

jh210003-NAH

GPU・CPU・ARM プロセッサに対する原子力 CFD アプリケーション用の 混合精度ポアソン解法

小野寺 直幸（日本原子力研究開発機構）

概要

エクサスケールスパコンを活用した原子力数値流体力学（CFD）解析の実現に向けて、圧力 Poisson 解法の混合精度演算の適用による高速化、および、GPU・CPU・ARM プロセッサ等の様々な演算機向けの最適化を実施した。原子力 CFD 問題であるバンドル体系に対する気液二相流体解析を実施した結果、前処理手法に単精度演算（fp32）を適用することで、倍精度と比較して 75%の計算時間で解析が可能であることが確認された。マルチプラットフォーム向けの実装では、マクロを用いて並列ループを置き換えることで、ブロック構造 AMR 格子版の Poisson 解法において、CPU・GPU 上での高い演算性能と性能移植性を実現した。一方で、A64FX での富士通コンパイラによる最適化では、GNU/Intel/NVIDIA のコンパイラと同様の最適化が行われない問題点が明らかとなった。左記の問題に対して、直交格子版 JUPITER において、手動での関数インライナ化等の最適化を実施することで、高い演算性能を実現した。

1. 共同研究に関する情報

(1) 共同研究を実施している拠点名

- ・東京工業大学学術国際情報センター
- ・東京大学情報基盤センター
- ・名古屋大学情報基盤センター

(2) 共同研究分野

■ 超大規模数値計算系応用分野

(3) 参加研究者の役割分担

- ・ 代表者（日本原子力研究開発機構）：
小野寺 直幸：コード開発・解析結果の評価
- ・ 副代表者（東京工業大学）：
青木 尊之：TSUBAME3.0 での最適化の助言
- ・ 協力者（日本原子力研究開発機構）：
井戸村 泰宏：結果の評価、Poisson 解法の助言
山下 晋：物理モデルの助言、結果の評価
河村 拓馬：計算結果の可視化
朝比 祐一：ディレクティブ・組み込みの助言
伊奈 拓也：CPU・GPU・A64FX 用の最適化
長谷川 雄太：GPU 最適化
杉原 健太：CFD モデルの構築
- ・ 協力者（東京大学）：
下川辺 隆史：BDEC での最適化の助言
- ・ 協力者（名古屋大学）：

大島 聡史：Flow での最適化の助言

2. 研究の目的

本研究は、エクサスケールスーパーコンピュータを活用した原子力数値流体力学（CFD）解析の実現に向けて、圧力 Poisson 解法の高速化・高度化を目的とする。具体的には、ブロック型適合細分化格子（AMR）法に適した前処理手法の高速化、および、様々な演算機での最適化に対応可能なプログラミングモデルの開発、を実施する。これを、日本原子力研究開発機構（JAEA）で開発している多相流体解析モデル JUPITER[参考文献 1]に適用することで、原子力 CFD 解析の更なる高速化・大規模化を実現する。

JUPITER コードは継続的に高速化されており、昨年度の JHPCN 課題では、GPU・CPU・ARM プロセッサ上にて動作するブロック型 AMR 格子に対応した、マルチグリッド（MG）前処理付き Poisson 解法を開発し、TSUBAME の GPU を用いた計算において、MG 法を適用しない P-CG 法と比較して 10 倍以上の高速化を達成した[参考文献 2]。一方で、図 1 に示すバンドル体系に対する多相流体解析[研究成果 3]には、非常に長い計算時間が必要であるため、更なる高速化が求められて

いる。

今年度の JHPCN 課題では、(a) 前処理手法に対する半精度・単精度浮動小数点数 (FP32, FP16) の適用、並びに (b) GPU・CPU・ARM プロセッサに対する最適化を実施することで、高速化および性能可搬性の両立を目指す。

以上の高度化により、原子力分野の CFD 解析で必要となるエクサスケールスパコンを用いたマルチスケール現象への非定常解析が初めて可能となり、様々な工学問題への発展に貢献できる。

