

令和6年度高性能汎用計算機高度利用事業
「富岳」成果創出加速プログラム
「データ駆動型材料研究を変革するデータ基盤創出」
成果報告書

令和7年5月30日
大学共同利用機関法人 情報・システム研究機構

吉田 亮

目次

1. 補助事業の目的	- 1 -
2. 令和6年度（報告年度）の実施内容	- 1 -
2-1. 当該年度（令和6年度）の事業実施計画.....	- 1 -
2-2. 実施内容（成果）	- 2 -
2-3. 活動（研究会の活動等）	- 6 -
2-4. 実施体制	- 6 -

補助事業の名称

「富岳」成果創出加速プログラム
データ駆動型高分子材料研究を変革するデータ基盤創出

体系的番号： JPMXP1020200314

1. 補助事業の目的

高分子物性自動計算システム RadonPy を用いて、データ駆動型材料研究に資する包括的な高分子物性データベースを構築する。

2. 令和6年度（報告年度）の実施内容

2-1. 当該年度（令和6年度）の事業実施計画

(1) 「高分子物性自動計算ソフトウェア RadonPy の開発」

引き続き、全原子古典分子動力学シミュレーション自動化ソフトウェア RadonPy の開発を推進する。令和5年度までに RadonPy に 24 種類の物性の自動計算アルゴリズムを実装してきた。令和6年度は、計算可能物性（溶媒和自由エネルギー、溶解度パラメータ、レオロジー物性、誘電特性など）と系（低分子化合物、架橋高分子、1 軸延伸配向構造、高分子溶液など）を拡張する。

(2) 「データ駆動型材料研究に資する体系的な高分子物性データベースの開発」

令和5年度までに約7万種類のアモルファスポリマーの物性データを蓄積してきた。令和6年度は、アモルファスポリマー、セルロース誘導体、架橋高分子、共重合体ポリマー、モノマー物性、温度依存性物性を含む包括的な高分子物性データベースを開発する。さらに、API やユーザーインターフェースを開発し、データベースを公開する。

(3) 「マテリアルズインフォマティクス基盤技術の創出とデータ駆動型材料研究の推進」

引き続き、RadonPy と高分子設計の機械学習アルゴリズム（確率的分子生成モデル、ベイズ推論に基づく逆解析、適応的実験計画法など）を統合したソフトウェア SPACIER の開発を推進する。令和5年度までにソフトウェアの実装はほぼ完了した。令和6年度は論文発表とソフトウェア公開を実施する。また、高熱伝導高分子液晶、光学系高分子、セルロース誘導体を対象に実証研究を推進する。

(4) 「プロジェクトの総合的推進」

本プロジェクトは、複数の国研・大学・企業が参画するコンソーシアム型共同事業として推進される。現時点で1国研・6大学・32企業に属する208名がコンソーシアムに参画している。このような大規模事業を円滑に進めていくために、令和6年度は運営会議とテクニカルミーティングを6-8回程度開催する。また、論文発表や学会発表を通じて、プロジェクトの研究成果を積極的に発信していく。特に令和6年度は、データベースの公開とコンソーシアム連名の論文発表に注力する。さらに、第四四半期を目途に成果発表と情報発信を目的とするシンポジウムを開催する。また、事業終了後のデータベースの維持・拡張のための制度設計と研究体制の整備に着手する。

2-2. 実施内容（成果）

高分子物性全自動計算ソフトウェア RadonPy と「富岳」を用いて世界最大規模の高分子物性データベースを開発し、データ駆動型高分子材料研究に資する体系的且つ包括的なデータプラットフォーム及び機械学習の基盤技術を整備した。

