Computational aspects of a scalable high-order discontinuous Galerkin atmospheric dynamical core

R.D. Nair, H.-W. Choi, H.M. Tufo

Institute for Mathematics Applied to Geosciences (IMaGe), Computational and Information Systems Laboratory, National Center for Atmospheric Research (NCAR), 1850 Table Mesa Drive, Boulder, CO 80305, USA

Department of Computer Science, University of Colorado, Boulder, CO 80309, USA

Global Modeling and Assimilation Office, Code 610.1, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA

Abstract

A new atmospheric general circulation model (dynamical core) based on the discontinuous Galerkin (DG) method is developed. This model is conservative, high-order accurate and has been integrated into the NCAR’s high-order method modeling environment (HOMME) to leverage scalable parallel computing capability to thousands of processors. The computational domain for this 3-D hydrostatic model is a cubed-sphere with curvilinear coordinates; the governing equations are cast in flux-form. The horizontal DG discretization employs a high-order nodal basis set of orthogonal Lagrange–Legendre polynomials and fluxes of inter-element boundaries are approximated with Lax–Friedrichs numerical flux. The vertical discretization follows the 1-D vertical Lagrangian coordinates approach combined with the cell-integrated semi-Lagrangian conservative remapping procedure. Time integration follows the third-order strong stability preserving explicit Runge–Kutta scheme. The domain decomposition is applied through space-filling curve approach. To validate the 3-D DG model in HOMME framework, a baroclinic instability test is used and the results are compared with those from the established models. Parallel performance is evaluated on IBM Blue Gene/L supercomputers.

1. Introduction

Atmospheric general circulation modeling (AGCM) has been going through radical changes over the past decade. There have been several efforts to develop new dynamical cores (a basic component of AGCM that describes the general circulation governed by adiabatic processes at resolved scales [33]) which rely on advanced numerical methods and non-conventional spherical grid systems. The primary reason for this trend is to exploit the enormous computing potential of the massively-parallel petascale systems that will dominate future high-performance computing (HPC). The discretization schemes for these new generation models are based on ”local” methods such as finite-volume or spectral-element methods, and employing spherical grid system such as the geodesic or cubed-sphere grid that is free from singularities [10,24,31,34]. These features are the essential ingredients for any new dynamical core with horizontal resolution of the order of a few kilometers and capable of scaling to several thousands of processors.

To meet the challenges of building a new generation AGCM, the National Center for Atmospheric Research (NCAR) has developed a computationally efficient and scalable atmospheric modeling framework known as the high-order method modeling environment (HOMME) [9]. HOMME primarily employs the spectral element (SE) method on a cubed-sphere [25] tiled with quadrilateral elements and it can be configured to solve the global shallow water [30] or the hydrostatic primitive equations [18,31]. Moreover, the shallow water version has adaptive mesh refinement capability [28]. Recently, the HOMME spectral element dynamical core has been shown to efficiently scale to 32,768 processors of an IBM Blue Gene/L (BG/L) system [2].

However, for climate and atmospheric applications, conservation of integral invariants such as mass and total energy is of significant importance, and for that the hydrostatic primitive equations should be cast in flux form. A major limitation of the current SE atmospheric model is that it is not inherently conservative, and local (element-wise) conservation is not obvious. In order to resolve conservation issues, we have included the discontinuous Galerkin (DG) shallow water model [22,21] to support the HOMME framework [8]. Recently, the DG 2-D shallow water model–extended to a baroclinic model based on 3-D hydrostatic primitive equations in flux form [3,23]. This paper details the computational aspects of the DG baroclinic model in the HOMME framework.

The major motivations for choosing a DG baroclinic model as the conservative option in HOMME are the following. The DG method [4,6], which is a hybrid technique combining the finite
element and finite volume methods, is inherently conservative and shares the same computational advantages as the SE method such as scalability, high-order accuracy, spectral convergence, and thus is a potential candidate for climate modeling. Besides, the element-based HOMME grid system is well-suited for implementing the high-order DG method. Parallelization in HOMME is effected through a hybrid MPI/OpenMP design and domain decomposition through the space-filling curve approach which is again amenable to DG implementation as demonstrated in [8]. Our contribution here is the application of known DG techniques to a large and complex problem.

The remainder of the paper is organized as follows: Section 2 describes the HOMME grid system and horizontal and vertical discretization of the 3-D model. The DG model is evaluated using a baroclinic instability test and results are compared with standard models in Section 3. Parallel implementation and performance are discussed in Section 4, followed by conclusions.

2. The 3-D discontinuous Galerkin model

Here we describe the development of a baroclinic DG model in the HOMME framework [3,23]. The governing equations are cast in flux form over the cubed-sphere with non-orthogonal curvilinear coordinates. This formulation enables the applications of conservation laws, a basic requirement for the system to be conservative. The vertical component of the 3-D atmosphere is treated with the HOMME grid system is based on the cubed-sphere geometry [22,25], where the sphere (globe) is decomposed into six identical regions by an equiangular central (gnomonic) projection of the faces of an inscribed cube (Fig. 1). This results in a nonorthogonal curvilinear \((x^1,x^2)\) coordinate system free of singularities for each face of the cubed-sphere, such that \(x^1,x^2 \in [-\pi/4,\pi/4]\). Each face of the cubed sphere is partitioned into \(N_x \times N_y\) rectangular non-overlapping elements so that the total number of elements spanning the surface of the sphere \(N_{elem} = 6 \times N_x^2\).

