AFRL-AFOSR-VA-TR-2020-0136

Towards Extreme-Scale Computing with High-Order Discontinuous Methods

Zhi Wang
UNIVERSITY OF KANSAS CENTER FOR RESEARCH INC.

07/31/2020
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)* 18-08-2020	**2. REPORT TYPE** Final Performance	**3. DATES COVERED** *(From - To)* 15 May 2016 to 31 Dec 2019

4. TITLE AND SUBTITLE
Towards Extreme-Scale Computing with High-Order Discontinuous Methods

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-16-1-0128

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Zhi Wang

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
UNIVERSITY OF KANSAS CENTER FOR RESEARCH INC.
2385 IRVING HILL RD
LAWRENCE, KS 660457563 US

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR RTA2

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-VA-TR-2020-0136

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Exa-scale computers are expected to appear in the next decade. State-of-the-art CFO algorithms used today will need a drastic overhaul to take advantage of the massively parallel exa-scale computers with mixed CPU and GPU architectures perhaps with hundreds of millions of compute cores. In the present project, we examine several critical issues important for scalability with high-order discontinuous methods, and develop algorithms balancing scalability, with accuracy, efficiency and robustness.
The last decade has seen significant progresses on adaptive high-order CFO methods in many aspects, including new formulations, efficient time integration algorithms, hp-adaptations, shock-capturing, high-order mesh generation and visualization. Their potential has been demonstrated in computational aeroacoustics, large eddy simulation (LES) and direct numerical simulation (DNS) of vortex dominated turbulent flows at moderate Reynolds numbers.

15. SUBJECT TERMS
exascale computing, asynchronous, local time-stepping

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			FAHROO, FARIBA
Unclassified	Unclassified	Unclassified	UU		**19b. TELEPHONE NUMBER** *(Include area code)* 703-696-8429

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Final Report Submitted to AFOSR
Grant No. FA9550-16-1-0128
For period of 05/15/2016 – 12/31/2019

Towards Extreme-Scale Computing with High-Order Discontinuous Methods

PI: Z.J. Wang
Spahr Professor of Aerospace Engineering
University of Kansas
2120B Learned Hall, Lawrence, KS 66049
Tel: 785-8642440
E-mail: zjw@ku.edu

July 30, 2020

Table of Contents

Accomplishments

1. Major Goals of the Project

Exa-scale computers are expected to appear in the next decade. State-of-the-art CFD algorithms used today will need a drastic overhaul to take advantage of the massively parallel exa-scale computers with mixed CPU and GPU architectures perhaps with hundreds of millions of compute cores. In the present project, we examine several critical issues important for scalability with high-order discontinuous methods, and develop algorithms balancing scalability, with accuracy, efficiency and robustness.

The last decade has seen significant progresses on adaptive high-order CFD methods in many aspects, including new formulations, efficient time integration algorithms, hp-adaptations, shock-capturing, high-order mesh generation and visualization. Their potential has been demonstrated in computational aeroacoustics, large eddy simulation (LES) and direct numerical simulation (DNS) of vortex dominated turbulent flows at moderate Reynolds numbers.

The main goal of the present project is to develop efficient and scalable time integration algorithms for large eddy simulation on extreme-scale computers using high-order methods. More specifically, we set to achieve the following objectives:

- Evaluate the performance of explicit and implicit schemes on extreme-scale super computers with scale-resolving benchmark problems
- Develop and test an implicit memory efficient p-multigrid solution approach
- Develop and test a matrix-free GMRES approach with and without a preconditioner
- Demonstrate the developed approaches for industrial large eddy simulation on extreme-scale supercomputers

