jh200041-NAH

Innovative Multigrid Methods II

Akihiro Fujii (Kogakuin University)

Abstract

This project is the first year of the three year project. The purpose of the project is to enhance performance scalability and to extend the applicability of the multigrid algorithm. We promote researches in the fields of Geometric Multigrid(GMG), Algebraic Multigrid(AMG) and Parallel-in-Space-Time(PinST). In the first year, we have conducted researches mainly on GMG with Sell-C-σ ILU smoother, AMG weak scaling test up to 10^{10} DOF, MGRIT scalability enhancement techniques, and application of a PinST algorithm to explicit schemes.

1 Basic Information

1.1 Collaborating JHPCN Centers

- The University of Tokyo(Oakforest-PACS, Oakbridge-CX)
- Hokkaido University(Grand Chariot)
- Nagoya University(FX1000)

1.2 Research Areas

- Very large-scale numerical computation

1.3 Roles of Project Members(*:Student)

- Akihiro Fujii (Kogakuin University) (PI) Administration, AMG, PinST
- Kengo Nakajima (The University of Tokyo) (Co-PI) Applications, GMG, AMG, PinST
- Matthias Bolten (University of Wuppertal) (Co-PI) GMG, AMG, PinST
- Takeshi Iwashita (Hokkaido University) AMG, PinST
- Yasuhiro Takahashi (Doshisha University) PinST
- Osni Marques (Lawrence Berkeley National Laboratory) GMG, AMG
- Ryo Yoda* (The University of Tokyo) PinST, AMG
- Akihiro Ida (JAMSTEC) GMG, AMG
- Masatoshi Kawai (The University of Tokyo) GMG, AMG
- Yen-Chen Chen* (The University of Tokyo) PinST
- Satoshi Oshima (Nagoya University) Code Parallelization, Profiling & Optimization
- Tetsuya Hoshino (The University of Tokyo) Code Parallelization, Profiling & Optimization, SELL-C-σ
- Toshihiro Hanawa (The University of Tokyo) Code Parallelization, Profiling & Optimization, Mesh Generation
- Gerhard Wellein (Friedrich-Alexander-University(FAU) of Erlangen-Nürnberg) SELL-C-σ
- Kenji Ono (Kyushu University) PinST
- Mikio Iizuka (Kyushu University) PinST
- Robert Speck (Juelich Supercomputing Centre) PinST
- Martin Schreiber (Technical University of Munich) PinST
- Christie Alappat (Friedrich-Alexander University Erlangen) PinST
- Atsuhiro Miyagi (Taisei Corpration) PinST

2 Purpose and significance of Research

A multigrid method is a hierarchical and scalable algorithm, which would be more and more important especially on an exascale systems. The method can solve linear and non-linear problems. It is widely used as one of the most efficient solvers for large sized practical problems in many fields. However, its hierarchical structure of the algorithm makes it difficult to achieve high performance on highly parallel many-core machine environment. At the coarse levels, it requires small sized (unstructured) sparse matrix calculations on many distributed computing nodes. As for the robustness, many problems are also left to be solved such as robust smoothers and coarse level generation methods.

Therefore, we will study the techniques to enhance the multigrid method for high efficiency and robustness. In addition, the multigrid method can be applied to parallel time integration, which is usually calculated by time-marching method. We have already proposed new approaches. In this project, we will enhance and apply them to new applications.

Our target areas are linear solvers and PinST methods as follows.

GMG and AMG: Efficiency and Robustness for Geometric and Algebraic MultiGrid

- Research on smoothers

A smoother is a key component of the multigrid algorithm for efficiency and robustness. We will study efficient implementation methods of multi-color smoothers to AMG and GMG. We will also consider to accelerate these smoothers by SIMD oriented sparse matrix data structures such as SELL-C-σ.

- Lower precision utilization

Lower precision accelerates the calculation, but needs more iteration for convergence. We will study how to use the lower precision calculations for efficient multigrid solvers.

