

# グラフニューラルネットワークと生成モデルを用いた 非晶質系動力学予測システム開発

芝 隼人（兵庫県立大学）

## 概要

ガラスは、乱雑な原子分子（粒子）の配置を保ったまま固まった状態の物質である。広範な種類の物質が、液体状態からの急冷によって、あたかも相転移が発生したかのように急激に固まってガラス状態に変化する。ガラスの構造は液体と同様にランダムな配置のまま固まった状態であり、さらにその構造が非常にゆっくりと緩和していく。このような乱雑な構造の背後にどのような特徴が隠されており、その特徴がガラスの構造変化・構造緩和のダイナミクスとどのように関係しているかを理解することが、ガラスの理解や応用にきわめて重要なことと考えられてきた。

本課題は、ガラスの理解に新たな突破口を開きつつある機械学習・深層学習によるガラスへのアプローチに焦点を当てており、特に課題代表者らが開発したグラフニューラルネットワーク (BOTAN) を基盤とした展開を目指すものである。昨年度に BOTAN を含む最新の機械学習モデルを比較するための共通のデータセット基盤を作成しコミュニティに提供したのに引き続き、今年度はは主に GNN による予測結果の物理的不自然さを研究や、学習に用いた温度とは異なる温度に対する動力学予測を可能とする「転移学習」の実装に進展があった。

## 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

- 東京大学 情報基盤センター
- 東京科学大学 学術国際情報センター
- 大阪大学 D3 センター
- mdx

### 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

- 芝 隼人（兵庫県立大学、代表者）  
研究取りまとめ・全般の実行

- 下川辺 隆史（東京大学、副代表者）  
データ駆動モデル開発・アクセラレータ利用検討
- 川崎 猛史（名古屋大学 → 大阪大学、副代表者）  
シミュレーション・物理モデリング
- 朝比 祐一（CEA Saclay）  
データ駆動モデル開発
- 華井 雅俊（東京大学）  
データ駆動モデル開発
- 別所 秀将（名古屋大学）  
シミュレーション・物理モデリング
- 西垣内 大祐（名古屋大学）  
シミュレーション・物理モデリング

## 2 研究の目的と意義

ガラス科学において中心的な課題の一つは、構造的特徴のない中でどのように構造緩和するか、すなわちどの空間的領域にある原子（粒子）が時間とともに構造的変化を示すかを予測することだが、この困難の中、2010年代半ば頃から、ガラスに潜む特徴的構造を機械学習によって自動的に抽出する教師なし学習手法、ガラスの構造と動力学を結びつける教師つき学習手法が開発され、従来よりも遥かに高い構造抽出性能・予測性能を達成するに至った。ガラスに関連した物質材料の研究において長時間シミュレーションに多大なコストがかかる現状の中、深層学習モデルを用いて少ない計算量でガラスの長時間状態を実現する代替計算方法、あるいはガラスの理解・設計に資する深層学習手法の改良・開発を行うのが、本研究課題の目的である。

本課題では、物理シミュレーションの中でも、最も長時間にわたるダイナミックレンジでの現象を示す「ガラス（非晶質固体）」の動力学予測を取り上げた。ガラスは、乱雑な原子分子（粒子）の配置を保ったまま固まった状態の物質である。広範な種類の物質が、液体状態からの急冷によって、あたかも相転移のように急激に固まってガラスとなる。ガラスの構造は液体と同じように乱雑なままであり、結晶の周期構造のような明確な秩序を有しない。この事情は、ガラス転移を特徴づけること、ガラスの内部の構造欠陥（の類似物）を抽出することを困難にしており、ガラスの理解および工学的応用に大きな障害となっている。ガラスに普遍的に見られる特徴として、ある乱雑な構造を持ったガラスを長時間で見ると、動きやすい領域と動きにくい領域とはモザイク状にコントラストが広がっていく現象が挙げられる。この物理的

起源が完全に解明されるには至っていないが、長年の研究を通じて、初期の（乱雑）構造が運動しやすい領域の空間的分布が決定づけているのではないか、という考え方が比較的定着している。すなわち、乱雑で一見特徴を抽出することが不可能と思われる原子分子の配置情報そのものに、どの原子分子（たち）が将来動きやすいのかの情報が埋め込まれているという考えが根強く支持されているのである。