The metric tensor \(G_{ij}\) associated with the transformation between the regular longitude–latitude \((\lambda, \theta)\) sphere with radius \(R\), and the cubed-sphere with curvilinear coordinates \((x^1,x^2)\) is defined as follows [22]:

\[
G_{ij} = A^tA, \quad A = \begin{bmatrix} R \cos \theta \partial \lambda / \partial x^1 & R \cos \theta \partial \lambda / \partial x^2 \\ R \partial \theta / \partial x^1 & R \partial \theta / \partial x^2 \end{bmatrix}
\]  

The matrix \(A\) is used for transforming the spherical velocity vector \((u,v)\) to the covariant \((u_1,u_2)\) and contravariant \((u^1,u^2)\) velocity vectors such that

\[
\begin{bmatrix} u \\ v \end{bmatrix} = A \begin{bmatrix} u^1 \\ u^2 \end{bmatrix}; \quad u^1 = G_{ij}u_j; \quad G = (G_{ij})^{-1}.
\]  

Note that the transformations (2) are free of singularities and the metric term \(G = \det(G_{ij})\) is identical for all cube faces.

2.2. Hydrostatic primitive equations on the cubed-sphere

A major feature of the baroclinic DG model is the 1D vertical Lagrangian coordinates [17,19,29]. The horizontal layers or surfaces of the atmosphere are treated as material (Lagrangian) surfaces in the vertical direction and free to move up or down with respect to a reference 1D vertical (Eulerian) coordinate as the time evolves. Over time, the Lagrangian surfaces deform and thus must be periodically remapped onto the reference coordinates. As a result the primitive equations can be recast in a conservative form without the vertical advection terms. The prognostic variables are pressure thickness \(\delta p\), covariant wind vectors \((u_1,u_2)\), potential temperature \(\theta\), and moisture \(q\). The flux form hydrostatic primitive equations, consisting of the momentum, mass continuity, thermodynamic, and moisture transport equations, can be expressed in curvilinear coordinates as follows [3,23]:

\[
\begin{align*}
\frac{\partial u_1}{\partial t} + \nabla \cdot \mathbf{E}_1 &= \sqrt{G}u_2(f + \zeta) - R_d \frac{\partial}{\partial x^1}(\ln p), \\
\frac{\partial u_2}{\partial t} + \nabla \cdot \mathbf{E}_2 &= -\sqrt{G}u_1(f + \zeta) - R_d \frac{\partial}{\partial x^2}(\ln p), \\
\frac{\partial}{\partial t}(\rho \theta) + \nabla \cdot (\mathbf{U} \theta) &= 0, \\
\frac{\partial}{\partial t}(\rho q) + \nabla \cdot (\mathbf{U} q) &= 0,
\end{align*}
\]

where

\[
\nabla \cdot \mathbf{E}_1 = \mathbf{E}_1 = (E,0), \quad \mathbf{E}_2 = (0,E), \quad E = \Phi + \frac{1}{\rho} (u_1u_1 + u_2u_2), \quad \mathbf{U} = (u^1,u^2), \quad \nabla \cdot \mathbf{U} = \sqrt{G} \rho \partial \theta / \partial \theta, \quad \Theta = \Theta(p_0/\rho), \quad \kappa = R_d/\rho C_p.
\]

where \(T\) is the temperature, \(E\) is the energy term, \(\zeta\) is the relative vorticity [21], \(\Phi = gh\) is the geopotential and \(f\) is the Coriolis parameter; \(p_0\) is a standard pressure, \(R_d\) is the gas constant and \(C_p\) is the specific heat at constant pressure. In order to close the system of Eqs. (3)–(7), we consider the hydrostatic equation as given in [1].

Fig. 1. The left panel shows a cubed-sphere grid tiled with \(N_x = 5\) elements so that the total number of elements \(N_{elem} = 6 \times 5 \times 5\). The central panel shows the computational grid where each of the elements is mapped onto a \(N_x \times N_y\) Gauss–Lobatto–Legendre (GLL) grid. The right panel shows the structure of the corresponding GLL grid with \(8 \times 8\) quadrature points \((N_x = 8)\) defined in the reference element \([-1,1] \oplus [-1,1]\).
\[ d\Phi = -C_\alpha \theta \, d\Omega. \]  

(8)

where \( \Omega = (p/p_0)^{r} \) is the Exner function.

2.3. Horizontal DG discretization

Here we adopt the established DG methods for the horizontal discretization of the baroclinic model. Since the vertical advection terms are absent in the continuous equations, the entire 3-D system can be treated as a vertically stacked shallow water (2-D) DG models [21], where the vertical levels are coupled only through the hydrostatic relation (8). Horizontal aspects of the discretization of the baroclinic model is very similar to that of the SW model, therefore, here we briefly outline the procedure. Details of the DG discretization of SW model as well as conservation laws on the cubed-sphere are given in [8,16,21].

