GPU Accelerated Paired Explicit Runge-Kutta Methods for High-Order Spatial Discretizations

Brian C. Vermeire
Department of Mechanical, Industrial, and Aerospace Engineering, Concordia University, Montreal, Canada
*brian.vermeire@concordia.ca

Abstract—The ability to perform unsteady scale-resolving simulations, such as Large Eddy Simulation (LES) and Direct Numerical Simulation (DNS), relies on accurate, efficient, and stable discretizations that are synergistic with modern high-performance computing architectures. In this paper we explore a combination of Graphical Processing Units (GPUs) combined with Paired Explicit Runge-Kutta (P-ERK) temporal discretization for high-order accurate LES/DNS solvers. The P-ERK approach is a fully explicit solver technology that allows different Runge-Kutta schemes with different numbers of active stages to be using in stiff and non-stiff regions of the domain. Results from LES of turbulent flow over an SD7003 airfoil demonstrate that speedup factors of 17.76 and 6.05 can be obtained from GPU acceleration and P-ERK, separately. Combining these yields speedup factors up to 112. This represents a significant two order of magnitude reduction in the computational cost of performing LES/DNS. Final qualitative and quantitative results will be provided for a range of test cases in the final presentation.

Keywords-component—paired; Runge-Kutta; GPU; high; order (key words)

I. INTRODUCTION

The ability to advance locally-stiff time-dependent systems of Partial Differential Equations (PDEs) relies on accurate, efficient, and stable space-time discretizations. In the context of scale resolving turbulent flow simulations, such as Direct Numerical Simulation (DNS) and Large Eddy Simulation (LES), this stiffness arises due to thin boundary layer elements, shockwaves, and reaction terms, amongst several others. Typically, this necessitates implicit temporal discretizations. However, implicit schemes on Central Processing Units (CPUs) require the solution of globally-coupled systems of non-linear equations, which results in a high cost per time step and limited parallel scalability.

To address these limitations, explicit formulations on modern hardware architectures, such as Graphical Processing Units (GPUs), are becoming increasingly popular. The advantage of this approach is even more pronounced for compact high-order spatial discretizations, such as the Discontinuous Galerkin (DG) and Flux Reconstruction (FR) approaches, which can exploit the arithmetic compute ability of GPUs without being memory bandwidth bound [5]. It has been demonstrated that this approach can be over an order of magnitude faster and, simultaneously, several orders of magnitude more accurate than conventional implicit Finite Volume (FV) formulations on unstructured grids [5]. However, available explicit schemes have conditional stability criteria. This limits the maximum achievable time step size increasing computational cost, particularly for high-Reynolds LES and DNS [5].

Recently, Paired Explicit Runge-Kutta (P-ERK) schemes have been proposed as an alternative to classical explicit methods [6]. The P-ERK approach introduces families of explicit Runge-Kutta methods. Members of the same family can be seamlessly paired with one another, but can have different numbers of active stages, where only these active stages require right hand side evaluations. In this manner, schemes with larger numbers of active stages can be used in stiff regions of the computational domain to improve stability, while schemes with fewer active stages can be used in non-stiff regions to reduce computational cost. By design, these schemes can be paired while maintaining global accuracy and conservation [6]. Importantly, CPU implementations of P-ERK using high-order FR/DG spatial discretizations have been shown to be 5-10 times faster than conventional explicit schemes [6].

It follows that GPU acceleration and P-ERK schemes have both independently demonstrated order of magnitude performance improvements over classical approaches, while either maintaining or improving accuracy. However, the utility
of combining these two acceleration strategies has yet to be demonstrated. The objective of this work is to explore whether these performance improvements can be combined in a multiplicative manner, significantly reducing computational cost.

The remainder of this paper is organized as follows. In Section II we introduce the methodology including governing equations, spatial discretization, and temporal discretizations. Section III presents numerical results for LES/DNS of turbulent flow over a sphere and SD7003 airfoil, including a computational cost comparison with classical schemes on CPUs. Finally, conclusions and recommendations for future work are presented in Section IV.