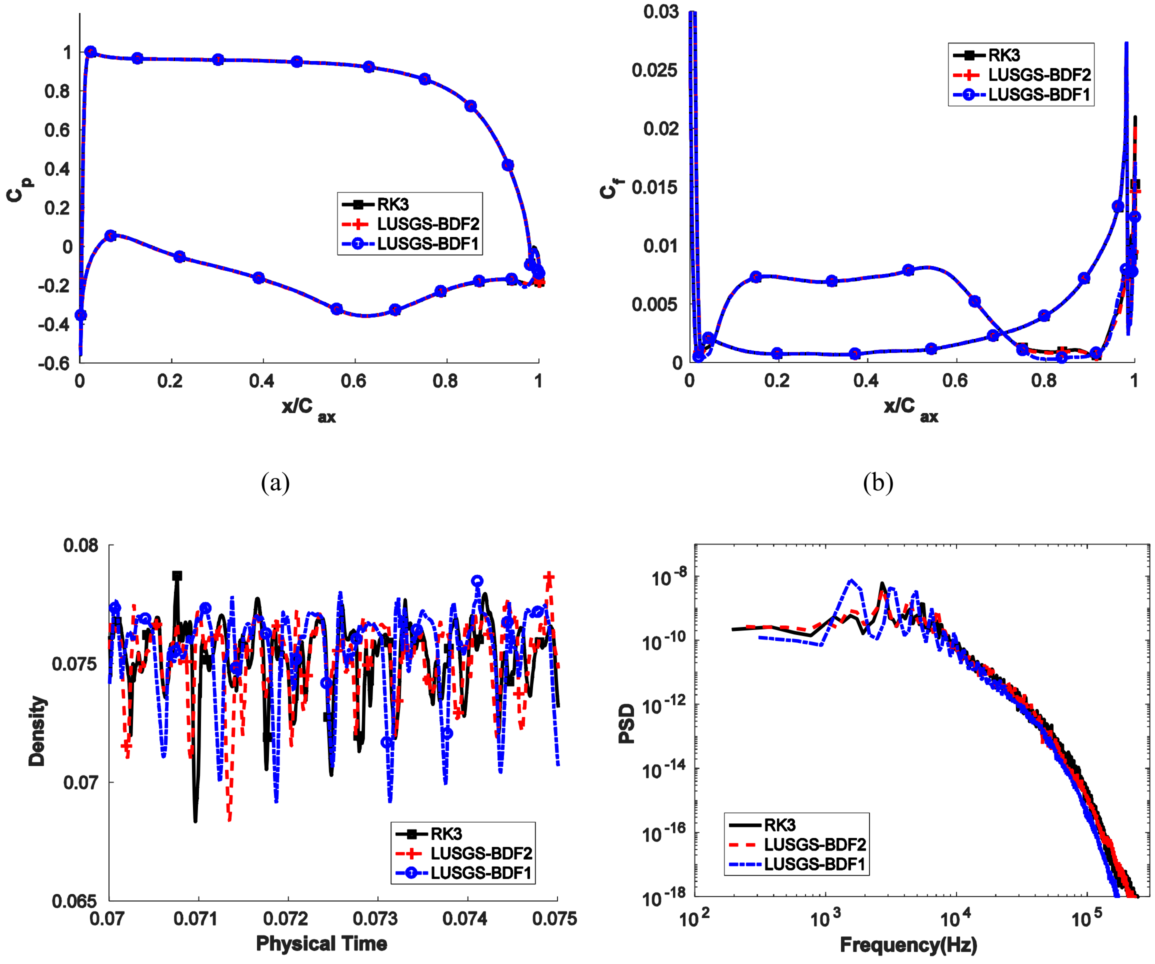
2. Accomplishments

2.1 Explicit vs implicit schemes

A detailed comparative study was performed to compare an explicit Runge-Kutta scheme with an implicit LU-SGS scheme for scale-resolving benchmark flow problems. All aspects relevant to scale resolving simulations are evaluated: time accuracy, iterative convergence, power-spectral density, Reynolds stresses, mean surface pressure coefficient and skin friction, mean lift and drag coefficients.

The evaluations were performed at both low and high-Reynolds numbers. Some typical results are shown in Figure 1. Major conclusions include:

- 2nd order accuracy with two-order iterative convergence is adequate in capturing the turbulent dynamics as well as all important parameters
- The implicit scheme is much faster than the explicit scheme from a factor 3 to over an order of magnitude for the benchmark problems. For unstructured meshes with nearly diminishing cell volumes, implicit schemes are mandatory.



(a) (b)

(c) (d)

Figure 1. Mean and instantaneous flow parameters for a benchmark problem

This study firmly establishes implicit approaches as the choice for industrial LES applications. Later, the implicit LU-SGS scheme was successfully implemented on DOE's leadership computers, and a strong scalability study was performed with excellent parallel efficiency as shown in Figure 2.

