tag{3}$$

is the source function where the nabla operator is defined as $\nabla = (\partial_x, \partial_y)^{\mathscr{T}}$, $\otimes$ denotes the tensor product operator, $\phi = g(h_S + h_B)$ is the geopotential height where $g = 9.80616\,\mathrm{m/s^2}$ is the gravitational constant, h_S is the free surface height of the fluid, $\phi_B = g h_B$, h_B is the bathymetry (e.g. bottom of the ocean) which we assume constant (i.e. in this work it does not vary in space or time), $\mathbf{U} = \phi\mathbf{u}$ is the momentum, $\mathbf{u} = (u, v)^{\mathscr{T}}$ is the velocity vector, $f = f_0 + \beta(y - y_m)$ is the Coriolis parameter, y_m is the position about which the beta-plane approximation is applied (the midpoint of the domain in the y direction), $k = (0, 0, 1)^{\mathscr{T}}$ is the unit normal vector of the x–y plane, and the term $\mathscr{I}_2$ is a rank-2 identity matrix. The vector $\boldsymbol{\tau}$ is the wind stress, ρ is the density of the water, H is a mean height, and the constant γ is the bottom friction.

2.1. Linearized continuous equations

Let us now decompose Equations (1)–(3) into their linear and nonlinear components. Splitting the geopotential height ϕ into the depth from mean sea level to the ocean bottom ϕ_B and the height from mean sea level to the water surface ϕ_S we then have $\phi(\mathbf{x}, t) = \phi_S(\mathbf{x}, t) + \phi_B(\mathbf{x})$ which we can now use to substitute into the equations to get

$$\frac{\partial \phi_S}{\partial t} + \nabla \cdot \mathbf{U} = 0 \tag{4}$$

$$\frac{\partial \mathbf{U}}{\partial t} + \nabla \cdot \left[\delta_{\mathrm{NL}} \left(\frac{\mathbf{U}\otimes\mathbf{U}}{\phi} + \frac{1}{2}\phi_S^2 \mathscr{I}_2 \right) + \phi_S \phi_B \mathscr{I}_2 \right] = -f(\mathbf{k}\times\mathbf{U}) + \phi_S \nabla \phi_B + \frac{\boldsymbol{\tau}}{\rho H} - \gamma \mathbf{U} \tag{5}$$

where δ_{NL} is a switch that retains the nonlinear terms when $\delta_{\mathrm{NL}} = 1$ and turns them off for $\delta_{\mathrm{NL}} = 0$.

Published in 2009 by John Wiley & Sons, Ltd.

DOI: 10.1002/fld

In the next few sections, we describe the SI method for the oceanic shallow water equations in the particular case that the DG method is used to represent spatial derivatives. As the reader will see, a pivotal component of the SI method is the construction of a linearized form of the continuous equations. Linearizing Equations (1)–(3) yields

$$\frac{\partial \phi_S}{\partial t} + \nabla \cdot \mathbf{U} = 0 \tag{6}$$

$$\frac{\partial \mathbf{U}}{\partial t} + \nabla \cdot (\phi_S \phi_B \mathcal{I}_2) = -f(\mathbf{k} \times \mathbf{U}) + \phi_S \nabla \phi_B + \frac{\boldsymbol{\tau}}{\rho H} - \gamma \mathbf{U} \tag{7}$$

which are obtained by setting $\delta_{\mathrm{NL}} = 0$ in Equations (4) and (5). It should be pointed out that the linear shallow water equations given in Equations (6) and (7) are not the same linear operator obtained for the equations in non-conservation form. For the SI time integration, it is of no consequence which linear operator is chosen as long as it contains the maximum waves in the system. However, in Section 5 we will need to make adjustments to the analytic solutions of the linearized shallow water equations in order to force them to satisfy Equations (4) and (5).

The maximum eigenvalue of the linear system given in Equations (6) and (7) is $\lambda_L = \sqrt{\phi_B}$ which is in fact the linearized eigenvalue of λ_{NL} obtained for the nonlinear system given in Equations (1)–(3). From Equations (6) and (7) we can define the linear operator as follows:

$$L = -\begin{pmatrix} \nabla \cdot \mathbf{U} \\ \nabla \cdot (\phi_S \phi_B \mathcal{I}_2) + f(\mathbf{k} \times \mathbf{U}) - \phi_S \nabla \phi_B + \gamma \mathbf{U} \end{pmatrix} \tag{8}$$

We will return to this linear operator in Section 4. Let us now describe the approximation of the spatial derivatives by the DG method. We need to know this information before we can construct the SI solution.

3. TRIANGULAR DG METHOD

In this section we describe the approximation of the spatial derivatives of the shallow water equations using the DG method on triangles. This includes: the choice of basis functions, integration, construction of the semi-discrete problem, and its corresponding matrix form.

3.1. Basis functions

To define the discrete local operators we begin by decomposing the domain Ω into N_e conforming non-overlapping triangular elements Ω_e such that

$$\Omega = \bigcup_{e=1}^{N_e} \Omega_e$$

The condition on grid conformity, however, is not required by the DG method; we only impose this condition to simplify the discussion.

To perform differentiation and integration operations, we introduce the non-singular mapping $\mathbf{x} = \Psi(\boldsymbol{\xi})$, which defines a transformation from the physical Cartesian coordinate system

$\mathbf{x}=(x,y)^{\mathscr{T}}$ to the local reference coordinate system $\boldsymbol{\xi}=(\xi,\eta)^{\mathscr{T}}$ defined on the reference triangle $\Omega_e=\{(\xi,\eta),-1\leqslant\xi,\eta\leqslant 1,\xi+\eta\leqslant 0,\}$.

Let us now represent the local elementwise solution $\mathbf{q}$ by an Nth-order polynomial in $\boldsymbol{\xi}$ as

$$\mathbf{q}_N(\boldsymbol{\xi})=\sum_{i=1}^{M_N}\psi_i(\boldsymbol{\xi})\mathbf{q}_N(\boldsymbol{\xi}_i)$$

where $\boldsymbol{\xi}_i$ represents $M_N=\frac{1}{2}(N+1)(N+2)$ interpolation points and $\psi_i(\boldsymbol{\xi})$ are the associated multivariate Lagrange polynomials. For the interpolation points $\boldsymbol{\xi}_i$ we choose the nodal sets based on the electrostatics [23] and Fekete [24] points; for simplicity we shall refer to these nodal sets collectively as Fekete points. We have already described the construction of the nodal basis functions in [9, 25] and, for the sake of brevity, omit this discussion here.

3.2. Integration

3.2.1. Area integrals. In order to complete the discussion of the local elementwise operations required to construct discrete spatial operators, we must describe the integration procedure required by the integral formulation of all Galerkin methods. For any two functions f and g the 2D (area) integration $\mathscr{I}_A$ proceeds as follows:

$$\mathscr{I}_A[f,g]=\int_{\Omega_e}f(\mathbf{x})g(\mathbf{x})\,\mathrm{d}\mathbf{x}=\sum_{i=1}^{M_Q}w_i^e\,|\,J^e(\boldsymbol{\xi}_i)\,|\,f(\boldsymbol{\xi}_\mathbf{i})g(\boldsymbol{\xi}_\mathbf{i})$$

where M_Q is a function of Q, that represents the order of the cubature approximation. For w_i and $\boldsymbol{\xi}_i$ we use the HO cubature rules for the triangle given in [26–29]; because we use order $2N$ integration, which is exact for this equation set, then neither spatial filters nor smoothing diffusion operators are used. In the present work we omit the use of slope limiters.

3.2.2. Boundary integrals. The DG method also requires the evaluation of boundary integrals, which is the mechanism by which the fluxes across element edges are evaluated and allows the discontinuous elements to communicate. For any two functions f and g the 1D (boundary) integration $\mathscr{I}_B$ proceeds as follows:

$$\mathscr{I}_B[f,g]=\int_{\Gamma_e}f(\mathbf{x})g(\mathbf{x})\,\mathrm{d}\mathbf{x}=\sum_{i=0}^{Q}w_i^s\,|\,J^s(\boldsymbol{\xi}_i)\,|\,f(\boldsymbol{\xi}_\mathbf{i})g(\boldsymbol{\xi}_\mathbf{i})$$

where Q represents the order of the quadrature approximation. Using Legendre–Gauss quadrature we can use $Q=N$ to achieve order $2N$ accuracy.

3.3. Tangent and normal vectors of the element edges

Below it will become evident that in order to construct a DG discretization requires knowing a bit about the element geometry. In continuous Galerkin methods, such as the finite/spectral element method, the only required information is the basis functions, metric terms, and cubature rules. The DG method requires all of this finite element-type information plus some finite volume-type information regarding the element edges and the element neighbors sharing these edges. However, the good news for the DG method is that regardless of the order of the basis function, N, each element only has three edge neighbors (this is true only for conforming grids). This is the process by which a DG element shares its local information with its neighbors.

3.4. Semi-discrete equations

Applying the DG discretization to Equation (1), and using Green's theorem yields the classical DG which we refer to as the *weak* form

$$\int_{\Omega_e} \left(\frac{\partial \mathbf{q}_N^{(e)}}{\partial t} - \mathbf{F}_N^{(e)} \cdot \nabla - S_N^{(e)} \right) \psi_i(\mathbf{x}) \, d\mathbf{x} = - \sum_{l=1}^{3} \int_{\Gamma_e} \psi_i(\mathbf{x}) \mathbf{n}^{(e,l)} \cdot \mathbf{F}_N^{(*,l)} \, d\mathbf{x} \tag{9}$$

where $F_N = F(\mathbf{q}_N)$ and $S_N = S(\mathbf{q}_N)$ with $\mathbf{F}$ and S given by Equations (2) and (3), respectively. Note that Equation (9) states that $\mathbf{q}_N$ satisfies the equation on each element Ω_e for all $\psi \in \mathscr{S}$ where $\mathscr{S}$ is the finite-dimensional space

$$\mathscr{S} = \{ \psi \in \mathscr{L}_2(\Omega) \colon \psi|_{\Omega_e} \in P_N(\Omega_e) \forall \Omega_e \}$$

P_N is the polynomial space defined on Ω_e and the union of these elements defines the entire global domain—that is, $\Omega = \bigcup_{e=1}^{N_e} \Omega_e$ with N_e representing the total number of triangular elements. It should be mentioned that in DG methods, the space $P_N - P_N$ can be used without violating the inf–sup (Ladyzhenskaya–Babuska–Brezzi) condition that must be observed by continuous Galerkin methods (such as the finite element method) in order to avoid the effects of spurious pressure modes.

