

jh250073

AI leveraged multiscale scheme for kinetic equations

安田 修悟（兵庫県立大学大学院 情報科学研究科）

概要

本研究では、流体力学やプラズマ物理などに現れる運動論方程式の数値計算における数値的剛性の課題に対し、AI 技術と HPC を融合した高速化手法の構築を目的とした。特に、既存の陰解法スキームの精度と構造を保持したまま、AI により内部反復計算を加速する「MIWS-Net (Macro Informed Warm Start Neural Network)」の開発を行った。2025 年度は、大阪大学 D3 センターの SQUID GPU 環境を用いて、モデル Boltzmann-BGK 方程式に対する GPU 実装および AI 援用型数値スキームの検証を実施した。その結果、1 次元衝撃波管問題において、従来の陰解法と同等の計算精度を維持したまま、ピカール反復回数を少なくとも約 3 倍、条件によっては一桁以上削減できることを確認した。

1 共同研究に関する情報

1.1 共同研究を実施した拠点名

- 大阪大学 D3 センター

1.2 課題分野

- 大規模計算科学課題分野

1.3 参加研究者の役割分担

- 安田 修悟（兵庫県立大学）：代表，研究統括・モデリング
- 沼田 龍介（兵庫県立大学）：副代表，プラズマ物理・モデリング
- Martin Frank（カールスルーエ工科大学 /KIT）：副代表，国際共同研究（NHR 連携）・AI モデル
- 初鳥 匡成（京都大学）：希薄気体・GPU コード開発
- 高田 滋（京都大学）：希薄気体・モデリング助言
- 小田 浩太郎（兵庫県立大学）：ソフトマ

ター・コード開発支援

- 宿南 湧真（兵庫県立大学）：大学院生，MIWS-NET・kineticEQ の開発・検証

下線は、2025 年度における主要な研究成果の主担当者を示す。

2 研究の目的と意義

本研究は、AI 技術と高性能計算（HPC）を融合することにより、流体力学、プラズマ物理、生物学、社会科学など多様な分野に応用される運動論方程式のモデリングおよび数値計算における計算効率の改善を目的とするものである。これに加えて、将来的には、分子・粒子レベルのミクロ相互作用を記述する衝突項モデルを、データ駆動型アプローチにより構築するための方法論的基盤を確立することも目的としている。

研究計画は以下の二つの柱から構成される。

- (1) 物理的・数学的性質を保存した衝突項サロ

ゲートモデルの構築：

運動論モデルの衝突項に対して、保存則や H 定理などの物理的・数学的性質を保持する AI サロゲートモデルの構築を行う。まず、標準的なボルツマン衝突項を対象として、数値計算データに基づくサロゲートモデルの構築手法を確立する。さらに、本手法を発展させることで、分子動力学計算や実験データなど、様々なマイクロ相互作用を記述するデータに基づく、データ駆動型衝突項モデルへの展開可能性について検討する。

- (2) 既存数値スキームの精度と骨格を保持した高速化技術の確立：

既存の数値計算スキームの精度および基本構造を保持したまま、AI を補助的に導入することで計算を加速する手法の確立を行う。特に、多様な運動論方程式の数値計算に共通して現れる数値的剛性に対して、個別のモデルやスキームに強く依存しない、汎用性の高い高速化手法の構築を目指す。

2025 年度は特に (2) に重点を置き、AI による反復計算の高速化手法の検証を行った。

本研究の学術的意義は、物理的・数学的制約を組み込んだ AI サロゲートモデルの構築と、運動論方程式に共通する構造に基づく AI 援用型数値計算手法の確立にある。AI サロゲートモデルの構築は、数学的な記述が困難な複雑なマイクロ相互作用に対しても、データ駆動型アプローチにより運動論モデルを構築するための新たな方法論につながるものである。これにより、自然科学から社会科学に至る様々な現象に対して、運動論に基づくモデリング研究の裾野を広げるとともに、多様な分野への応用可能性を広げる点で社会的意義も有する。一方、AI 援用型数値計算手法の確立は、既存の理論およ

び数値計算スキームを活用しつつ性能向上を実現する AI-HPC 融合技術の実装可能性を示す点に学術的意義がある。

3 当拠点公募型研究として実施した意義

本研究は、大規模 GPU 計算資源を必要とする運動論シミュレーションおよび AI モデルの開発を対象としており、代表者の所属機関単独では実施が困難な計算規模を伴う。

