

jh230026

GPU 並列計算による高分子材料系シミュレーションの高速化技法の検討

萩田克美（防衛大学校）

概要

本研究では、高分子材料系のシミュレーションにおける GPU 活用技術の高度化として、シングル GPU からマルチノードのマルチ GPU までを利用するコード開発から利用ノウハウ形成までを中心に、自動的な利用法（パラメータ最適化を含む）の確立などを含む HPC 研究を進めた。これまでに開発や改良を進めてきた独自コードを公開する論文出版を行うとともに、MD アプリの活用による成果を含めた高分子材料研究の論文出版を行った。GPU コード開発として、OpenACC と cuFFT のマルチ GPU 利用の検討も進め、シングル GPU 利用のコードとの性能比較を行い、現状のターゲットサイズでの有効性を評価した。また、Python のパッケージを積極活用し、作業の自動化／省人化／効率化を進めた。水中 PEG ネットワークの負のエネルギー弾性に関する検討では、ノイズに埋もれたデータから、統計的な解析で有意なデータを引き出す形式を見いだした。AlphaFold2 の活用などの新しい方向性に関するノウハウ形成も実現した。この3年間の JHPCN 課題で集中検討した GPU 活用技術を通じ、多くの成果を得ることができた。

1. 共同研究に関する情報

(1) 共同利用・共同研究を実施している拠点名

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
大阪大学 サイバーメディアセンター
mdx

(2) 課題分野

大規模計算科学課題分野

(3) 共同研究分野 (HPCI 資源利用課題のみ)

超大規模数値計算系応用分野

(4) 参加研究者の役割分担

参加研究者	役割
萩田克美(防衛大)	代表、技術検討
村島隆浩(東北大)	副代表、情報提供
藤原 進(京都工繊大)	議論、技術検討

仙田康浩(山口大)	議論、技術検討
作道直幸(東京大)	議論、技術検討
畷山多加志(名大)	議論、情報提供
片桐孝洋(名大)	議論、情報提供
伊達 進(阪大)	議論、情報提供
埴 敏博(東大)	議論、情報提供
川波竜太(京都工繊大)	議論、技術検討
増本晃太郎(京都工繊大)	議論、技術検討
橋 拓実 ティモシー(京都工繊大)	議論、技術検討
藤井新之輔(山口大)	議論、技術検討

2. 研究の目的と意義

(研究目的)

マルチノードのマルチ GPU を駆使して高分子物理や高分子材料のシミュレーション実施環境を整備・構築することが目的である。そのために、コード開発から効率的利用に関するノウハウ構築や情報共有を実施目的としている。

(継続課題：研究計画全体の目的と今年度の目的)
2021 年度および 2022 年度は、独自 GPU コード開発とともに、GROMACS や LAMMPS 等の既存の MD アプリを利用し、マルチノードのマルチ GPU を効率よく利用する技術検討も進めた。

2022 年度中間報告以降の新たな進展として、アミノ酸配列からの構造予測ツール AlphaFold2 の活用や、Amber の GPU 利用などの検討を進め、HOOMD-blue をノード内 GPU 並列で利用する環境整備を行った。これらを踏まえ、本年度は「①MD 計算アプリを GPU スパコンで効率よく実施する技法やツールの開発、アプリの改造」と「②高分子材料実験データの解析用アプリの開発と普及に向けた取り組み」の 2 つのテーマで、GPU スパコンで効率よく実施する技法やツールの開発やアプリの改造を行う。

3. 当拠点の公募型共同研究として実施した意義

最新の GPGPU の HPC としての効率的利用のためには、最新の技術情報を欠かすことができない。マルチノードでのマルチ GPU の効率的な並列利用には、プログラムコードの開発に並び、通信ハードウェアに適したシステムソフトウェアの利用ノウハウが極めて重要である。現に、ライブラリのバージョンアップ等が性能を大きく左右するため、最新の技術的分析と対処が常に重要であった。このような状況から、JHPCN センター教員との密な情報共有と対応検討は、本課題の遂行に欠かすことができず、本課題の根本的意義であると言える。