2.2 P-multigrid solution approach

The p-multigrid method is motivated by the same principles of the standard h-multigrid method, that is, to efficiently eliminate low and high-frequency error modes in the solution by using multiple levels of discretization and fine-coarse corrections. Instead of using different meshes in the solution representation as it is the case of h-multigrid, p-multigrid uses different orders of approximation for the solution in the same mesh. The p-multigrid framework fits naturally into unstructured grids and high-order methods such as DG and FR/CPR since major operations like prolongation and restriction become local to the grid cells. In the context of unsteady simulations, multigrid methods can be used together with a dual time stepping technique to accelerate the simulation.

The approach has been shown to be effective in speeding up the solution convergence while dramatically reducing memory requirement. A typical performance plot is shown in Figure 3, which shows that a p-multi-grid approach can have larger time steps than the implicit LU-SGS approach, and can be more efficient than LU-SGS.

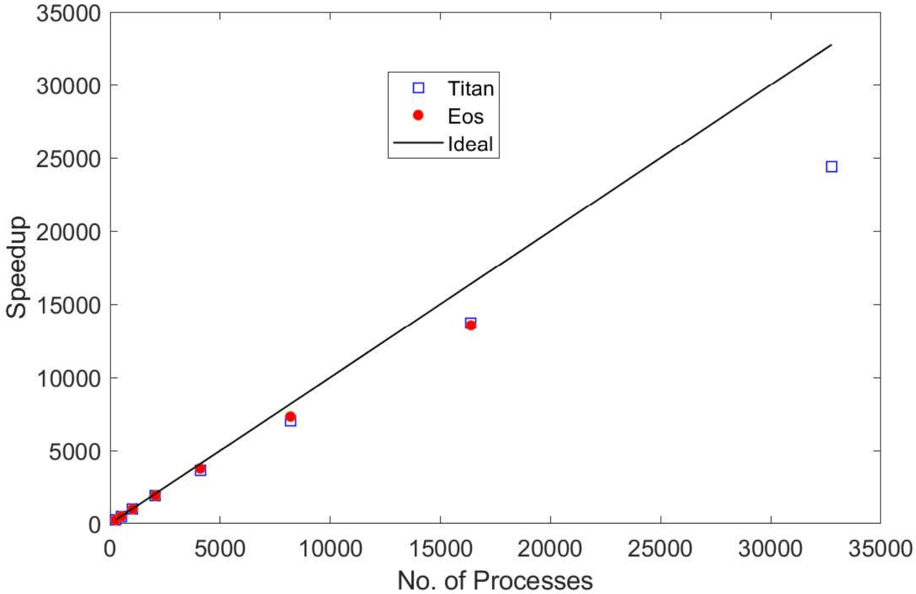


Figure 2. Strong scalability study of the implicit LU-SGS algorithm performed on Titan and Eos with a total of 27 million DOFs/equation at p = 2 (3rd order accuracy)

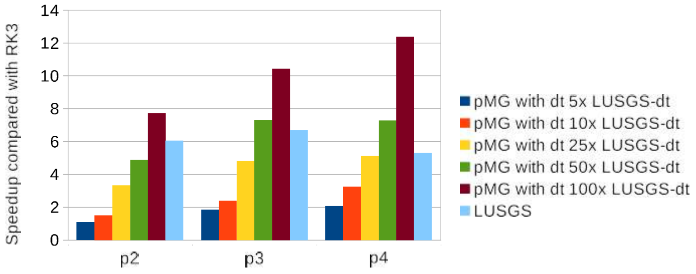


Fig. 3 Unsteady simulation performance for the unsteady viscous flow over a cylinder varying the physical time step. Speedup compared to a simple explicit RK3 unsteady simulation.

2.3 Matric-free GMRES approach for unsteady flow computation

The matrix-free (MF) implementation of the GMRES approach replaces the matrix vector multiplication with an efficient finite difference formula. The large sparse-matrix does not need to be formed. For stiff problems, this approach needs an effective pre-conditioner to achieve convergence. Most preconditioners require the generation of the Jacobian matrix, which is large for high $p > 2$.