(1) 高分子物性自動計算ソフトウェア RadonPy の開発

RadonPy は、本プログラムで開発された全原子古典分子動力学（以下、MD: molecular dynamics）に基づく高分子物性計算機実験を全自動化する世界初の Python オープンソースソフトウェアである（Hayashi et al., *npj Comput Mater* **8**, 222 (2022), 引用数 75)。高分子の繰り返し単位の化学構造、重合度、温度等の計算条件を入力とし、配座探索、電荷計算、力場パラメータの割当、ポリマー鎖の生成、シミュレーションセルの構築、平衡化・非平衡 MD 計算、物性計算など、MD 計算の全工程を完全に自動実行する。最新版には、熱物性、機械特性、光学特性を含む 34 種類の物性の自動計算アルゴリズムを実装した。繰り返し単位と諸条件を指定すれば、ホモポリマー、ランダム・ブロック・交互共重合ポリマー、架橋ポリマーのシミュレーションセル構築や物性計算を簡単に実行できる。その他にも、アモルファス・結晶状態の物性計算、単軸延伸配向による加工プロセス、非平衡 MD による熱輸送特性やガラス転移温度、交流電場を印加したもとの誘電特性（Maeda et al., *Polymer J* **57**, 665–677 (2025)）など、様々なポリマーシステムに対する自動計算ワークフローを実装した（**図 1**）。

産学連携コンソーシアム：様々な系に対する計算機実験を自動するプリセット群を開発

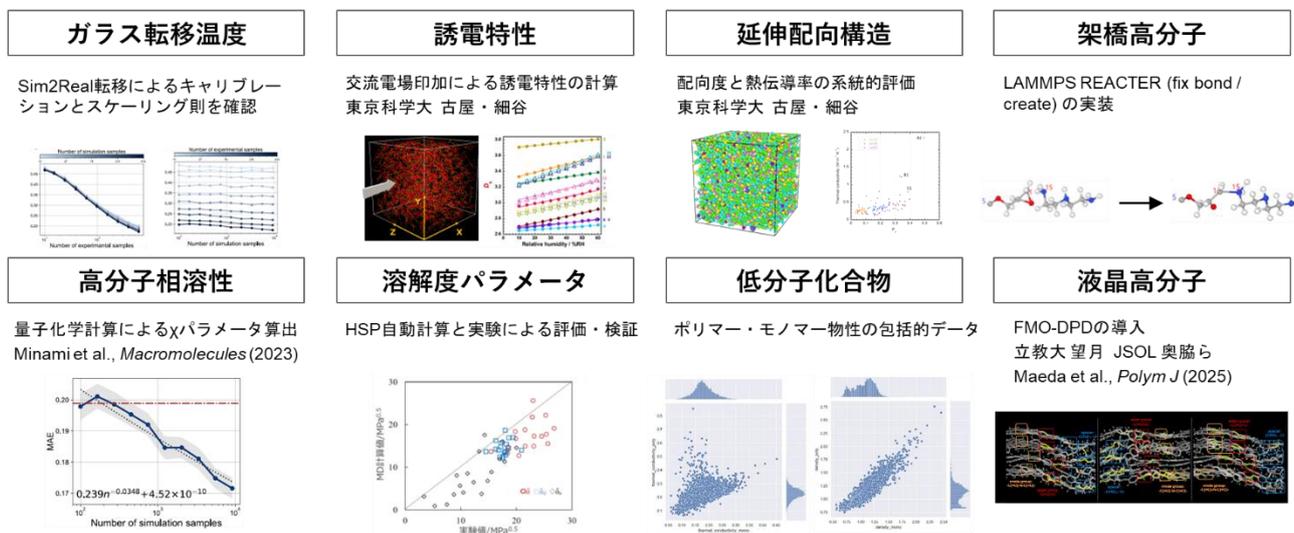


図 1. RadonPy に実装されている様々な自動計算機実験ワークフロー

(2) データ駆動型材料研究に資する体系的な高分子物性データベースの開発

本研究では、「富岳」の大量のノードを用いて多数のポリマーの物性計算を並列に実行し、最終的に 10 万種類以上の分子骨格を包含するオープンデータベースを構築する。令和 6 年度までに、約 13 万件のデータ生産を完了した。データベースの規模は世界最大となる。我々は、開発プロセスを加速するた