4. 前年度までに得られた研究成果の概要

2022 年度の JHPCN 課題では、2021 年度の検討に引き続き、プログラム開発に加えて、マルチノードのマルチ GPU を駆使して高分子物理や高分子材料のシミュレーション実施環境を整備・構築する狙いを念頭に、MP-SRP 法を中心にして、次の 3 つの研究テーマについて検討した。「①独自 GPGPU コードの作成」「②パラメータの機械的探索の自動化の検討」「③汎用高分子アプリで

の GPU 利用の高度化」特に、2022 年度は、2021 年度をより発展させ、③のテーマを設定し研究成果創出に重点を置いた。それとともに、①の検討を一時中断する研究戦略の変更を行った。

項目①では、粒子系 GPGPU コードを整備し、拡張を検討した。それとともに、マルチノード・マルチ GPU の効率的な利用について検討した。また、OpenACC 等を活用し、解析コード等での GPU 活用技術を検討した。MP-SRP 法の計算については、架橋ネットワーク等におけるトポロジー効果や、環状鎖と線状鎖の混合系における鎖交差禁止の効果を評価する研究を実施した。予想以上に周辺研究が進んだ結果、最適化した CPU コードのみで論文出版が実現し、MP-SRP モデルの独自 GPU コード開発の優先度が相対的に低下した。コード開発よりも他の成果創出を優先させる戦略の変更をした。一方で、OpenACC を活用した独自コード開発としては、SPRING-8 で計測した 2 次元小角散乱データから、三次元構造モデルを構築する PM-2DpRMC 法コードの整備を進め、コード公開と論文出版に向けた作業も進めた。

項目②については、python スクリプトとの連携による解析やシミュレーションの計算制御の自動化、省人化、効率化の検討を進めた。機械的探索用の python スクリプトを作成し、「線状鎖のメルトで熱力学状態を満たす MP-SRP 法のパラメータを試行的に探索した後に、鎖交差禁止の実現を Trefoil knot で評価する」繰り返し作業を効率化した。加えて、python の Topoly パッケージを活用した数学的解析手法を活かし、Trefoil knot のポリエチレン鎖メルトの結晶化挙動について、United-Atom 模型を用いた研究を進めた。さらに、MP-SRP 法よりも確実に鎖の交差を禁止するモデルである Kremer-Grest 模型について、実験と計算のパラメータマッピングの探索の自動化/効率化に関する検討を開始した。

項目③は、「Gromacs, HOOMD-blue, LAMMPS の活用深化」と「Amber や AlphaFold 2 などの生体高分子 MD ツールの活用検討」を進めた。

5. 今年度の研究成果の詳細

プログラム開発に加えて、マルチノードのマルチ GPU を駆使して高分子物理や高分子材料のシミュレーション実施環境を整備・構築する狙いを念頭に、下記の研究テーマを実施した。

- [研究テーマ 1] MD 計算アプリを GPU スパコンで効率よく実施する技法やツールの開発、アプリの改造
- [研究テーマ 2] 高分子材料実験データの解析用アプリの開発と普及に向けた取り組み

上記に加えて、高分子材料の研究者や学生への情報普及として、オンライン勉強会を実施し、OpenACC 利用や東大・名大・阪大のスパコン利用のノウハウを交換した。

[研究テーマ 1] MD 計算アプリを GPU スパコンで効率よく実施する技法やツールの開発、アプリの改造

①基板に拘束された高分子系の粗視化 MD 計算

HOOMD-blue は、ノード内の直接型高速通信機構を活用した高速化もされている GPU アプリであり、粗視化 MD や DPD 計算に対応している。HOOMD-blue は高速であるものの、LAMMPS に比べ、変形などの計算機能が実装されておらず、機械的性質の評価には物足りない。一方で、HOOMD-blue は、python がインターフェースなので、複数のシミュレーション機構が接続した「マルチフィジックス計算」の素地があるとみなせる。最も簡単な「マルチフィジックス計算」として、固定粒子壁に挟まれた高分子材料の変形に対する応答を調べる接着・破壊やトライボロジーの問題がある。

