Parallelization of MPDATA on Multicore Architectures with GPU Accelerators Using Load Balancing and Autotuning Techniques

Lukasz Szustak, Krzysztof Rojek, and Roman Wyrzykowski
Czestochowa University of Technology
Dabrowskiego 73, 42-201 Czestochowa, Poland
{lszustak, krojek, roman}@icis.pcz.pl

Abstract. The multidimensional positive defined advection transport algorithm (MPDATA) is among the most time-consuming components of the EULAG model.

The main goal of this paper is to investigate tenets of optimal parallel formulation of MPDATA on modern supercomputing architectures. Novel multi-, many-core and GPU hardware architectures require not only a different philosophy of memory management than traditional massively parallel supercomputers, but also a comprehensive look at load balancing in the heterogeneous co-processing computing model. New strategies for memory and computing resources management allow us to ease memory bounds and optimally exploit the theoretical floating point efficiency of new architectures.

In this paper, the algorithmic parameter optimization of MPDATA using the autotuning technique is applied to provide the performance portability across various GPU platforms (NVIDIA and AMD). This method assumes generating a variety of parametrized code variants for a given algorithm, and determining best performing ones. Such a heuristic is combined with a performance model-driven approach to limit the search space.

When adapting MPDATA to modern hybrid architectures, consisting of GPU and CPU cores, the first challenge is to provide high performance for each system's hybrid component, taking into account their properties. The second challenge concerns data partitioning and load balancing across the system's heterogeneous resources.

In our work, we used such standards for multicore and GPGPU programming as OpenCL and OpenMP. The achieved performance results on hybrid of GPU and CPU outperform results achieved on a single GPU or CPU.