The flux form of Eq. (3)-(7) can be written in the following compact form:

\[ \frac{\partial}{\partial \Omega} \begin{bmatrix} U \nabla_c F(U) = S(U), \end{bmatrix} \]

(9)

where \( U = [u_1, u_2, \ldots, \Delta x \theta, \Delta y \theta, \Delta z \theta] \) denotes prognostic variables (state variable vector), \( F(U) \) is flux function, and \( S(U) \) is the source term.

The computational domain \( \mathcal{D} \) is the surface of the cubed-sphere, tiled with \( N_{elem} \) rectangular elements \( \{(\mathcal{D})\} \) such that \( \mathcal{D} = \cup_{i=1}^{N_{elem}} \mathcal{D}_i \). Since the DG spatial discretization considered here is the same for every element on \( \mathcal{D} \), it is only necessary to discuss the discretization for a generic element \( \mathcal{D}_i \) with the boundary \( \partial \mathcal{D}_i \).

Let \( \tilde{U}_i \) be an approximate solution of (9) and \( \phi_n \) a test function, both defined in a finite dimensional vector space \( \mathcal{V}_h(\Omega) \). Then the weak form of (9) can be formulated by integration by parts [6] such that

\[ \int_{\mathcal{D}_i} \tilde{U}_i \phi_n \, d\Omega = \int_{\mathcal{D}_i} F(U_i) \cdot \nabla \phi_n \, d\Omega + \int_{\partial \mathcal{D}_i} F(U_i) \cdot \nu \phi_n \, d\Omega, \]

(10)

where \( \nu \) is the unit outward-facing normal vector on the boundary \( \partial \mathcal{D}_i \). Note that the solution \( U_h \) at the boundary is discontinuous, and this issue is resolved by employing an approximate Riemann solver for the flux term \( F(U_i) \cdot \nu \) in (10). A Riemann solver essentially replaces the analytic flux \( F(h) \cdot \nu \) by a numerical flux \( \tilde{F}(U_i; U_j) \).

The numerical flux provides the only mechanism by which adjacent elements interact. The flux computation is local to the element boundaries, and which has an important role in the parallel communication (MPI) between adjacent elements.

A variety of approximate Riemann solvers are available with varying complexity [32], however, for our application the Lax–Friedrichs numerical flux formulas found to an excellent choice in terms of its efficiency and accuracy [21,8]. The Lax–Friedrichs formula is given by

\[ \tilde{F}(U_i; U_j) = \frac{1}{2} \left[ (F(U_i) + F(U_j)) \cdot \nu - s(U_i - U_j) \right], \]

(11)

where \( U_i \) and \( U_j \) are the left and right limits of the discontinuous function \( \tilde{U}_i \) evaluated at the element interface, \( s \) is the upper bound for the absolute value of eigenvalues of the flux Jacobian \( F(U) \) in the direction \( \nu \). In many ways the system (3)-(7) behaves like a multilayer shallow water system. However, for the shallow water system on the cubed-sphere, the local maximum values of \( x \) in \( x^1 \) and \( x^2 \)-directions for each element \( \mathcal{D}_i \) are defined by \( x^1 = \max \left( |x^1| + \sqrt{C^1} \right) \) and \( x^2 = \max \left( |x^2| + \sqrt{C^2} \right) \), respectively (see, Nair et al. [21]). The baroclinic model is solved for finite number of vertical layers and \( x \) is computed for each layer as needed in (11).

An important aspect of the DG discretization is the choice of an appropriate set of basis functions (polynomials) that span \( \nu \). In order to evaluate the integrals in (10) efficiently and accurately, orthogonal polynomial based basis set is usually employed. Recently, Levy et al. [16] compared the efficiency of high-order nodal and modal basis functions (Legendre polynomial) for DG formulation, and showed that nodal version is computationally far more efficient than the modal version while both methods produce almost equivalent solutions. For the present study, we adopt the nodal basis set for the DG discretization as described in [8].

The nodal basis set is constructed using Lagrange–Legendre polynomials \( h_i(x), x \in [-1, 1] \), with roots at the Gauss–Lobatto quadrature points. The basis functions are defined by

\[ h_i(x) = \frac{(x - 1)(x + 1)I_{i0}(x)}{N(N + 1)I_{N+1}(x)}(x - x_i), \]

(12)

where \( I_{i0}(x) \) is the Legendre polynomial of degree \( N \). In order to take advantage of efficient quadrature rules, new independent variables \( x_1^i, x_2^i \in [-1, 1] \) are introduced in such a way that \( x_1^i, x_2^i \in [-1, 1] \).

The high-order accuracy of the HOMME spectral element grids is directly related to the Gauss–Lobatto–Legendre (GLL) quadrature points. For a given order of accuracy \( N_h \) of the SE scheme, every element in the computational domain is mapped onto the unique GLL grid, is often referred to as the reference element, bounded by \([-1, 1] \) \( \times [-1, 1] \) with \( N \times N \) GLL grid points. In Fig. 1, central and right panels shows the computational domain with \( N_{elem} = 216 \) elements and the GLL grid with \( N_{elem} = 8 \), respectively. The DG baroclinic model exploit this grid structure and built onto the HOMME framework [23].