大阪大学 D3 センターの SQUID を利用することで、多 GPU 環境における運動論ソルバの実装および AI モデルの学習・推論の統合的検証が可能となった。また、GPU 間通信やメモリ配置を考慮した実装により、高次元問題への拡張に向けた基盤的知見も得ることができた。特に、本研究で対象とする運動論方程式は、物理空間に加えて速度空間も離散化する必要があるため、計算規模およびメモリ消費が極めて大きく、高性能 GPU 環境を用いた大規模並列計算が不可欠である。そのため、本拠点が提供する大規模 GPU 計算資源は、本研究の実施において本質的な役割を果たした。

さらに、本研究を通じて、AI モデルと数値シミュレーションを統合した GPU 実装に関する知見を蓄積することができ、将来的な高次元運動論問題や AI-HPC 融合型シミュレーション技術への展開に向けた重要な基盤形成にもつながった。

したがって、本拠点の計算資源は本研究の遂行に不可欠であり、AI と HPC の融合に関する実証研究として重要な役割を果たした。

4 前年度までに得られた研究成果の概要

該当なし。

5 今年度の研究成果の詳細

本年度は、モデルボルツマン方程式に対する GPU 実装および AI 援用型数値スキームの開発を中心に、国際共同研究の推進も含めて研究を実施した。以下に、2025 年度に得られた主な研究成果を示す。

- (i) **MIWS-Net の開発と実証:** 巨視量の更新量を予測するニューラルネットワーク (MIWS-Net) を構築し、陰解法におけるピカール反復回数を少なくとも約 3 倍削減し、条件によっては一桁以上削減できることを確認した。
- (ii) **GPU 実装による高速化:** A100 GPU を用いた実装により、32 コア CPU (OpenMP) 実装に対して最大 123 倍の高速化を達成した。
- (iii) **ソフトウェア実装:** Python/PyTorch を基盤とし、計算コアを C++/CUDA で実装した運動論シミュレーションパッケージ「kineticEQ」を開発した。
- (iv) **国際共同研究の推進:** カールスルーエ工科大学 (KIT) の Martin Frank 教授らと連携し、AI を用いた運動論モデルの高度化に関する共同研究を実施した。特に、研究者の相互訪問を通じて、アルゴリズム設計およびモデル構築に関する議論を深化させた。
- (v) **国際ワークショップの開催:** 2025 年 12 月に兵庫県立大学において「Workshop on AI and Simulation (<https://u-hyogo-gsis.org/topics/workshop20251203/>)」を開催し、AI とシミュレーション科学の融合に関する最新研究の共有と議論を行った。本ワークショップには KIT を含む国内外の研究者

が参加し、今後の共同研究の基盤形成に寄与した。

以下では、2025 年度に得られた研究成果について詳述する。

5.1 機械学習ウォームスタート予測子による陰的運動論ソルバーの高速化

5.1.1 はじめに

希薄気体の衝撃波や蒸発・凝縮流など、界面近傍で強い非平衡が生じる気体の輸送現象に対しては、分子の衝突過程を考慮した気体分子運動論 (ボルツマン方程式) による記述が必要となる。またボルツマン方程式は連続体極限において、通常の流体力学の基礎方程式であるナビエ=ストークス方程式を導出できることから、流体力学のメゾスコピック理論としても重要な意義を有する。

一方で、ボルツマン方程式やそのモデル方程式 (運動論モデル) では、時間および空間変数に加えて、気体分子の速度も独立変数として扱う必要があるため、運動論モデルの数値計算は、連続体モデルと比較して計算コストが大きくなる。さらに、連続体極限近傍では、分子の衝突過程を記述する衝突項の緩和時間が、マクロな時間スケールに比べて非常に短くなるため、数値的な剛性も問題となる。

数値的剛性に対する代表的な対処法として陰解法が用いられるが、運動論モデルは衝突項に複雑な非線形項を含むため、単純な陰解法では各時間ステップにおいて非線形問題を解くためのピカール反復が必要となる。しかし、その反復回数は連続体極限において著しく増加するため、陰解法のみでは数値的剛性を十分に緩和することが難しい。