In the boundary integral of Equation (9), $\mathbf{n}$ is the outward pointing unit normal vector of the element edge Γ_e and $\mathbf{F}_N^*$ is the Rusanov numerical flux (although other fluxes are also possible)

$$\mathbf{F}_N^{(*,l)} = \tfrac{1}{2} [\mathbf{F}_N(\mathbf{q}_N^{(e)}) + \mathbf{F}_N(\mathbf{q}_N^{(l)}) - |\lambda^{(l)}| (\mathbf{q}_N^{(l)} - \mathbf{q}_N^{(e)}) \mathbf{n}^{(e,l)}]$$

where $\lambda^{(l)} = \max(|U^{(e)}| + \sqrt{\phi^{(e)}}, |U^{(l)}| + \sqrt{\phi^{(l)}})$ with $U^{(e,l)} = \mathbf{u}^{(e,l)} \cdot \mathbf{n}^{(l)}$ being the normal component of velocity with respect to the edge Γ_e, and the superscripts e and l represent the element e and its three edge neighbors l. The normal vector $\mathbf{n}^{(e,l)}$ is defined as pointing outward from the element e to its edge neighbor l.

Integrating Equation (9) by parts once more yields the *strong* form

$$\int_{\Omega_e} \psi_i(\mathbf{x}) \left(\frac{\partial \mathbf{q}_N^{(e)}}{\partial t} + \nabla \cdot \mathbf{F}_N^{(e)} - S_N^{(e)} \right) d\mathbf{x} = \sum_{l=1}^{3} \int_{\Gamma_e} \psi_i(\mathbf{x}) \mathbf{n}^{(e,l)} \cdot (\mathbf{F}_N^{(e)} - \mathbf{F}_N^{(*,l)}) \, d\mathbf{x} \tag{10}$$

which, although mathematically equivalent to the weak form, yields different numerical solutions. Based on a previous study (see Giraldo [25]) we use the strong form exclusively in this paper. We understand that the naming convention of *strong* versus *weak* is somewhat unusual, but we point out that we use this convention here in order to remain consistent with the terminology in our previous work (e.g. [9, 25, 30, 31]). It should be understood that the weak and strong forms mentioned here both represent weak formulations of the continuous problem (perhaps weak form and strong weak form are more apt).

3.5. Matrix form of the semi-discrete equations

Substituting the polynomial approximation

$$\mathbf{q}_N = \sum_{i=1}^{M_N} \psi_i \mathbf{q}_i$$

into Equation (10) we can now write the semi-discrete system as

$$\int_{\Omega_e} \psi_i \psi_j \, \mathrm{d}\mathbf{x} \frac{\partial \mathbf{q}_j^{(e)}}{\partial t} + \int_{\Omega_e} \psi_i \nabla \psi_j \, \mathrm{d}\mathbf{x} \cdot \mathbf{F}_j^{(e)} - \int_{\Omega_e} \psi_i \psi_j \, \mathrm{d}\mathbf{x} S_j^{(e)}$$
$$= \sum_{l=1}^{3} \int_{\Gamma_e} \psi_i \psi_j \mathbf{n}^{(e,l)} \, \mathrm{d}\mathbf{x} \cdot (\mathbf{F}^{(e)} - \mathbf{F}^{(*,l)})_j \tag{11}$$

Next, note that by defining the following element matrices

$$M_{ij}^{(e)} = \int_{\Omega_e} \psi_i \psi_j \, \mathrm{d}\mathbf{x}, \quad \mathbf{M}_{ij}^{(e,l)} = \int_{\Gamma_e} \psi_i \psi_j \mathbf{n}^{(e,l)} \, \mathrm{d}\mathbf{x}, \quad \mathbf{D}_{ij}^{(e)} = \int_{\Omega_e} \psi_i \nabla \psi_j \, \mathrm{d}\mathbf{x}$$

allow us to write Equation (11) in the following matrix form as:

$$M_{ij}^{(e)} \frac{\partial \mathbf{q}_j^{(e)}}{\partial t} + (\mathbf{D}_{ij}^{(e)})^{\mathscr{T}} \mathbf{F}_j^{(e)} - M_{ij}^{(e)} S_j^{(e)} = \sum_{l=1}^{3} (\mathbf{M}_{ij}^{(e,l)})^{\mathscr{T}} (\mathbf{F}^{(e)} - \mathbf{F}^{(*,l)})_j \tag{12}$$

where the superscript e denotes an element-based evaluation and l denotes edge-based evaluations. Next, using the approach described in [9] for eliminating the mass matrix, we write

$$\widehat{\mathbf{D}}^{(e)} = (M^{(e)})^{-1} \mathbf{D}^{(e)}, \quad \widehat{\mathbf{M}}^{(e,l)} = (M^{(e)})^{-1} \mathbf{M}^{(e,l)}$$

which then allows us to write Equation (12) as follows:

$$\frac{\partial \mathbf{q}_i^{(e)}}{\partial t} + (\widehat{\mathbf{D}}_{ij}^{(e)})^{\mathscr{T}} \mathbf{F}_j^{(e)} - S_i^{(e)} = \sum_{l=1}^{3} (\widehat{\mathbf{M}}_{ij}^{(e,l)})^{\mathscr{T}} (\mathbf{F}^{(e)} - \mathbf{F}^{(*,l)})_j \tag{13}$$

Equation (13) is the form that we shall use in the construction of the explicit and SI discretizations.

3.6. Boundary conditions

In all the test cases we only consider no-flux boundary conditions; we will extend our model to more general boundary conditions in future work. The no-flux boundary conditions are enforced by the virtue of the statement

$$\mathbf{n} \cdot \mathbf{u} = 0$$

at the boundaries. Thus, we seek to eliminate the normal component of the velocity to the no-flux boundary without altering the tangential component. The tangent vector to a boundary is defined as $\mathbf{t} = \mathbf{k} \times \mathbf{n}$, which is equal to $\mathbf{t} = -n_y \mathbf{i} + n_x \mathbf{j}$. Thus, we solve the following 2×2 system:

$$\begin{pmatrix} n_x & n_y \\ -n_y & n_x \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ u_T \end{pmatrix}$$

where $u_T = \mathbf{t} \cdot \mathbf{u}$ is the tangential component of velocity. This boundary condition is imposed only weakly through the boundary integrals in Equation (13); that is, it only comes in through the Rusanov flux.

4. TIME-INTEGRATOR

In Section 3 we described the approximation of the spatial derivatives using the DG method. We are now in a position to describe the approximation of the time derivatives. Let us begin with the description of the explicit time integration followed by the SI time-integration methods. We use a method of lines approach for both methods.

4.1. Explicit method

In order to advance the solution in time while retaining some level of HO accuracy we use third-order strong stability preserving (SSP) RK methods (see [32] and [33]). For completeness we define them now. Let us write the semi-discrete (in space) equations as follows:

$$\frac{\partial \mathbf{q}}{\partial t} = R(\mathbf{q})$$

where

$$R(\mathbf{q}) = -(\widehat{\mathbf{D}}_{ij}^{(e)})^{\mathscr{T}} \mathbf{F}_j^{(e)}(\mathbf{q}) + S_i^{(e)} + \sum_{l=1}^{3} (\widehat{\mathbf{M}}_{ij}^{(e,l)})^{\mathscr{T}} (\mathbf{F}^{(e)}(\mathbf{q}) - \mathbf{F}^{(*,l)}(\mathbf{q}))_j \tag{14}$$

and $\mathbf{q}$ is in fact $\mathbf{q}^{(e)}$, that is, the solution within each element e.

The SSP temporal discretization of this semi-discrete equation is

$$\text{for } k = 1, \ldots, S: \quad \mathbf{q}^k = \alpha_0^k \mathbf{q}^n + \alpha_1^k \mathbf{q}^{k-1} + \alpha_2^k \mathbf{q}^{k-3} + \beta^k \Delta t\, R(\mathbf{q}^{k-1})$$

where $\mathbf{q}^0 = \mathbf{q}^n$, $\mathbf{q}^S = \mathbf{q}^{n+1}$, S denotes the number of RK stages, and the coefficients α and β are given in Table I. The attraction of SSP methods is that they can be written as convex combinations of Euler steps; this then means that any property proven for Euler's method will be retained by the HO SSP methods such as stability in a total variation diminishing/bounded sense.

Table I. Coefficients for the explicit strong stability preserving third-order Runge–Kutta methods.

Method	k	α_0	α_1	α_2	β
RK3	1	1	0	0	1
	2	$\frac{3}{4}$	$\frac{1}{4}$	0	$\frac{1}{4}$
	3	$\frac{1}{3}$	$\frac{2}{3}$	0	$\frac{2}{3}$
RK34	1	1	0	0	1
	2	0	1	0	$\frac{1}{2}$
	3	$\frac{2}{3}$	$\frac{1}{3}$	0	$\frac{1}{6}$
	4	0	1	0	$\frac{1}{2}$
RK35	1	1	0	0	0.377268915331368
	2	0	1	0	0.377268915331368
	3	0.355909775063327	0.644090224936674	0	0.242995220537396
	4	0.367933791638137	0.632066208361863	0	0.238458932846290
	5	0	0.762406163401431	0.237593836598569	0.287632146308408

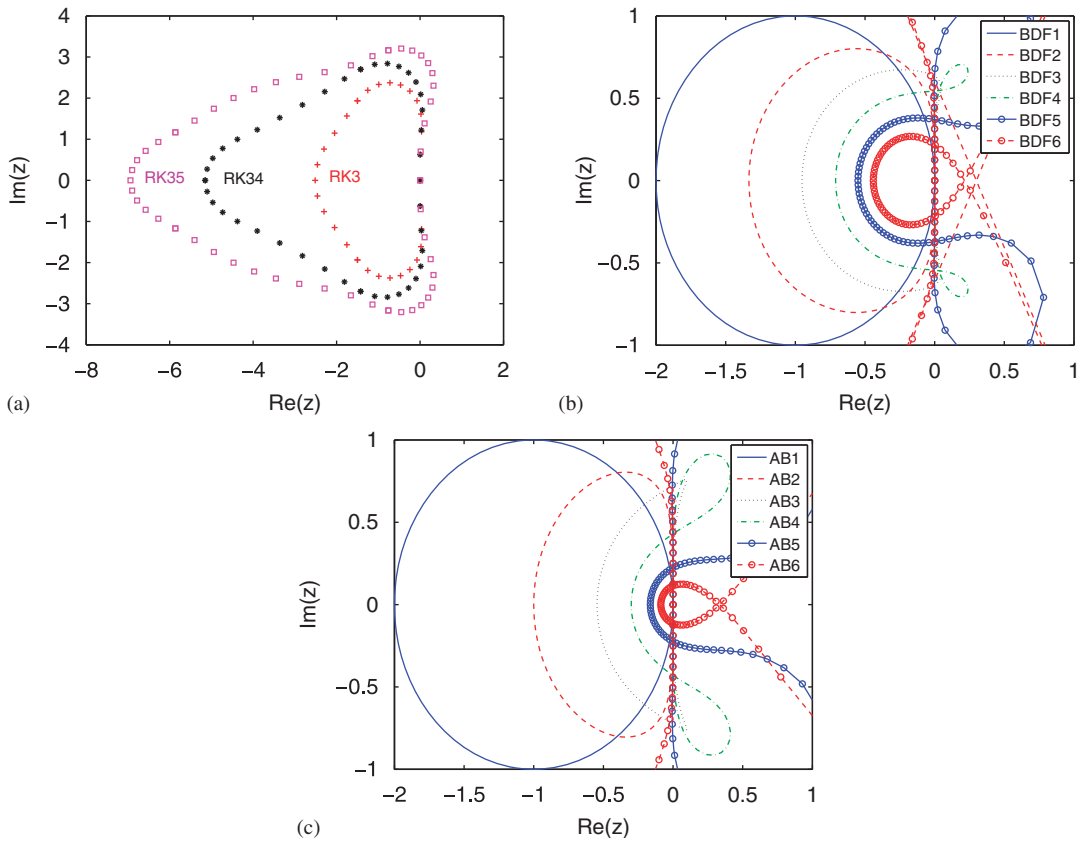

Figure 1. The stability regions for: (a) explicit strong stability preserving third-order Runge–Kutta methods, RK3 (three stage), RK34 (four stage), and RK35 (five stage); (b) explicit backward difference formulas of order $1 \leqslant K \leqslant 6$; and (c) explicit Adams–Bashforth methods of order $1 \leqslant K \leqslant 6$.