本研究課題では、形状を持つ基板に接着したポリマーネットワークの剥離挙動に対し、化学的吸着と機械的接着の要因分析を目指し、粗視化 MD 計算を実施した。この研究に先立ち、自然酸化したシリコン基板に接着するエポキシ樹脂の全原子

MD 計算を行い、基板のナノ形状がシア変形で顕著な効果を発揮することを全原子 MD 計算で明らかにし、Langmuir 誌[論文 1]で論文出版した。

図 1(a)に示すように、y 方向に薄い擬二次元的な周期境界条件下で、粒子で構成される壁を、z 方向に一軸延伸する計算を実施した。周期境界条件の箱の大きさは、 $142\sigma \times 17.75\sigma \times 100\sigma$ である。これは、壁の形状を考慮する場合、システムサイズが大きくなるためである。このような擬二次元的な状況ならば、single GPU 活用した HOOMD-blue で、1 つのケースが 2 日程度で計算可能であり、広いパラメータ空間における全容把握には有用である。ポリマーネットワークは、図 1(b)に示すプレポリマーの末端をランダム架橋することで得た。形状を持つ基板は、図 1(c)のような張出形状を含む粒子壁を系統的に作成した。広範なパラメータスタディの結果として、図 1(d)と(e)のように、見た目にも顕著な違いがある領域を見出した。詳細を検討し、Macromolecules 誌 [論文 6]で論文出版した。

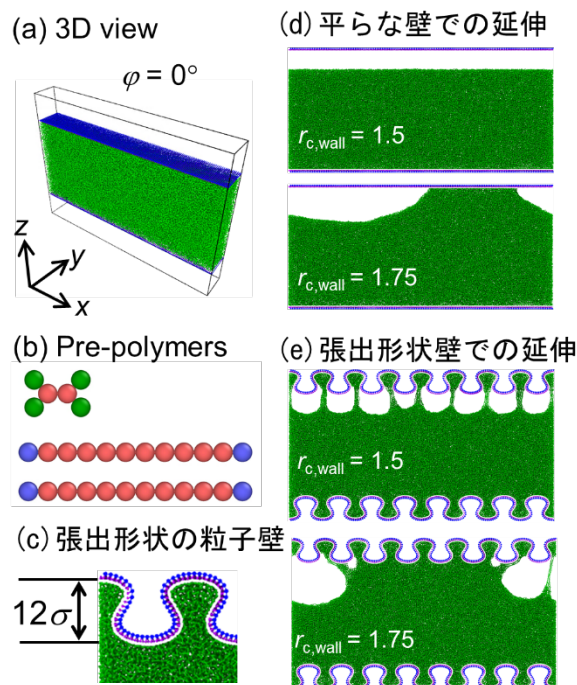


図 1 粒子壁からの高分子網目の延伸破壊

3 次元的なポイド形状や破壊の問題を検討する大規模系の計算実現のために、複数の GPU を並列利用する HOOMD-blue の利用手法の検討を進

めている。まずは、y 方向を 8 倍にした $142\sigma \times 142\sigma \times 100\sigma$ をターゲットとした。Nsight Systems を利用したプロファイル分析などでコード改良を進めている。