めに、2 国研、9 大学、37 企業に属する約 260 名が参画するコンソーシアムを形成し、多様なポリマーシステムに対する自動計算アルゴリズムの開発とデータ生産を共同で実施した。この前人未踏の包括的データベースは、高分子材料の広大な未踏領域に関する極めて重要な科学的知見をもたらした。例えば、複数物性のトレードオフが形成するパレート境界がどこにあるのか、さらに経験的に知られている到達可能限界を超える大量の仮想ポリマーが同定された。他にも、次世代通信システムへの応用が期待される低誘電ポリマーのデータ集合 (Maeda et al., *Polymer J* **57**, 665–677 (2025))、延伸配向されたポリマーの熱輸送特性、PFAS (Per- and Polyfluoroalkyl Substances) 代替候補材料、生分解性ポリマーの開発用に生成されたセルロース誘導体やポリエステル、ポリカーボネート、低分子化合物の材料物性、高屈折光学高分子 (Nanjo et al., *npj Comput Mater* **11**, 16 (2025), プレリリース)、高分子・有機溶媒の相溶性 (Aoki et al., *Macromolecules* **56**, 14, 5446–5456 (2023))、溶解度パラメータ、ラダーポリマー、共役高分子、架橋高分子など、多岐に渡るデータ集合を構築した。

本研究では、Sim2Real 転移学習という方法論を導入し、計算機実験の大量データと限られた実験データを統合的に解析することで実験物性に対する高い汎化能力を持つ予測器を構築する。計算物性データベースを用いてモデルを事前学習し、限られた実験データを用いて現実世界の予測タスク向けに微調整 (ファインチューニング) する。このような Sim2Real 転移で得られたモデルは、実験データだけで学習されたモデルでは到達できない高度な予測能力を発揮することが知られている。

Sim2Real 転移学習の理論解析 (Mikami et al., *Machine Learning and Knowledge Discovery in Databases*, 477–492 (2023).) によれば、ファインチューニングされたモデルの実験物性に対する予測性能は、計算データベースのサイズ n の増加に伴い、べき乗則 $\text{prediction error} = Dn^{-\alpha} + C$ に従って単調に改善することが示されている。データ数 n の増加に対する改善率 α が大きく、転移ギャップ C が小さいデータベースが望ましいと考えられる。転移ギャップは、データベース拡大によって到達可能な限界性能であり、計算物性データベースの将来価値を表す指標となる。我々は、本研究は、RadonPy データベースで訓練されたモデルが多様な現実系に対して転移学習の強いスケーラビリティを持つことを実証した (Minami et al., *npj Comput Mater*, in press) (図 2)。スケーリング挙動の解析は、目標精度に到達するために必要なデータ数や、到達可能な性能限界を見積もることに役立つ。また、スケーリング挙動が収束した場合には、それ以上のデータ生産を停止し、計算資源を他のプロジェクトに再分配するという意思決定が可能になる。さらに本研究では、スケーリング挙動の解析に基づく実験計画の策定や、物理実験と計算機実験の最適な資源配分を決定することが可能であることを示した。これまでにさまざまな計算物性データベースが開発されてきたが、スケーリング則の観点からその有用性を定量的に示した例は報告されていない。