In Figure 1, we show the stability regions of these methods for the equation

$$\frac{\mathrm{d}q}{\mathrm{d}t} = \lambda q$$

which is a proxy for an advection-dominated equation; a reasonable approximation for the shallow water equations. Figure 1(a) shows that the stability region of RK35 is larger than those for RK34 and RK3; Figure 1(b) shows the stability region of the explicit backward difference formulas (BDF) of orders $K \leqslant 6$; and Figure 1(c) shows the stability region of the explicit Adams–Bashforth (AB) methods of orders $K \leqslant 6$. We are interested specifically in the region along the imaginary axis because this is the region that is most important to advection-dominated problems; this would not be the case for diffusion-dominated problems which are beyond the scope of this work. In [34] all three RK methods (as well as the BDF methods) were studied for the Navier–Stokes equations and it was determined that RK35 is indeed the most efficient of the third-order methods. We have found similar results here for the shallow water equations; although we have not analyzed the AB

methods in detail, from Figure 1 one can see that the stability regions of the AB methods are very similar to those of the BDF methods meaning that if the RK methods are more efficient than the explicit BDF methods, then they will be more efficient than the AB methods as well. Based on the results of these studies, we use RK35 (as representative of the best possible explicit method) for the comparisons with the SI methods which we now describe.

4.2. SI method

To extend the size of the time step we use a generalized SI method of order K. Let us write Equations (1)–(3) in the following compact vector form:

$$\frac{\partial \mathbf{q}}{\partial t} = R(\mathbf{q})$$

where $\mathbf{q} = (\phi_s, \mathbf{U}^{\mathscr{T}})^{\mathscr{T}}$ and $R(\mathbf{q})$ is defined in Equation (14). This system can be represented by the equivalent form as

$$\frac{\partial \mathbf{q}}{\partial t} = \{R(\mathbf{q}) - \delta_{\mathrm{SI}} L(\mathbf{q})\} + [\delta_{\mathrm{SI}} L(\mathbf{q})] \tag{15}$$

where $L(\mathbf{q})$ is the linear approximation to R given in Equation (8) and contains the gravity wave terms (i.e. the fastest waves in this system, at least for subcritical flow). In Equation (15) the curly brackets denote explicit time integration, whereas the square brackets denote implicit time integration. Note that the variable δ_{SI} is a switch that yields a fully explicit method for $\delta_{\mathrm{SI}} = 0$ and the SI method for $\delta_{\mathrm{SI}} = 1$.

As was done in [35–37] we now write the time discretization in the general form

$$\mathbf{q}^{n+1} = \sum_{k=0}^{K-1} \alpha_k \mathbf{q}^{n-k} + \gamma \Delta t \sum_{k=0}^{K-1} \beta_k [R(\mathbf{q}^{n-k}) - \delta_{\mathrm{SI}} L(\mathbf{q}^{n-k})] + \gamma \Delta t \, \delta_{\mathrm{SI}} L(\mathbf{q}^{n+1}) \tag{16}$$

where K denotes the order of the time-integrator. To simplify the discussion of the SI formulation, let us now introduce the following auxiliary variables:

$$\mathbf{q}_{tt} = \mathbf{q}^{n+1} - \sum_{k=0}^{K-1} \beta_k \mathbf{q}^{n-k}$$

$$\mathbf{q}^E = \sum_{k=0}^{K-1} \alpha_k \mathbf{q}^{n-k} + \gamma \Delta t \sum_{k=0}^{K-1} \beta_k R(\mathbf{q}^{n-k})$$

$$\widehat{\mathbf{q}} = \mathbf{q}^E - \sum_{k=0}^{K-1} \beta_k \mathbf{q}^{n-k}$$

which then allows us to write Equation (16) as

$$\mathbf{q}_{tt} = \widehat{\mathbf{q}} + \lambda L(\mathbf{q}_{tt}) \tag{17}$$

where $\lambda = \gamma \Delta t \, \delta_{\mathrm{SI}}$. In Table II we list the coefficients for the BDF of orders $K = 1, \ldots, 6$. In Figures 2(a) and (b) we show the stability regions of the implicit BDF methods. The closed loops are the instability regions of the implicit BDF methods, in Figure 2(b). For the shallow water

Table II. Coefficients for the backward difference formulas of orders $K=1,\ldots,6$.

	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$
α_0	1	$\frac{4}{3}$	$\frac{18}{11}$	$\frac{48}{25}$	$\frac{300}{137}$	$\frac{360}{147}$
α_1	0	$-\frac{1}{3}$	$-\frac{9}{11}$	$-\frac{36}{25}$	$-\frac{300}{137}$	$-\frac{450}{147}$
α_2	0	0	$\frac{2}{11}$	$\frac{16}{25}$	$\frac{200}{137}$	$\frac{400}{147}$
α_3	0	0	0	$-\frac{3}{25}$	$-\frac{75}{137}$	$-\frac{225}{147}$
α_4	0	0	0	0	$\frac{12}{137}$	$\frac{72}{147}$
α_5	0	0	0	0	0	$-\frac{10}{147}$
γ	1	$\frac{2}{3}$	$\frac{6}{11}$	$\frac{12}{25}$	$\frac{60}{137}$	$\frac{60}{147}$
β_0	1	2	3	4	5	6
β_1	0	-1	-3	-6	-10	-15
β_2	0	0	1	4	10	20
β_3	0	0	0	-1	-5	-15
β_4	0	0	0	0	1	6
β_5	0	0	0	0	0	-1

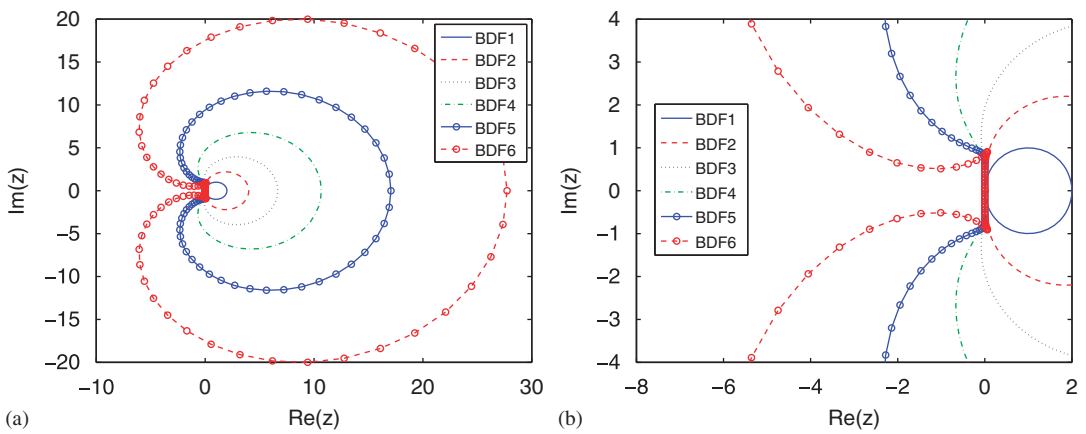

Figure 2. The stability regions for the implicit backward difference formulas of order $K=1,\ldots,6$ for: (a) large values of z and (b) small values of z.

equations, we are interested in the region near the imaginary axis (for $\mathrm{Re}(z)=0$). The BDF methods of order $K\leqslant 2$ are A-stable (stable in the entire left-hand side of the imaginary axis). Only $K=1$ is L-stable (the stability function goes to zero as $|z|\to\infty$); however, we are only interested in A-stable methods.

Let us now describe the SI method in terms of the governing equations. Note that the operator R referenced above is the same operator described for the explicit method. However, let us now write the full expression given in Equation (17) in terms of the operator L. Substituting the linear

operator defined in Equation (8) into Equation (17) results in the following system:

$$\phi_{tt} = \widehat{\phi} - \lambda \nabla \cdot \mathbf{U}_{tt}$$

$$\mathbf{U}_{tt} = \widehat{\mathbf{U}} - \lambda[\nabla \cdot (\phi_{tt}\phi_B \mathcal{I}_2) + f(\mathbf{k} \times \mathbf{U}_{tt}) - \phi_{tt}\nabla\phi_B + \gamma \mathbf{U}_{tt}]$$

where we have retained the continuous spatial operators for clarity. At this point, there is no difference between the SI formulation of a continuous Galerkin (e.g. finite elements) and a DG method. The differences arise through the method selected for the discretization of the spatial operators (see [35–37] for details on SI methods for continuous Galerkin methods for the shallow water equations). Replacing the continuous spatial operators with the DG discrete representations yields

$$\phi_{tt}^{(e)} = \widehat{\phi}^{(e)} - \lambda \left[(\widehat{\mathbf{D}}^{(e)})^{\mathscr{T}} \mathbf{U}_{tt}^{(e)} - \sum_{l=1}^{3} (\widehat{\mathbf{M}}^{(e,l)})^{\mathscr{T}} (\mathbf{U}_{tt}^{(e)} - \mathbf{U}_{tt}^{(*,l)}) \right]$$

$$\mathbf{U}_{tt}^{(e)} = \widehat{\mathbf{U}}^{(e)} - \lambda (\widehat{\mathbf{D}}^{(e)})^{\mathscr{T}} (\phi_{tt}\phi_B \mathcal{I}_2)^{(e)} - \lambda \sum_{l=1}^{3} (\widehat{\mathbf{M}}^{(e,l)})^{\mathscr{T}} [(\phi_{tt}\phi_B \mathcal{I}_2)^{(e)} - (\phi_{tt}\phi_B \mathcal{I}_2)^{(*,l)}]$$

$$-\lambda[f(\mathbf{k} \times \mathbf{U}_{tt}^{(e)}) - \phi_{tt}^{(e)} \nabla \phi_B + \gamma \mathbf{U}_{tt}^{(e)}]$$

where the flux values are specifically defined as

$$\mathbf{U}_{tt}^{(*,l)} = \tfrac{1}{2}[\mathbf{U}_{tt}^{(e)} + \mathbf{U}_{tt}^{(l)} - |\lambda_L| \mathbf{n}^{(l)} (\phi_{tt}^{(l)} - \phi_{tt}^{(e)})]$$

and

$$(\phi_{tt}\phi_B \mathcal{I}_2)_{tt}^{(*,l)} = \tfrac{1}{2}[(\phi_{tt}\phi_B \mathcal{I}_2)^{(e)} + (\phi_{tt}\phi_B \mathcal{I}_2)^{(l)} - |\lambda_L| \mathbf{n}^{(l)} (\mathbf{U}_{tt}^{(l)} - \mathbf{U}_{tt}^{(e)})]$$

These equations can now be solved as a coupled system of linear equations for ϕ_{tt} and $\mathbf{U}_{tt}$. We use a GMRES solver with Jacobi preconditioning to solve this system. While this choice of preconditioner is not optimal, the resulting iterative method is nonetheless robust and efficient in terms of computational time and memory requirements since no global matrix problem ever needs to be stored; this, of course, is true for all implicit methods solved iteratively.

