

9-2 Exascale CFD Simulations on Fugaku and Summit

— Development of Mixed-Precision Matrix Solvers for Many-Core CPUs and GPUs —

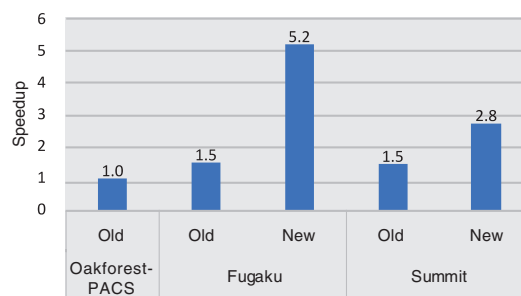
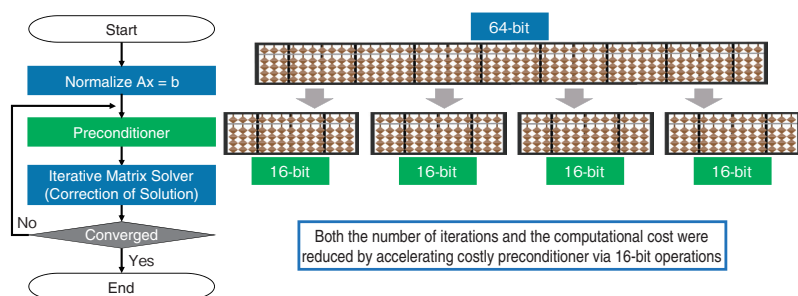


Fig.9-4 Schematic of a mixed-precision matrix solver

An iterative matrix solver for the linear system $Ax = b$ is shown. The preconditioner is accelerated by computing four operations at once using 16-bit operations (green), which are newly supported on Fugaku. The remaining part of the iterative solver is computed using 64-bit operations (blue). The picture shows a Japanese abacus, which is divided into four parts.

Fig.9-5 Computational performance of a mixed-precision matrix solver

Speedups of matrix solvers in plasma CFD simulations (about 100 billion grids) are shown for 1440 CPUs on Oakforest-PACS and Fugaku and 1440 GPUs on Summit.

Computational fluid dynamics (CFD) simulations on exascale supercomputers (i.e., supercomputers capable of processing 10^{18} FLOPS), such as Fugaku, have been widely used in various scientific and engineering fields, including nuclear engineering. Computational algorithms for exascale computers are essential for processing such extreme-scale CFD simulations efficiently. In this work, a mixed-precision matrix solver, vital to CFD simulations, was developed in collaboration with Riken. The solver was developed using 16-bit operations, which are newly supported on the many-core CPU-based computer Fugaku (Riken). The solver was then implemented on the GPU-based computer Summit (Oak Ridge National Laboratory). It was demonstrated that the solver enables high-performance computing on exascale computers based on many-core CPUs and GPUs.

CFD simulations often compute a linear system $Ax = b$, which is given by discretizing the time development of fluid models. Here, a matrix A is an $N \times N$ sized sparse matrix given by the CFD scheme, b represents a vector given by variables at the current time step, x represents the solution vector for variables in the next time step, and N represents the number of grids. This type of large sparse matrix is normally computed using iterative solvers, which require several thousand iterations per time step in extreme-scale CFD simulations. The convergence may be improved by using a preconditioner, which converts the problem by multiplying the system by a preconditioning matrix M , so that the converted problem $M^{-1}Ax = M^{-1}b$ has better mathematical properties. Here, M is an approximate solution of A that can be represented in arbitrary precision. We constructed a mixed-precision solver, in which the preconditioner is computed in 16-bit operations, while the remaining part of the iterative solver is processed in 64-bit operations, which are the standard

precision in scientific computing; the proposed solver is shown schematically in Fig.9-4. Although 16-bit operations have been widely used in machine learning on GPUs, their use on CPUs was first supported on Fugaku. 16-bit operations can accelerate computation because four times the operations can be processed at once compared with 64-bit operations. Additionally, the problem $Ax = b$ was normalized to avoid accuracy degradation, which is a common issue when using 16-bit operations, and the number of iterations was reduced by an order of magnitude.

We then conducted plasma CFD simulations with about 100 billion grids using 1440 processors on Fugaku and Summit. In Fig.9-5, the resulting computing performance is compared with the conventional many-core CPU-based computer Oakforest-PACS (Univ. Tokyo/Univ. Tsukuba). Here, the performance of CFD simulations is mainly determined by the memory bandwidth; both Fugaku and Summit have a memory bandwidth approximately two times higher than Oakforest-PACS. Further, the performance improvement of the developed solver exceeded that of the hardware. Using the proposed methodology, Fugaku and Summit solved the system approximately 5.2 times and 2.8 times faster, respectively, whereas the conventional solver showed only 1.5 times speedup. These results demonstrate the high performance of the developed solver on many-core CPUs and GPUs.

This work was supported by MEXT as the Program for Promoting Research on the Supercomputer Fugaku “Exploration of Burning Plasma Confinement Physics” (hp200127) and Oak Ridge Leadership Computing Facility (OLCF) Director’s Discretion Project “Exascale CFD Simulations at JAEA (CSC367)”.

(Yasuhiro Idomura)

Reference

Idomura, Y. et al., Acceleration of Fusion Plasma Turbulence Simulations Using the Mixed-Precision Communication-Avoiding Krylov Method, Proceedings of International Conference for High Performance Computing, Networking, Storage and Analysis (SC), online, U. S. A., vol.1, 2020, p.1318–1330.