- Acceleration technique with file IO optimization

Parallel linear solvers and mesh generation routines often need to read and write distributed matrix files. Especially when the number of computing nodes are increased, optimization of them is needed. We will investigate efficient usage of the burst buffer functionality of supercomputers for mesh generation or sparse linear solvers.

- Evaluation with weak scaling and large sized problems

Our project already has spent several years to implement multigrid solvers such as GMG and AMG. We will analyze the performance of the solvers, and check the space for improvement.

PinST: Parallel in Space Time

- New approaches for PinST

We are developing new approaches such as Time segmented correction (TSC) method and MGRIT preconditioning. We will enhance the algorithms and evaluate them.

- Application of PinST to Explicit time integration

We consider explicit time integration is an important area for parallel time integration. We will apply multigrid methods to various explicit practical simulation and increase case studies.

3 Significance as JHPCN Joint Research Project

Multigrid methods are scalable and used in many fields. Our research project has original codes and algorithms as written in Research purpose. Therefore, research papers and codes from the project will enhance the efficiency of the multigrid solver, and will help many researchers exploit parallelism in time direction.

Our research focuses on hierarchical algorithms and their performance on supercomputers. Thus, availability of supercomputers with different kinds of architectures helps us verify the codes we are developing. In addition, a JHPCN joint research project offers collaborative research opportunity with JH-

PCN members who have expertise knowledge in various application fields. We are sure that this JHPCN joint research project promotes the international collaborative activity with JHPCN members.

4 Outline of Research Achievements up to FY2019

5 Details of FY2020 Research Achievements

This project is set as a 3-year project, and this final report is for the first year. For GMG and AMG optimization, we have checked SELL-C-σ data structure for sparse matrix in [Nakajima-SELL1, Nakajima-SELL2]. We also investigated the communication optimization using Pipelined CG method for AMG solver in [Yoda-Pipeline]. We have checked our AMG solver's scaling performance with large sized problems with 10^{10} unknowns [Fujii-AMG]. As for PinST, we studied Parareal method for explicit time marching scheme [Chen-PinST]. MGRIT's scalability enhancement with 3 techniques is investigated in [Yoda-PinST2,3]

In the following subsections, research results are introduced with graphs in detail.

5.1 GMG with SELL-C-σ ILU smoother

In the work of [Nakajima-SELL2], authors introduced SELL-C-σ to the MGCG solver of pGW3D-FVM, and evaluated the performance of the solver with various types of OpenMP/MPI hybrid parallel programing models on OFP using up to 1,024 nodes. Because SELL-C-σ (Fig. 2) is suitable for wide-SIMD architecture, improvement of the performance over the sliced ELL(Fig. 1) was more than 20%, which is shown in Fig. 3. This is one of the first examples of SELL-C-σ applied to forward/backward substitutions in ILU-type smoother of multigrid solver. Furthermore, effects of IHK/McKernel has been investigated in Fig 4, and it achieved 11% improvement on OFP with 1,024 nodes. This is just a preliminary work, and we are going to implement SELL-C-σ to MGCG solvers with hCGA and AM-hCGA for more efficient computations at larger number of nodes. Further investigations of effects of IHK/McKernel on performance of OpenMP/MPI hybrid parallel programming models are also conducted. According to the recent work by the authors, reducing number of threads at coarser level may improve the performance of multigrid method with multiple threads[Nakajima-SELL1]. Investigation on automatic selection of the optimum number of threads and the optimum method for matrix storage is the important topic for the future work.

