

令和3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

「データ駆動型高分子材料研究を変革するデータ基盤創出」

成果報告書

令和4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3年度（報告年度）の実施内容.....	- 1 -
2-1. 当該年度（令和3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）.....	- 7 -
2-4. 実施体制.....	- 7 -

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

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3年度（報告年度）の実施内容

2-1. 当該年度（令和3年度）の事業実施計画

(1) 「高分子物性データベースの開発」

RadonPy の「富岳」への移植・最適化を実施する。「富岳」における古典分子動力学ソフトウェア LAMMPS の並列化効率を測定し、最適な計算スキームを確立する。さらに、各物性や系に対し、プロジェクト内でタスクフォースを立ち上げ、最適な計算条件やパラメータを検討・決定し、この標準プリセットを用いてデータ生産を実施する。当該年度第四半期を目途に、協力機関・連携機関との連携体制を確立し、「富岳」を利用してデータ生産を本格始動する。

(2) 「高分子物性自動計算アルゴリズム・ソフトウェアの開発」

自動計算コードの対象物性や系を拡張する。

(3) 「マテリアルズインフォマティクスの基盤技術の研究」

高分子物性自動計算コードと高分子設計の機械学習アルゴリズムを統合したベイズ最適化による自動分子設計のプログラムを開発する。

(4) 「プロジェクトの総合的推進」

プロジェクトを円滑に進めていくために、定期的に運営会議とテクニカルミーティングを開催し、参画機関の間で意見交換と情報交換を行っていく。論文発表・学会発表やホームページでの情報公開を通じて、プロジェクトの研究成果を積極的に発信していく。また、年度中に本プロジェクトの成果発表と情報発信を目的とする成果報告会を開催する。

2-2. 実施内容（成果）

高分子物性自動計算システム RadonPy を用いて、データ駆動型研究に資する包括的な高分子物性データベースを構築した。実施内容の詳細を以下に示す。

(1) 「高分子物性データベースの開発」

全原子分子動力学 (MD) シミュレーションによる高分子物性計算を全自動化する Python ライブラリ RadonPy を用いて、世界初となる包括的な高分子物性データベースを創出する。ポリマーの繰り返し単位の化学構造を入力し、力場の割り当て、初期構造の生成、エラー処理、平衡・非平衡 MD 計算による物性評価までの全工程を完全に自動化する (図 1)。本プロジェクトでは、富岳の大量のノードを用いて多数のポリマーの物性を高速に計算し、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。

2021 年 9 月にプロジェクトが始動し、まずは RadonPy の富岳への移植とチューニングを試みた。し

かしながら、x86 系に比べると A64FX では RadonPy の一部モジュール (量子化学計算ソフトウェア Phi4 による電荷計算) の計算性能が著しく低下することが判明した。そこで計算対象のポリマー集合の電荷計算を外部スパコン (自然科学研究機構 岡崎共通研究施設 計算科学研究センター NEC LX2U-Twin2 サーバ 406Rh-2) で実行し、その他の計算を富岳で実行するための機能拡張を行い、パイプラインを構築した。

また、LAMMPS の計算性能も x86 系に比べて 4 倍ほど低下することが判明したため、高度情報科学技術研究機構 (RIST) の協力の下で LAMMPS のチューニングを実施した。このような過程を経て 2022 年 1 月中旬頃にデータ生産を本格始動する体制が整った。

2022 年 1 月以降、当該年度の割当資源である 1,080,000 ノード時間を使用し、計 11,181 アモルファスポリマーの 15 物性の計算を実施した (2021 年度 KPI 目標値 10,000 ポリマーに到達)。

図 2 に示すように、大量のポリマーの複数物性の同時分布を観測することに成功した。これにより、複数物性のパレート面の位置が特定されて、パレート面を構成するポリマー群の構造的特徴に関する体系的な知見を得るに至った。熱伝導率の計算では、パレート境界を超えるような特異的な分子骨格が同定された。通常のアモルファスポ

