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

On the GPU-enabling techniques for Finite Elements CFD codes

L. Gasparino[1], G. Houzeaux[2], F. Spiga[3] and O. Lehmkuhl[4]

[1] BSC, lucas.gasparino@bsc.es
[2] BSC, guillaume.houzeaux@bsc.es
[3] NVIDIA, fspiga@nvidia.com
[4] BSC, oriol.lehmkuhl@bsc.es

Keywords: *Compressible Flow, LES, DNS, CFD, GPU computing, HPC, High-order scheme, Low dissipation algorithm, Finite Elements, TVD FEM*

The present work covers the subject of GPU implementation methodologies for a high-order (low dissipation) Finite Elements algorithm for LES and DNS simulations of turbulent compressible fluid flows. The aim is to demonstrate how much impact different implementations have on performance, and how different techniques and approaches (code optimizations) help truly enable a GPU approach under the CFD/FE framework. To achieve our goals, we compare 2 different parallelization strategies, namely cell-per thread and cell-per-threadblock, in 2 different test cases, involving Q1, Q2 and Q3 3D element types, as well as single vs. double precision studies. The cell-per-threadblock model will be presented in 2 forms, one focused on reducing per-thread work, the other focused on reducing GPU memory usage at the cost of increased computational intensity. The test cases to be used for assessing the selected algorithms and techniques are:

- A 3D shock tube (Sod's test);
- A compressible Taylor-Green model without shock interactions;

The selected algorithm to carry out these tests is a low dissipation Continuous Galerkin model, namely an Entropy Viscosity Finite Elements algorithm, which involves use of a linear system solver to handle a consistent mass matrix, a set of stabilization terms driven by a non-linear "entropy viscosity" parameter, explicit time integration through an RK4 scheme, and a "prediction-correction" technique in which the prediction is used to establish values for the stabilization parameter.

By this work's conclusion, we expect to paint a better picture of how a low dissipation FE algorithm can be efficiently implemented on a heterogeneous architecture, taking as much advantage of a high value GPU as possible.

REFERENCES

[1] O.C. Zienkiewicz, R.L. Taylor, P. Nithiarasu *The finite element method for Fluid Dynamics*. 7th Edition, Elsevier, 2014.

[2] J. L. Guermond, R. Pasquetti, B. Popov, Entropy Viscosity method for nonlinear conservation laws. *J. of Comp. Phys.*, Vol. **230**, Issue **11**, pp. 4248–4267, 2011.

[3] J. L. Guermond, R. Pasquetti, A correction technique for the dispersive effects of mass lumping for transport problems. *Comp. Meth. in Appl. Mech. and Eng.*, Vol. **253**, pp. 186–198, 2013.

[4] K. Ljungkvist, Matrix-free Finite-Element computations on Graphics Processors with adaptively refined unstructured meshes. *Proc. 25th HPC Symp.*, pp. 1–12, 2017.

[5] M. Harris, Unified Memory for CUDA beginners. NVIDIA Developer blog, 2017.

[6] N. Sakharnykh, Maximizing Unified Memory Performance in CUDA. NVIDIA Developer blog, 2017