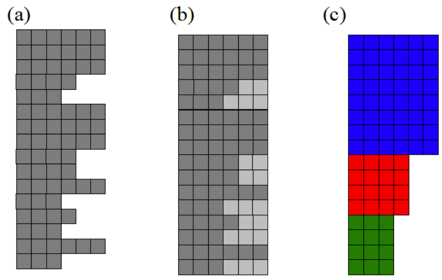

Fig. 1 Storage Format for Sparse Matrices, (a) CRS (Compressed Row Storage, only non-zero components are stored), (b) ELL (Ellpack-Itpack) (components colored in light gray are set to 0), (c) Sliced ELL

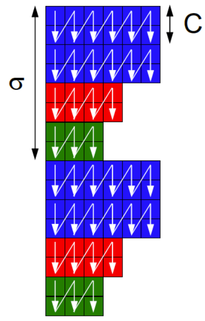

Fig. 2 SELL-C-σ, C: chunk size, σ: sorting scope, C=2 and σ=8 in this case

5.2 AMG with AM-hCGA for large sized problems

In [Fujii-AMG], authors investigates the effectiveness of AM-hCGA to AMG solver. AM-hCGA with independent aggrega-

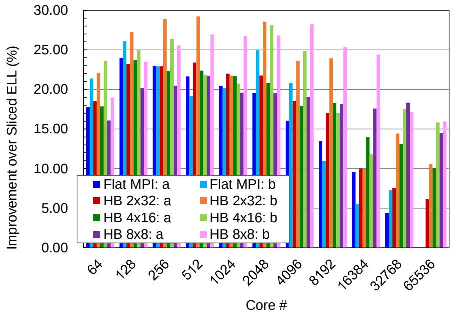

Fig. 3 Performance Improvement of MGCG solver by SCS-a (:a) and by SCS-b (:b) over *Sliced ELL*

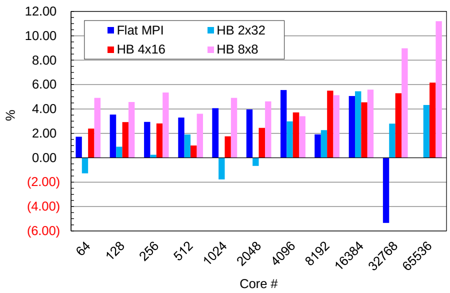

Fig. 4 Performance Improvement of MGCG solver with SCS-b by IHK/McKernel

tion(CGA_LU) is compared with various coarse grid generation strategies (w/o_LU, LU, Coupled_LU). w/o_LU, LU, and Coupled_LU corresponds with independent aggregation with iterative coarsest smoother, independent aggregation with direct coarsest solver, and coupled aggregation with direct coarsest solver, respectively. The problem is the Poisson equation with heterogeneous change of the distribution coefficient between 10^{-5} to 10^{5}. This problem is the similar setting with pGW3D-FVM. Fig. 5 has iteration number and convergence time for each problem DOF between 10^{7} and 10^{10}. Solid and dashed line corresponds to the total time and setup time, respectively. The figure shows the high scalability of AM-hCGA strategy(CGA_LU).

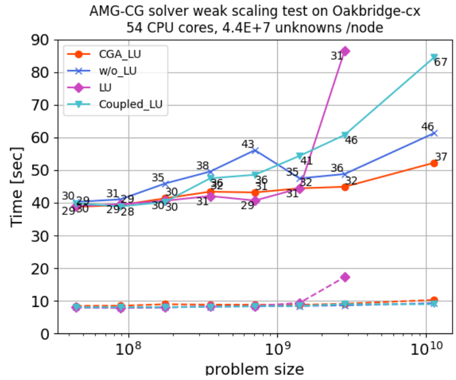

Fig. 5 Weak scaling performance and multi-level setup strategies

The solver performance is compared with PETSc-gamg with default parameter setting in Fig. 6. The PETSc-gamg solver can solves up to 10^{9} DOF matrix on the Oakbridge-CX . Thus comparison is done between 10^{7} DOF and 10^{9} DOF problems. Fig. 6 shows our code reaches convergence twice faster than PETSc-gamg with default setting.