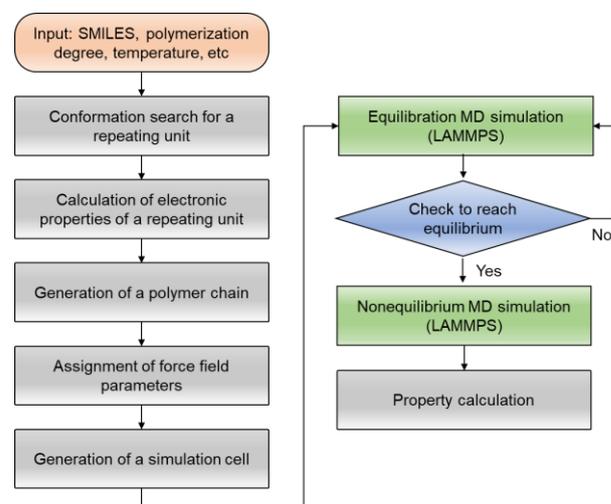


図 1. RadonPy の高分子物性計算のワークフロー

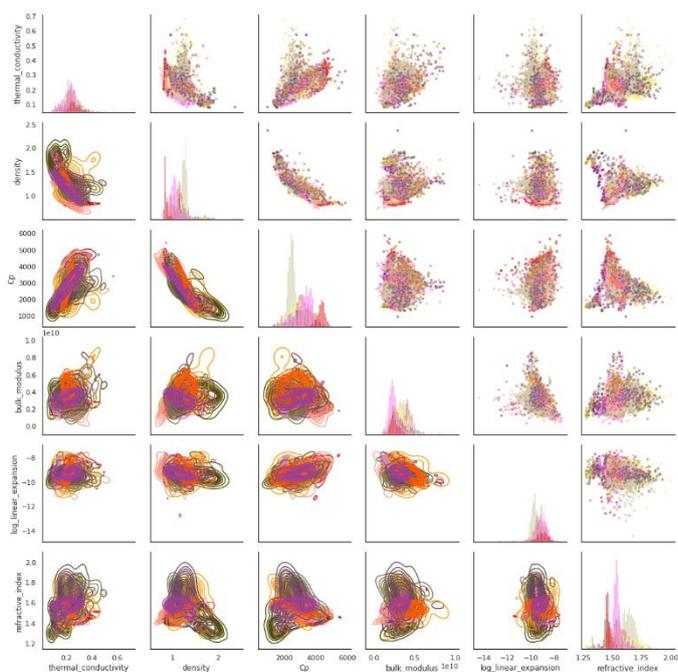


図 2. 複数物性の同時分布の観測とパレート境界の同定

リマーの熱伝導率は高くても 0.2-0.3 W/mK ほどであることが知られているが、計算されたポリマーの一部に熱伝導率が 0.5 W/mK を超えるようなものが含まれていた。これらのポリマーの構造的特徴を解析した結果、分子骨格の剛直性、水素結合可能なユニットが高密度で存在すること、水素結合や dipole-dipole 相互作用を介したメカニズムがアモルファスポリマーの高熱伝導化を実現していることが明らかになった (プレプリント Hayashi et al. arXiv:2203.14090 (2022))。

(2) 「高分子物性自動計算アルゴリズム・ソフトウェアの開発」

2021 年度はアモルファスポリマーの 15 物性の自動計算を実装し、プレプリントの公開と RadonPy のファーストリリースを行った。

論文 : Hayashi, Y., Shiomi, J., Morikawa, J., Yoshida, R., RadonPy: automated physical property calculation using all-atom classical molecular dynamics simulations for polymer informatics. arXiv preprint. arXiv:2203.14090 (2022).

DOI: <https://doi.org/10.48550/arXiv.2203.14090>

コード (GitHub) : <https://github.com/RadonPy/RadonPy>

現在実装されている物性リストは、以下の通りである。

- Thermal conductivity
- Thermal diffusivity
- Density
- Radius of gyration
- Specific heat capacity Cp
- Specific heat capacity Cv
- Compressibility (isothermal)
- Isentropic compressibility
- Bulk modulus (isothermal)
- Isentropic bulk modulus
- Self-diffusion coefficient
- Thermal expansion coefficient
- Linear expansion coefficient
- Dielectric constant (static)
- Refractive index

RadonPy の入力変数は、ポリマーの繰り返しユニットを表す SMILES (化合物の 1 次構造を表す文字列) と重合度である。重合度は、ポリマー鎖の原子数が約 1,000 になるように調整する。繰り返しユニットをランダムウォーク法で連結し、ランダムコイル状の高分子鎖を生成する。この高分子鎖を 10 本複製し、互いに重ならないようランダムに配置・回転し、アモルファスポリマーの初期構造を生成する。MD 計算に必要な力場割り当てなどの設定も全自動化した。系の構造緩和のための平衡化 MD を LAMMPS で実行する。平衡化計算の終了は、エネルギー、密度、原子座標の平均二乗変位などの収束状況から判定する。この平衡化 MD から密度や比熱などの物性値を算出する。さらに、平衡化後のアモルファス構造に対して、熱伝導 MD を実行する。熱伝導 MD の結果からアモルファス構造の熱伝導率と熱拡散係数、一軸延