上述の理由により、ガラス物質の理解そして活用の際に、「（乱雑な）構造からの動力学をどのように予測できるのか」ことは一番大きな問いであり続けてきた。これに対して、隣接配向秩序をもとに構成された構造秩序パラメータの適用や、準安定粒子配置に対する基準振動解析による低周波振動モードの計算など、構造と動力学との関係を抽出する試みが多く行われてきたが、決定的な方法は長く見出されていない状況と言える。この状況の中で、近年、機械学習・深層学習手法による動力学分布の予測がこの問題に対して非常に有効であることが認知され、この10年ほどでさまざまな手法が提示された。とりわけ、有限温度環境下においてガラスを構成しているそれぞれの粒子がどの程度、構造緩和に寄与する長時間運動を示すのかを定量的に予測するタスクにおいて、グラフニューラルネットワーク（GNN）を用いた教師付き学習が他の（学習モデルを含む）手法を凌駕して最高の性能をもつことが Google DeepMind のグループによる研究 [V. Bapst *et al.*, Nature Physics **16**, 448 (2020)] により示され、大きな関心呼んだ。この後も、本課題メンバーの芝、華井、下川辺のメンバーによって提案された BOTAN など、大幅に予測精度の向上した機械学習モデルが複数提案され、いずれも物理理論に基づく（シミュレーションなど時間積分によらない）手法の予測精度を大きく超えてい

る状況である。

ガラスに関連した物質材料の研究において長時間シミュレーションに多大なコストがかかる現状の中、深層学習モデルを用いて少ない計算量でガラスの長時間状態を実現する代替計算方法、あるいはガラスの理解・設計に資する深層学習手法の改良・開発を行うのが、本研究課題の目的である。

### 3 当拠点公募型研究として実施した意義

本課題は東大情報基盤センターの芝と下川辺により開発された GNN モデル、“BOnd TArgeting Network” (BOTAN) を発展的に展開させることを意図して 2023 年に開始された研究である。BOTAN を含む GNN モデルや、物理的忠実性の高い MLP モデル、解釈可能モデルなどさまざまに提案されている中、種類の学習モデルの開発には主に物理系・材料科学系の研究者が参入しているものの、特に国外ではデータ科学・高性能計算の専門家が本格的にテーマに参入している状況ではないと言える。

今回課題では BOTAN の中心的開発者の芝、GPU コンピューティングの専門家である下川辺、ガラス物理を長年専門とする川崎、そして生成モデルを含む最新の機械学習手法に知見を持った華井や朝比らをメンバーとした学際的体制により、GPU の先鋭的利用、モデルの効果的な実装が可能な体制である。

### 4 前年度までに得られた研究成果の概要

2023 年度課題においては、関連機械学習モデルの現状と将来の発展方向性のレビュー記事を完成し投稿され、1 年以上の査読期間を経た

が、2025 年 1 月に Nature Reviews Physics にて出版された [1]。このレビュー記事の作成においては、(i) ガラスにおいて重要な特徴的構造の抽出、動力学の予測などへの機械学習利用の現状をサマリーする、(ii) アルゴリズムやサンプリングなど、将来発展の方向性を提示する、とともに、(iii) 各種の機械学習モデルの予測精度・予測内容を共通の基盤で多角的に比較し、機械学習モデルの予測性能について「ベンチマーキング」を確立することを目的として行った。特に課題代表者である芝の直接的な貢献によって、以下の成果が得られた。

1. BOTAN 論文 (Shiba *et al.*, 2023) とともに芝・下川辺らが公開に供した、3 次元 Kob-Adndersen Lennard-Jones ガラス形成液体基準データセットとして用い、分野で提案された主要機械学習モデルの持つ予測性能、転移学習性能などを整理し、新たにデータベース化した。結果は、一般の研究者の検証や利用に役立ててもらえるように Zenodo リポジトリ <sup>\*1</sup> 上に公開された。
2. 前項の作業の過程で、「固有構造」(Inherent Structure) と通称される状態を学習の入力として使用することで、予測性能が多くのモデルで向上することが明らかとなった。ガラスの遅い緩和とは、概念的には多粒子座標の関数であるエネルギー地形における多数の準安定状態の極小を徐々に降っていくダイナミクスとしてとらえることができる。通常、有限温度ガラスでは、熱ゆらぎの影響があるため各粒子が運動中心の周辺で熱運動しているため極小値のところからわずかにずれることになる。鞍点 (準

