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

Alya towards Exascale: efficient finite element assembly on GPUs for LES

Herbert Owen [1], Oriol Lehmkuhl[2], Guillaume Houzeaux[3], Guillermo Oyarzun[4], Georg Hager[5], Gerhard Wellein[6] and Dominik Ernst[3]

[1] Barcelona Supercomputing Center, herbert.owen@bsc.es
[2] Barcelona Supercomputing Center, oriol.lehmkuhl@bsc.es
[3] Barcelona Supercomputing Center, guillaume.houzeaux @bsc.es
[4] Barcelona Supercomputing Center, guillermo.oyarzun@bsc.es
[5] Erlangen National High Performance Computing Center (NHR@FAU), georg.hager@fau.de
[6] Erlangen National High Performance Computing Center (NHR@FAU), gerhard.wellein@fau.de
[7] Erlangen National High Performance Computing Center (NHR@FAU), dominik.ernst@fau.de

Key Words: *Finite element, CFD, LES, Exascale, GPU.*

On the path to Exascale, most codes will need significant changes to adapt to new architectures. Currently, GPUs are the most mature among such architectures. This work describes the improvements we have recently introduced in the High-Performance Computing code Alya within the EoCoE-II project. We shall focus on incompressible flow problems for Large Eddy Simulation. The momentum equation is treated explicitly while the pressure is solved implicitly. A fractional step scheme is used to enable elements that do not satisfy the inf-sup condition [1].
Except in cases with mesh deformation, the Laplacian matrix for the pressure remains fixed during the whole simulation. Therefore, the two main kernels are calculating the right-hand side term for the momentum equation and the solution of a linear system for the pressure. In this work, we shall concentrate on the improvements we have obtained in the GPU implementation of the first kernel. Some small comments on the algebraic multigrid strategy will also be presented.
The simple steps that have been followed to obtain more than an order of magnitude reduction in computational time for the right-hand side term assembly compared to our previous GPU implementation [2] will be presented by a finite element CFD specialist. The GPU implementation is based on OpenACC, and it is therefore quite friendly. However, it has allowed us to reach 50% of the maximum floating-point performance on an A100 Nvidia GPU. Finally, while optimizing the GPU version, improvements to the CPU implementation have also been obtained. A comparison between the current CPU and GPU performances will show that we are more energy efficient on the GPU than on the CPU nowadays.

REFERENCES

[1] O. Lehmkuhl, G. Houzeaux, H. Owen, G. Chrysokentis and I. Rodriguez A low-dissipation finite element scheme for scale resolving simulations of turbulent flows, *Journal of Computational Physics*, 2019.
[2] R. Borrell et al., Heterogeneous CPU/GPU co-execution of CFD simulations on the POWER9 architecture: Application to airplane aerodynamics, *Future Generation Computer Systems*, 2020.