2.1. 研究の意義

(a) 混合精度を用いた前処理手法の高速化

JUPITER で取り扱う Poisson 方程式の疎行列は 7 本の帯を持つブロック対角行列であるため、Red-Black (RB) -SOR 法等の前処理手法ではメモリアクセスが計算速度のボトルネックとなる。本課題では、FP32・FP16 を用いた前処理手法を開発することで、メモリアクセスを削減する。更に、これを昨年度の JHPCN 課題で提案したキャッシュを再利用した反復解法 (CR-SOR) に適用することで、更なる高速化を目指す。以上の、メモリアクセスの最適化およびキャッシュを利用した前処理手法の開発により、様々な工学問題に対する疎行列反復解法の高速化に貢献することが可能である。

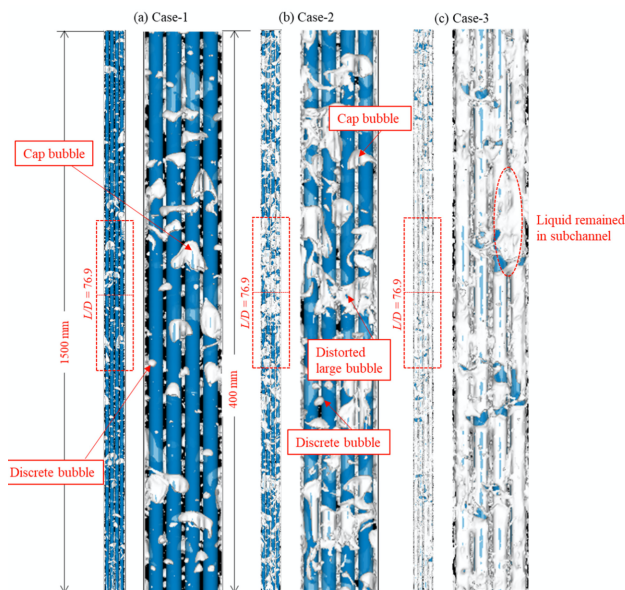


図 1 4×4 バンドル体系での様々な流動形式に対する気液多相流体解析結果。

(b) GPU・CPU・ARM プロセッサの最適化に対応したプログラミングモデル

エクサスケールスパコンの演算器の候補として GPU・CPU・ARM が挙げられ、これらのプロセッサ間の計算性能の移植性は大きな研究課題となっている。本研究では、ブロック型 AMR 法を適用した Poisson 解法が局所的なメモリアクセスによる演算から構成されることに着目し、キャッシュを明示的に利用し、マルチプラットフォームに対応したプログラミングモデルの構築を目指す。具体的には、C++ のプリミティブな記述、OpenMP 等のディレクティブ、およびメタプログラミングによるコード生成、のそれぞれの方法に対して性能可搬性を調査する。最終的には、GPU・CPU・ARM に対する最も効率的なプログラミングモデルを明らかとし、JUPITER コードに適用する。

(c) 原子力工学分野の CFD に対する大規模計算

原子力工学分野の多相流体解析は、構造物に起因した乱流と気液界面を捉えたマルチスケール解析が高コストになるため、これまでは経験則に基づいて混相流をモデル化した流体解析が主であった。本研究では、AMR 格子に基づく Poisson 解法の高速度化により、気液界面を直接捉えた高解像度の非定常解析の実現を目指す。具体的には、[参考文献 3] で実施したバンドル体系に対する解析を、様々な実験条件にて実施し、実験結果の再現を目指す。このような多相流体解析は幅広い工学問題に適用可能であることから、原子力分野の CFD ソフトウェアの発展だけでなく、実験手段の代替として工学分野に与える影響は大きい。

3. 当拠点公募型研究として実施した意義

本研究プロジェクトは、ブロック型 AMR 法を適用した Poisson 解析を高速化し、最新のスーパーコンピュータ上にて大規模な原子力 CFD 解析を実現することが目標である。この実現には、エクサスケールスーパーコンピュータの有力な候補である、GPU・CPU・ARM プロセッサの利用および最適化の知見が必須となる。そこで、それらを有する東工大・東大・名大と共同研究を実施するこ