RadonPyデータベースのサイズ n の増加 → 様々な現実系タスクにおいて転移モデルの予測性能がスケーリング

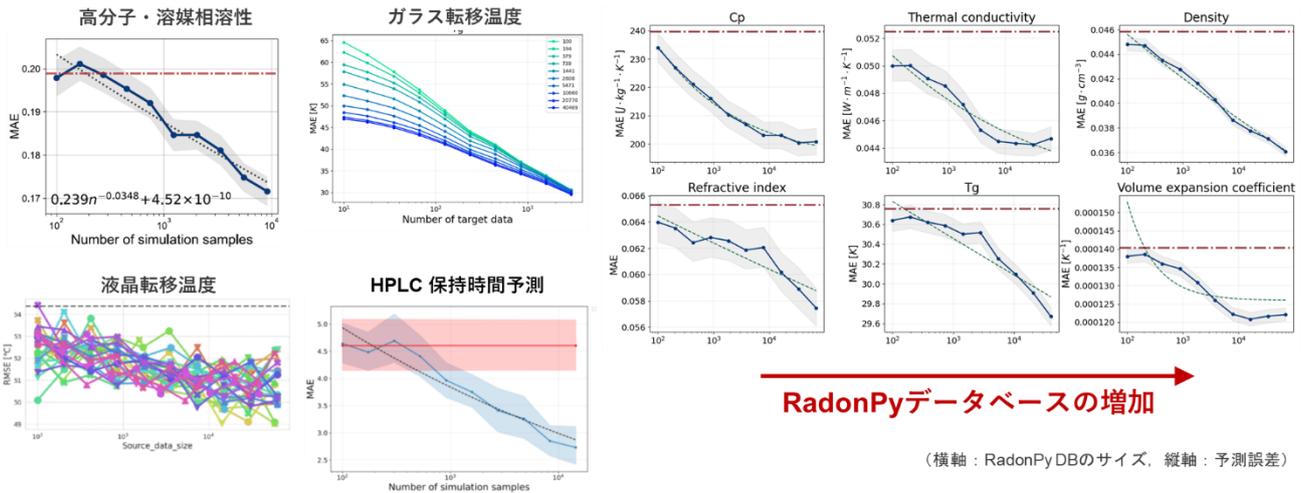


図 2. RadonPy データベースを用いた Sim2Real 転移学習とスケーリング則

(3) マテリアルズインフォマティクス基盤技術の創出とデータ駆動型材料研究の推進

データの外の世界を予測する方法を学ぶ AI 技術

材料科学の究極の目標は、データが存在しない未踏領域から新しい材料を発見することである。しかし、機械学習の予測は一般に内挿的であり、その有効性は既存のデータ分布に近い範囲に限られる。さらに、材料研究ではデータ取得コストが非常に高いため、十分な学習データを確保することが難しく、必然的にデータの範囲外、すなわち外挿的領域の探索が求められる。この課題を解決するために、我々は E2T (extrapolative episodic training) という機械学習アルゴリズムを開発した (Noda et al., *Commun Mater* **6**, 36 (2025)) (図 3)。E2T では、手元のデータセットから人工的に生成した大量の外挿的タスクを用いて、メタ学習器という特殊なモデルを訓練する。その結果、モデルは外挿的な予測を実現するための汎用的な“学習方法”を自律的に学習する。本研究では、E2T を RadonPy データベースを用いた材料特性の予測タスクに適用し、学習データには存在しない構造的特徴を持つ材料の特性についても高精度な予測を実現できることを実証した。また、大量の外挿的タスクを経験したモデルは、未知ドメインに少量のデータを追加するだけで迅速に予測能力を獲得できることも明らかにした。

機械学習と分子シミュレーションを融合した高分子材料自動設計ツール SPACIER の開発 — 高性能光學用高分子の発見

機械学習と RadonPy の分子シミュレーションを融合した高分子材料設計ツール SPACIER を開発した (Nanjo et al., *npj Comput Mater* **11**, 16 (2025), プレリリース) (図 3)。マテリアルズインフォマティクスの分野では、実験データの不足や機械学習の内挿的予測の限界を克服するために、第一原理計算や分子動力学計算などの計算機シミュレーションを統合した材料設計ツールが開発されている。しかしながら、高分子材料系の計算機シミュレーションは、計算コストの高さや計算自動化の技術的課題が依然として大きな障壁となっており、機械学習・シミュレーション統合型の材料設計ツールの開発は進んでいない。