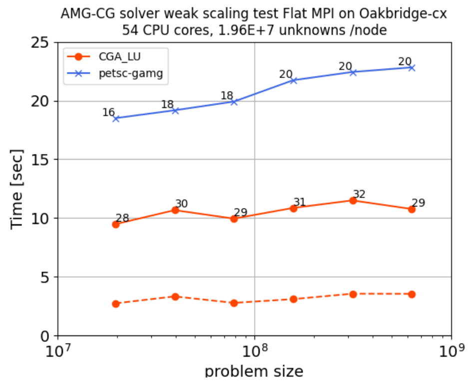

Fig. 6 Performance comparison with PETSc-gamg

5.3 MGRIT scalability enhancement with three techniques

As with the traditional multigrid method for spatial problems, MGRIT's scalability deteriorates due to increased communication costs on the coarsest level with high parallelism.

In this research[Yoda-PinST2,3], authors apply three techniques to MGRIT in order to enhance the scalability. The first one is coarse-grid agglomeration (CGA), which shrinks the number of active processes on the coarse grid. The second one is the use of approximate solves on the coarsest-level. The third one is the use as an MGRIT preconditioner proposed in previous work.

Numerical experiments were conducted for the two-dimensional time-dependent Stokes problem on a staggered grid. Fig. 7 shows the effect of coarse-grid agglomeration in strong scaling of temporal parallelism. We can see that the performance deterioration of simple MGRIT due to the increased cost of solving the coarsest problem. In contrast, MGRIT with CGA works with higher parallelism efficiently. Authors observed that CGA provides a well-scaling property for MGRIT. Next, authors investigate the effect of the combination of three techniques. The approximate solves provided a similar convergence behavior compared with the direct method. Also, the MGRIT preconditioner improved the convergence. The combination of the three techniques achieved a decrease of the run-time more than four times compared with simple MGRIT. Future works investigate the effect of the coarsening ratio and number of levels and apply time-dependent Oseen problems.

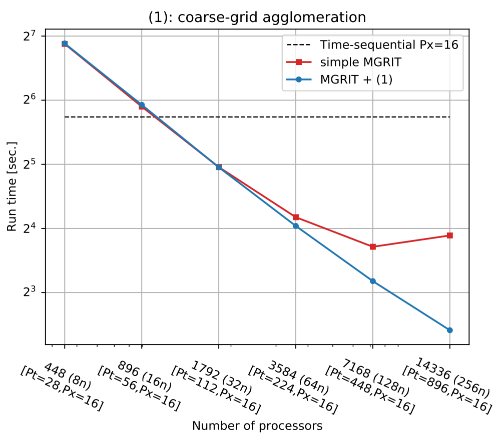

Fig. 7 Strong scaling experiment of coarse-grid agglomeration

5.4 Parallel-in-time method for explicit schemes

In this research[Chen-PinST], authors propose a novel parallel-in-time method for explicit schemes. The method aims to exploit the most performance while using explicit time-marching schemes as base solvers.

The proposed method constructs multiple layers of solvers with different time step size, similar to the MGRIT method. Each coarse level picks out time points uniformly from its upper fine level. In order to satisfy the CFL condition, which is necessary while working with explicit schemes, each coarse level coarsens both space and time grid with the same ratio. Moreover, the method defines relaxation base on the Parareal algorithm using the coarsest level and a finer level. This way the explicit schemes solves more efficiently compare to the FCF-relaxation used by the MGRIT method. From coarse to fine levels, results from each Parareal algorithm serves as an initial guess for the next level, which efficiently reduces the computational cost and iteration number.

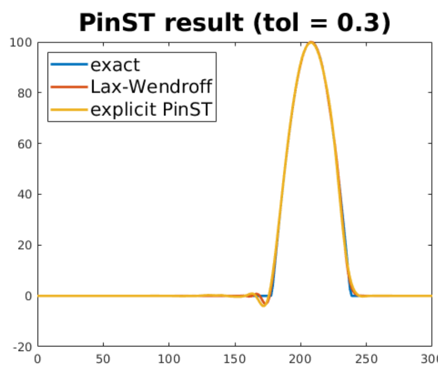

Fig. 8 Numerical result of the parallel-in-time method compare to the result of the sequential explicit scheme (Lax-Wendroff method).