とで、初めて研究課題が達成できる。

4. 前年度までに得られた研究成果の概要

本課題の基盤となる JUPITER コードは、研究協力者の山下と共に実験の代替に向けた検証を進めた。また、研究代表者の所属するシステム計算科学センターでは、Poisson 解法の CPU・GPU 向けの高高速化として、MG 法に基づく Poisson 解法および省通信手法など、計算アルゴリズムおよび最適化を長年行ってきた[参考文献 2,4-6]。昨年度の JHPCN 課題の成果として、ブロック AMR 法を適用した格子に対して、ブロック間およびブロック内の格子間の依存関係を階層的な参照により解決し、さらにブロック内の格子データを GPU キャッシュ (shared memory) に読み込み再利用することで、収束性と高速計算を両立した前処理手法 (CRMG-CG 法) を提案した。CRMG-CG 法では

CG 法に MG 法前処理を適用し、更に MG 法のスムーザとして CR-SOR 法を使用する。図 2 に 8×8 バンドル体系に対する Poisson 解法の収束履歴を示す。従来の RB-SOR 法前処理付き CG 法 (P-CG 法) では、収束まで約 1300 回の反復が必要なのに対して、MG 法前処理を適用した MG-CG 法では、約 200 回、提案手法である CRMG-CG 法では約 100 回と、大幅に収束性が改善した (図 2) [参考文献 2]。一方で、この強スケーリング性能では、依然として Poisson 解法の前処理が計算時間の大部分を占めており、更なる高速化が望まれる (図 3)。今年度の JHPCN 課題では、混合精度演算の適用による前処理手法の高速化を目指した。

マルチプラットフォームに対する実績として、OpenACC や Kokkos 等の最新のライブラリを用いた成果が挙げられる[参考文献 7]。この知見を基に、Poisson 解法および JUPITER-AMR のマルチプラットフォーム向けの設計方針がたてられた。

参考文献

- [1] S. Yamashita, T. Ina, Y. Idomura, and H. Yoshida, "A numerical simulation method for molten material behavior in nuclear reactors," Nuclear Engineering and Design, vol. 322, pp. 301 - 312, 2017.
- [2] N. Onodera, Y. Idomura, Y. Hasegawa, S. Yamashita, T. Shimokawabe, and T. Aoki, "GPU Acceleration of Multigrid Preconditioned Conjugate Gradient Solver on Block-Structured Cartesian Grid", HPC Asia 2021, 2021.
- [3] A. Ono, S. Yamashita, T. Suzuki, and H. Yoshida, "Numerical simulation of two-phase flow in 4x4 simulated bundle", Mechanical Engineering Journal, Vol. 7, No. 3, Paper No. 19-00583, pp 1 - 12, 2020
- [4] Y. Idomura, T. Ina, Y. Ali, and T. Imamura, "Acceleration of fusion plasma turbulence simulations using the mixed-precision communication-avoiding Krylov method",