For unsteady problems, however, the small time-step needed by a scale resolving simulation significantly improves the conditioning of the resultant implicit system. As a result, the MF-GMRES may converge without a preconditioner. If we couple the MF-GMRES approach with an dual time-step time integration, we obtain an more efficient time marching approach than the explicit Runge-Kutta approach.

For the DNS of a benchmark flow problem over the T106C turbine blade, we obtained speedups on the order of a factor 5 comparing with an explicit scheme, while on the high-lift configuration, the speed up is over an order of magnitude because the time-step size can be two orders of magnitude larger than the explicit scheme.

2.4 Demonstrations on extreme-scale computers

The explicit, implicit LU-SGS and the GMRES approaches have all been successfully ported to DOE’s leadership computers, Titan and Eos. In addition, the explicit scheme and the GMRES approach have also been implemented on GPU clusters, and successfully run on the fastest computer in the world (in 2019), Summit. A strong scalability study was performed on Summit with a mesh of 36.7 million tetrahedral elements. Up to 5,000 V100 GPU cards were used in the study. Two techniques to improve the parallel efficiency were compared with the baseline MPI implementation. With CUDA-aware and non-blocking MPI implementations, the parallel efficiency is significantly improved, as shown in Figure 4.

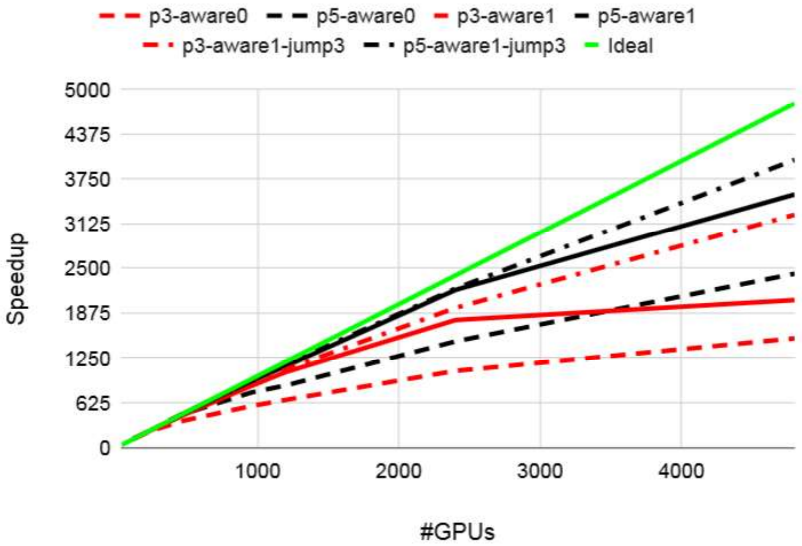
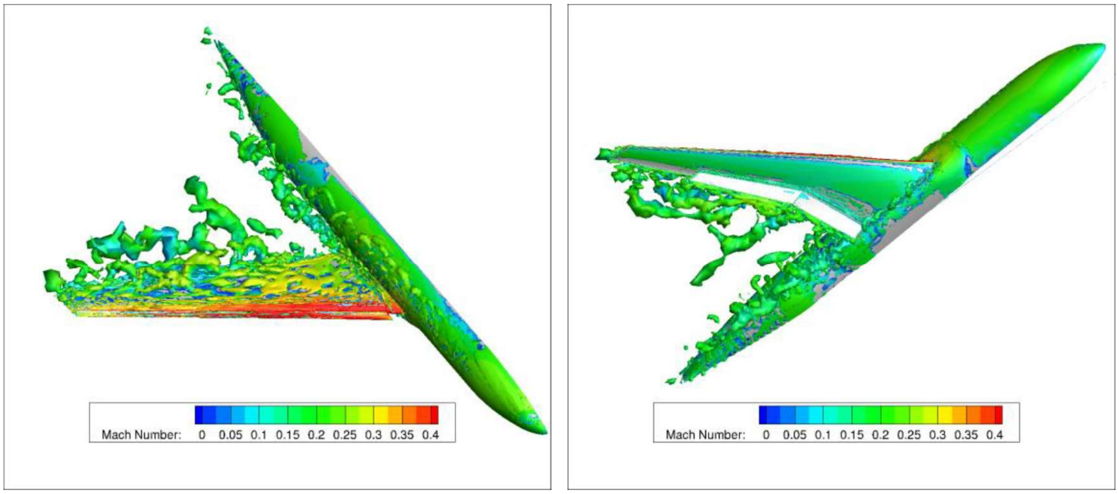


Fig. 4 Strong scalability study with GPU simulations at Summit for a mesh with 36.7M tet elements.