In standard SI methods (e.g. see [35–37]), upon writing the fully discrete system the goal is then to apply a block LU decomposition in order to reduce the vector system of equations into an equivalent scalar system of equations; for first-order systems of equations, the resulting problem is in fact quite similar to a Helmholtz equation—the resulting problem is the Schur complement. Constructing the *Schur complement* for general DG polynomial spaces and boundary conditions remains an open problem. Thus far, we only know how to construct the Schur complement for collocated DG formulations (where the interpolation points coincide with the integration points) for a specific class of boundary conditions (see [15] and [16] for the solution to this problem for the Navier–Stokes equations). However, co-located interpolation and integration points currently exist for the quadrilateral (see, e.g., [15, 29, 30]) but they do not exist for the triangle.

A few additional comments regarding the SI discretization are in order. Since only the gravity waves (pressure gradient) are discretized implicitly, with the Rossby waves (advection operator) discretized explicitly, the model will be unconditionally stable with respect to the gravity waves for any time-step size as long as the conditional stability with respect to the Rossby waves is satisfied

by the explicit part of the SI method. This is the reason why we show both the explicit and implicit stability regions of the BDF methods in Figures 1(b) and 2. Of course one could also choose to discretize all of the terms implicitly (including the nonlinear advection operator)—this is known as the fully implicit method. The reason for choosing the SI method over the fully implicit method is that in doing so we only need to contend with a linear matrix problem; for the fully implicit method, we would have to solve a nonlinear matrix problem, that while not impossible, requires many more iterations for convergence (outer Newton loops, plus inner Krylov loops, [38]). For most types of the shallow water problems that we are considering, the SI method should be more efficient than the fully implicit method; however, this is by no means guaranteed and should be explored in detail. There are scenarios such as highly nonlinear flows, typical of the run-up of a tsunami wave, where the fully implicit method would outperform the SI method. Both approaches need to be considered and we hope to show comparison between the two methods in future work.

5. NUMERICAL EXPERIMENTS

For the numerical experiments, we use the normalized L^2 error norm

$$\|h_S\|_{L^2} = \sqrt{\frac{\int_\Omega (\phi_{\text{exact}} - \phi)^2 \, \mathrm{d}\Omega}{\int_\Omega \phi_{\text{exact}}^2 \, \mathrm{d}\Omega}}$$

computed at the cubature points to judge the accuracy of the methods. To compute the Courant number the elements are decomposed into their HO grid points (which are in fact the Fekete points) and these grid points form a fine grid that we refer to as the HO cells. The velocities and grid spacings are then defined at the centers of these cells. Using these definitions the Courant number is then given as

$$\text{Courant number} = \max \left(\frac{C \Delta t}{\Delta s} \right)_{HO}^{e} \quad \forall e \in [1, \ldots, N_e] \tag{18}$$

where $C = U + \sqrt{\phi}$ is the characteristic speed, $U = \sqrt{\mathbf{u} \cdot \mathbf{u}}$ is the magnitude of the velocity, and $\Delta s = \sqrt{\Delta x^2 + \Delta y^2}$ is the grid spacing. In addition, note that the Courant number based on the advection is given by Equation (18) with $C = U$.

We use the symbol n_r to refer to the refinement level of the grid. This variable n_r represents the number of quadrilateral subdivisions in each of the Cartesian directions. For example, $n_r = 1$ corresponds to n_r^2 quadrilaterals and $2n_r^2$ triangles; the factor of two is required since each quadrilateral is subdivided into two triangles. Examples of square domains with $n_r = 1$, $n_r = 2$, and $n_r = 4$ are shown in Figure 3.

5.1. Description of the test cases

We now describe the test cases and their solutions. It should be noted that all the tests presented below require no-flux boundary conditions at all four walls. In addition, since the linearization of the shallow water equations in conservation form is different from the linear form of the shallow water equations in non-conservation form, we have modified the analytic solutions simply by writing the solutions in terms of the conservation variables.

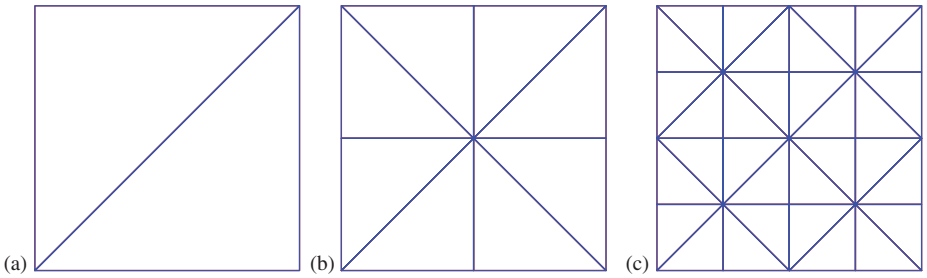

Figure 3. Grid refinement for the structured triangular grids with: (a) $n_r = 1$; (b) $n_r = 2$; and (c) $n_r = 4$.

5.1.1. Linear standing wave. This problem involves the transient solution of a linear inviscid standing wave without rotation that *sloshes* within a square basin of unit depth. From [39] we take the analytic solution as

$$\phi_S(\mathbf{x}, t) = \cos \pi x \cos \pi y \cos \sqrt{2}\pi t$$

$$U(\mathbf{x}, t) = \frac{1}{\sqrt{2}} \sin \pi x \cos \pi y \sin \sqrt{2}\pi t$$

$$V(\mathbf{x}, t) = \frac{1}{\sqrt{2}} \cos \pi x \sin \pi y \sin \sqrt{2}\pi t$$

with $(x, y) \in [0, 1]^2$, $t \in [0, 2]$, $\phi_B = 1$, $g = 1$. The source function, S, in Equation (1) is zero and the flux tensor is linearized.

5.1.2. Linear kelvin wave. This problem involves the transient solution of the linearized inviscid equations with rotation. From [6] we use the analytic solution

$$\phi_S(\mathbf{x}, t) = \exp\left(-\frac{y^2}{2}\right) \exp\left(-\frac{(x+5-t)^2}{2}\right)$$

$$U(\mathbf{x}, t) = \exp\left(-\frac{y^2}{2}\right) \exp\left(-\frac{(x+5-t)^2}{2}\right)$$

$$V(\mathbf{x}, t) = 0$$

with $f_0 = 0$, $\beta = 1$, $y_m = 0$, and $(x, y) \in [-10, 10] \times [-5, 5]$, $t \in [0, 10]$, $\phi_B = 1$, and $g = 1$.

5.1.3. Nonlinear Rossby soliton wave. This problem describes an equatorially trapped Rossby soliton wave [40]. The soliton wave starts off in the center of the domain. It then moves westward along the equator without changing its shape. The asymptotically derived analytic solution is given by

$$\phi_S(x, y, t) = h^{(0)} + h^{(1)}$$

$$U(x, y, t) = \phi(h^{(0)} + u^{(1)})$$

$$V(x, y, t) = \phi(v^{(0)} + v^{(1)})$$

where the superscripts (0) and (1) denote the zeroth- and first-order asymptotic solutions of the shallow water equations, respectively, and $\phi=\phi_S+\phi_B$ where $\phi_B=1$. The zeroth- and first-order terms are given by

$$h^{(0)}=\eta\left(\frac{-9+6y^2}{4}\right)\mathrm{e}^{-y^2/2}$$

$$u^{(0)}=\frac{\partial\eta}{\partial\xi}(2y)\mathrm{e}^{-y^2/2}$$

$$v^{(0)}=\eta\left(\frac{3+6y^2}{4}\right)\mathrm{e}^{-y^2/2}$$

and

$$h^{(1)}=c^{(1)}\eta\frac{9}{16}(-5+2y^2)\mathrm{e}^{-y^2/2}+\eta^2\Phi^{(1)}(y)$$

$$u^{(1)}=c^{(1)}\eta\frac{9}{16}(3+2y^2)\mathrm{e}^{-y^2/2}+\eta^2U^{(1)}(y)$$

$$v^{(1)}=\frac{\partial\eta}{\partial\xi}\eta V^{(1)}(y)$$

where $\eta(\xi,t)=A\,\mathrm{sech}^2B\xi$, $\xi=x-ct$, $A=0.771B^2$, $B=0.394$, and $c=c^{(0)}+c^{(1)}$, where $c^{(0)}=-\frac{1}{3}$ and $c^{(1)}=-0.395B^2$. Finally, the remaining terms are given by

$$\begin{pmatrix}\Phi^{(1)}(y)\\ U^{(1)}(y)\\ V^{(1)}(y)\end{pmatrix}=\mathrm{e}^{-y^2/2}\sum_{n=0}^{\infty}\begin{pmatrix}\varphi_n\\ u_n\\ v_n\end{pmatrix}H_n(y)$$

where $H_n(y)$ are the Hermite polynomials and φ_n, u_n, v_n are the Hermite series coefficients given in [40]. The Coriolis parameter is given by $f(y)=y$ where $(x,y)\in[-24,+24]\times[-8,+8]t\in[0,10]$ and $g=1$.

We include this analytic solution for completeness, but one cannot use this test for validating the exponential convergence of the method because the analytic solution is only first order. However, this solution can be used to check the phase speed of the soliton wave simulated by the numerical model.