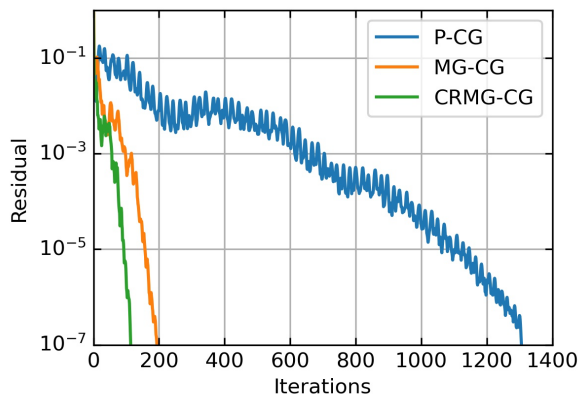


図 2 8×8 バンドル体系に対する Poisson 解法の収束履歴。1024×1024×3072 格子相当。

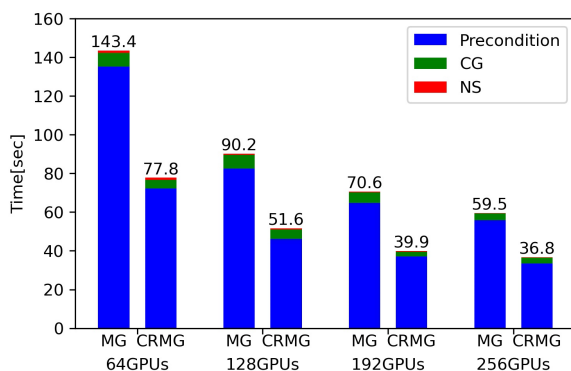


図 3 8×8 バンドル体系に対する Poisson 方程式の強スケーリング性能

Proceedings of SC 2020, p. 1318 – 1330, 2020

- [5] Y. Ali, N. Onodera, Y. Idomura, T. Ina, et al., “GPU Acceleration of Communication Avoiding Chebyshev Basis Conjugate Gradient Solver for Multiphase CFD Simulations”, ScalA'19 workshop in SC19, pp. 1 - 8, 2019.
- [6] Y. Idomura, T. Ina, S. Yamashita, N. Onodera, et al., “Communication Avoiding Multigrid Preconditioned Conjugate Gradient Method for Extreme Scale Multiphase CFD Simulations”, ScalA'18 workshop in SC18, pp. 17 - 24, 2018.
- [7] Y. Asahi, G. Latu, V. Gradgirard, and J. Bigot, “Performance Portable Implementation of a Kinetic Plasma Simulation Mini-App”, WACCPD in SC19, pp.1 - 23, 2019

5. 今年度の研究成果の詳細

5.1 混合精度を用いた前処理手法の高速化

今年度の研究成果として、TSUBAME の GPU を用いて、2020 年度に開発した MG 法に基づく Poisson 解法に対して、混合精度の前処理手法の実装および性能測定を実施した。図 4 に 5×5 バンドル体系の計算条件および主流方向・スパン方向断面の速度分布・流体率分布の瞬時値を示す。計算格子として、直交格子の 384×384×6144 (Leaf 数は 48×48×768) 相当を設定した。境界条件として、バンドル下部の領域に水・空気の流入境界を、バンドル上部に流出境界を設定した。

図 5 に上記の解析の収束履歴を示す。Poisson 解法として、P-CG 法および CRMG-CG 法を比較した。CRMG-CG 法の前処理として、倍精度計算 (fp64)、単精度計算 (fp32)、単精度計算・半精度通信 (fp32-fp16) を採用した。(A) の収束履歴より、P-CG 法の 900 回に対して、CRMG-CG 法では 100 回へと収束性を劇的に改善した。また、CRMG-CG 法の混合精度を用いた 2 つの条件では、いずれも倍精度と同様の収束履歴となることを確認した。(B) の計算時間の比較では、前処理に単