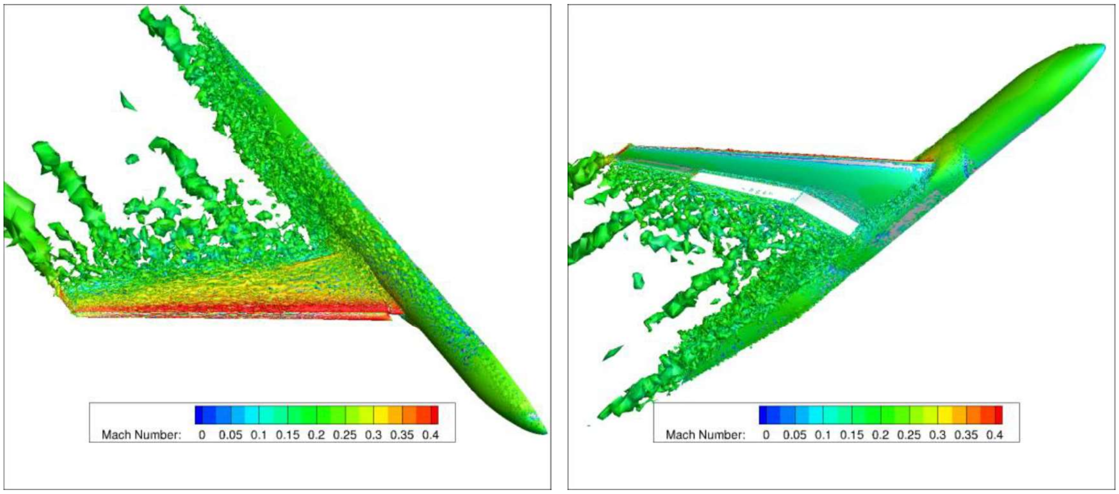
In addition, we continue to push the boundaries by performing a LES without a wall-model for the NASA CRM high-lift configuration. The Reynolds number based on the mean aerodynamic chord (MAC = 7.005m) is 3.26 million, and the AOA is 16° with an incoming Mach number of 0.2. This simulation was performed on Titan. A p-refinement study was carried out for p = 1, 2 and 3. A p-independent prediction of the lift and drag coefficients has been achieved. The results are compared with experimental data in Table 1. The converged lift is within 0.3% and the drag is within 1.4% of the experimental data. The iso-surfaces of the Q-criterion are displayed in Figure 5 for different polynomial degrees. Note that p2 and p3 captured significantly more flow features.

Table I. Predicted lift and drag coefficients with comparison to experimental data

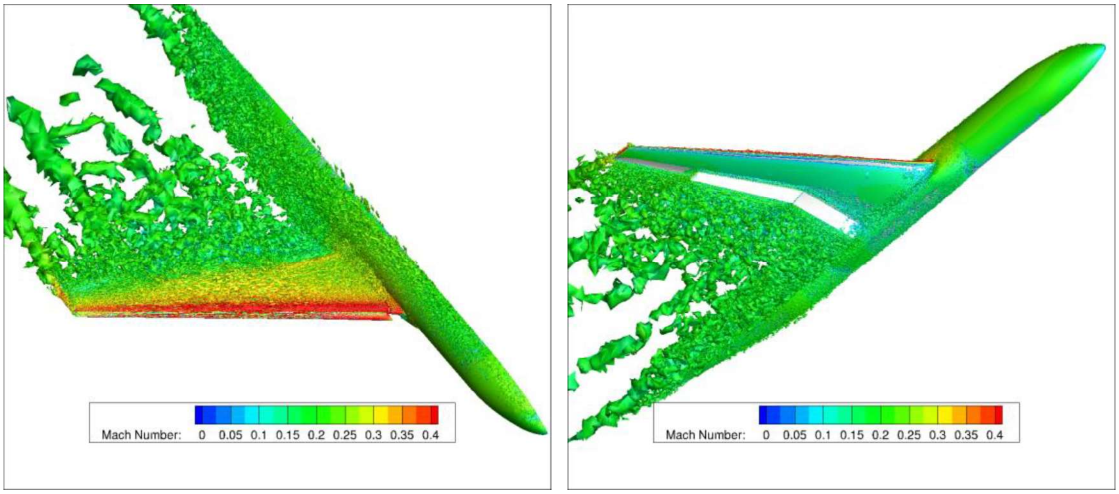
	C_L	C_D
P = 1	2.020	0.293
P = 2	2.411	0.282
P = 3	2.413	0.283
Experiment	2.479	0.252
Experiment (corrected)	2.405	0.287