②水中の高分子の全原子 MD 計算

水中の高分子のナノ力学（ナノスケールでの機械的性質）を扱うために、大規模となる全原子 MD 計算を効率よく実施する技法を検討した。

②-1 Gromacs を用いた水中 PEG ネットワークの負のエネルギー弾性に関する研究

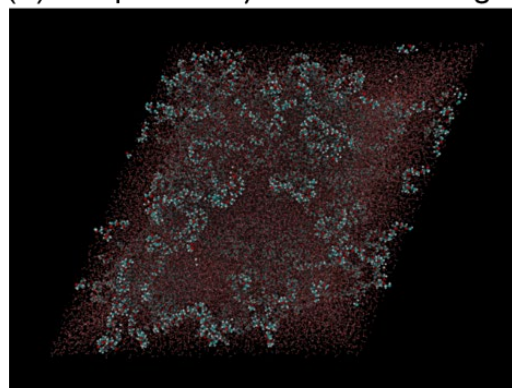
水中の PEG ネットワーク（ダイヤモンド格子構造）について、負のエネルギー弾性の再現を確かめるために、シア変形下の応力の温度依存性に関するプロダクトラン計算を HPCI 課題で実施した。図 2(a) のようにして 360ns の計算で求めたシア歪 γ 毎の応力値 σ_{xy} は、弾性率 G ($\sigma_{xy} = G\gamma$) を精度良く求めるには計算時間が不十分であった。弾性率 G の温度依存性で $G = A(T - T_E)$ の T_E がゼロより大きい場合に、負のエネルギー弾性の状態となる。この負のエネルギー弾性の検証には、さらに長時間の計算追加か、新たな工夫が必要である。

近年、MD データ解析においても、機械学習や統計（次元圧縮等）の活用が期待されている。この物理的な意味は、着目する物理量の数値精度のみにだけ着目すれば良いことと同じである。理論の立場からは、仮に平衡状態に近い典型的な構造が得られていれば、比較的短時間で物理量が評価できる。一般に、ランダムなネットワーク構造に対して平衡状態に近い典型的な構造を MD 計算で得ることは計算が難しい問題である。この点は、綺麗なダイヤモンド格子構造を計算対象とすれば、問題が軽減できる。本問題では、 γ と温度 T に対して得られる σ_{xy} から、 $\sigma_{xy} = A\gamma(T - T_E)$ に対する多変量回帰として、 T_E を決定する問題にして T_E のみに着目することとした。

図 2 (b) は、シミュレーション長を変えて、多変量回帰の問題として T_E を決定した結果である。体積一定条件の場合、200ns 程度の計算でも T_E

が 200K 付近であり、予測としては悪くなさそうである。圧力一定条件の場合、320ns 程度の計算で予測値が収束していた。水中の PEG ネットワークに、ゲスト鎖やイオンなどを混合した場合に、負のエネルギー弾性指標 T_E に与える影響を調べる研究に有効である。なお、本研究成果は、Macromolecules 誌[論文 4]で論文出版した。

(a) Snapshot at $\gamma = 40\%$ for 200g/L



(b) T_E 評価値の run-length 依存性

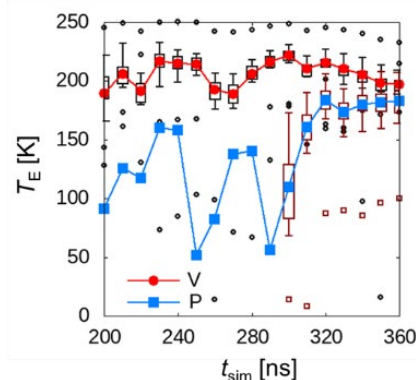


図 2 水中の PEG ネットワークの負のエネルギー弾性指標 T_E の評価。V (P) : 体積 (圧力) 一定条件。

②-2 AlphaFold2 によるコラーゲン 3 重ヘリックスでナノダンパー構造を実現する探索的設計

2022 年度の JHPCN 課題において、3 本のコラーゲンの複合体構造について、図 3(a) に示すような興味深い準安定構造が得られた。偶然の発見を活かし、このような構造を安定構造で実現するアミノ酸配列を AI 的に探索する問題を立案した。なお、コラーゲンのアミノ酸配列は、基本 PPG が繰り返されたものであり、種によりわずかに異なる。(P: Proline, G: Glycine)

