Parallelization of Variable Preconditioned Krylov Subspace Method using Multi-GPU

Soichiro Ikuno *

Abstract

In recent years, commodity Graphics Processing Units (GPUs) have been obtained large computation power since applications like 3D games needs a realistic visualization. Furthermore, GPU architectures have been changed from fixed operation to flexible organization for programmability; therefore, GPUs are capable of scientific computing more than the specific graphics operation. For example, GeForce GTX 480 can perform up to 1.35 TFLOPS by using single precision and 672 GFLOPS by using double precision. This performance is about a performance of 13 times of Core i7 930 by using double precision.

The communication of data between GPUs is needed for using multi-GPU because the VRAM is physically separated, and the data should be communicated by using Message Passing Interface (MPI) through the main memory. Meanwhile, GPU Direct v2.0 has been supported from newest CUDA tool kit 4.0 [1]. In the GPU Direct v2.0, the peer-to-peer memory access and peer-to-peer memory copy between GPUs can be used through the PCI Express x16. By applying this technology, the data can be communicated directly, and it can reduce the communication time.

The variable preconditioning method has been developed as a new preconditioning strategy [2]. The variable preconditioned generalized conjugate residual (VPGCR) method has two nested iterations for GCR and variable preconditioning for GCR are called as outer-loop and inner-loop, respectively. In the preconditioned procedure, the residual equation is solved to determine the preconditioner for the outer-loop.

The variable preconditioned GCR method has the sufficient condition for convergence [3]. The residual of the problem converges if the relative residual norm of inner-loop satisfies the inequality in each steps. That is to say, residual equation can be solved roughly by using some iterative method with only a few iteration, and a stationary iterative method such as Gauss-Seidel method, is adopted for variable preconditioning procedure, generally. In the present study, Jacobi Over Relaxation (JOR) method is adopted because of the parallelization. From above reasons, we apply the hybrid scheme that uses single precision and double precision operations to VP Krylov subspace method using multi-GPU.

Result of computation shows that VPCG with mixed precision on GPU demonstrates significant achievement than that of ICCG on CPU. VPCG with mixed precision on GPU is about 66 times faster than that of ICCG on CPU.

References

[1] http://www.nvidia.com/object/cuda_develop.html

[2] H. A. Van der Vorst and C. Vuik, Numerical Linear Algebra with Applications Vol.1, **4** (1994) p. 369.

[3] K. Abe and S. L. Zhang, Int. J. Numer. Anal. Model. **22** (2005) p. 147.

*Tokyo University of Technology, 1404-1 Katakura, Hachioji, Tokyo 192-0982, Japan.