(a) p = 1



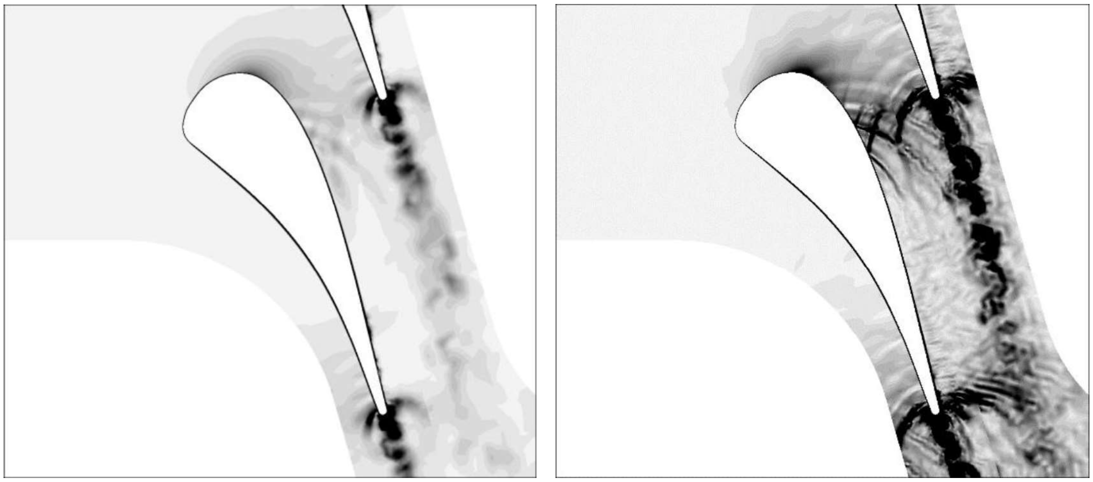
(b)p = 2



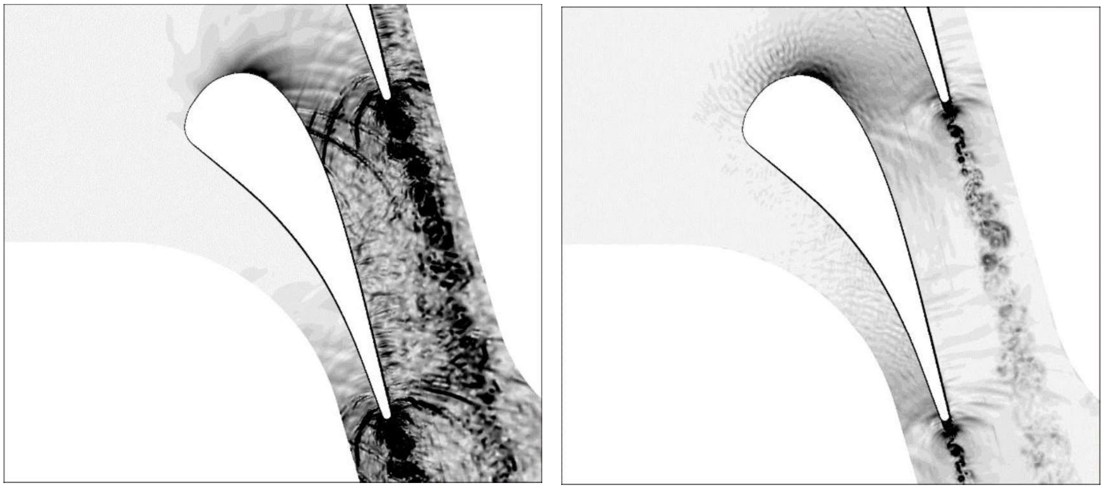
(c) p = 3

Figure 5. Iso-surfaces of the Q-criterion colored by the Mach number on both the pressure and suction side of the wing

