The 8th European Congress on Computational Methods in Applied Sciences and Engineering
ECCOMAS Congress 2022
5-9 June 2022, Oslo, Norway

Efficient implementation of a high-order compressible Navier-Stokes equations solver running on Graphics Processing Units

Fernando Gisbert[1], Adrián Sotillo[1] and Jesús Pueblas[1]

[1] ITP Aero, Simulation Technologies Department
Francisca Delgado 9, 28830 Alcobendas, Spain
e-mail: {fernando.gisbert, adrian.sotillo, jesus.pueblas}@itpaero.com

Keywords: *Compressible Navier-Stokes, GPU, High-order, Flux Reconstruction, Parallelism*

This paper presents the implementation of a compressible Navier-Stokes equations solver that runs on multiple GPUs. The solver, known as Mu^2s^2T[1], uses a flux reconstruction[2] high-order spatial discretization, and it is advanced in time using an explicit Runge-Kutta method. The numerical resolution of computing-intensive problems such as the one presented here increasingly relies on using hardware architectures such as GPUs. However, two main hurdles prevent the widespread adoption of these computing platforms. On the one hand, it is not straightforward to unleash their computational power, often due to very stringent memory access requirements. That is especially sensitive in memory-bounded solvers like most CFD codes, where the proper memory access pattern has enormous implications in terms of computing efficiency. On the other hand, the very own nature of the GPU as an accelerator rather than a CPU processor with direct access to the computer memory makes the data transfer between the GPU and the computer memory a bottleneck. It has not any consequence in problems that fit into one single GPU, but it produces a substantial impact if many GPUs must exchange data, which is the usual scenario when solving large enough problems in parallel. We address both topics in this paper. The implementation of the most time-consuming parts of the fluxes evaluation will be detailed. The intensive use of the GPU local memory has proven crucial to the efficient use of the hardware. The design of parallel communications also plays a key role in improving the efficiency of the solver. We have implemented a pipelined non-blocking data transfer between GPUs, overlapping communication and computation as much as possible and therefore minimizing the penalties associated with transferring data between GPU and CPU memories. The solver efficiency will be demonstrated by running a series of benchmark cases. The execution time measurements will be compared against similar freely available CFD solvers. Finally, we will present cases of industrial relevance that have been simulated using this tool, highlighting its benefits in terms of computational cost and energy consumption.

REFERENCES

[1] Gisbert, F., and Bolinches-Gisbert, M., and Pueblas, J. and Corral, R. Efficient implementation of Flux Reconstruction schemes for the simulation of compressible

viscous flows on Graphics Processing Units. 10th International Conference on Computational Fluid Dynamics, 2018

[2] Huynh, H. A Flux Reconstruction Approach to High-Order Schemes Including Discontinuous Galerkin Methods 18th AIAA Computational Fluid Dynamics Conference, 2007