本研究では、ベイズ最適化や能動学習などの適応実験計画に基づき、RadonPy の計算機実験を循環的に実行しながら高分子材料を設計するソフトウェア SPACIER を開発した。さらに、SPACIER に実装された多目的最適化アルゴリズムを用いて、屈折率とアッペ数のトレードオフが形成する経験的限界線を超える光学用高分子を設計し、その合成を実験的に実現した。

SPACIER ベイズ最適化 + RadonPy 計算機実験 + 高分子生成 AI を統合した高分子材料自動設計アルゴリズム

E2T 学習データの範囲外の予測の方法を学習するメタ学習アルゴリズム。大量の外挿的タスクを経験したメタ学習器は分布外汎化能力を獲得できることを実証

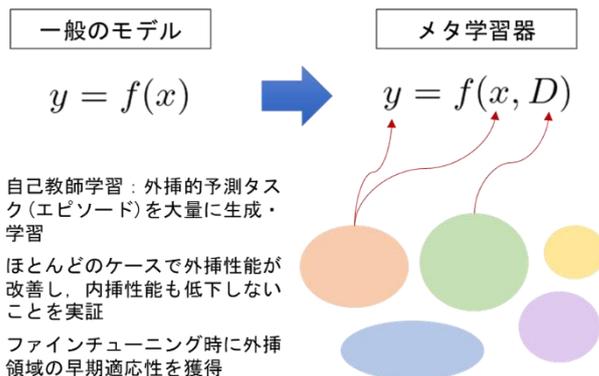
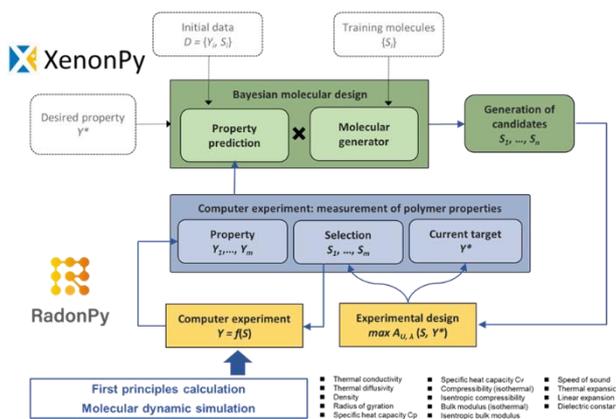


図 3. 機械学習アルゴリズムとソフトウェア開発

高分子液晶の予測と発見

高分子の繰り返し単位から液晶転移の有無を予測する機械学習モデルを開発した (Maeda et al., *npj Comput Mater*, in press)。実証研究では、機械学習アルゴリズムを用いて、芳香族ポリイミドの大規模仮想ライブラリから7個の候補を設計し、ポリマー重合を実施した。合成されたポリイミド樹脂はいずれもスメクチック相を形成し、熱伝導率が $1.0 \text{ W}/(\text{m} \cdot \text{K})$ を超えることを確認した。これらは機械学習により予測・発見された初めての高分子液晶である。

(4) プロジェクトの総合的推進

現時点において、統計数理研究所を含む2国研・9大学・37企業に属する約260名が本プロジェクトに参画している。多数の参画者による複数の共同研究プロジェクトを円滑に進めるために、定例ミーティング(年5回)や研究交流会(合宿形式)、Slackでの情報交換、ソフトウェア講習会等を介して密に情報・技術共有を行ってきた。2024年5月には、統計数理研究所が主催する産学連携シンポジウム「統計数理を活用して未踏物質空間を切り拓く」(開催場所:統計数理研究所、参加者352名)を開催した。また、日本化学会春季年会の特別企画として、第四四半期に成果発表と情報発信を目的とするシンポジウム「AI×自動計算×自動実験による化学研究のデジタル革新」(開催場所:関西大学、参加者の推定約250名)を開催した。本シンポジウムは日本化学会春季年会のハイライトシンポジウムに選定され(40件中採択1件)、記者会見を行った。