5.1.4. Linear Stommel problem. The linear Stommel problem [41] is the exact steady-state solution of the linearized inviscid equations with rotation, wind stress, and bottom friction. The analytic solution to this problem can be written as

$$\phi_S(\mathbf{x})=\left[C_0\beta\frac{L}{\pi}\cos\left(\frac{\pi y}{L}\right)+f(y)\psi(x,y)+\frac{\gamma\pi}{L}\cos\left(\frac{\pi y}{L}\right)\left(\frac{C_1}{\lambda_1}\mathrm{e}^{\lambda_1x}+\frac{C_2}{\lambda_2}\mathrm{e}^{\lambda_2x}\right)\right]$$

$$U(\mathbf{x})=\phi\left(-\frac{\pi}{L}\widehat{\psi}(x)\cos\left(\frac{\pi y}{L}\right)\right)$$

$$V(\mathbf{x})=\phi\left((C_1\lambda_1\mathrm{e}^{\lambda_1x}+C_2\lambda_2\mathrm{e}^{\lambda_2x})\sin\left(\frac{\pi y}{L}\right)\right)$$

where $\psi(x,y)$, $\widehat{\psi}(x)$, C_0, C_1, C_2, λ_1, and λ_2 are given in the Appendix. The constants required to completely define the solution are $f_0=1\times10^{-4}$, $\beta=1\times10^{-11}$, $y_m=L/2$, $\gamma=1\times10^{-6}$, $g=10$, $\rho=1000$, $\tau=0.2$, $H=1000$, $L=1\times10^6$, and $\phi_B=gH$. The models are integrated between 200 and 400 days in order to reach steady state. We regard steady state as the condition when the error norms cease to decrease.

5.1.5. Nonlinear Riemann problem. We used the nonlinear Riemann problem previously (see [9]) in order to validate the spatial operators of our DG model and its slope limiters. We follow the outline of the problem presented in Toro [42]. The source function S is set to zero; this leaves a balance between the time rate of change of the conservation variable $\mathbf{q}$ and the divergence of the flux tensor. Following [42] we use

$$\phi_S(x,y,0)=\begin{cases}2g & \text{if } r\leqslant R\\ 0 & \text{if } r>R\end{cases}$$

with $\phi_B=0.5g$, $\mathbf{U}(x,y,0)=0\ \forall(x,y)\in[-20,20]^2$, where $r=\sqrt{x^2+y^2}$, $R=2.5$, and $t\in[0,0.4]$. The cylindrical wave is positioned initially at the origin and moves outward for increasing time t.

5.2. Comparison of the explicit and SI models

In the following sections we compare the accuracy and efficiency of the explicit RK35 method with the SI BDF methods of order $K\leqslant6$. For all simulations, the largest Courant number shown for RK35 represents the maximum Courant number allowed by this method. The smallest Courant number shown for the SI BDF methods represents the maximum Courant number allowed by the explicit BDF methods; the only exception is the time-step convergence study that we now describe. All Courant number limits reported are determined numerically.

5.2.1. Time-step convergence study. The first study we conduct is the rate of convergence of explicit and implicit time-integrators. For this study we use the linear standing wave problem with $n_r=1$ refinement level, corresponding to two triangular elements, and 14th order ($N=14$) polynomials. For this resolution, the best possible normalized h_S L^2 error norm that can be achieved by the model is 1×10^{-9} which we obtained experimentally as $\Delta t\to0$; this we consider to be the *exact numerical solution*.

In Figure 4 we show the results for the RK35 and BDF methods; Figure 4(a) shows the results for the free surface height, h_S, and Figure 4(b) the results for the x-momentum variable, U. The maximum time step used for RK35 is the maximum allowed by the method. For the BDF methods of order $K\leqslant4$, the smallest time steps correspond to the explicit methods, whereas the last few points correspond to the implicit methods. For the BDF methods of order $K\geqslant5$, all the simulations are obtained with the implicit methods; because these methods are HO in time, they achieve the exact numerical solution for time steps much larger than those allowed by the explicit method.

We define the time rate of convergence as

$$\text{rate}=abs\left(\sum_{i=1}^{N_T-1}\frac{\log[\text{error}_{\Delta t_{i+1}}/\text{error}_{\Delta t_i}]}{\log[\Delta t_i/\Delta t_{i+1}]}\right)$$

where Δt_i are the N_T time-step values. Figure 4(a) shows that for h_S, RK35 is indeed formally third-order accurate, and that the BDF methods achieve their theoretical values of order K. Figure 4(b)

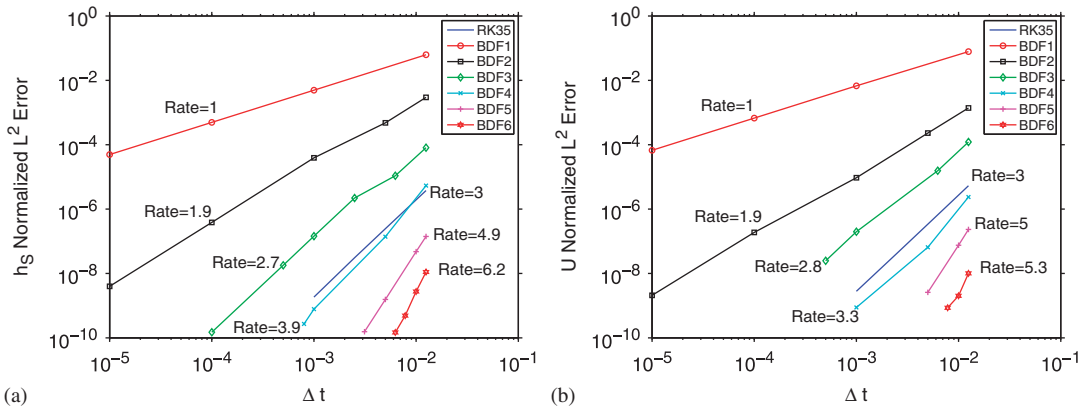

Figure 4. Linear standing wave. The normalized L^2 error as a function of time step for various time-integrators for: (a) h_S and (b) U. All runs use $n_r = 1$ and $N = 14$.

shows that for U, the theoretical convergence rates are achieved for all the methods except BDF4 and BDF6, which show a slightly lower value; these convergence rates are also identical for the y-momentum, V. Because the convergence rates between all the three variables are so similar, from here on, we exclude the momentum and only show error norms for the free surface height.

It should be mentioned that for the explicit and all implicit methods exact mass conservation (to within machine double precision) is achieved, not just for this test case but for all cases discussed below. Let us now examine the accuracy and efficiency of the SI BDF time-integrators for various test cases and compare them with the explicit RK35 time-integrator.

5.2.2. Linear standing wave. In Figures 5(a) and (b) we show the L^2 error norms and the wallclock time as functions of Courant number for various time-integrators for the linear standing wave problem. In these simulations we use 10th-order polynomials ($N = 10$) with 72 triangular elements (corresponding to $n_r = 6$). Figure 5(a) shows that the RK35 explicit method is more accurate than the SI BDF methods of order $K \leqslant 3$. For $K \geqslant 4$ the BDF methods are more accurate. Figure 5(b) shows that all of the BDF methods are more efficient than the explicit RK35 method for the same Courant number. Furthermore, the BDF methods of order $K \leqslant 6$ allow larger Courant numbers than the explicit RK35 method and the lower-order BDF methods are more efficient than the HO BDF methods.

The small Courant numbers reported for BDF6 in Figure 5(b) need to be explained. To understand these results, let us begin by discussing the stability regions of the implicit BDF methods in Figures 2(a) and (b). Along the imaginary axis all the implicit BDF methods are stable for large z; this can be observed in Figure 2(a) for $|\mathrm{Im}(z)| > 5$ for BDF4, for example. For a smaller range of z the BDF methods of orders $K \geqslant 3$ become unstable; only the BDF methods of order $K \leqslant 2$ are A-stable (stable along the entire left-hand plane including the imaginary axis). This can be observed quite clearly in Figure 2(b) for $\mathrm{Re}(z)=0$ where the BDF methods of order $K \geqslant 3$ become unstable for $|\mathrm{Im}(z)| > 1$, whereas the BDF methods of order $K \leqslant 2$ remain stable for all values of $|\mathrm{Im}(z)|$.

Once the first instability region is reached for the BDF methods of order $K \geqslant 3$, we stop increasing the Courant numbers resulting in the small Courant numbers reported in Figure 5. Note that we do this for all of the simulations throughout this paper.

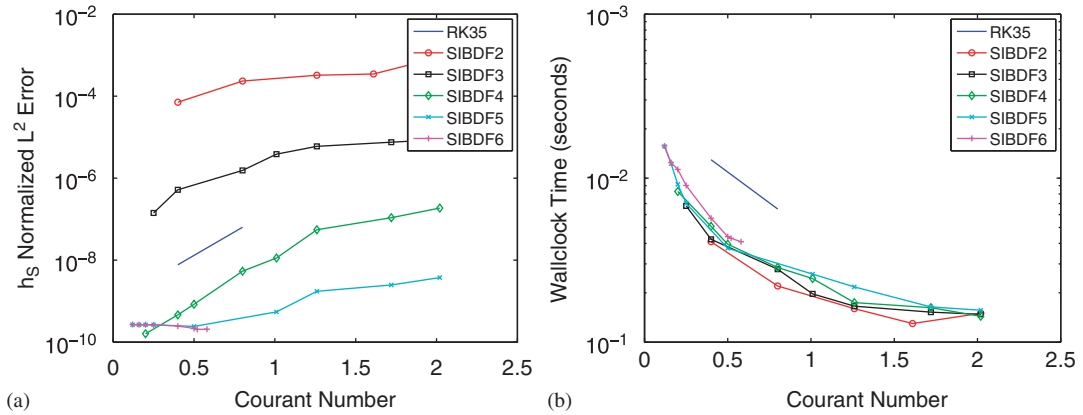

Figure 5. Linear standing wave: (a) the normalized h_S L^2 error and (b) wallclock time as the functions of Courant number for various time-integrators. All runs use $n_r = 6$ and $N = 10$.

5.2.3. Linear Kelvin wave. In Figures 6 and 7 we show the normalized h_S L^2 error norms and the wallclock time as functions of Courant number for various time-integration methods for the linear Kelvin wave problem. In these simulations we vary the polynomial orders and number of elements to maintain a constant number of grid points (8448 points). Figure 6 shows that the RK35 explicit method is more accurate than the SI BDF methods of order $K \leqslant 3$; for $K \geqslant 4$ the BDF methods are more accurate. This result holds for all polynomial orders that we studied. Figure 6 also shows that the error for all the methods decreases as we increase the polynomial order of the DG method. This trend is most noticeable for BDF4 where the error curves become more horizontal as the polynomial order is increased. This means that the truncation errors between the time and spatial discretization methods are balanced and are sufficiently accurate so that the error remains flat for increasing time step; this is particularly noticeable for 10th-order polynomials ($N = 10$) where the error remains flat beyond $CFL > 1$ (Figure 6(d)).

Figure 7 shows that all of the SI BDF methods are more efficient than the explicit RK35 method for the same Courant numbers; this is true for all polynomial orders. Furthermore, Figure 7 shows that the computational cost of the methods increases with polynomial order.