影響も慎重に注意する必要があることがわかる。高速な M2.SSD を搭載した手元 PC を含めて、上手な使い方が肝要であるとの教訓が得られた。

手元 PC 環境 (Ryzen 9 7950X、7300MB/s read の SSD(WD_BLACK SN850X)、RTX3090) でのベンチマーク評価を実施したところ、(1) 410 秒、(2) 1160 秒であった。当面は、最新 PC 環境による検討が合理的と判断することができた。

次に、計算や AI によるアミノ酸配列の探索に関する試行的探索を実施した。図 3(b)のように、折れ曲がり誘発するアミノ酸配列を導入した 1 本の鎖について、AlphaFold2 による構造予測結果を調べた。ここで、折れ曲がり誘発部位を G のみで表現し、挿入する G の数を 3,6,9 と変えた。また、2つの折れ曲がり誘発部位の間の PPG の連鎖数を変えた。挿入する G の数が、3 の場合は、折れ曲がり誘発できなかった。6 と 9 の場合は、折れ曲がり誘発し、3重ヘリックスを形成する長さを制御できた。一方で、G の代わりに P とすると折れ曲がり誘発できなかった。この違いは、Proline の骨格の N 原子は2つの C 原子と接続していることに由来すると考えられる。

(Proline 以外のアミノ酸は、骨格の N 原子に 1 つの C 原子が接続する構造である。) G の代わりに、側鎖の構造が異なる A, L, R, K の 6 連鎖を挿入した場合を調べ、この折れ曲がり部分の形成に側鎖の影響は小さいことを確かめた。

図 3(b)のような部分的に 3重ヘリックス構造を持つ鎖を 3本複合させると、図 3(a)のような構造を自発的に形成すると予想される。一方で、図 3(b)のようなシーケンスの鎖 3本の複合体について、AlphaFold2 で構造予測すると、単純な 3重ヘリックスとなった。これは、単純な 3重ヘリックスと図 3(a)のような構造間のエネルギー差が低いと考えられる。親水性や側鎖電荷を微妙に変えるようにアミノ酸配列を探索設計することで、図 3(a)の構造を取りやすくと期待される。図 3(b)の構造を 3本複合させた構造の大規模 MD 計算による直接評価は、今後の課題である。

[研究テーマ 2] 高分子材料実験データの解析用アプリの開発と普及に向けた取り組み

SPring-8 で計測した 2次元小角散乱データ (図 4(a)) から、三次元構造モデル (図 4(b)) を構築する PM-2DpRMC コードは、OpenACC+cuFFT で GPU を利用するコードとして、最初に Single-GPU 用のコードを作成した。