The approximate solution \( \tilde{U}_h \) belongs to the finite dimensional space \( \mathcal{V}_h(\Omega) \) expanded in terms of a tensor-product of the Lagrange basis functions defined at the GLL points such that

\[ \tilde{U}_h(x_1^i, x_2^i) = \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{U}_h^i \phi_i(x_1^i) \phi_j(x_2^i) h_i(x_1^i) h_j(x_2^i). \]

(13)

The test function \( \phi_n \) also follows the same representation. By substituting the expansion (13) for \( \tilde{U}_h \) and \( \phi_n \) in the weak formulation (10), exploiting the orthogonality of the polynomials, simplifications leads to the following system of ordinary differential equations [8,21]:

\[ \frac{d\tilde{U}_h}{dt} = \mathbf{L} h, \quad \tilde{U}_h(0) = \tilde{U}_h(0) \times \Omega. \]

(14)

The above ODE can be solved with an explicit time integration strategy such as the third-order strong stability preserving Runge–Kutta (SSP–RK) scheme by Gottlieb et al. [11]:

\[ \tilde{U}_h^{n+1} = \tilde{U}_h^n + \Delta t \mathbf{L} \tilde{U}_h^n, \]

\[ \tilde{U}_h^{n+1} = \tilde{U}_h^n + \frac{3}{4}\Delta t \tilde{U}_h^n + \frac{1}{4}\left[ \tilde{U}_h^n + \Delta t \mathbf{L} \tilde{U}_h^n \right]. \]

(15)

where the superscripts \( n \) and \( n + 1 \) denote time levels \( t \) and \( t + \Delta t \), respectively. The above three-stage explicit R–K scheme is robust and popularly used for DG methods [5,6]. Currently, the DG baroclinic model employs SSP-RK (15) for time integration. Unfortunately this scheme is very time-step restrictive and therefore it may not be efficient for very long term climate-scale integration. Development of efficient time integration scheme is ongoing research.

2.4. Vertical discretization

The Lagrangian vertical coordinate was first introduced by Starr [29] in the early days of atmospheric modeling. Recently, Lin [17] reintroduced the “vertically Lagrangian” coordinates in finite-volume (based on a piecewise parabolic method) dynamical core
and demonstrated its application with practical climate simulations [7]. The “control-volume approach” employed in [17] for the pressure gradient terms avoids the explicit usage of these terms in the momentum equations. However in the DG formulation, such an approach is not consistent with the high-order horizontal DG discretization and the pressure gradient terms are required in the momentum Eqs. (3) and (4) as used in [29].

The vertical discretization based on the 1-D vertical Lagrangian coordinates of [29] follows evolve and remap approach. Fig. 2 shows the schematic illustration of the vertical Lagrangian coordinates, where the smooth curves represent the Lagrangian surfaces and dashed lines indicate the corresponding Eulerian reference surfaces. At initial time Lagrangian surfaces coincide with the reference surfaces, however, as time evolves the Lagrangian surface are free to move up or down with respect to the reference layers. A terrain following Lagrangian vertical coordinate can be constructed by treating any reference Eulerian coordinate as a material surface. For the present study we employ the terrain following ‘eta’ coordinates \( \eta = \eta(p, p_s) \) [27], which is a hybrid of pressure and normalized pressure coordinates \( \eta = p/p_s \) (where \( p_s \) is the surface pressure), as the vertical reference coordinates.

Over time, the Lagrangian vertical coordinates deforms and must be periodically remapped onto a reference coordinate. The hydrostatic atmosphere is vertically divided into a finite number of \( (N_{lev}) \) pressure intervals or pressure thicknesses. The pressure thickness \( \delta p \) is defined as difference between the pressure interfaces such that \( \delta p_k = p_{k+1/2} - p_{k-1/2} \), where the suffix ‘\( k \pm 1/2 \)’ denotes the interfaces (half-levels) and staggered with respect to the full-level k as shown in Fig. 2.

The state variables \( U_k = [u_k, u_s, Ap, \Delta \rho \theta, \Delta \rho q]^T \) are predicted at the full-levels (dotted line in Fig. 2) for \( k = 1, 2, \ldots, N_{lev} \). With these notations, pressure at the top of the atmosphere is \( p_{top} = p_{1/2} \), the surface pressure is given by \( p_1 = p_{top} = p_{1/2} + \sum_{k=2}^{N_{lev}} \delta p_k \), and the surface geopotential (includes topography) \( \Phi_1 = \Phi_{1/2} \). Fig. 3 schematically shows the vertical staggering of the variables, for an element with 5 \( \times 5 \) GLL grid points.

At every time step \( \delta p_k \) is predicted at model levels \( k = 1, 2, \ldots, N_{lev} \) and then used to determine the pressure at the interfaces \( k \pm 1/2 \) by summing the pressure thickness from top \( (p_{top}) \) to the layer in question:

\[
p_{k+1/2} = p_{top} + \sum_{l=1}^{k} \delta p_l, \quad k = 1, 2, \ldots, N_{lev}. \tag{16}
\]

The geopotential at the interfaces is obtained by using the discretized hydrostatic relation (8), \( \delta \Phi_k = -C_p \delta \rho \delta H_k \), and summing the geopotential height from bottom \( (\Phi_k) \) to top, i.e.,

\[
\Phi_{k-1/2} = \Phi_k + \sum_{l=k}^{N_{lev}} \delta \Phi_l, \quad k = N_{lev}, \quad N_{lev} = 1, \ldots, 1. \tag{17}
\]

The discretized momentum equations require \( p \) and \( \Phi \) at full-levels, and they are interpolated from corresponding values given at half-levels by Eqs. (16) and (17) using a method described in [27]. Note that \( p_{top} \) and \( \Phi_k \) serve as the top and bottom (surface) boundary conditions, respectively, for the vertical discretization.