2-3. 活動（研究会の活動等）

- 2024年5月23日 統計数理研究所産学連携シンポジウム「統計数理を活用して未踏物質空間を切り拓く」（開催場所：統計数理研究所）参加者352名
- 2025年3月26日 日本化学会春季年会 企画シンポジウム「AI×自動計算×自動実験による化学研究のデジタル革新」（開催場所：関西大学）（年会のハイライトシンポジウムに選定。40件中1件採択、3/13に記者会見）参加者約250名（推定）

2-4. 実施体制

業務項目	担当機関	担当責任者
(1) 高分子物性自動計算ソフトウェア RadonPy の開発	東京都立川市緑町 10-3 大学共同利用機関法人 情報・システム研究機構 統計数理研究所	吉田 亮
	東京都目黒区大岡山 2-12-1-S1-20 東京科学大学 物質理工学院	古屋 秀峰
(2) データ駆動型材料研究に資する体系的な高分子物性データベースの開発	東京都立川市緑町 10-3 大学共同利用機関法人 情報・システム研究機構 統計数理研究所	吉田 亮
	東京都目黒区大岡山 2-12-1-S1-20 東京科学大学 物質理工学院	古屋 秀峰
(3) マテリアルズインフォマティクス基盤技術の創出とデータ駆動型材料研究の推進	東京都立川市緑町 10-3 大学共同利用機関法人 情報・システム研究機構 統計数理研究所	吉田 亮
	東京都目黒区大岡山 2-12-1-S1-20 東京科学大学 物質理工学院	古屋 秀峰
(4) プロジェクトの総合的推進	東京都立川市緑町 10-3 大学共同利用機関法人 情報・システム研究機構 統計数理研究所	吉田 亮

別添1 学会等発表実績

1. 学会誌・雑誌等における論文掲載

No.	掲載した論文（発表題目）	発表者氏名	発表した場所（学会誌・雑誌名等）	発表した時期
1	Advancing extrapolative predictions of material properties through learning to learn using extrapolative episodic training	Kohei Noda, Araki Wakiuchi, Yoshihiro Hayashi, Ryo Yoshida	Communications Materials 6: 36	2025年 2月
2	Smectic liquid crystalline poly(ester imide)s with low dielectric dissipation factors for high-frequency applications	Hayato Maeda, Yucheng Liang, Ryohei Hosoya, Rika Marui, Erina Yoshida, Yuqian Chen, Kan Hatakeyama-Sato, Yuta Nabae, Shiori Nakagawa, Junko Morikawa, Masatoshi Tokita, Ririka Sawada, Shinji Ando, Yoshihiro Hayashi, Ryo Yoshida, Hidemine Furuya, Teruaki Hayakawa	Polymer Journal 57	2025年 2月
3	SPACIER: on-demand polymer design with fully automated all-atom classical molecular dynamics integrated into machine learning pipelines	Shun Nanjo, Arifin, Hayato Maeda, Yoshihiro Hayashi, Kan Hatakeyama-Sato, Ryoji Himeno, Teruaki Hayakawa, Ryo Yoshida	npj Computational Materials 11: 16	2025年 1月
4	Shotgun crystal structure prediction using machine-learned formation energies	Chang Liu, Hiromasa Tamaki, Tomoyasu Yokoyama, Kensuke Wakasugi, Satoshi Yotsuhashi, Minoru Kusaba, Artem R. Oganov, Ryo Yoshida	npj Computational Materials 10: 298	2024年 12月
5	Comprehensive experimental datasets of quasicrystals and their approximants	Erina Fujita, Chang Liu, Asuka Ishikawa, Tomoya Mato, Koichi Kitahara, Ryuji Tamura, Kaoru Kimura, Ryo Yoshida, Yukari Katsura	Scientific Data 11: 1211	2024年 11月
6	Machine learning-based prediction of the electron energy distribution function and electron density of argon plasma from the optical emission spectra	Fatima Jenina Arellano, Minoru Kusaba, Stephen Wu, Ryo Yoshida, Zoltán Donkó, Peter Hartmann, Tsanko V. Tsankov, Satoshi Hamaguchi	Journal of Vacuum Science and Technology A 42: 053001	2024年 7月
7	[プレプリント] Bayesian Kernel Regression for Functional Data	Minoru Kusaba, Megumi Iwayama, Ryo Yoshida	arXiv:2503.13676	2025年 3月投稿