Finally we also compared the present high-order hpMusic solver and a commercial flow solver (Fluent) on the same family of meshes for a LES benchmark problem. We demonstrated that to achieve a similar accuracy, the 3rd order flux reconstruction scheme is more than an order magnitude faster than the 2nd order finite volume scheme in Fluent. A comparison of computational Schleren obtained with hpMusic and Fluent is shown in Figure 6. Note clearly that the p2 results with less degrees of freedom are much more accurate than the Fluent results. The computed surface mean pressure and heat transfer are compared with experimental data in Figure 7. Again the high-order schemes produced with better agreement with experimental data.



(a) hpMusic p1 (coarse mesh, 3.0M DOFs)　(b) hpMusic p2 (coarse mesh, 9.6M DOFs)



(c) hpMusic p3 (coarse mesh, 22.0M DOFs)　(d) Fluent (fine mesh, 28.7M DOFs)

Figure 6. Comparison of instantaneous Schlieren ($|\nabla\rho|C/\rho$) distributions (20 levels between 0 and 3.38)

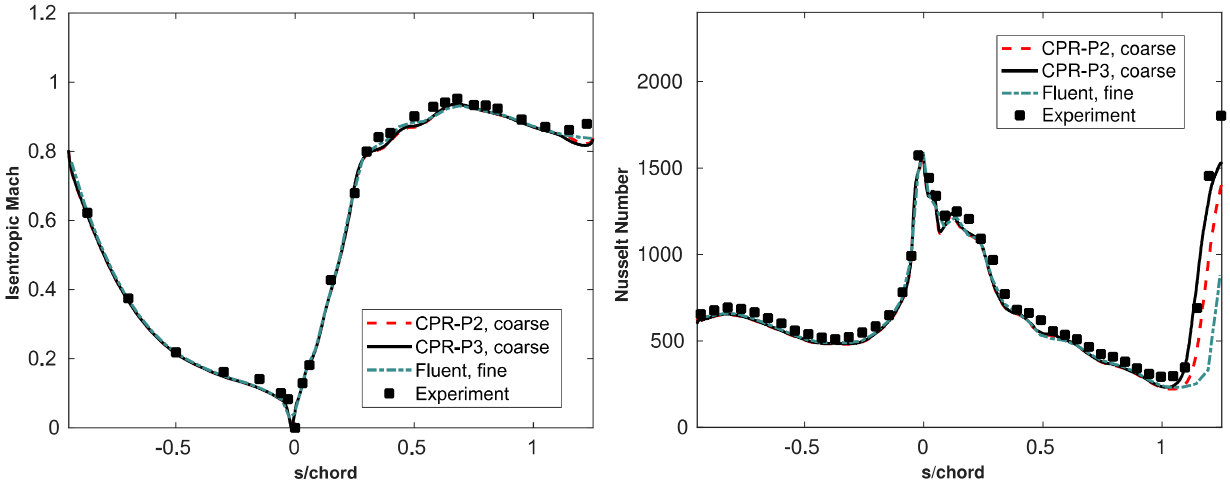


Figure 7. Comparison of mean surface isentropic Mach number and heat transfer between Fluent and hpMusic predictions and experimental data

3. Training and Professional Development Opportunities

The grant provided training opportunities for the following PhD students:

- Jia, Feilin (obtained PhD in May 2019)
- Duan, Zhaowen (obtained PhD in December 2019)
- Alhawwary,Mohammad (obtained PhD in May 2020)
- Jourdan, Eduardo
- Rahmani, Salman

4. Dissemination of Research Results

The main approach to disseminate the research results is to publish journal and conference papers, invited seminars, and oral presentations in conferences. Here is the list journal and conference publications partially supported by the grant:

Journal Publications:

- Alhawwary, M., & Wang, Z.J. On the mesh resolution of industrial LES based on the DNS of flow over the T106C turbine. Advances in Aerodynamics 1, Article number: 21 (2019).
- Jia, F., Wang, Z.J., Bhaskaran, R., Paliath, U., & Laskowski, G. Accuracy, efficiency and scalability of explicit and implicit FR/CPR schemes in large eddy simulation. Computers and

Fluids Volume 195, 15 December 2019, 104316.