<sup>\*1</sup> <https://zenodo.org/records/10118191>

安定状態) から鞍点へと長時間をかけて遷移するガラスのダイナミクスと対応づけるには、熱振動下でのダイナミクス準安定状態＝エネルギー極小を満たす配置の方が「捉えやすい」入力であることが示唆されることが言える。

なお、熱的な粒子配置の状態から（最急降下法などによる準安定構造へのクエンチを行わずに）、直接粒子の運動度を予測するタスクにおいては、BOTAN は他の（教師つき学習の）モデルを凌駕してトップの予測性能に位置していることも明らかにされた。BOTAN はグラフの辺 (edge) 上で 2 粒子間の距離変化を学習することにより、隣接粒子間で生じる弾性歪みの度合いが予測可能なモデルに訓練され、その情報が、学習モデルの内部で背後に隠された固有構造を「推定する」のに寄与していると考えられる。

3. 従来、ガラスの運動の予測については 3 次元のガラス形成液体に対する予測が実施されてきたが、今回、2 次元 Kob-Andersen 液体のシミュレーション結果の提供を Gerhard Jung 博士（モンペリエ大学 [当時]）から受け、各機械学習モデルによる検証を行った。2 次元の場合にも BOTAN が良好な予測性能を発揮することが示された。
4. 転移学習に対するテストとして、ある温度で訓練されたモデルが、他の温度のガラス形成液体の動力学にどの程度有効であるかをベンチマークした。この結果、BOTAN など GNN を含む多層パーセプトロンを構成要素とした深層学習手法に、一定の優位性が認められた。

## 5 今年度の研究成果の詳細

以下、申請書した研究項目に概ね沿う形で記載する。

### (A) 動的多体相関を学習できる GNN モデルの開発

今年度課題においては、これまでの GNN が部分的に物理的に不自然な予測を与えることに着目してモデルの改良を実施している。レビュー原稿 [1] の執筆過程で、BOTAN を含む GNN は損失関数に単純な平均自乗誤差または平均絶対誤差を用いているために、入力された各粒子の局所構造と動力学の間に過剰なフィッティングをかけてしまうという問題が認識されるに至った。実際には粒子の構造再配置が活発な領域は空間的に広がりを持っているのであるが、GNN においては近接する粒子同士の運動活性の相関を無視して、個別の粒子に合わせすぎた予測をしばしば行ってしまう。

昨年度から、課題参加者の別所が主体となって、動力学の協調性の再現ができないことに対するペナルティを表現した GlassMLP モデル (G. Jung *et al.*, 2023) の損失関数を BOTAN に対して実装し、検証進めてきた。この際、予測結果を単にピアソン係数のような統計尺度で見のではなく、予測結果から得られる構造関数や、時間を経たあとの粒子の運動距離がどのように分布しているか（動的不均一）、などといった物理量を通じて比較し、その予測の妥当性を検討した。その結果は粒子の運動距離の分布の確率密度分布や、動的不均一性を特徴付ける量を見た時に、分布が学習元のデータセット（シミュレーションデータ）により近いものに改善した。まだ現段階では投稿予定であるためデータは本報告書に掲載しないが、国内学会発表において芝が行っている [4]。

また、さらに頂点 (node) と辺 (edge) に対して embedded vector を埋め込んだ GNN モデルである BOTAN を、さらに系全体を特徴付ける全体量 (global) に対する embedded vector も付与したモデルも開発した。目的は、異なる温度や異なる時間を単一のモデル上で訓練可能に改造し、異なる条件の結果の予測をより妥当にする「転移学習」を改良することである。BOTAN において global 特徴量として用いるのは、入力例えば温度、出力量に系全体のダイナミクスを特徴づける何らかの量である。本年度課題では、グラフニューラルネットワークライブラリ PyTorch Geometric 上での実装までは完成しているが、具体的なモデル訓練と予測結果の検討は 2025 年度課題で取り組む予定である。