### 2.5. Conservative vertical remapping

For the baroclinic model, the velocity fields \( (u_1, u_2) \), the moisture \( q \), and total energy \( (E_x) \) are remapped onto the reference Eulerian coordinates using the 1-D conservative cell integrated semi-Lagrangian (CISL) method of Nair and Machenhauer [29]. The temperature field \( \theta \) is retrieved from the remapped total energy \( E_x \) as discussed in [17]. The vertical remapping process is described as below, which requires a reference 1-D grid. A variety of vertical coordinates systems are available for remapping, however, we choose the hybrid pressure coordinate \( \eta = \eta(p, p_s) \) as the reference coordinate [27]. The values \( \eta \) monotonically increases form the top \( (\eta_{top}) \) of the atmosphere to the earth’s surface level \( \eta_1 \), as shown in Fig. 4.

Consider a 1-D vertical domain \( l \equiv [\eta_{top}, \eta_1] \) spanned by two monotonic grid lines, \( \eta_{k-1/2} \) and \( \eta_{k+1/2} \), where \( k = 1, 2, \ldots, N_{lev} + 1 \), representing Lagrangian (source) and Eulerian (target) grids, respectively. There exist a one-to-one correspondence between the grid points on these two grids. The grid spacing is defined by \( \delta \eta_k = \eta_{k+1/2} - \eta_{k-1/2} \) and \( \delta \eta_k^* = \eta_{k+1/2}^* - \eta_{k-1/2}^* \), for \( k = 1, 2, \ldots, N_{lev} \), such that \( \eta_1 - \eta_{top} = \sum_{k=1}^{N_{lev}} \delta \eta_k = \sum_{k=1}^{N_{lev}} \delta \eta_k^* \).

Let \( \phi(\eta) \) be a generic density function for \( \eta \in L \), representing any of the variables to be remapped. By remapping we transfer the average values of \( \phi(\eta) \) defined in the Lagrangian grid to that of \( \phi(\eta^*) \) in the Eulerian grid with mass conservation as a constraint. Fig. 4 schematically depicts the vertical Lagrangian and Eulerian grids over the interval \([\eta_{top}, \eta_1]\), and the relative positions of the grid.
A natural question arises with the treatment of Lagrangian coordinates is how frequent the remapping need to be performed? Our experience suggests that the remapping frequency is in general \( \Delta t \), where \( \Delta t \) is the model time step, and this estimate is consistent with [17].

3. Numerical test

There are only a few idealized test-cases available for evaluating 3-D atmospheric dynamical cores. Most of them rely on the forced solution obtained by adding strong diffusion or explicit forcing terms. Typically such test-cases require long-term integrations for converged solution [12] and hence not suitable for debugging. The baroclinic wave instability test proposed by Jablonowski and Williamson [13] fills this gap, and is an excellent tool for debugging and performing preliminary validation new dynamical cores. The initial conditions for the test are quasi-realistic and defined by analytic expressions which are a steady-state solution of the hydrostatic primitive equations. Baroclinic waves are triggered in the northern hemisphere when using the initial conditions with an overlaid perturbation. This test may be integrated up to 30 model days, and has the predictability limit up to approximately 16 days [13]. We consider this particular test for evaluating the DG dynamical core.

3.1. Initial Conditions

The surface pressure is initially given as \( p_s = 1000 \) hPa and the initial spherical wind velocities are defined as follows. We follow the notations as used in [13], where initial conditions are prescribed in horizontal longitude–latitude \((\lambda, \varphi)\) coordinates and the vertical coordinate \( \eta = \eta(p, p_s) \).

The balanced initial wind velocity \((u, v, w)\) is given by

\[
\begin{align*}
  u(\lambda, \varphi, \eta) &= u_0 \cos^2 \eta_s \sin^2(2\varphi), \\
  v(\lambda, \varphi, \eta) &= 0,
\end{align*}
\]

where \( u_0 = 35 \, \text{m/s}, \eta_s = (\eta - \eta_0)\pi/2, \eta_0 = 0.252, \eta \in [0, 1] \). The characteristic atmospheric temperature profiles are given as

\[
T(\eta) = \begin{cases}
  T_0 \left( \frac{\eta'}{\eta_0} \right)^2 & \text{for } \eta_0 \geq \eta \geq \eta_1, \\
  T_0 \left( \eta' + \Delta T(\eta_0 - \eta)^2 \right) & \text{for } \eta < \eta_1,
\end{cases}
\]

where \( T_0 = 288 \, \text{K}, \Gamma = 0.005 \, \text{K/m}, R_d = 287 \, \text{J/kg K}, g = 9.806 \, \text{m/s}^2, \Delta T = 4.8 \times 10^7 \, \text{K}, \) and \( \eta_1 = 0.2 \) is the tropopause level. By defining the following dummy functions:

\[
A_c = \begin{cases}
  -2 \sin^6 \varphi \left( \cos^2 \varphi + \frac{1}{3} \right) + \frac{10}{63}, & \text{for } \eta_0 \geq \eta \geq \eta_1, \\
  \frac{8}{3} \cos^4 \varphi \left( \sin^2 \varphi + \frac{2}{3} \right), & \text{for } \eta < \eta_1
\end{cases}
\]

the potential temperature is given as follows:

\[
\Theta(\lambda, \varphi, \eta) = \left( \frac{3}{4} \right) \sin \eta_s \cos^2 \eta_s \left( 2u_0 \cos^2 \eta_s A_c + B_c \right) / P(\lambda, \varphi, \eta),
\]

where \( P(\lambda, \varphi, \eta) = \left( \frac{\rho}{\rho_0} \right)^{\gamma/\kappa} \) is the Exner function, \( \alpha = 6.371229 \times 10^7 \) m denotes the radius of the earth and \( \Omega = 7.29212 \times 10^{-5} \) s\(^{-1}\) is the earth's angular velocity. The surface geopotential given as

\[
\phi_s(\lambda, \varphi) = u_0 \cos^2 \frac{\pi}{2} (\eta_s - \eta_0) \left( u_0 \cos^2 \frac{\pi}{2} (\eta_s - \eta_0) \right) A_c + B_c,
\]

where \( \eta_0 = 1 \).
The perturbation field required to be added to the zonal wind field (21) for triggering baroclinic waves is given by

\[ u'(\lambda, \varphi, \eta) = u_p \exp \left( -\left( \frac{r}{R} \right)^2 \right) \]

(28)

with maximum amplitude \( u_p = 1 \) m/s, radius \( R = a/10 \) and the great-circle distance \( r \), which is given by

\[ r = a \arccos(\sin \varphi_c \sin \varphi + \cos \varphi_c \cos \varphi \cos(\lambda - \lambda_c)) \]

(29)

where \((\lambda_c, \varphi_c) = (\pi/9, 2\pi/9)\) is the center of the perturbation geographically located at \((20^\circE, 40^\circN)\).

3.2. The model simulation results

The horizontal resolution of the model is controlled by two parameters which are number of elements on a face of the cubed-sphere along any coordinate direction \( (N_e) \) as well as the...
number of 1-D quadrature points on an element \( (N_v) \). Several choices are possible for the horizontal resolution with \( N_e \) and \( N_v \) combinations (also known as “h–p refinement” in SE literature). For this study we choose 26 full vertical levels \( (N_{lev} = 26) \) with \( \eta_{top} \approx \) 0.002515 (approximately 2.5 hPa) [27]. The remapping frequency \( (R_f) \) is in general \( \Delta t c(10) \), for the baroclinic instability test \( R_f \) is set to 18. However, we note that this choice of \( R_f \) is not optimal, it is dependent on the horizontal resolution as well as the forcing; more research is needed in this regard and will be a topic for future study.

**Fig. 6.** Orthographic projection of the simulated surface pressure in hPa (top panels) and the temperature fields in K at the 850 hPa pressure level (bottom panels) at day 8; using the DG baroclinic model (panels (a) and (c)) and the reference global spectral model (panels (b) and (d)). The horizontal resolution of the models are approximately 1.4° at the equator. Both models employ 26 vertical levels \( (N_{lev} = 26) \).
Fig. 7. Simulated surface pressure (hPa) at day 11 for the baroclinic instability test with the DG model (left), the NCAR global spectral model (central) and the FV dynamical core (right). The horizontal resolution is approximately 0.7° for the DG and global spectral model, while the FV model employs a resolution of 0.5° × 0.625°.

Since our primary focus here is the computational aspects of DG baroclinic model, we exclude the detailed analysis (including statistics) for the 30 day simulation with various resolutions as recommended in [13]. However, we compare the DG baroclinic model results with that of the established climate models in the NCAR Community Atmospheric Model (CAM) framework [7]. The reference models used are global spectral transform model and a finite-volume (FV) model, both rely on regular latitude–longitude global grids. However, the cubed-sphere has a quasi-uniform grid structure and over the polar regions it has ‘coarse’ resolution as compared to the latitude–longitude grids, for a given set of grid points. Moreover, the GLL grid structure of the elements results in a highly non-uniform computational grid. Because of the inconsistency of the grid resolutions of these two grid systems [15], we compare an average grid resolution at the equator of the cubed-sphere by using the approximate relation \( 90° / \left( N_e \left( N_v - 1 \right) \right) \), with that of the regular spherical grid in polar coordinates. Nevertheless, with this setup the polar regions of the latitude-longitude sphere get approximately 30% more grid points as opposed to the polar panels of the cubed-sphere. For visualization, the data from the cubed-sphere grid are interpolated onto the latitude–longitude grid.

Fig. 5 demonstrates the evolution of surface pressure \( (p_s) \) as a result of the triggering of baroclinic waves in the northern hemisphere. Panels (a)–(c) in Fig. 5 show early development of the cyclogenesis until simulation day 4, where the surface pressure deviation is less than 2 hPa. The generation of typical gravity wave perturbations at early hours of integration is visible in panel (b). By day 4 a well-defined wave train is established, and by days 6–9 a significant deepening of the highs and lows take place [13]. Panel (d) of Fig. 5 shows the surface pressure at day 6. Approximate horizontal resolution used for the simulation is 1.6° at the equator, for that we use the cubed-sphere combination \( N_e = 8 \) and \( N_v = 8 \).