II. METHODOLOGY

A. Governing Equations

In this paper we solve the compressible Navier-Stokes equations

$$\frac{\partial \vec{u}}{\partial t} + \nabla \cdot \vec{F} = 0,$$

where \(\vec{u}\) is the vector of conserved variables

$$\vec{u} = \begin{bmatrix} \rho \\ \rho u_i \\ \rho e \end{bmatrix},$$

where \(\rho\) is the density, \(u_i\) is a component of the velocity, and \(\rho e\) is the total energy. The total flux \(\vec{F} = \vec{F}_e - \vec{F}_v\) is the inviscid Euler fluxes

$$\vec{F}_{e,j}(\vec{u}) = \begin{bmatrix} \rho u_j \\ \rho u_i u_j + \delta_{ij} \rho \frac{\partial u_i}{\partial x_j} \\ u_j (\rho E + p) \end{bmatrix},$$

where pressure is obtained from

$$p = (\gamma - 1) \rho \left( E - \frac{1}{2} u_i u_i \right),$$

where \(\gamma = c_p/c_v\) is the ratio of specific heats, \(c_p\) is the specific heat at constant pressure, and \(c_v\) is the specific heat at constant volume, and the viscous fluxes are

$$\vec{F}_{v,j}(\vec{u}, \nabla \vec{u}) = \begin{bmatrix} 0 \\ \tau_{ij} \\ -q_j - u_i \tau_{ij} \end{bmatrix},$$

where

$$q_j = -\frac{\mu}{\text{Pr}} \frac{\partial}{\partial x_j} \left( E + \frac{p}{\rho} - \frac{1}{2} u_i u_i \right),$$

\(\text{Pr}\) is the Prandtl number, and \(\mu\) is the viscosity. The viscous stress tensor is

$$\tau_{ij} = \mu \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \frac{2}{3} \frac{\partial u_k}{\partial x_k} \delta_{ij} \right).$$

For all cases considered in this study \(\gamma = 1.4\) and \(\text{Pr} = 0.71\).

B. Spatial Discretization

Following the FR the solution for a scalar conserved variable \(\vec{u}(\vec{x}, t)\) is represented on each element [1], [4]

$$\vec{u}(\vec{x}, t) \approx \vec{u}^h(\vec{x}, t) = \sum_{i=1}^{N_e} \vec{u}^h_i(\vec{x}, t),$$

where \(\vec{u}^h_i(\vec{x}, t)\) is the global approximation of the solution and \(\vec{u}^h_i(\vec{x}, t)\) is a continuous polynomial representation on one of \(N_e\) elements, and the summation notation represents a direct summation over all elements. The approximate polynomial solution on each element is

$$\vec{u}^h_i(\vec{x}, t) = \sum_{j=1}^{N_s} \vec{u}_{i,j}(t) \phi_{s,i,j}(\vec{x}),$$

where \(\vec{u}_{i,j}(t)\) is the solution at one of \(N_s\) nodal points in red in Figure 1 and \(\phi_{s,i,j}(\vec{x})\) is its nodal basis function.

Following the flux reconstruction approach [4] and its extension to simplex element types [1], [2], the physical conservation law that must be satisfied on each element is

$$\frac{\partial \vec{u}^h_i}{\partial t} + \nabla \cdot \vec{F}^h_i + \vec{\delta}_i = 0,$$

where \(\vec{F}^h_i = F^h_i(\vec{u}^h_i, \nabla \vec{u}^h_i)\) and \(\vec{\delta}_i\) is a correction field on the element that is in the same polynomial space as the solution. This correction field is analogous to the divergence of the penalty functions introduced in the original FR scheme for tensor product elements [4]. Finally, applying the conservation law at each of the solution points

$$\frac{d \vec{u}^h_{i,j}}{dt} + \nabla \cdot \vec{F}^h_{i,j}\bigg|_{\vec{x}_{i,j}} + \delta_{i,j} = 0,$$

where \(\vec{x}_{i,j}\) is the corresponding solution point location and, following the FR formulation

$$\delta_{i,j} = \frac{1}{|\Omega_i|} \sum_{f \in S} \sum_{l} \alpha_{i,j,f,l} [\vec{F}]_{i,j,f,l} S_f,$$