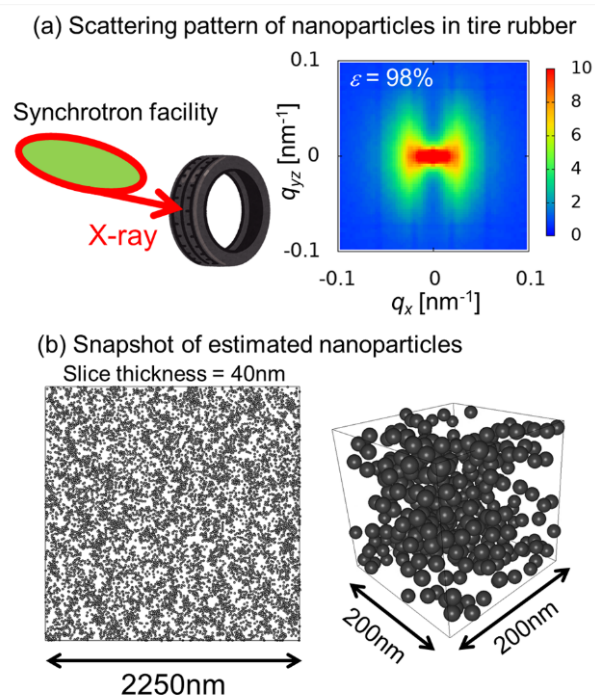


図 4 SPring-8 の 2次元散乱像からのゴム中ナノ粒子の 3次元凝集配置のモデリング

このコード公開論文は Comput. Phys. Commun.誌[論文 3]で論文出版した。応用した研究は Macromolecules 誌[論文 2]で論文出版した。本年度は、より大規模なシステムサイズ (現在の 8~64 倍) を扱うため、cufftXt でのマルチ GPU-3D-FFT の実施と、GPU メモリ上のデータの OpenACC による足し込み計算のタスク並列と cufftXtLib のデータ管理フレームワークを活用 (想定外利用) したノード内での高速な Gather などによるノード内マルチ GPU のコードを検討し始めた。現状の解析におけるターゲットサイズでは、メリットを見いだしにくい状況であり、当面は、シングル GPU での利用が合理的と判断できる結果が得られた。

[追加テーマ] MDX の GPU 環境を活用した検討

申請書にも記載したとおり、MDX の環境を利用した PM-2DpRMC コードの検証や、コードの普及に向けた共同作業環境として利用できることを最初に確かめた。

次に、MDX を広い目的で試行的な利用が可能な環境であることから、急速に発達している LLM などの技術を、試行・検証することとした。MDX は、クラウドとして柔軟な環境設定が可能であり、LLM などの技術を、試行・検証する環境として優れている。最近、言語処理による材料の研究として、CHEM-BERT や ChemBERTa/ChemBERTa-2 など、AlphaFold2 以外にもさまざまなものが提案され、研究が盛んに行われている。

一般に、GPU による MD 計算は、演算密度は高いがメモリ要求や IO 要求が小さい特性がある。一方で、LLM や AlphaFold2 などの AI 処理では、(学習を除き) 演算密度が低い一方でメモリや IO への要求は高い。直感的には、性能特性の凸凹がうまく合えば、限られた計算資源の効率を高めることが可能と考えられる。

MDX では、GPU 搭載のインスタンスを利用できることから、LLM 学習の動作環境の情報収集/確認を行うとともに、MD 計算と AI 処理の共存実行の影響を調べることにした。

MDX の 8GPU 環境では、Stability AI の StableBeluga2 (70B モデル) を、3.8 tokens/s の性能で実行可能であった。RTX4090 搭載の手元 PC での 0.05 tokens/s と比べると、実用性能として差は明確である。最初に、LLM 利用と Gromacs 計算の同時実行の影響について調べた。StableBeluga2 (70B モデル) のデータをロードした GPU において、Gromacs の計算を実施した。LLM 側での演算利用が少ないため、Gromacs の計算性能に与える影響は顕著ではなかった。

次に、LLM 学習と Gromacs 計算の同時実行による演算資源競合の影響を調べた。LLM 学習は、Megatron-DeepSpeed を用いた LLM 学習を実施した。表 2 に示すように、単独実行時と同時実行

時の性能を評価した。単独実行の性能を 1 とすると、同時実行では、 $0.476+0.481=0.957$ となり、同時実行は不利であることを確かめた。

表 2 LLM 学習と Gromacs 計算の性能計測結果

	samples/s	ns/days
LLM 学習のみ	50.148	---
同時実行	23.883 (47.6%)	29.98 (48.1%)
Gromacs のみ	---	60.28

6. 進捗状況の自己評価と今後の展望

2021 年度から 3 年間の検討で、GPU を活用した高分子材料系シミュレーションの高速化技法を大いに進展させることができた。多くの面で、当初計画したよりも、予想以上の進捗であった。まず、粗視化 MD 計算における HooMD-blue コードの活用や、United-Atom 模型や全原子模型での Gromacs の GPU 活用により、複数の論文成果[論文 3,5,6]を得ることができた。次に、長年開発や改良を進めてきた PM-2DpRMC 法による SPring-8 データからの 3 次元構造モデリングについて、GPU コードの開発によって、実材料に対する実用的な研究がはじめて可能になった。応用研究の論文出版[論文 2]とともに、コード公開の論文出版[論文 4]まで実現した。さらに、AlphaFold2 の活用などの新しい研究手法に関する予備検討も進めることができた。このように、当初よりも予想以上の進捗が得られた。一方で、MP-SRP 法の独自 GPU コードの開発については、CPU コードによる研究の予想以上の進展による相対的なニーズ減により、開発中断の判断をすることとなった。MP-SRP 法の活用の次のステップである Kremer-Grest 模型でのパラメータマッピング検討については、[論文 7,8]を論文出版し、順調に進捗している。