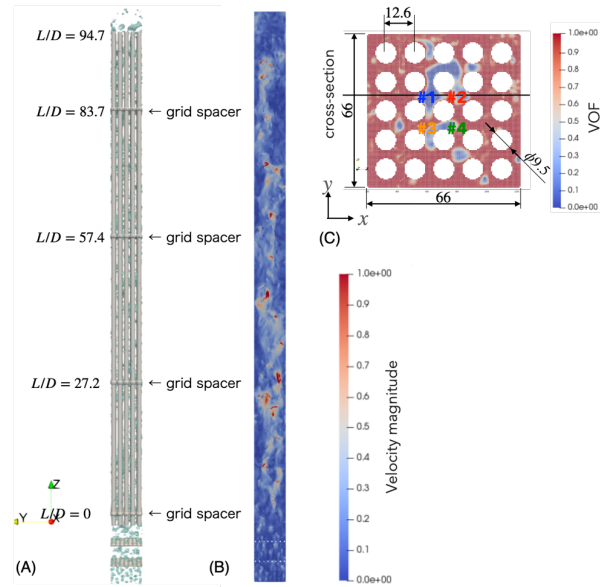
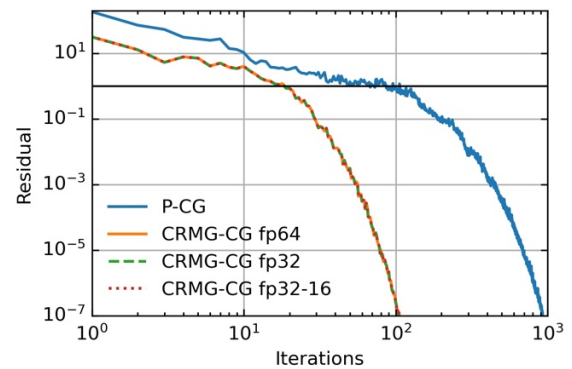
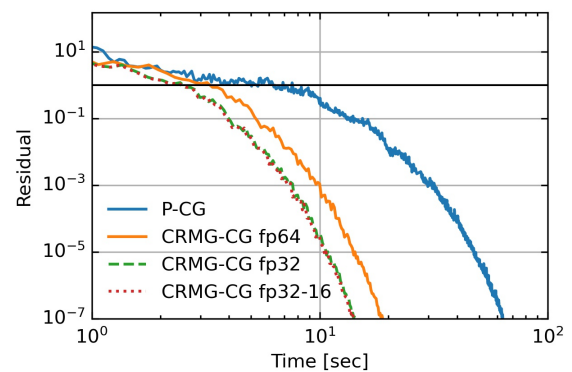


図 4 (A) 5×5 バンドル計算体系での気液海面、(B) バンドル間断面の速度分布、(C) スパン方向断面の流体率、の可視化



(A) 反復回数 - 収束残差



(B) 計算時間 - 収束残差

図 5 P-CG 法 (青線) および CRMG-CG 法 (倍精度計算 fp64: 橙線、単精度計算 fp32: 緑破線、単精度計算-半精度通信 fp32-16: 赤点線) の収束履歴

精度を採用することで、倍精度に対して 75%程度までコストを削減した。一方で、通信に半精度を用いた条件 (fp32-16) では、単精度の条件とほぼ同じコストとなり、期待通りの高速化が実現されなかった。また、前処理に半精度計算 (float16、もしくは bfloat16) を採用した条件での計算も実施したが、Poisson 解法が収束しない結果となった。

5.2 マルチプラットフォーム向け Poisson 解法の開発

GPU・CPU・ARM プロセッサに対応したマルチプラットフォーム向けの開発として、2020 年度課題に引き続いて、C++および CUDA を用いて実装されている JUPITER-AMR の移植・最適化を進めた。ここでは、2020 年度に実施した OpenACC や Kokkos 等の最新のライブラリを用いた知見[参考文献 7]を基に、GPU に対応した CUDA コードから CPU コードを作成する場合には、マクロを用いて並列化が可能なループ構造を置き換えていく開発方針を設定した。具体的には、GPU 最適化に対応した CUDA カーネルのブロック・スレッド文と、CPU 最適化に対応した OpenMP のループ構造に対する指示文 (#pragma omp parallel for 等) を切り替えることが可能なフレームワークを作成した。上記の実装により、Intel CPU および NVIDIA GPU において高い計算性能と移植性を両立した。一方で、A64FX においては、ブロック型 AMR 格子のデータ構造では、満足な性能が得られなかった。上記を共同研究先の富士通株式会社と共に調査したところ、A64FX ではソフト・ハードウェア的に性能が得られないことが判明した。以下に最適化に必要な条件や問題点を挙げる。まず、A64FX の性能を引き出すためには SVE (Scalable Vector Extension) の有効化が重要であり、富士通コンパイラ (trad モード) の利用が必須となる。また、高い性能を達成するためには、for ループに対して、ソフトウェアパイプラインおよび SIMD 最適化の有効化が必須となる。しかしながら、A64FX において以下の問題が発生した。

- ・ 富士通コンパイラでは、GNU/Intel/NVIDIA コ

ンパイラで最適化される for ループ内の関数呼び出しに対応しておらず、ソフトウェアパイプラインおよび SIMD 最適化が阻害される。関数をインライン関数に置き換えても、最適化されないため、手動での関数のインライン化が必要となる。