Figure 8 shows the h_S error norms and wallclock time for 10th-order polynomials ($N = 10$) for very large Courant numbers. Figure 8(b) shows that all of the SI BDF methods are more efficient than the explicit RK35 method for the same Courant numbers. The BDF methods allow larger Courant numbers with BDF2 allowing a Courant number almost one order of magnitude larger than the RK35 method.

5.2.4. Rossby soliton wave. Figure 9(a) shows the wallclock time as a function of Courant number for various time-integration methods for the nonlinear Rossby soliton wave problem. In Figure 9(b) we show the same simulations but for smaller Courant numbers. Recall that for this test we only have a first-order solution that is not sufficiently accurate for performing a convergence study; however, we can use it to determine whether the solitons are moving with the proper wave speed. We consider a numerical solution to be accurate if it agrees exactly with the semi-analytic solution in terms of the location of the highest peak of the solitons.

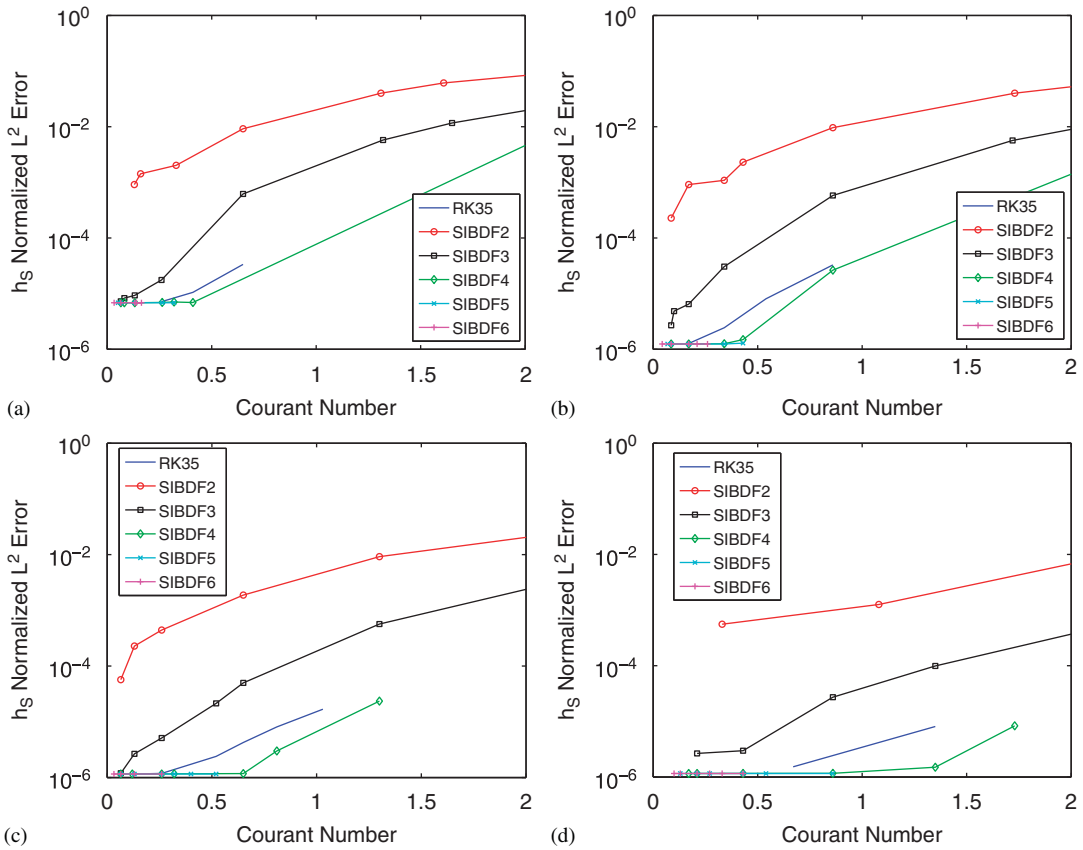

Figure 6. Linear Kelvin wave. The normalized h_S L^2 error as a function of Courant number for various time-integrators for: (a) $N=4$; (b) $N=6$; (c) $N=8$; and (d) $N=10$. All runs use 160×80 grid points.

In these simulations we use eighth-order polynomials ($N=8$) with 384 triangular elements (corresponding to $n_r^x=24$ and $n_r^y=8$). Figure 9 shows that the SI BDF methods $K<6$ are more efficient than the explicit RK35 method for the same Courant numbers. Additionally, the SI BDF methods $K\leqslant 3$ admit larger Courant numbers than the explicit RK35 method. Since this case is nonlinear, both the explicit and implicit BDF methods are used in tandem to solve the problem. Therefore in this test case, the stability regions of both the implicit and explicit BDF methods are relevant. Looking at the stability region of the explicit BDF methods given in Figure 1(b) we note that the BDF5 and BDF6 have particularly small stability regions that result in the small Courant numbers reported in Figure 9 for these methods. The BDF methods of order $K\leqslant 4$ have larger stability regions in both the explicit and implicit forms and is the reason why these methods perform more efficiently than the HO BDF ($K\geqslant 5$) and RK35 methods.

5.2.5. Linear Stommel problem. Figure 10(a) shows the wallclock time as a function of Courant number for various time-integration methods for the linear Stommel problem; in Figure 10(b) we show a closeup of the same simulations for smaller Courant numbers. For this case we have a

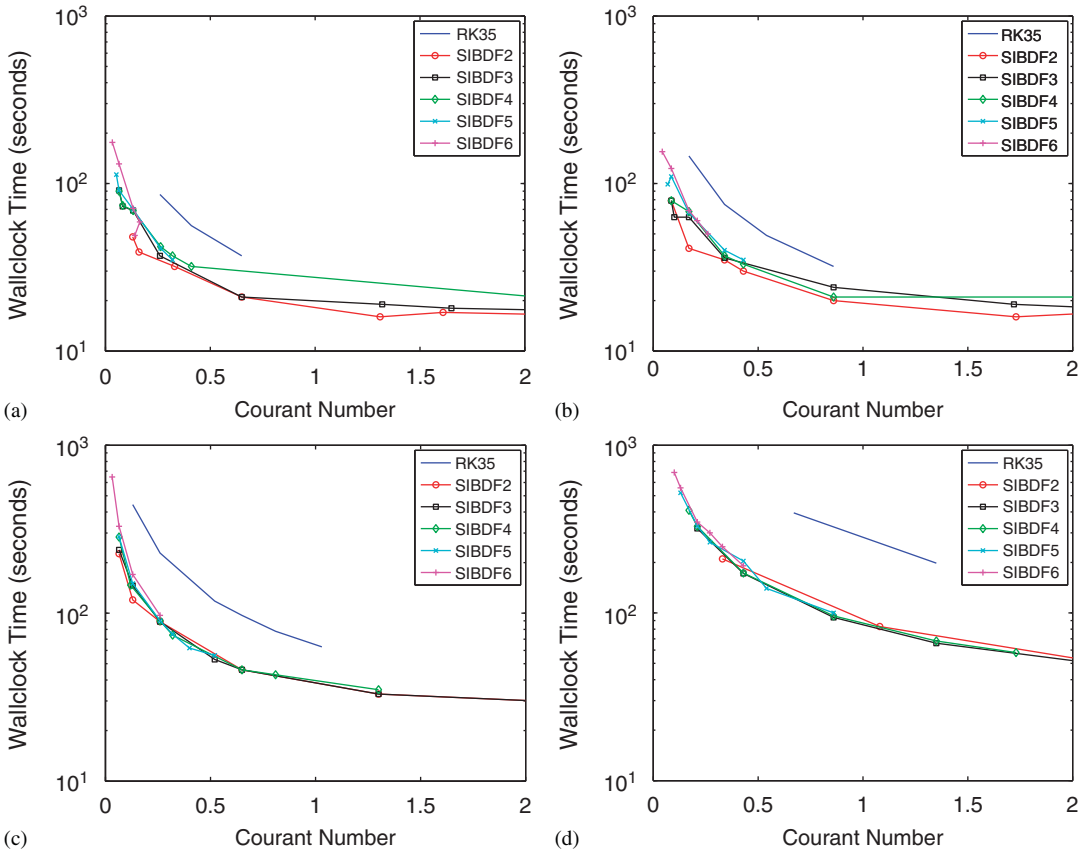

Figure 7. Linear Kelvin wave. The wallclock time as a function of Courant number for various time-integrators for: (a) $N=4$; (b) $N=6$; (c) $N=8$; and (d) $N=10$. All runs use 160×80 grid points.

steady-state analytic solution and so the accuracy of the time-integrator only plays a small role. The accuracy of the model is completely dependent on the polynomial order of the DG method; the only role that the time-integrator has is to maintain the stability of the solution while doing so as efficiently as possible.

In these simulations we use eighth-order polynomials ($N=8$) with 32 triangular elements (corresponding to $n_r=4$). Figure 10(a) shows that the BDF methods $K\leqslant 2$ allow larger Courant numbers than the other methods. More importantly, Figure 9(b) shows that all of the BDF methods are more efficient than the explicit RK35 method for the same Courant numbers. In addition, for this test case, the implicit BDF methods admit as large a Courant number as the explicit RK35 method. The fact that the implicit BDF methods are more efficient than the explicit RK35 even for the same Courant number is impressive especially since the implicit BDF methods require much more machinery to solve the problem. Recall that implicit/SI methods require the use of iterative solvers (in this case GMRES) and preconditioners (in this case Jacobi preconditioning) in order to solve the resulting linear matrix problem. Even with all of this machinery, the implicit methods are more efficient than an explicit RK method—this does not seem possible at first glance. The reason

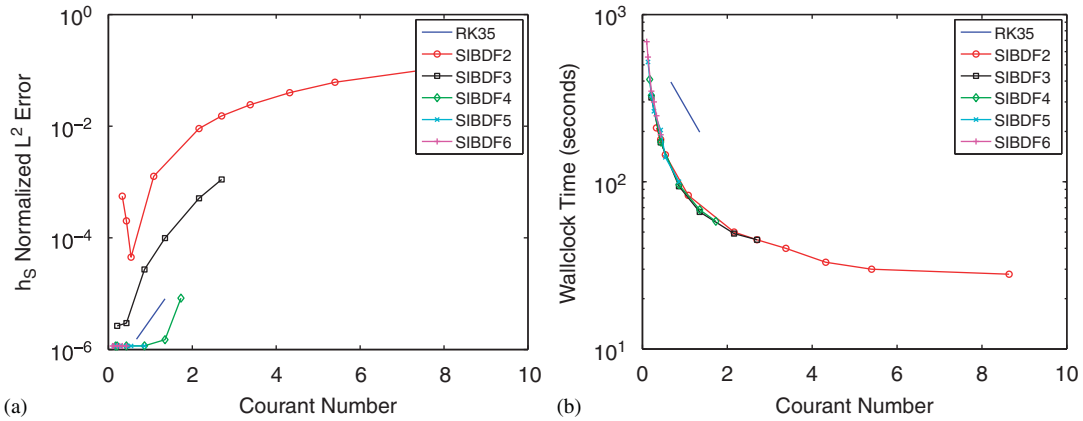

Figure 8. Linear Kelvin wave. (a) the normalized h_S L^2 error and (b) wallclock time as the functions of Courant number for various time-integrators. All runs use $n_r^x = 16$, $n_r^y = 8$, and $N = 10$.