Fig. 6 shows the simulated pressure and temperature at 850 hPa pressure surface at day 8 by the DG baroclinic model and the NCAR global spectral model. At day 8 the baroclinic instability waves in surface pressure are well developed and the wave patterns in temperature fields are clearly visible. The horizontal resolution of the NCAR reference model is set to be T85, i.e., \( 128 \times 256 \) latitude–longitude grid (excluding the pole points) with a horizontal resolution of 1.4° [13]. To match with this resolution approximately at the equator, we choose \( N_e = 12 \) and \( N_v = 6 \) for the DG model. Note that with this configuration DG model employs fewer number of grid points as compared to global spectral model. The surface pressure shown in Fig. 6a by the DG model is smooth and free from spectral ringing (spurious oscillation) as opposed to the reference model shown in Fig. 6b. The baroclinic wave patterns over the temperature fields are well captured by the DG model (Fig. 6c).

The DG model has been tested for convergence by varying \( N_e \) or \( N_v \) (for brevity results are not shown). At a resolution of 0.7° with \( N_e = 26 \) and \( N_v = 6 \), the solution attains convergence and which is consistent with the comparison results given in [13]. Fig. 7 shows the surface pressure at day 11 for the DG model (left panel) and the reference models (central and right panels). The reference models used are NCAR global spectral model with resolution T170 \((256 \times 512 \) grid) and the FV model with horizontal resolution 0.5° × 0.625° \((361 \times 576 \) grid). All these models simulated strikingly similar pressure patterns even though the underlying numerics are entirely different. Fig. 8 shows the evolution of minimum pressure for 16 days of model simulation by the DG and reference models (using the same resolution as discussed above). The minimum surface pressure is almost identical until day 9 and this behavior changes slightly when the baroclinic instability intensifies. The minimum \( p_s \) for the DG model is more similar to that of the FV model until day 13. DG results again agree with the reference solutions.
4. Parallel implementation

4.1. Domain decomposition

The salient feature of the cubed-sphere geometry is its grid structure which provides a quasi-uniform grid spacing over the sphere. The cubed-sphere face edges are discontinuous, however they are physically connected by using a set of local transformations provided in [22]. Fig. 9 shows the topological (logical) connectivity of the six sub-domains which span the cubed-sphere. Since every sub-domain has the identical square pattern it is rather easy to tile the entire sphere with rectangular elements as shown in Fig. 1. In other words, for the cubed-sphere the computational domain is naturally decomposed into \( N_{\text{elem}} \) number of non-overlapping rectangular elements.

The parallel implementation of HOMME is based on a hybrid MPI/OpenMP approach. HOMME parallel environment is designed to support the element based methods such as spectral element and discontinuous Galerkin methods. The natural way to parallelize the element based method is to allocate a chunk of elements to each processor. Each element only needs information from adjacent elements, so the domain decomposition reduces to a standard graph partitioning problem using algorithm based on either Hilbert space-filling curve approach by Sagan [26], Dennis et al. [8] or METIS [14].

The approach generates best partitions when \( N_{\text{elem}} = 2^\ell 3^m 5^n \), where \( \ell, m, \) and \( n \) are integers. The first step to partitioning the computing grid involves the mapping of the 2-D surface of the cubed-sphere into a linear array. Fig. 10 illustrates the Hilbert space-filling curve and elements when \( N_{\text{elem}} = 24 \). Then the second step involves partitioning the linear array into \( P_x \) contiguous groups, where \( P_x \) is the number of MPI tasks. The space-filling curve partitioning creates contiguous groups of elements and load-balances the computational and communication costs. For more detail communication frameworks of HOMME, see [8].

To perform parallel computing experiments, we employ the 2048 processors of a IBM Blue Gene/L (BG/L) supercomputer at NCAR. The configuration of these systems is summarized in Table 1. A Message Passing Interface (MPI) job for IBM BG/L machine can be run in co-processor mode (i.e., a single MPI task runs on each compute node) or in virtual-node mode (i.e., two MPI tasks are run on each compute node). To determine sustained MFLOPS per processor, the number of floating-point operations per time step was measured for the main DG time stepping loop using hardware performance counters for IBM supercomputer.

- IBM Blue Gene/L system uses libmpihpm library and its link and code examples are given as follows:
  
  ```
  add -I$(BGL_LIBRARY_PATH) -lmpihpm_f -lbgl_perfctr.rts ...
  call trace_start()
  call dg3d_advance()
  call trace_stop()
  ...
  ```

4.2. Parallel performance

To assess parallel performance we consider baroclinic instability test [13]. Note that all writing and printing functions are turned off during performance evaluations. Table 2 summarizes the dynamical core configurations for 3-D model used in parallel performance experiments (note that the R–K time steps used are not optimal).