where \(|\Omega_i|\) is the element volume, \(f\) is one of the number of faces on the element surface \(S\), \(l\) is one of the flux points shown blue in Figure 1, \(\alpha_{i,j,f,l}\) is a constant lifting coefficient, \([\vec{F}]_{i,j,f,l}\) is the difference between a common Riemann flux at the flux point and the value of the internal flux, and \(S_f\) is the area of the face. Depending on the specification of these lifting coefficients, a number of different energy stable schemes can be obtained for general element types, including the Spectral Difference (SD), Spectral Volume (SV), and Discontinuous Galerkin (DG) methods. In this study we use lifting coefficients based on the nodal basis functions to recover the DG method [1], [4]. The solution and flux points are located at tensor products of Gauss points, the common inviscid fluxes are recovered using a Rusanov approximate Riemann solver, and the common viscous fluxes are recovered using the second method of Bassi and Rebay (BR2).
This yields the following general Butcher tableau
\begin{equation}
\begin{aligned}
c_1 &= 0. \\
c_2 &= a_{2,1} \quad 0 \\
c_3 &= a_{3,1} \quad a_{3,2} \quad 0 \\
c_i &= a_{i,1} \quad a_{i,2} \quad a_{i,3} \quad \ldots \quad a_{i,s-1} \quad 0 \\
&\quad b_1 \quad b_2 \quad b_3 \quad \ldots \quad b_{s-1} \quad b_s
\end{aligned}
\end{equation}

where each \( c_i \) is known explicitly through Equation (22). Hence, for every additional stage we add to such a tableau, we add one additional unknown stage coefficient. Finally, to reduce the number of stage derivative evaluations to \( e \) for a general P-ERK \( s,e,2 \) scheme we set
\begin{equation}
a_{i+1,i} = 0, \quad 2 \leq i \leq s - e + 1,
\end{equation}
which gives a scheme that has \( e - 2 \) unknown Butcher tableau coefficients.

### III. NUMERICAL RESULTS

#### A. Airfoil

The first demonstration case is for turbulent flow over an SD7003 airfoil at a Reynolds number \( Re_c = 60,000 \) where
\begin{equation}
Re_c = \frac{u_\infty c}{\nu},
\end{equation}
where \( u_\infty \) is the free stream velocity, \( c \) is the airfoil chord length, and \( \nu \) is the viscosity, and a Mach number \( Ma = 0.2 \) where
\begin{equation}
Ma = \frac{u_\infty}{c_\infty},
\end{equation}
where \( c_\infty \) is the free stream speed of sound.

The computational mesh is composed of 116,200 hexahedral element with a span of \( 0.2c \) and outer boundary that extends to \( 20c \) normal to the airfoil surface and \( 40c \) in the downstream
Figure. 2. Contours of Q-criteria colored by velocity magnitude for the SD7003 example case.

direction. The boundary conditions are specified as no-slip adiabatic on the airfoil surface, Riemann invariant on the outer faces, and periodic in the spanwise direction. Computational runtime assessments were performed using a $p_s = 5$ solution polynomial degree on full CPU and GPU nodes on the Narval cluster from the Digital Research Alliance of Canada. The GPU nodes contained 4x NVIDIA A100 GPUs, and the CPU nodes contained 2x AMD Rome 7532 32 core processors. Each performance run was performed using both RK44 and P-ERK schemes.

Computational performance comparisons demonstrate that switching to P-ERK instead of RK44 improves computational performance by a factor of 6.31 and 6.68 for CPU and GPU runs, respectively. Furthermore, on a node-for-node basis, using P-ERK on GPUs was 17.76 times faster than P-ERK on CPUs. Combined, this means that switching from RK44 on CPUs to P-ERK on GPUs yields a 112 times performance improvement on a node-for-node basis.

IV. CONCLUSIONS

Based on these preliminary results, speedup factors in excess of 100 can be obtained by switching from RK44 schemes on CPUs to P-ERK schemes on GPUs, on a node-for-node basis. This comes from a combination of computational performance provided via GPU acceleration, and increased global time step size via P-ERK schemes.

Final results for several different test cases will be provided in the final presentation. Additional performance and scaling results will be discussed, and quantitative and qualitative comparisons with reference LES/DNS results will be provided.

REFERENCES