- ・ 富士通コンパイラでは、GNU/Intel/NVIDIA コンパイラで最適化される for ループ内の条件分岐に対応しておらず、手動にてマスク処理による条件分岐の書き換えが必要となる。
- ・ ブロック構造 AMR 格子のデータ構造では、for ループ内に格子のインデックス取得用 (リストアクセス用) の関数が含まれるため、その手動インライン化が必要となる。また、リストアクセスによる性能劣化が CPU・GPU と比較して大きいことが確認された。
- ・ CRMG-CG 前処理手法は、ループ内に複雑な処理を含むため、上記の最適化を実施しても、ソフトウェアパイプラインおよび SIMD 最適化が有効とならなかった。

結局、最適化後の CRMG-CG 法の前処理手法の性能として、TSUBAME の CPU (Intel Broadwell)、GPU (NVIDIA P100) ではそれぞれルーフラインモデルの 14%、37% という妥当な実行性能が得られたのに対し、A64FX では 1% の実行性能しか得られず、高速化が実現されなかった。

一方で、直交格子版の JUPITER では、反復改良法と ILU 法を組合せた混合精度前処理を適用した P-

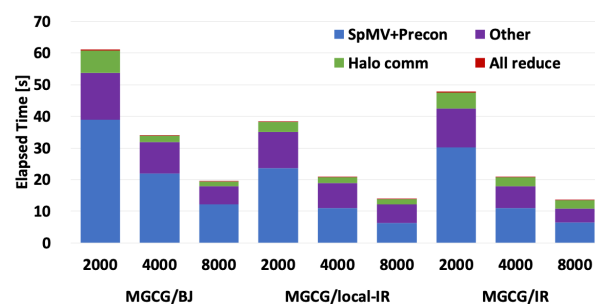


図 5 直交格子版 JUPITER の富岳の A64FX を用いた様々な Poisson 解法に対する強スケーリング性能。

CG 法および MG-CG 法を開発することで、A64FX 向けの最適化に成功した。直交格子版の JUPITER は 8000 台の A64FX に対しても、良い強スケーリング結果を達成し、研究成果を高性能計算に関する国際会議である SC21 のワークショップ ScalA21 にて発表した (図 5) [研究成果 1]。

5.3 バンドル体系の気液二相流体解析

原子力工学分野の CFD に対する大規模計算として図 5 に示すバンドル体系の気液二相流体解析を実施した。計算条件として、ブロック構造 AMR 格子版の JUPITER-AMR に 0.58mm 解像度 (直交格子の $128 \times 128 \times 2,048$ 相当) を設定すると共に、10 秒間 (約 500,000 ステップ) の解析を実施した。統計量として、サブチャンネル内のボイド率の確率分布を、直交格子版 JUPITER (1mm 格子解像度) および実験結果 [Ren et al., Meas. Sci. Technol., 2018] と比較した (図 6)。実験の流動形式として、ボイド率が 0 付近の確率分布が最も高い気泡流 (bubbly flow) が観測されているが、ボイド率が 0 (流路内が流体のみ) の確率分布が実験の約 0.22 と比較して、JUPITER-AMR は 0.36、JUPITER は 0.55 となり、過大評価する結果となった。一方、確率分布の大半を占めるボイド率が

0 から 0.1 の領域では、JUPITER が実験値を過小評価しているのに対して、JUPITER-AMR が実験値とよく一致している。ボイド率が 0.1 以上においては、JUPITER-AMR が実験値の確率分布を過大評価しており、この改善が今後の課題と考えている。以上より、GPU を用いた JUPITER-AMR の開発により、CPU を用いた従来の直交格子版 JUPITER と比較して、高解像度かつ高速な解析を実現し、実験結果を高精度に再現できることが示された。

6. 今年度の進捗状況と今後の展望

2021 年度は、(a) 混合精度を用いた前処理手法の高速化、(b) マルチプラットフォームに対応した Poisson 解法の開発、および (c) 原子力工学分野の CFD に対する大規模計算、を実施した。

