

Evaluating Discontinuous Galerkin Methods on Many-core Architectures

T. Warburton^a and A.Klöckner^b

^a *Department of Computational and Applied Mathematics, Rice University*

^b *Courant Institute of Mathematical Sciences, New York University*

Modern many-core processing units, including graphics processing units (GPU), presage a new era in on-chip massively parallel computing. The advent of processors with $\mathcal{O}(1000)$ floating point units (FPU) raises new issues challenging conventional measures of “optimality” of numerical methods. The ramp up in FPU counts for each new generation of GPU over the past four years has been accompanied by a slower increase in the the memory capacity of the GPU. For example, a few hundred US dollars currently buys a parallel computer that is capable of performing $\mathcal{O}(4 \cdot 10^{12})$ floating point operations per second, but only of reading $\mathcal{O}(5 \cdot 10^{10})$ values from memory per second. From the point of view of numerical analysis, this means that the traditional approach of comparing optimality of alternative numerical methods based on their floating point operation count per degree of freedom has become mostly irrelevant. Claims of optimality derived from this measure therefore need to be reevaluated and the formulation of numerical methods in general need to be revisited given the changing computational landscape.

In 2009 we demonstrated that the nodal discontinuous Galerkin time-domain method for computational electromagnetics can achieve a high percentage of peak performance of NVIDIA GPUs [1]. In subsequent articles we demonstrated that it is possible to improve time-stepping efficiency of these methods using local time-stepping methods on GPUs [2, 3]. We also demonstrated scalability when using multiple GPUs in a workstation [4] and on over 400 GPUs in a larger scale cluster [5]. Finally we also created a new variant of the discontinuous Galerkin methods that enables the use of curvilinear elements to accurately model non-planar domain boundaries within the restricted memory budget of current GPUs [6]. The most challenging aspect of this effort is not in fact the implementation but rather the analysis of the new method. We have reported initial analysis results in [7].

The presentation will touch on several important and inter-linked issues that impacted the development of discontinuous Galerkin based solvers for a moving many-core architecture target. We will discuss on-chip scalability, multi-GPU scalability, inter-generational GPU scaling, and how we modified the solver memory requirements. Finally we will discuss programming tools that we are currently developing to enhance the productivity of programmers engaged in this type of implementation task.

References

- [1] A. Klöckner, T. Warburton, J. Bridge, and J.S. Hesthaven. *Nodal discontinuous Galerkin methods on graphics processors*. *J. Comp. Phys.*, **228**:7863–7882, 2009.
- [2] Nico Gödel, Steffen Schomann, Tim Warburton, and Markus Clemens, *GPU Accelerated Adams-Bashforth Multirate Discontinuous Galerkin Simulation of High Frequency Electro-magnetic Fields*, *IEEE Transactions on Magnetics*, **46**:8, Pages 2735–2738, 2010.
- [3] S. Schomann, N. Gödel, T. Warburton, and M. Clemens, *Local Time-stepping Techniques Using Taylor Expansion for Modeling Wave Propagation with Discontinuous Galerkin FEM*, *IEEE Transactions on Magnetics*, **46**:8, Pages 3504–3507, 2010.
- [4] N. Gödel, N. Nunn, T. Warburton, and M. Clemens, *Accelerating Multi GPU Based Discontinuous Galerkin FEM Computations for Electromagnetic Radio Frequency Problems*, *Applied Computational Electromagnetic Society Journal*, **25**:4 pp. 331–338, 2010.

- [5] Carsten Burstedde, Omar Ghattas, Michael Gurnis, Tobin Isaac, Andreas Klöckner, Georg Stadler, Tim Warburton, and Lucas C. Wilcox, Extreme-Scale AMR. ACM IEEE SC Conference Series, 2010.
- [6] T. Warburton, *A low storage curvilinear discontinuous galerkin time-domain method for electromagnetics*, 2010 URSI International Symposium on Electromagnetic Theory (EMTS), pp. 996–999, IEEE, 2010.
- [7] T. Warburton, *Aspects of the a priori convergence analysis for the low storage curvilinear discontinuous Galerkin method*, extended abstract submitted to Oberwolfach Workshop on the Theory and Applications of Discontinuous Galerkin Methods, MFO Technical Report 1208a, 2012.