#### (B) ガラスの剪断破壊予測システムの開発

剪断変形のように外部から対称性が破られるような条件が発生するもとで、脆弱なスポットの周辺で伸長と縮小が直行するような形で発生する。このような異方的な運動の予測は大きな課題であるということができる。我々は 2023 年度までにかから 2023 年度前半にかけて、ガラスの標準模型である Kob-Andersen (KA) LJ 粒子系を利用して、「非熱的準静剪断変形」と呼ばれる方法で剪断変形のシミュレーションを多数実施し、データセットを作成した。これは、微小な剪断変形を加えるごとに粒子配置を最急降下法などでエネルギー極小配置とすることにより、常に温度ゼロ状態（「固有構造 (inherent structure)」と呼ばれる）を保ったまま変形を続ける手法である。2023 年度に、inherent structure を入力とする機械学習モデルで、予測精度が BOTAN に匹敵するものが多数提案され、BOTAN のアーキテクチャ拡張方針の見直しに取り組んだ。具体的には剪断によって生じるホットスポットの把握を GNN

のエッジデコーダーでの距離変化学習に任せ、ノード上では 3 次元の変位場を学習させる、という方法をとっている。ノードに向かってメッセージ伝搬させることで、ノードが本来学習しにくいコヒーレントな運動パターンを抽出できるようにさせるのが、BOTAN の思想を受け継いだ部分である。

我々は様々な損失関数を試しながら、この新たな BOTAN を拡張した GNN モデルに粒子変位場を学習させる手法の開発を試み、この中で、頂点 (node) 上のデータ (変位ベクトル) に対する損失関数として、変位ベクトル同士の単純な差を取り、その差分のベクトルの大きさの平均を損失関数として使用した。この結果、粒子の入れ替わりがおきやすいより脆弱な「欠陥」の部分からある程度離れた空間領域では、予測された変位場の向きの予測が比較的妥当になされていることが確認された。一方、「欠陥」の近傍では、変位の大きさはある程度合わせられても、変位の方向が合いにくい状況である。この理由は、ガラスの粒子配置変化の素過程が本質的にランダム性を帯びる、という純粹に物理的な理由によるものと考えられる。この拡張されたモデルは更なる改善が必要な状況である。

モデリングの現状については、国内研究会講演 [5] にて発表を実施した。

#### (C) 生成モデルによるガラス動力学のサンプリング

本研究課題のもう一つの生成モデルの適用は今年度の実装までは至っていないが、2024 年になって GlassMLP の開発者の Gerhard Jung らが先行して、正規化フローモデルによるモンテカルロ遷移ルールを実装することにより、長時間サンプリング手法を実装した [G. Jung *et al.*, (2024).]。論文の中では平衡熱力学量として定積比熱を取り上げて、その計算効率を種々の

既存のモンテカルロ計算と比較がなされた。生成モデルサンプリングのソフトウェアとしては Boltzmann Generator (F. Noé, 2019) の公開コードを使用しており、またデータセットについては、カスタマイズ途上であるために非公開とされている。これらを参考に、データセットの改良について検討を進めたが、具体的実装については、まだこれからの段階にとどまっている。

#### (D) BOTAN および関連 GNN の各種 GPU 上での性能評価

申請計画書中、計画部分 (A) で触れた性能評価については、別項目として成果を記載する。前年度の成果論文 [1] で整理された論点の一つに、教師つき学習の範囲に限ってもモデル規模の多彩さが挙げられる。この論文では、代表的なものとして BOTAN (モデルアーキテクチャは DeepMind により提案された GNN と等価)、少数パラメーターで記述子・損失関数を非線形なガラス動力学を再現するために調整した GlassMLP (G. Jung *et al.*, 2024)、一般化された結晶結合角秩序への線形回帰モデル (E. Boattini *et al.*, 2021; R. Alkemade *et al.*, 2024)、SE(3) 同変 GNN (F. S. Pezzicoli *et al.*, 2024) を取り上げ、これらの間での性能の比較が実施された。これらのうち、GNN を利用している 2 者が突出して重みパラメーター数が多く、特に訓練に際して学習時間とデータ量に対する要求が強い。GNN は多パラメーターモデルでありそのコストが短所ではあるものの、簡易的な記述子による入力でも使用可能である、汎化可能なモデルであり物理的条件の異なる場合の推論予測に用いることができる (転移学習) などの長所も大きく、この立場からみたときに学習・推論実行の性能向上は重要な課題に位置づけられる。