(a) の前処理手法の高速化の成果として、前処理に単精度を採用することで、倍精度と同じ収束履歴を維持しつつ 75% 程度までコストを削減することに成功した。一方で、気液二相流体解析のような問題に対しては fp16 や bfloat16 等の半精度を利用した前処理手法では、収束しないことが確認された。現状では事前計算により、前処理手法の演算精度を選択しているが、今後は行列の条件数等から、それらを定める手法の開発が必要であると考えている。

(b) のマルチプラットフォーム向けの開発では、並列処理の可能なループに対して、マクロを用いたフレームワークを新たに開発することで、Intel CPU と NVIDIA GPU 間の高い性能移植性を実現した。一方で、A64FX においては、ソフトウェア・ハードウェアの問題により、非常に多くの書き換えによる開発コストの増加と、ブロック構造 AMR 格子では致命的な性能劣化が引き起こされることが確認された。今後の展望としては、GPU・CPU 版の JUPITER-AMR の開発と、A64FX 版の直交格子版 JUPITER の開発を分離して進めていく予定である。また、現状の富士通コンパイラが GNU/Intel/NVIDIA のコンパイラと比較して、非常に多くの問題点を抱えていることに対して、今後の改善を期待している。

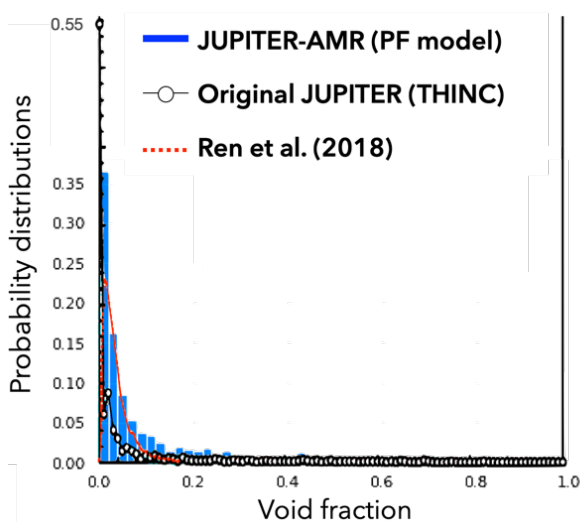


図 6 サブチャンネル内のボイド率の確率分布。青線：JUPITER-AMR (0.58mm 解像度)、白丸：直交格子版 JUPITER (1mm 解像度)、赤線：実験結果

最後に (c) の原子力工学分野の CFD 解析に関しては、GPU に対応した JUPITER-AMR の開発により、CPU を用いた直交格子版 JUPITER の倍の解像度である 0.58mm 格子の解析を実現し、実験結果をより高い精度で再現することに成功した。一方で、ボイド率が 0.1 以上の領域においては、実験結果を過大評価している。この原因の一つとして、格子幅程度に接近した気泡の非物理的な合体が挙げられ、気液界面捕獲手法の改良により改善できると考えている。

2022 年度の JHPCN 課題「原子力气液二相流体解析における界面捕獲手法の高度化」では、開発した GPU 版 JUPITER-AMR を基に、共同研究相手である東工大 青木研究室の有するそれぞれの気泡に対して独立の変数を与えるマルチフェーズフィールド法の知見を基に、その合体を制御するモデルを構築することで多相流体解析の高精度化を進め、バンドル体系の解析のさらなる高精度化と様々な流動形式の再現を目指す予定である。

7. 研究業績一覧

(1)国際会議プロシーディングス (査読あり)

T. Ina, Y. Idomura, T. Imamura, S. Yamashita, and N. Onodera, "Iterative methods with mixed-precision preconditioning for ill-conditioned linear systems in multiphase CFD simulations", ScalA21 @ SC21 (11/19, Online)

(2)国内会議発表 (査読なし)

小野寺 直幸、井戸村 泰宏、朝比 祐一、長谷川 雄太、下川辺 隆史、青木 尊之、”ブロック型適合細分化格子での Poisson 解法の混合精度演算による高速化”、日本計算工学会第 26 回計算工学講演会 (5/26-28、オンライン)