World Academy of Science, Engineering and Technology
International Journal of Computer and Systems Engineering
Vol:16, No:03, 2022

Parallel Pipelined Conjugate Gradient Algorithm on Heterogeneous Platforms

Authors : Sergey Kopysov, Nikita Nedozhogin, Leonid Tonkov

Abstract : The article presents a parallel iterative solver for large sparse linear systems which can be used on a heterogeneous platform. Traditionally, the problem of solving linear systems does not scale well on multi-CPU/multi-GPUs clusters. For example, most of the attempts to implement the classical conjugate gradient method were at best counted in the same amount of time as the problem was enlarged. The paper proposes the pipelined variant of the conjugate gradient method (PCG), a formulation that is potentially better suited for hybrid CPU/GPU computing since it requires only one synchronization point per one iteration instead of two for standard CG. The standard and pipelined CG methods need the vector entries generated by the current GPU and other GPUs for matrix-vector products. So the communication between GPUs becomes a major performance bottleneck on multi GPU cluster. The article presents an approach to minimize the communications between parallel parts of algorithms. Additionally, computation and communication can be overlapped to reduce the impact of data exchange. Using the pipelined version of the CG method with one synchronization point, the possibility of asynchronous calculations and communications, load balancing between the CPU and GPU for solving the large linear systems allows for scalability. The algorithm is implemented with the combined use of technologies: MPI, OpenMP, and CUDA. We show that almost optimum speed up on 8-CPU/2GPU may be reached (relatively to a one GPU execution). The parallelized solver achieves a speedup of up to 5.49 times on 16 NVIDIA Tesla GPUs, as compared to one GPU.

Keywords : conjugate gradient, GPU, parallel programming, pipelined algorithm
Conference Title : ICPP 2022 : International Conference on Parallel Processing
Conference Location : Rome, Italy
Conference Dates : March 03-04, 2022