伸 MD の結果からヤング率やポアソン比などの物性値を算出する。これらの計算条件は、高分子物性データベース PoLyInfo (文献から抽出した物性値データベース) の 1,070 ポリマーの物性値との系統的な比較検証に基づいて決定された (図 3)。

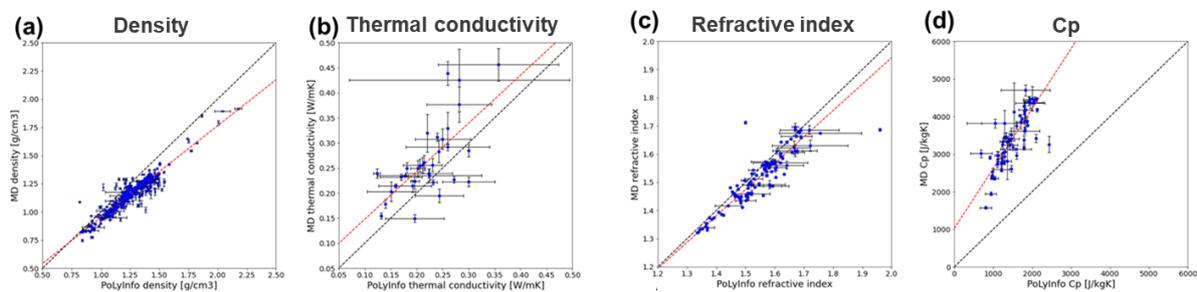


図 3. PoLyInfo の実験値と RadonPy の計算物性との系統的な比較検証

(3) 「マテリアルズインフォマティクスの基盤技術の研究」

高分子物性自動計算コードと高分子設計の機械学習アルゴリズムを統合したベイズ最適化による自動分子設計プログラムの開発に着手した。本プロジェクトでは、主な計算対象を仮想ポリマーに定める。高分子の骨格を以下の 20 種類に分類し、機械学習アルゴリズム (Ikebata et al. J Comput Aided Mol Des. 31(4):379-391 (2017)) で各ポリマークラスの分子生成モデルを構築した。ここで、既存のポリマーの化学構造を用いて機械学習のモデルを訓練し、既存分子に現れる頻出パターン (フラグメントや結合ルールなど) を模倣した構造生成器を構築した。このように多様な分子骨格を有する計 1,778,039 個の候補分子からなる包括的な仮想ライブラリを作製した (図 4)。

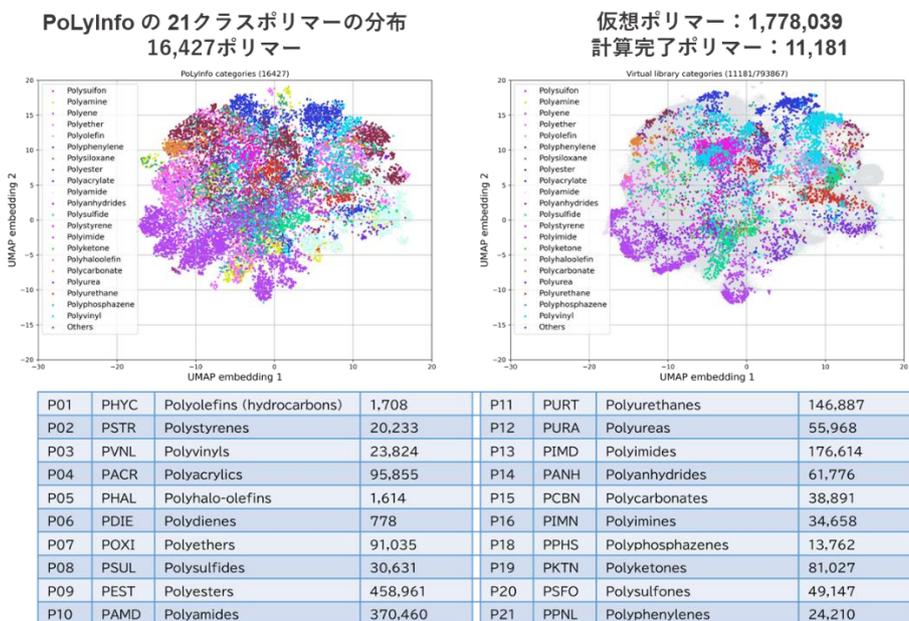


図 4. 機械学習アルゴリズムによる 20 種類の分子骨格を包含するポリマー仮想ライブラリ

また、所望の特性を持つ分子を設計するための、機械学習アルゴリズムの開発を実施し、プレプリントとコードを公開した。

論文：Zhang Q, Liu C, Wu S, Yoshida R. Bayesian sequential stacking algorithm for concurrently designing molecules and synthetic reaction networks. ChemRxiv. (2022). DOI: 10.26434/chemrxiv-2022-5qpv9