8	[プレプリント] Discovery of liquid crystalline polymers with high thermal conductivity using machine learning	Hayato Maeda, Stephen Wu, Rika Marui, Erina Yoshida, Kan Hatakeyama-Sato, Yuta Nabaе, Shiori Nakagawa, Meguya Ryu, Ryohei Ishige, Yoh Noguchi, Yoshihiro Hayashi, Masashi Ishii, Isao Kuwajima, Felix Jiang, Xuan Thang Vu, Sven Ingebrandt, Masatoshi Tokita, Junko Morikawa, Ryo Yoshida	ChemRxiv 10.26434/chemrxiv-2024-tj786	2024年 8月投稿
9	[プレプリント] Scaling Law of Sim2Real Transfer Learning in Expanding Computational Materials Databases for Real-World Predictions	Shunya Minami, Yoshihiro Hayashi, Stephen Wu, Kenji Fukumizu, Hiroki Sugisawa, Masashi Ishii, Isao Kuwajima, Kazuya Shiratori, Ryo Yoshida	arXiv:2408.04042	2024年 8月投稿

2. 国際会議・シンポジウムにおける口頭・ポスター発表

No.	発表した成果（発表題目、口頭・ポスター発表の別）	発表者氏名（所属機関）	発表した場所（学会名等）	発表した時期
1	自動シミュレーションによるデータ駆動型高分子材料研究の最前線、口頭	林 慶浩	日本化学会 第105春季年会	2025年 3月
2	データ駆動型高分子材料研究を変革する Sim2Real マテリアルズインフォマティクス、口頭	林 慶浩	高分子学会千葉地域活動若手セミナー	2025年 3月
3	Sim2Real Machine Learning in Data-Driven Materials Research、口頭	Ryo Yoshida	ISI-ISM-ISSAS 2025 Joint Conference	2025年 3月
4	適応的実験計画と高分子物性自動計算の融合：高屈折率・高アッペ数高分子の発見、ポスター	南條 舜	第4回「富岳」成果創出加速プログラム研究交流会	2025年 2月
5	自動分子シミュレーションを用いた高分子材料における Sim2Real マテリアルズインフォマティクス、口頭	林 慶浩	第4回「富岳」成果創出加速プログラム研究交流会	2025年 2月
6	高分子材料データベース PoLyInfo と外部データベースの統合解析、口頭	吉田 亮	NIMS データ中核拠点 (MDPF) 技術開発・共用部門オープンセミナー	2025年 2月
7	データ駆動型材料研究におけるデータ基盤の在り方：AI活用の視点から、口頭	吉田 亮	マテリアル戦略総合シンポジウム 2025 (MatISS 2025)	2025年 1月