The numerical experiment solves an one-dimensional advection equation of a sine wave with the proposed parallel-in-time method. Fig. 8 shows the result comparison of the proposed method with relative tolerance 0.3 and the sequential result using

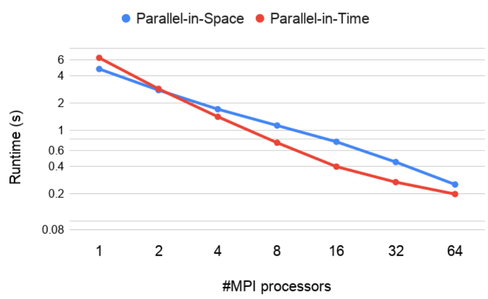

Fig. 9 Runtime comparison of parallel-in-space and parallel-in-time.

the Lax-Wendroff method; and Fig. 9 shows the runtime comparison using spatial parallelization and the proposed parallel-in-time method. The results show that the proposed method could converge very fast to a similar result and the largest error concentrates at the discontinuous points of the original wave. This result shows that parallel-in-time methods could achieve similar parallelize performance in the time dimension compare to the space dimension even for explicit schemes.

6 Progress during FY2020 and Future Prospects

This is a 3-year project starting from FY.2020. As a first year, we took over previous MG project, and enhance the research results described in the Section 5. Things are generally going well.

Research plan in FY.2021 is the further optimization and enhancement in each items listed in Sec. 2.

7 List of Publications and Presentations

Proceedings of International Conferences (Refereed)

Proceedings of International Conferences (Non-refereed)

[Chen-PinST] Chen, Y.C., Nakajima,K., A parareal-based parallel-in-time method for explicit time-marching schemes, IPSJ SIG Technical Report, 2020-HPC-175-20(Online, July 31, 2020).

[Yoda-PinST] 依田凌，中島研吾，Matthias Bolten，藤井昭宏，時空間行列に対する GMRES 上限解析への SAMA の適用，第 29 回単独研究会（SWoPP2020）日本応用数理学会「行列・固有値問題の解法とその応用」研究部会第 29 回研究会（in Japanese）

[Nakajima-SELL1] 中島研吾，メニィコアクラスタ向け並列多重格子法，情報処理学会研究報告（2020-HPC-176-6）（第 176 回 HPC 研究会）（オンライン，2020 年 9 月 15 日）(in Japanese)

[Nakajima-SELL2] Nakajima,K., Gerofi,B., Ishikawa, Y., Horikoshi, M., Efficient Parallel Multigrid Method on Intel Xeon Phi Clusters, ACM proceedings of IXPUG(Intel Extreme Performance Users Group) in conjunction with HPC Asia 2021, January 2021. (in press)

[Yoda-Pipeline] 依田凌，中島研吾，藤井昭宏，Oakbridge-CX における Pipelined CG 法への AMG 前処理の適用，日本応用数理学会，2020 年度年会（オンライン，2020 年 9 月 9 日）（in Japanese）

[Fujii-AMG] 藤井昭宏，田中輝雄，大規模線形問題における代数的多重格子法の粗格子集約手法の有効性評価，情報処理学会研究報告（2020-HPC-177-9），（第 177 回 HPC 研究会）（オンライン，2020 年 12 月 21 日）(in Japanese)

[Yoda-PinST2] 依田 凌，中島 研吾，Matthias Bolten，藤井 昭宏，時間発展 Stokes 方程式に対する粗格子集約を用いた Multigrid Reduction in Time の適用，情報処理学会研究報告（2020-HPC-178-3），（第 178 回 HPC 研究会）（オンライン，2021 年 3 月 15 日）（in Japanese）

[Yoda-PinST3] Yoda, R., Bolten, M., Nakajima, K. and Fujii, A., Multigrid Reduction in Time using Coarse-grid Agglomeration, approximate solves and Krylov acceleration, 20th Copper Mountain Conference On Multigrid Methods (Online, March 31, 2021)

Published library and relating data

Other (patents, press releases, books and so on)