今年度課題においては、以前 2022 年に実

施した BOTAN プロトタイプのベンチマーキングを拡張する形で、新たに東工大 (現: 東京科学大学) TSUBAME4.0 にも導入された NVIDIA H100 GPU を使用して、BOTAN および SE(3)-GNN の学習速度 (GPU 性能) に対するベンチマークを新たに実施し [3]、SWOPP 会議にて発表を行なった。特に、実行時間のうち H100 GPU のテンソルコアを使用している時間の割合は、BOTAN で 29.3%, SE(3)-GNN で 17.8% であった。プロファイリングによれば、これらの加速は FP8 と FP16 による低精度演算を活用した行列積演算 (GEMM) の加速のみであり、それら以外は全て TensorCore でない通常の演算器の使用により計算されていることになる。今後、TensorCore による加速がない部分の演算加速が大きな課題である。

また、東大情報基盤センターより AMD MI210 GPU (PCIe) を搭載した計算機環境へのアクセスを提供を受け、MI210 における実行性能評価についても実施した。BOTAN については、2022 年に測定した NVIDIA A100GPU と同水準の学習時間、H100 との比較ではおよそ 2.5 倍前後の時間を要する結果となった。演算加速は、H100 におけると同様に、行列積の matrix engine による加速によっており、AMD MI300X など、より最新の GPU または APU における学習性能評価も計画していたが、実機へのアクセスが達成できなかったため今回は見送ることとなった。

なお、SE(3) 同変 GNN については、開発者である Francois P. Landes 博士 (パリ南大学) により、BOTAN データセットを読み込めるように調整されたコードの提供にご協力いただいた (本成果報告書の執筆時点では、公開コードがリポジトリ公開されている <sup>\*2</sup>)。

<sup>\*2</sup> <https://zenodo.org/records/10805522>

## 6 今年度の進捗状況と今後の展望

(A) 動的多体相関の学習モデルについては、申請当時のアイデアを着実に実装を進めるとともに計算が進められている。申請書にしるした「GNN の性能評価」と合わせて、想定通りの 100% の達成度であると考えている。特に、先述した global feature を積極的に GNN に取り入れた拡張を BOTAN に対して行うことで異なる温度条件のガラスの間での転移学習の予測性能向上が実現すれば、まもなく結果を早急に原著論文として投稿する予定である。

(B) 「ガラスの剪断変形予測システムの開発」については、BOTAN 特有のエッジ出力側での距離変化の学習機構の導入を行うことで、変位の異方的な発生を部分的に再現することはできている。それ自体は GNN の新しい利用法を実現しているものになっているが、ガラスの準静的剪断変形の問題を考えるという立場からは、粒子ごとの動力学予測の精度は、もともと BOTAN など GNN で扱っていたタスクほどの性能は出ていない。粒子同士が入れ替わるスポット周辺でのダイナミクスの学習に足りない要因を考慮したモデルの改善が必要であり、成果をまとめるには多少のアイデアの改善が必要である。2025 年度課題にこの内容を引き継ぎ、幾何学的深層学習のアイデアの併用などによる解決を図っていく方針である。

(C) 生成モデルについては、本課題のチームにおいてはこの年度内には研究担当者（分担者）が十分に確保できなかった状況で、想定したようなペースでの進捗が得られていない。2025 年継続課題の中では十分数の課題担当者を確保できる状況が予定されているため、改めてデータセット作成とモデル開発を進めて

いるところである。

また、そのほかにレビュー記事 [1] を、大幅に内容を取捨選択をし日本の読者向けに紹介した日本語の雑誌用記事を、商業出版誌「固体物理」に投稿し、8月号に記事掲載された [6]。また、BOTAN のモデルを一般のユーザーが訓練する際に、最もわかりにくいと思われる辺での pretraining（事前学習）の役割について論じた解説記事を、計算工学会会誌に投稿した [2]。

(以上)