Fig. 11 demonstrates parallel performance, where 2048 processors of a IBM Blue Gene/L machine sustains between 248 and 257 MFLOPS per processor with coprocessor mode and sustains between 220 and 247 MFLOPS per processor with virtual-node mode for 1° horizontal resolution (i.e., \( N_{\text{elem}} = 1944 \)). For 0.5° horizontal resolution (i.e., \( N_{\text{elem}} = 7776 \)), 2048 processors of a IBM Blue Gene/L machine sustains between 247 and 261 MFLOPS per processor with coprocessor mode and sustains between 232 and 248 MFLOPS per processor with virtual-node mode. The CPU cost in wall clock hours is shown in Fig. 12.

Table 3 summaries the percentage of peak performance for strong scaling results for 2048 processors of a IBM Blue Gene/L system. Two thousand and fourty-eight processors of a IBM Blue Gene/L machine sustains 9.5% and 9.3% of peak performance for coprocessor and virtual-node modes, respectively.

---

**Table 1**

Comparison between co-processor and virtual-node modes of IBM Blue Gene/L supercomputers

<table>
<thead>
<tr>
<th>Mode type</th>
<th>Coprocessor (CO)</th>
<th>Virtual-node (VN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clock cycle</td>
<td>0.7 GHz</td>
<td>0.7 GHz</td>
</tr>
<tr>
<td>Memory/proc</td>
<td>0.5 GB</td>
<td>0.25 GB</td>
</tr>
<tr>
<td>Tasks/node</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

---

**Table 2**

Summary of dynamical core configurations for Baroclinic Instability Test used in 3-D DGM model on HOMME

<table>
<thead>
<tr>
<th>Horizontal resolution</th>
<th>Vertical resolution ((N_{\text{lev}}))</th>
<th>Number of elements ((N_{\text{elem}}))</th>
<th>Time step ((dt))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1°</td>
<td>26</td>
<td>1944</td>
<td>6 s</td>
</tr>
<tr>
<td>0.5°</td>
<td>26</td>
<td>7776</td>
<td>3 s</td>
</tr>
</tbody>
</table>

---

![Fig. 9. The logical orientation of the six faces (sub-domains) of the cubed-sphere.](image9.png)

![Fig. 10. A mapping of a Hilbert space-filling curve for \( N_{\text{elem}} = 24 \) cubed-sphere grid.](image10.png)
A conservative high-order discontinuous Galerkin (DG) baroclinic model has been developed and implemented in NCAR's high-order method modeling environment (HOMME). The 3-D DG model rely on hydrostatic primitive equations cast in flux-form. The computational domain is the singularity-free cubed-sphere geometry. The DG discretization uses high-order nodal basis set of Lagrange–Legendre polynomials and fluxes of inter-element boundaries are approximated with Lax–Friedrichs numerical flux. The vertical discretization follows the 1-D vertical terrain-following Lagrangian coordinates approach. Consequently, the 3-D atmosphere is divided into a finite number of Lagrangian 2-D horizontal surfaces (layers) and DG discretization is performed for each layer. Over time, the Lagrangian surfaces are subject to deformation, and therefore the state variables are periodically remapped onto a reference coordinate. The vertical remapping is performed with the conservative cell-integrated semi-Lagrangian method.

The time integration of the model rely on the explicit third-order strong-stability preserving Runge–Kutta scheme. The DG dynamical core is validated using a baroclinic instability test and results are compared with that of the NCAR global spectral model and a finite-volume model available in community atmospheric modeling framework. The DG model successfully simulated baroclinic instability growth and the results are comparable with that of the reference models. Idealized tests show that the new model simulation is accurate, smooth and free from ‘spectral ringing’ (spurious oscillations) as opposed to the global spectral model.

However, the DG baroclinic model employs explicit Runge–Kutta time integration with moderate time step size resulting in a lower integration rate. This may not be adequate for long-term climate simulations. New time integration schemes for the baroclinic model is under investigation. This includes semi-implicit, implicit–explicit Runge–Kutta and the Rosenbrock class of time integrators. The vertical Lagrangian treatment is an elegant approach by which the 3D model can be treated as vertically-stacked shallow water (2D) models. Nevertheless, the remapping frequency used for the present study is somewhat ad hoc, and more rigorous studies are required in this regard.

Parallel experiments are tested on 2048 processors of a IBM Blue Gene/L supercomputer at NCAR. Conservative 3-D DG baroclinic model sustains 9.3% peak performance (i.e., 261 MFLOPS/processor) for IBM Blue Gene/L’s coprocessor mode and sustains 8.9% peak performance (i.e., 248 MFLOPS/processor) for IBM Blue Gene/L’s virtual-node mode. The preliminary performance study with DG model is encouraging and comparable to that of the SE baroclinic model in HOMME. Extending the new DG model to a fully-fledged climate model with realistic physics packages and scaling up to $O(100\, K)$ processors is our ultimate goal.

5. Conclusion

Acknowledgements

The Authors would like to thank Prof. Christiane Jablonowski for the baroclinic instability test comparison data, Dr. John Dennis for the performance study support and Dr. Peter Lauritzen for the internal review of the manuscript. This project is supported by the NSF sponsorship of the NCAR and DOE awards #DE-FG02-04ER63870 and #DE-FG02-07ER64646. Computer time is provided through the NSF MRI Grants CNS-0421498, CNS-0420873, and CNS-0420985, and through the IBM Shared University Research Program.

References