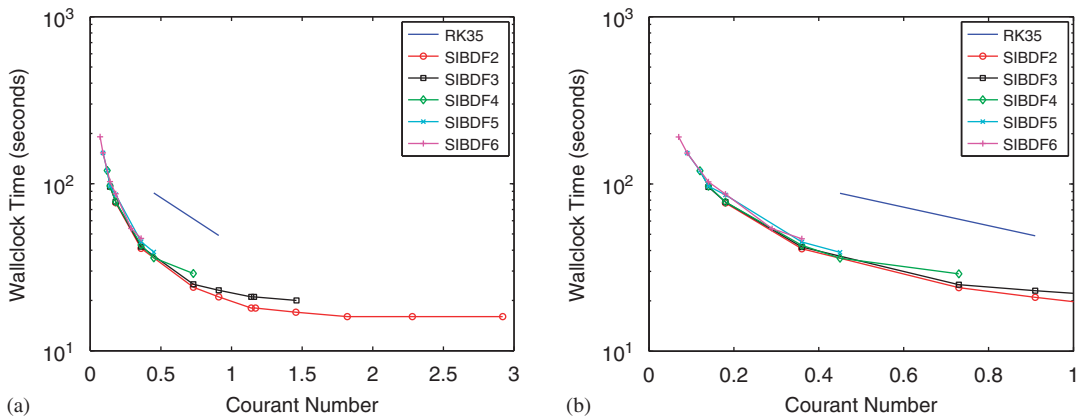

Figure 9. Rossby soliton wave. The wallclock time as a function of Courant number for various time-integrators. Figure (a) shows the results for large Courant numbers, while (b) shows them for small Courant numbers. All runs use $n_r^x = 24$, $n_r^y = 8$, and $N = 8$.

for these surprising results is simple: for Courant numbers less than 1 the implicit BDF methods require fewer than 5 GMRES iterations to converge; recall that the RK35 method requires five stages. Thus, at this range of Courant numbers the implicit BDF methods are more efficient than RK35 with respect to operation count which translates to smaller wallclock times. For the larger Courant number values, the number of iterations is greater than 5, but the larger time steps (hence fewer time-integration loops) compensate for the extra costs incurred with respect to operation count.

One final comment is in order. Since this test case is linear and the solution is steady state, we could have used infinitely large Courant numbers for the implicit BDF methods $K \leqslant 2$. We have only chosen to report the maximum Courant numbers that maintained stability for the nonlinear

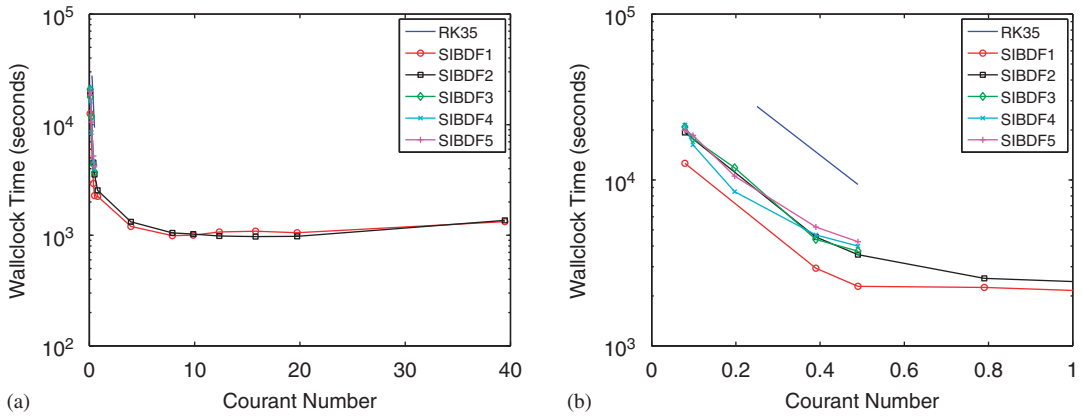

Figure 10. Linear Stommel problem. The wallclock time as a function of Courant number for various time-integrators. Figure (a) shows the results for large Courant numbers, while (b) shows them for small Courant numbers. All runs use $n_r = 4$ and $N = 8$.

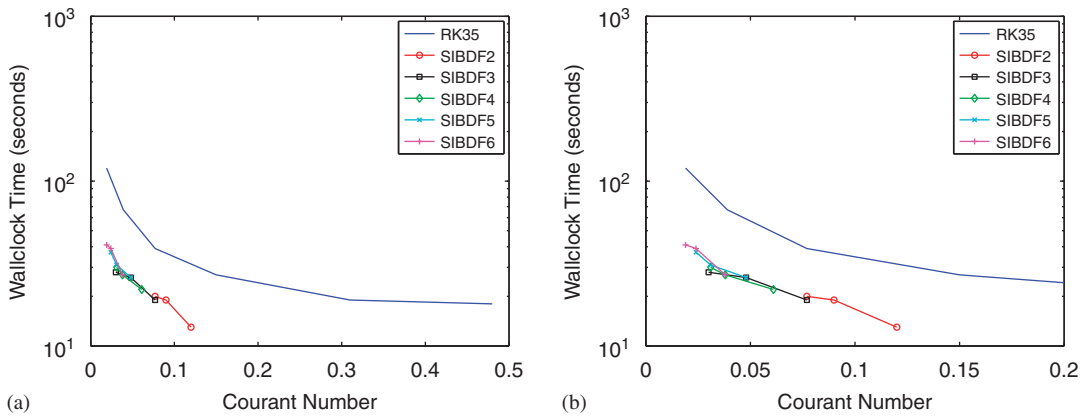

Figure 11. Riemann problem. The wallclock time as a function of Courant number for various time-integrators. Figure (a) shows the results for large Courant numbers, while (b) shows them for small Courant numbers. All runs use $n_r = 100$ and $N = 1$.

Stommel problem. Since we do not have an analytic solution to the nonlinear Stommel problem, we then use the linear problem to ensure that we are achieving L^2 errors of 1×10^{-6}, which is the *exact numerical solution* for eighth-order polynomials with $n_r = 4$ (see [9]). This means that the results reported in Figure 10 are also representative of the types of efficiency gains offered by the SI BDF methods for nonlinear problems. However, it should be noted that the reason why such large Courant numbers can be used in this test case has to do with the disparity between the speed of the gravity waves (the height of the fluid) and the Rossby waves. For the Stommel problem, the gravity waves are much faster than the Rossby waves and this will be the case for all deep ocean flows. Let us now discuss a type of flow problem for which the SI method is not well suited.

5.2.6. Nonlinear Riemann problem. In Figures 11(a) and (b) we show the wallclock time as a function of Courant number for various time-integrators for the nonlinear Riemann problem. Note that we only report the explicit RK35 and explicit BDF methods. The results for this test show that the explicit BDF methods of order $K \leqslant 4$ compete with RK35 in terms of efficiency. The stability regions of the BDF methods of order $K \geqslant 5$ are too small and, while faster than RK35 for a given Courant number, cannot compete with the maximum Courant number admitted by RK35. Let us now discuss why we do not show results for the SI BDF methods.

We cannot use the SI BDF methods for this case because the Rossby waves are faster than the gravity waves. This means that the linearization used to construct Equations (6) and (7) is no longer valid. The linearization used in the current SI formulation assumes that ϕ_B is much greater than ϕ_S, which is not true for the Riemann problem (as is evident by the initial conditions where $\phi_B = 0.5$ and $\max(\phi_S) = 2 = \mathrm{m}^2/\mathrm{s}$). We show the result of the Riemann problem only to point out the limitation of our current approach. Let us now discuss some possible solutions to this dilemma.

Defining the Froude number as the ratio of propagation speeds of Rossby (call them R) and gravity (call them G) waves, then if the flow is subcritical (i.e. $G > R$) then the fix to the problem is relatively simple. Instead of linearizing about a constant state, say ϕ_B, we linearize instead about the known state at the current time step (i.e. $\phi_S^n + \phi_B$, where n denotes the current time level). This presents few changes to the current SI approach. On the other hand, if the flow is supercritical (i.e. $R > G$) then nothing in the SI machinery can improve the efficiency since the terms responsible for the fastest waves in the system are discretized explicitly in time. In this case, the simplest solution, given the methodology described in this paper, is to switch from the SI to the explicit methods, which is achieved by setting the parameter $\delta_{\mathrm{SI}} = 0$ in the code—this, of course, has to be done with the additional constraint that the time step be changed in order to satisfy the explicit stability region of the BDF methods.

Another approach is to discretize the equations fully implicitly in time which then requires the solution of a nonlinear matrix problem. We prefer the first approach for its simplicity and to this end we are developing tools to automate the selection of the time step as well as the value of the switch δ_{SI}. Our analysis has shown that not all K order methods are created equally. For example, taking all the results collectively shows that the RK35 method behaves most like the BDF4 method and so the optimal combination would be to use the SI BDF4 as long as the linearization is valid and then switching to the explicit RK35 when the linearization breaks down or supercritical flow is encountered. The value of such a hybrid solution strategy can be appreciated by considering the SI time integration of a tsunami wave beginning in the middle of the deep ocean. As the wave approaches the coastline, the SI linearization breaks down and the flow becomes supercritical which then requires the code to switch to explicit mode. We hope to report the results of such simulations in the near future.

6. CONCLUSIONS

We present a HO family of SI time-integration methods based on backward difference formulas (BDF) for the triangular DG method as applied to the oceanic shallow water equations. We use an HO DG method defined on unstructured triangular elements that is especially useful when attempting to resolve the complex geometry resulting from the representation of coastlines in coastal ocean models. In this work, we have extended the explicit in time HO DG method that

was shown to be exponentially convergent (for smooth problems) to SI in time. The SI BDF time-integrators of order $K \geqslant 4$ are shown to yield better accuracy than the third-order explicit RK method. Furthermore, the BDF methods of order $K \leqslant 4$ require far less wallclock time to deliver these solutions. We show that the SI BDF methods, even without any optimization and without the use of sophisticated preconditioners, yield better efficiency than the explicit RK method. We expect that with the aid of preconditioning and reduction of the implicit problem to its Schur Complement will result in further speedup of the SI method compared with the fastest explicit methods. We have shown that in order to reap the full benefits of HO DG methods require developing HO time-integrators; however, the BDF methods may not be the best choice beyond $K > 4$ due to their limited implicit stability regions (for $K \geqslant 5$). To ameliorate this limitation we plan on extending these ideas to additive RK methods that will then allow the use of HO time integration without the stability limitations suffered by the BDF methods. In future work we plan on adding lateral diffusion, variable bathymetry, and wetting and drying algorithms in order to perform tsunami, storm surge, and inundation simulations.