この種の課題に対して、運動論モデルと関連する巨視量方程式を援用した加速法 (例えば、HOLO スキーム [W. T. Taitano, *et. al.*,

J. Comput. Theor. Trans. (2014)]) や漸近保存型スキーム (Asymptotic Preserving 法 [S. Jin, SIAM J. Sci. Comput. (1999)]) など、様々な高度な数値スキームの開発・研究が活発に行われている。しかし、これらの数値スキームは、個々の運動論モデルの理論的な解析によって個別に導出されるものであり、一般的な運動論モデルに対する汎用性のある手法とは言い難い。

そこで本研究では、運動論モデルの一般的な陰解法スキームに対して、ニューラルネットワークを導入し、内部ピカール反復を加速する Warm-Start 予測子を組み込む手法を提案する。これにより、既存の陰解法の構造を維持したまま低コストで数値的剛性の影響を緩和することができるようになる。

5.1.2 Boltzmann-BGK 陰解法と Picard 反復

本節では、物理空間 1 次元、速度空間 1 次元の BGK 方程式に対する陰解法スキームとその内部ピカール反復について解説する。

BGK モデル方程式

物理空間 1 次元 (x)、速度空間 1 次元 (v) の BGK 方程式は以下のように書ける。

$$\frac{\partial f}{\partial t} + v \frac{\partial f}{\partial x} = \frac{f_M - f}{\tau}, \quad (1)$$

ここで、 $f(t, x, v)$ は速度分布関数、 $f_M(t, x, v)$ は局所マクスウェル分布、 $\tau(t, x)$ は衝突項緩和時間である。

局所マクスウェル分布 $f_M(t, x, v)$ と緩和時間 $\tau(t, x)$ は、速度分布関数 $f(t, x, v)$ のモーメントから次のように定義される巨視量 (数密度 $n(t, x)$ 、流速 $u(t, x)$ 、温度 $T(t, x)$)、

$$n(x, t) = \int f(x, v, t) dv, \quad (2a)$$

$$n(x, t)u(x, t) = \int v f(x, v, t) dv, \quad (2b)$$

$$n(u^2 + T) = \int v^2 f(x, v, t) dv. \quad (2c)$$

を用いて、以下のように求められる。

$$f_M(x, v, t) = \frac{n(x, t)}{\sqrt{2\pi T(x, t)}} \exp\left(-\frac{(v - u(x, t))^2}{2T(x, t)}\right), \quad (3)$$

$$\tau(x, t) = \frac{\tilde{\tau}}{n(x, t) \sqrt{T(x, t)}}. \quad (4)$$

ここで、 $\tilde{\tau}$ は、時空間に依存しない定数パラメータである。

陰解法スキーム

BGK 方程式 (1) の有限体積法に基づく陰解法スキームは次のように書かれる。

$$f_{i,j}^{k+1} = f_{i,j}^k - \frac{\Delta t}{\Delta x} v_j \left(\hat{f}_{i+\frac{1}{2},j}^{k+1} - \hat{f}_{i-\frac{1}{2},j}^{k+1} \right) + \frac{\Delta t}{\tau_i^{k+1}} \left(f_{M,i,j}^{k+1} - f_{i,j}^{k+1} \right). \quad (5)$$

ここで、上付き添字 k は時間ステップ、下付き添字 i, j はそれぞれ、空間メッシュと速度空間メッシュを表す。また、右辺第 2 項にあるハット “ $\hat{\cdot}$ ” は有限体積セル界面での数値フラックスを表す。

内部ピカール反復

陰解法スキーム (5) の解 $f_{i,j}^{k+1}$ は、Algorithm1 のピカール反復によって求める。ここで、収束判定は許容誤差 tol を用いて次のように行う。

$$\left\| \mathbf{W}_i^{(z+1)} - \mathbf{W}_i^{(z)} \right\|_{\infty} \leq \text{tol}. \quad (6)$$

5.1.3 MIWS-Net: Macro Informed Warm Start Neural Network

MIWS-Net は、BGK 方程式の非線形性が巨視量に依存する点に着目し、ピカール反復の初期値を“巨視量空間”で直接改善するための Warm-Start 予測子モデルである。具体的には、セルごとの入力として $\mathbf{W}^k = (n^k, (nu)^k, T^k)$ とスカラー条件 $(\Delta t, \tilde{\tau})$ を与え、出力として $\Delta \mathbf{W}$ (次時刻に向けた巨視量の更新量) を予測