8	Sim2Real マテリアルズインフォマティクス：高分子材料データの乏しさを分子シミュレーションにより克服する、口頭	林 慶浩	信州大学 先鋭材料研究所 データ駆動型 AI セミナー	2025 年 1 月
9	Sim2Real 機械学習による未踏物質空間の開拓、口頭	吉田 亮	東京科学大学 物質・情報卓越教育院「新産業創出最前線」	2025 年 1 月
10	Conformal Prediction とベイズ最適化、口頭	南條 舜*, 吉田 亮	第 55 回 IBISML 研究会	2024 年 12 月
11	Scalability of Growing Computational Database in Sim2Real Materials Informatics、口頭	Ryo Yoshida	The 15th International Symposium on Atomic Level Characterizations for New Materials and Devices 2024 (ALC 2024)	2024 年 11 月
12	マテリアルサイエンスを次世代技術でどう進化させるか、パネルディスカッション	吉田 亮	第 18 回材料系ワークショップ ～次世代技術が拓く物質・材料開発の未来：演算加速器・生成 AI・量子コンピュータ～	2024 年 10 月
13	高分子物性自動計算ソフトウェア RadonPy と SIM2REAL 転移学習のスケールリング則、口頭	吉田 亮	第 18 回材料系ワークショップ ～次世代技術が拓く物質・材料開発の未来：演算加速器・生成 AI・量子コンピュータ～	2024 年 10 月
14	産学連携コンソーシアムによる Sim2Real マテリアルズインフォマティクス基盤の共同開発、口頭	林 慶浩*, 篠田 恵子, 南條 舜, 野口 瑤, 高橋 愛子, ウステファン, 吉田 亮	第 73 回高分子討論会	2024 年 9 月
15	SIM2REAL マテリアルズインフォマティクス：限られたデータの壁を乗り越える、口頭	吉田 亮	一般社団法人ゲノムテクノロジー研究会 第 8 回バイオインフォマティクス分科会「データサイエンス」	2024 年 9 月
16	高分子材料における Sim2Real 転移学習とスケールリング則の観測、口頭	林 慶浩*, 南 俊匠, 南條 舜, 高橋 愛子, 吉田 亮	2024 年度統計関連学会連合大会	2024 年 9 月
17	Development of Quasicrystals Datasets and Application to Machine Learning、ポスター	Erina Fujita	Materials Informatics: Accelerating Materials Research and Design with Artificial Intelligence	2024 年 8 月

18	Industry-Academia Consortium for Co-creating Polymer Property Database、ポスター	Aiko Takahashi	Materials Informatics: Accelerating Materials Research and Design with Artificial Intelligence	2024年 8月
19	Transferability and Scalability of Growing Computational Database in Sim2Real Materials Informatics、口頭	Ryo Yoshida	Materials Informatics: Accelerating Materials Research and Design with Artificial Intelligence	2024年 8月
20	SIM2REAL マテリアルズインフォマティクス：限られたデータの壁を乗り越える、口頭	吉田 亮	データ同化における深層学習の活用	2024年 6月
21	Co-creation of polymer properties database through industry-academia collaboration and SIM2REAL materials informatics、口頭	Ryo Yoshida	Next-Generation Electronic Materials & Advanced Thermal Management Technologies	2024年 6月
22	マテリアルズインフォマティクスの現状と展望、口頭	吉田 亮	「統計数理を活用して未踏物質空間を切り拓く」産学連携シンポジウム	2024年 5月
23	データ駆動型高分子設計による新規高分子の発見、口頭	Stephen Wu	「統計数理を活用して未踏物質空間を切り拓く」産学連携シンポジウム	2024年 5月
24	自動分子シミュレーションによる高分子材料の計算物性データベース創出と Sim2Real 転移学習、口頭	林 慶浩	「統計数理を活用して未踏物質空間を切り拓く」産学連携シンポジウム	2024年 5月
25	石油由来プラスチックに代わる生分解性ポリマーを探索する、口頭	篠田 恵子	「統計数理を活用して未踏物質空間を切り拓く」産学連携シンポジウム	2024年 5月
26	機械学習×第一原理計算で結晶構造を予測する、口頭	Chang Liu	「統計数理を活用して未踏物質空間を切り拓く」産学連携シンポジウム	2024年 5月
27	機械学習で準結晶を予測・発見する、口頭	草場 稜	「統計数理を活用して未踏物質空間を切り拓く」産学連携シンポジウム	2024年 5月
28	閉会挨拶：マテリアルズインフォマティクス研究推進センターについて、口頭	吉田 亮	「統計数理を活用して未踏物質空間を切り拓く」産学連携シンポジウム	2024年 5月