今後も引き続き、この 3 年間の JHPCN 課題を通じて整備した GPU 利用ノウハウを活用し、省人化や効率化したプロダクトランによる研究成果の創出に集中していく予定である。

7. 研究業績

(1) 学術論文 (査読あり)

- [1] K. Hagita, T. Miyata, H. Jinnai, Molecular Dynamics Simulations on Epoxy/Silica Interfaces Using Stable Atomic Models of Silica Surfaces. *Langmuir* **2023**, 39, 125733.
- [2] K. Hagita, T. Tominaga, R. Yamamoto, T. Sone, Structural changes of aggregated filler particles in elongated rubbers through two-dimensional pattern reverse Monte Carlo modeling. *Macromolecules* **2023**, 56, 4457–4467.
- [3] K. Hagita, S. Nagahara, T. Murashima, T. Sakai, N. Sakumichi, All-atom molecular dynamics simulations of poly(ethylene glycol) networks in water for evaluating negative energetic elasticity. *Macromolecules* **2023**, 56, 8095–8105.
- [4] K. Hagita, T. Tominaga, Two-dimensional pattern reverse Monte Carlo analysis of nanoparticles in polymer matrices using a combination of OpenACC and cuFFT. *Comput. Phys. Commun.* **2024**, 295, 108971.
- [5] K. Hagita, T. Yamamoto, H. Saito, E. Abe, Chain-level Analysis of Reinforced Polyethylene through Stretch-Induced Crystallization. *ACS Macro Letters* **2024**, 13, 247–251.
- [6] K. Hagita, T. Murashima, T. Miyata, H. Jinnai, River-Meander-Curved Surface Model for Simulations of Crosslinked Polymer Adhesions. *Macromolecules* **2024**, 57, 3862–3872.
- [7] K. Hagita, T. Murashima, Coarse-grained molecular dynamics model of AB diblock copolymers composed of blocks with different Lennard–Jones parameters forming lamellar structures. *Polymer* **2024**, 304, 127732.
- [8] K. Hagita, T. Murashima, Coarse-grained molecular dynamics model of conformationally asymmetric AB diblock copolymers forming lamellar structures. *Comput. Mat. Sci.* **2024**, in press.

(2) 国際会議プロシーディングス (査読あり) なし

(3) 国際会議発表 (査読なし)

- [1] K. Hagita, Systematic analysis of adhesion and delamination mechanisms at polymer-inorganic interfaces via coarse-grained molecular dynamics, CREST-NIST joint mini-symposium, 2024/3/15, ワシントン DC

(4) 国内会議発表 (査読なし)

- [1] 萩田克美, 村島隆浩, 長原颯太, 酒井崇匡, 作道直幸, 水中 PEG ネットワークが示す負のエネルギー弾性の全原子 MD 解析, 日本物理学会第 78 回年次大会, 2023/9/19, 東北大学
- [2] 萩田克美, 村島隆浩, ブロックコポリマーの相分離構造を自発的に形成する Kremer-Grest 粗視化 MD モデルの検討, 日本物理学会 2024 年春季大会, 2024/3/18, オンライン

(5) 公開したライブラリなど

- [1] K. Hagita, T. Tominaga, Two-dimensional pattern reverse Monte Carlo analysis of nanoparticles in polymer matrices using a combination of OpenACC and cuFFT. *Comput. Phys. Commun.* **2024**, 295, 108971.

(6) その他 (特許, プレスリリース, 著書等) なし