Algorithm 1 Fully implicit BGK step with Picard iteration

1. 初期化: $f^{(0)} \leftarrow f^k$ とし, $f^{(0)} \mapsto \mathbf{W}^{(0)} = (n, (nu), T)$ を計算する.
 2. for $z = 0, 1, \dots, N_{\max} - 1$:
 - (a) $f^{(z)} \mapsto \mathbf{W}^{(z)} = (n, (nu), T)$ を計算し, $\tau^{(z)}, f_M^{(z)}$ を構成する.
 - (b) $\{A_j(\mathbf{W}^{(z)}), \mathbf{b}_j(\mathbf{W}^{(z)})\}$ を組み立て, $A_j f_{\cdot, j}^{(z+1)} = \mathbf{b}_j$ (速度ごとの三重対角) を batched GTSV で解き, $f^{(z+1)}$ を得る.
 - (c) $f^{(z+1)} \mapsto \mathbf{W}^{(z+1)}$ を計算し, 収束条件 (6) を満たせば停止する.
 3. $f^{k+1} \leftarrow f^{(z+1)}$.
-

する. CNN により空間局所構造 (衝撃波近傍の勾配・局在) を捉え, ResNet により安定な残差学習として $\Delta \mathbf{W}$ を学習することで, 平坦領域での不要な摂動を抑えつつ, 支配的な非線形モードを先に潰すことを狙う. 推論は時間ステップごとに 1 回であり, 分布関数が物理次元と速度次元を持つことに比べて, MIWS-Net は物理次元のみの推論であるので計算コストは通常線形ソルバが支配的となるよう設計可能である. そして, 推論によって得られた $\Delta \mathbf{W}$ に基づいてピカル反復の初期値を与え, 既存解法のフローのまま反復削減を行えることが特徴である (Algorithm2).

Algorithm 2 MIWS-Net warm-start for fully implicit BGK step

1. 入力: \mathbf{W}^k と条件 $(\Delta t, \tilde{\tau})$.
 2. 推論: $\Delta \mathbf{W} \leftarrow \text{MIWS-Net}(\mathbf{W}^k; \Delta t, \tilde{\tau})$.
 3. warm-start 巨視量: $\mathbf{W}^{(0)} \leftarrow \mathbf{W}^k + \Delta \mathbf{W}$.
 4. warm-start 初期分布: $f^{(0)} \leftarrow f_M(\mathbf{W}^{(0)})$ として初期値を構成する.
 5. 以降の Picard 反復は Algorithm1 と同一とする.
-

5.1.4 性能評価と結果

一次元 Sod shock tube 問題を, 空間分割数 $n_x = 1024$, 速度空間分割数 $n_v = 256$, 時間

刻み幅 $\Delta t = 5 \times 10^{-4}$, ステップ数 100steps , 緩和時間パラメータ $\tilde{\tau} = 5 \times 10^{-7}$ の条件で数値計算し, 性能評価結果を Table.1 および Fig.1 に示す. ここで, “Base” は Algorithm1, “Warm-Start” は Algorithm2 の結果を表す.

表 1 は, Picard 収束許容値 tol を変化させた場合の総反復回数 iters_b (Base), iters_w (Warm-Start) と, その比 $\text{sp}_{it} = \text{iters}_b / \text{iters}_w$ を示す. $\text{tol} = 10^{-6}$ において $\text{sp}_{it} = 3.33$ が得られ, Warm-Start により反復回数が大きく削減されることが確認できる.

図 1 (左軸) は 1 ステップ当たり平均計算時間を示し, 反復回数の削減に対応して Walltime も同程度の倍率で短縮されており, 本設定では MIWS-Net の推論コストは Picard 反復計算に比べて十分小さいことが分かる. また Fig.1 (右軸) は, Base の $\text{tol} = 10^{-8}$ を参照解として最終ステップにおける巨視量の差の L_∞ ノルムを示す. Warm-Start は $\text{tol} = 10^{-6}$ 程度まで参照解に対して同程度, あるいはそれ以下の誤差に収束しており, 反復削減と精度の両立が確認できる. なお Walltime の計測は NVIDIA A30 GPU 上で行った.