- Feilin Jia, Jeremy Ims, and Z. J. Wang, James Kopriva and Gregory M. Laskowski, Evaluation of Second- and High-Order Solvers in Wall-Resolved Large-Eddy Simulation, AIAA JOURNAL, Vol. 57, No. 4, April 2019.
- M. Alhawwary and Z.J. Wang, Fourier analysis and evaluation of DG, FD and compact difference methods for conservation laws, Journal of Computational Physics, Volume 373, 15 November 2018, Pages 835-862.
- Z.J. Wang, Y. Li, F. Jia, G.M. Laskowski, J. Kopriva, U. Paliath, R. Bhaskaran, Towards industrial large eddy simulation using the FR/CPR method, Computers and Fluids, Volume 156, 12 October 2017, Pages 579-589.
- C. Zhou, L. Shi, Z.J. Wang, Adaptive high-order discretization of the Reynolds-averaged Navier-Stokes equations, Computers and Fluids 159 (2017) 137–155.
- Z.J. Wang and Y. Li, A mathematical analysis of scale similarity, Communications in Computational Physics, Vol. 21, No. 1 (2017) pp. 149-161

Conference Publications:

- Z.J. Wang and Salman Rahmani, Towards Wall-Resolved Large Eddy Simulation of CRM and JSM High-Lift Configurations, AIAA 2020-0775, Jan 2020 (https://doi.org/10.2514/6.2020-0775).
- Mohammad Alhawwary and Z.J. Wang, DNS and LES of the flow over the T106C turbine using the high-order FR/CPR method, AIAA 2020-1572, Jan. 2020 (https://doi.org/10.2514/6.2020-1572).
- Eduardo Jourdan and Z.J. Wang, Efficient Implementation of the FR/CPR Method on GPU Clusters for Industrial Large Eddy Simulation, AIAA 2020-3031, 8 Jun 2020 (https://doi.org/10.2514/6.2020-3031).
- Jourdan, E., & Wang, Z.J. (2019). A Study of p-multigrid Approach for the High Order FR/CPR Method. In AIAA-2019-3711.
- Jourdan, E., Wang, Z.J., & Azevedo, J.L.F. (2019). An Evaluation of p-multigrid and Low

Speed Preconditioning with the High Order FR/CPR Method. In AIAA-2019-0102.

- M.A. Alhawwary, Z.J. Wang, A study of DG methods for diffusion using the combined-mode analysis (AIAA 2019-1157) AIAA Scitech 2019 Forum, 2019, 10.2514/6.2019-1157
- M.A. Alhawwary, Z.J. Wang, Comparative Fourier Analysis of DG, FD and Compact Difference schemes (AIAA 2018-4267), 2018 Fluid Dynamics Conference, 2018, 10.2514/6.2018-4267.
- F. Jia, J. Ims, Z.J. Wang, J. Kopriva, G.M. Laskowski, An Evaluation of a Commercial and a High Order FR/CPR Flow Solvers for Industrial Large Eddy Simulation (AIAA 2018-0827), 2018 AIAA Aerospace Sciences Meeting, 2018, 10.2514/6.2018-0827.
- F. Jia, Z.J. Wang, R. Bhaskaran, U. Paliath, G.M. Laskowski, Accuracy, Efficiency and Scalability of Explicit and Implicit FR/CPR Schemes in Large Eddy Simulation, AIAA-2017-3096.
- Y. Li and Z.J. Wang, A convergent and accuracy preserving limiter for the FR/CPR method (AIAA 2017-0756) 55th AIAA Aerospace Sciences Meeting, 2017.