Performance modeling and optimization for 3D Lattice Boltzmann simulations on highly parallel on-chip architectures: GPUs vs. Multi-Core CPUs.

J. Habich*, T. Zeiser, G.Hager, G. Wellein

Regional Computing Center Erlangen
Martenstr. 1, D-91052 Erlangen
e-mail: hpc@rrze.uni-erlangen.de

ABSTRACT

Recent developments in GPU-based CFD computing, e.g. [1], [2], [3] show that optimized flow solvers can be up to an order of magnitude faster than current standard two-socket x86-type servers. We use a lattice Boltzmann method based flow solver kernel, which is proven to perform well both on CPUs and GPUs. We study the impact of the most important optimization steps on both architectures and compare with appropriate performance models. Furthermore a multi-core aware temporal blocking strategy for CPUs is demonstrated, which increases cache reuse and further narrows the performance gap to GPUs. Going beyond a single compute node we evaluate the potential of multi-node GPU clusters for our CFD solver. Basic estimates in table 1 show that the vast single GPU performance is put into perspective for parallel GPU computations if the required data transfers via PCIe and InfiniBand are considered as well. Finally we show the performance of an implemented standalone GPU solver as well as a hybrid CPU/GPU solver on single compute nodes and on distributed CPU/GPU compute nodes.

Steps	GTX 280 (∼ 300 MLUPS)		Intel Xeon 5550 node (∼ 100 MLUPS)
Compute Time	0.8 ms		2.6 ms
PCIe: 5 GB/s	(a)	0.3 ms	–
(2 GB/s)	(b)	(0.9 ms)	–
IB: 2.4 GB/s	(I)	0.7 ms	0.7 ms
Total Time: a + I	1.8 ms → 145 FluidMLUPS/s		3.3 ms → 79 FluidMLUPS/s
b + I	2.4 ms → 109 FluidMLUPS/s		3.3 ms → 79 FluidMLUPS/s

Table 1: Estimates for a hybrid single precision LBM implementation using cluster compute nodes equipped with GPUs and InfiniBand network interconnects. The speedup when using GPUs instead of CPUs ranges from 1.3 to 1.8 depending on the strength of the communication network.

References

[1] J. Tölke and M. Krafczyk. *TeraFLOP computing on a desktop PC with GPUs for 3D CFD.* Int. J. Comput. Fluid Dyn., 22(7):443–456 ,2008.

[2] W. Lie, X. Wei, and A. Kaufmann. *Implementing lattice Boltzmann computation on graphics hardware.* The Visual Computer, 19(7-8):444–456, 2003

[3] J. Habich, T. Zeiser, G. Hager, G. Wellein *Speeding up a Lattice Boltzmann Kernel on nVIDIA GPUs.* In Proceedings of The First International Conference on Parallel, Distributed and Grid Computing for Engineering 2009 (Pécs, Hungary, April 6-8, 2009)