表 1 Total Picard iterations (Base and Warm start). Here $\text{sp}_{it} = \text{iters}_b / \text{iters}_w$.

| tol | iters _b | iters _w | sp _{it} |
|-----------|--------------------|--------------------|------------------|
| 10^{-4} | 150493 | 113602 | 1.32 |
| 10^{-5} | 347869 | 133962 | 2.60 |
| 10^{-6} | 504556 | 151724 | 3.33 |
| 10^{-7} | 586876 | 184022 | 3.19 |
| 10^{-8} | 609104 | 225637 | 2.70 |

5.1.5 kineticEQ

本研究で開発したソルバは kineticEQ としてパッケージ化している. kineticEQ は, 主に運動論方程式を対象とした GPU ソルバであり, Python/ PyTorch を基盤として C++/CUDA

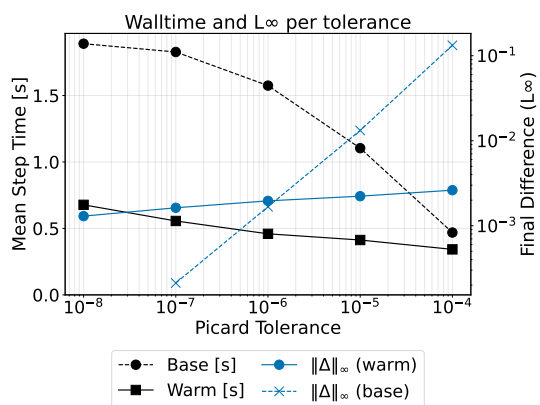


図 1 Mean step time (Base and Warm start) per Picard tolerance and final L_∞ differences.

拡張により計算コアを実装している。

6 進捗状況の自己評価と今後の展望

本年度は、当初計画に基づき、モデルボルツマン方程式に対する GPU 実装の高度化および AI 援用型数値計算加速手法の開発を中心に研究を推進した。特に、AI を用いて既存数値スキームの反復計算を高速化する「AI 援用型加速計算スキーム」の構築については、当初想定を上回る進展が得られたと自己評価している。

本研究では、巨視量更新量を予測するニューラルネットワーク MIWS-Net (Macro Informed Warm-start Neural Network) を構築し、陰解法におけるピカル反復の初期値生成へ適用した。その結果、既存スキームの保存則や数値精度を維持したまま、反復回数を平均的に約 3 倍以上削減し、条件によっては一桁以上の削減効果を確認した。この成果は、AI 導入型数値計算で懸念される「計算精度や信頼性の低下」を招くことなく、「既存スキームの保存則・数値安定性・計算精度を維持したまま加速のみを行う」という構造保存型アプローチの有効性を示すものである。

また、GPU 計算基盤についても大きな進展があった。BGK モデルに対する GPU ソルバ kineticEQ の整備を進め、A100 GPU を用いた計算では、CPU 実装に対して 100 倍規模の高速化を達成した。加えて、複数 GPU 環境を想定した通信重畳やデータ配置の検討を進めることで、将来的な大規模並列化へ向けた設計指針を得た。これらの成果により、本研究で提案する AI 援用型加速手法を、大規模・高次元運動論計算へ適用するための基盤技術が整備されつつある。

一方で、当初計画に含まれていた「衝突項そのものを学習する AI サロゲートモデル」の構築については、現時点では論文発表可能な段階には至っていない。特に、衝突演算子に対して保存則や H 定理などの物理的拘束条件を満たしつつ、高精度かつ高効率な近似を実現することは依然として難易度が高く、モデル設計およびデータ生成の両面で課題が残されている。しかしながら、本年度はこの課題に対する重要な前進として、KIT (Karlsruhe Institute of Technology) の Martin Frank 教授グループとの国際共同研究体制の構築を進めた。特に、KIT 側で開発が進められている衝突項サロゲートモデル RelaxNet に関する知見共有を開始し、今後の共同研究に向けた研究基盤を整備したことは大きな成果である。これにより、将来的には、日本側で開発している AI 援用型加速計算スキームと、KIT 側の AI サロゲート技術を統合し、高速かつ物理拘束条件を満たす新しい運動論ソルバの実現が期待される。

今後は、まず MIWS-Net を 2D2V 問題へ拡張し、多次元・大規模問題における有効性を検証する。また、マルチ GPU 並列化および通信最適化を進めることで、高次元運動論計算に対する AI-HPC 融合型計算基盤の高度化を進める予定である。さらに、KIT との国際共同研究

を本格化させ、RelaxNet をはじめとする衝突項サロゲートモデルとの統合を進めることで、「既存スキームを加速する AI」と「衝突演算子そのものを近似する AI」を組み合わせた次世代型運動論ソルバの構築を目指す。