APPENDIX

The analytic solution of the *linear Stommel problem* can be obtained by considering the linearized momentum equation as follows:

$$\nabla\phi + f(\mathbf{k}\times\mathbf{u}) + \gamma\mathbf{u} = \boldsymbol{\tau}$$

If we define the Coriolis term as $f(y) = f_0 + \beta(y - L/2)$ and the wind stress as

$$\boldsymbol{\tau} = -\frac{\tau_0}{\rho H}\cos\left(\frac{\pi y}{L}\right)\mathbf{i} + 0\mathbf{j}$$

then taking the curl yields

$$\gamma\nabla^2\psi + \beta\frac{\partial\psi}{\partial x} = -\frac{\tau_0\pi}{\rho H L}\sin\left(\frac{\pi y}{L}\right)$$

where we have written the velocity field in terms of the streamfunction as $u = -\partial\psi/\partial y$ and $v = \partial\psi/\partial x$. Assuming a separation of variables solution of the type $\psi(x, y) = \widehat{\psi}(x)\sin(\pi y/L)$ yields the following second-order ordinary differential equation (ODE) for x

$$\gamma\frac{\partial^2\widehat{\psi}}{\partial x^2} + \beta\frac{\partial\widehat{\psi}}{\partial x} - \gamma\left(\frac{\pi}{L}\right)^2\widehat{\psi} = -\left(\frac{\pi}{L}\right)\frac{\tau_0}{\rho H}$$

This ODE tells us that we need to seek solutions of the type $\widehat{\psi}(x) = Ce^{\lambda x} + C_0$ that, after substituting into the ODE, gives the two roots

$$\lambda_{1,2} = \frac{-\dfrac{\beta}{\gamma} \pm \sqrt{\left(\dfrac{\beta}{\gamma}\right)^2 + 4\left(\dfrac{\pi}{L}\right)^2}}{2}$$

with

$$C_0 = \frac{\tau_0}{\gamma \rho H} \frac{L}{\pi}$$

Imposing zero streamfunction boundary conditions at the domain boundaries yields the final solution of the streamfunction as

$$\psi(x, y) = (C_0 + C_1 e^{\lambda_1 x} + C_2 e^{\lambda_2 x}) \sin\left(\frac{\pi y}{L}\right)$$

where

$$C_1 = C_0 \frac{1 - e^{\lambda_2 L}}{e^{\lambda_2 L} - e^{\lambda_1 L}} \quad \text{and} \quad C_2 = -C_0 \frac{1 - e^{\lambda_1 L}}{e^{\lambda_2 L} - e^{\lambda_1 L}}$$

ACKNOWLEDGEMENTS

The first author (FXG) gratefully acknowledges the support of the Office of Naval Research through program element PE-0602435N and the Naval Postgraduate School through a research initiative grant.

REFERENCES

1. Alevras D. Simulations of the Indian Ocean tsunami with realistic bathymetry using a high-order triangular discontinuous Galerkin shallow water model. *Masters Thesis*, Naval Postgraduate School, 2008.
2. Schwanenberg D, Köngeter J. A discontinuous Galerkin method for the shallow water equations with source terms. In *Discontinuous Galerkin Methods*, Cockburn B, Karniadakis GE, Shu C-W (eds). Springer: Heidelberg, 2000; 289–309.
3. Li H, Liu RX. The discontinuous Galerkin finite element method for the 2d shallow water equations. *Mathematics and Computers in Simulation* 2001; **56**:171–184.
4. Aizinger V, Dawson C. A discontinuous Galerkin method for two-dimensional flow and transport in shallow water. *Advances in Water Resources* 2002; **25**:67–84.
5. Dupont F, Lin CA. The adaptive spectral element method and comparisons with more traditional formulations for ocean modeling. *Journal of Atmospheric and Oceanic Technology* 2004; **21**:135–147.
6. Eskilsson C, Sherwin SJ. A triangular spectral/hp discontinuous Galerkin method for modelling 2D shallow water equations. *International Journal for Numerical Methods in Fluids* 2004; **45**:605–623.
7. Remacle JF, Frazão SS, Li XG, Shephard MS. An adaptive discretization of shallow–water equations based on discontinuous Galerkin methods. *International Journal for Numerical Methods in Fluids* 2006; **52**:903–923.
8. Kubatko EJ, Westerink JJ, Dawson C. HP discontinuous Galerkin methods for advection dominated problems in shallow water flow. *Computer Methods in Applied Mechanics and Engineering* 2006; **196**:437–451.
9. Giraldo FX, Warburton T. A high-order triangular discontinuous Galerkin oceanic shallow water model. *International Journal for Numerical Methods in Fluids* 2008; **56**:899–925.
10. Dolejsi V, Feistauer M. A semi-implicit discontinuous Galerkin finite element method for the numerical solution of inviscid compressible flow. *Journal of Computational Physics* 2004; **198**:727–746.
11. Dolejsi V, Feistauer M, Hozman J. Analysis of semi-implicit DGFEM for nonlinear convection–diffusion problems on nonconforming meshes. *Computer Methods in Applied Mechanics and Engineering* 2007; **196**:2813–2827.
12. Feistauer M, Kucera V. On a robust discontinuous Galerkin technique for the solution of compressible flow. *Journal of Computational Physics* 2007; **224**:208–221.
13. Feistauer M, Dolejsi V, Kucera V. On the discontinuous Galerkin method for the simulation of compressible flow with wide range of Mach numbers. *Computing and Visualization in Science* 2007; **10**:17–27.
14. Dolejsi V. Semi-implicit interior penalty discontinuous Galerkin methods for viscous compressible flows. *Communications in Computational Physics* 2008; **4**:231–274.
15. Restelli M. Semi-Lagrangian and semi-implicit discontinuous Galerkin methods for atmospheric modeling applications. *Ph.D. Thesis*, Politecnico di Milano, 2007.

16. Restelli M, Giraldo FX. A conservative discontinuous Galerkin semi-implicit formulation for the Navier–Stokes equations in nonhydrostatic mesoscale modeling. *SIAM Journal on Scientific Computing* 2009; **31**:2231–2257.
17. Robert A, Henderson J, Turnbull C. An implicit time integration scheme for baroclinic models of the atmosphere. *Monthly Weather Review* 1972; **100**:329–335.
18. Arcas D, Titov V. Sumatra tsunami: lessons from modeling. *Surveys in Geophysics* 2006; **27**:679–705.
19. Geist EL, Titov V, Arcas D, Pollitz FF, Bilek SL. Implications of the 26 December 2004 Sumatra–Andaman earthquake on tsunami forecast and assessment models for great subduction zone earthquakes. *Bulletin of the Seismological Society of America* 2007; **97**:249–270.
20. Kowalik Z, Knight W, Logan T, Whitmore P. The tsunami of 26 December 2004: numerical modeling and energy considerations. *Pure and Applied Geophysics* 2007; **164**:379–393.
21. Wang X, Liu PL-F. Numerical simulations of the 2004 Indian Ocean tsunamis—coastal effects. *Journal of Earthquake and Tsunami* 2007; **1**:273–297.
22. Harig S, Chaeroni W, Pranowo S, Behrens J. Tsunami simulations on several scales: comparison of approaches with unstructured meshes and nested grids. *Ocean Dynamics* 2008; **58**:429–440.
23. Hesthaven JS. From electrostatics to almost optimal nodal sets for polynomial interpolation in a simplex. *SIAM Journal on Numerical Analysis* 1998; **35**:655–676.
24. Taylor MA, Wingate BA, Vincent RE. An algorithm for computing Fekete points in the triangle. *SIAM Journal on Numerical Analysis* 2000; **38**:1707–1720.
25. Giraldo FX. High-order triangle-based discontinuous Galerkin methods for hyperbolic equations on a rotating sphere. *Journal of Computational Physics* 2006; **214**:447–465.
26. Stroud AH. *Approximate Calculation of Multiple Integrals*. Prentice-Hall Publishing: New Jersey, 1971.
27. Cools R, Rabinowitz P. Monomial cubature rules since Stroud: a compilation. *Journal of Computational and Applied Mathematics* 1993; **48**:309–326.
28. Lyness J, Cools R. A survey of numerical cubature over triangles. *Applied Mathematics* 1994; **48**:127–150.
29. Cools R. Monomial cubature rules since Stroud: a compilation—part 2. *Journal of Computational and Applied Mathematics* 1999; **112**:21–27.
30. Giraldo FX, Hesthaven JS, Warburton T. Nodal high-order discontinuous Galerkin methods for the spherical shallow water equations. *Journal of Computational Physics* 2002; **181**:499–525.
31. Giraldo FX, Restelli M. A study of spectral element and discontinuous Galerkin methods for the Navier–Stokes equations in mesoscale nonhydrostatic atmospheric modeling: equation sets and test cases. *Journal of Computational Physics* 2008; **227**:3849–3877.
32. Cockburn B, Shu C-W. Runge–Kutta discontinuous Galerkin methods for convection-dominated problems. *Journal of Scientific Computing* 2001; **16**:173–261.
33. Spiteri RJ, Ruuth SJ. A new class of optimal high-order strong-stability-preserving time discretization methods. *SIAM Journal on Numerical Analysis* 2002; **40**:469–491.
34. De Luca TJ. Performance of hybrid Eulerian–Lagrangian semi-implicit time-integrators for nonhydrostatic mesoscale atmospheric modeling. *Master's Thesis*, Naval Postgraduate School, 2007.
35. Giraldo FX, Perot JB, Fischer PF. A spectral element semi-Lagrangian (SESL) method for the spherical shallow water equations. *Journal of Computational Physics* 2003; **190**:623–650.
36. Giraldo FX. Semi-implicit time-integrators for a scalable spectral element atmospheric model. *Quarterly Journal of the Royal Meteorological Society* 2005; **131**:2431–2454.
37. Giraldo FX. Hybrid Eulerian–Lagrangian semi-implicit time-integrators. *Computers and Mathematics with Applications* 2006; **52**:1325–1342.
38. Knoll DA, Keyes DE. Jacobian-free Newton–Krylov methods: a survey of approaches and applications. *Journal of Computational Physics* 2004; **193**:357–397.
39. Iskandarani M, Haidvogel DB, Boyd JP. A staggered spectral element model with application to the oceanic shallow water equations. *International Journal for Numerical Methods in Fluids* 1995; **20**:393–414.
40. Boyd JP. Equatorial solitary waves. Part 3: westward-travelling modons. *Journal of Physical Oceanography* 1985; **15**:46–54.
41. Stommel H. The westward intensification of wind-driven ocean currents. *Transactions of the American Geophysics Union* 1948; **29**:202–206.
42. Toro E. *Shock-Capturing Methods for Free-Surface Shallow Flows*. Wiley: New York, 2001; 245.