コード (GitHub) : <https://github.com/qi-zh/Seq-Stack-Reaction>

論文：Iwayama M, Wu S, Liu C, Yoshida R. Functional output regression for machine learning in materials science. ChemRxiv. (2022). 10.26434/chemrxiv-2022-vbnnh

コード (GitHub) : https://github.com/yoshida-lab/XenonPy/blob/master/samples/kernel_neural_network.ipynb

Zhang らは、機能性分子と合成経路を同時に設計するための機械学習アルゴリズムを開発した (図 5)。この研究では、ベイズ推論の枠組みの中で、分子構造と合成経路の同時設計問題を定式化した。設計変数は、反応ネットワークにおける反応物のセットとそのネットワークトポロジーからなる。設計空間は、購入可能な反応物質のすべての組み合わせからなるため非常に大きく、数百万以上のオーダーとなることもありえる。また、設計された反応ネットワークは、単純な多段階の直線的な反応経路にとどまらず、あらゆるトポロジーをとりうる。この組合せ問題を解決するために、単一ステップの反応を逐次的に積み上げて合成反応ネットワークを再帰的に設計する逐次モンテカルロアルゴリズムを開発した。Iwayama らは、関係型変数を予測する機械学習のアルゴリズムを開発した (図 6)。材料科学の教師あり学習では、分子の光吸収スペクトルや物性の温度依存性曲線、材料組織の電子顕微鏡画像のように出力変数が超高次元で且つ学習に利用できるデータの量が限られているケースが多い。一方、このような少数データに基づく多次元出力変数の教師あり学習を対象とした体系的な研究はほとんど存在しない。本研究では、材料データを主な適用対象として、敵対的生成学習モデルと関数型出力変数のカーネル回帰に基づく多次元出力変数の予測手法を開発した。

- ✓ 順方向予測: 反応物 → 生成物 → 物性
- ✓ 逆方向予測: 逐次モンテカルロ法で商用化合物から所望の特性を有する生成物を合成する反応物の組み合わせを探索

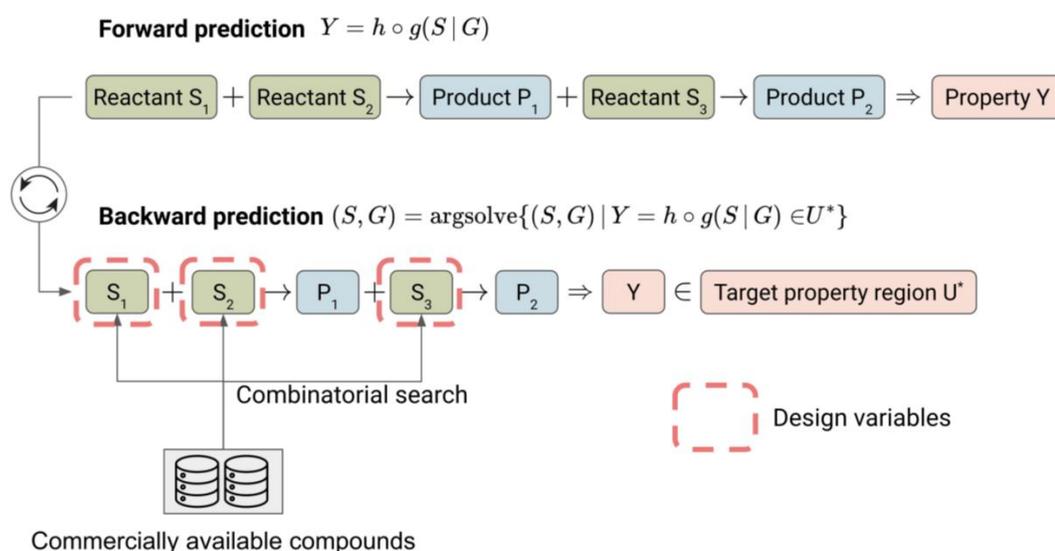


図 5. 所望の特性を持つ分子とその合成経路を予測する機械学習アルゴリズム

- ✓ 着眼点：材料研究の多くの問題は、関数出力変数の回帰問題に帰着
- ✓ 目的：問題発掘とモデリング・推定の汎用的方法論の確立
- ✓ 目標：少数データに適用可能

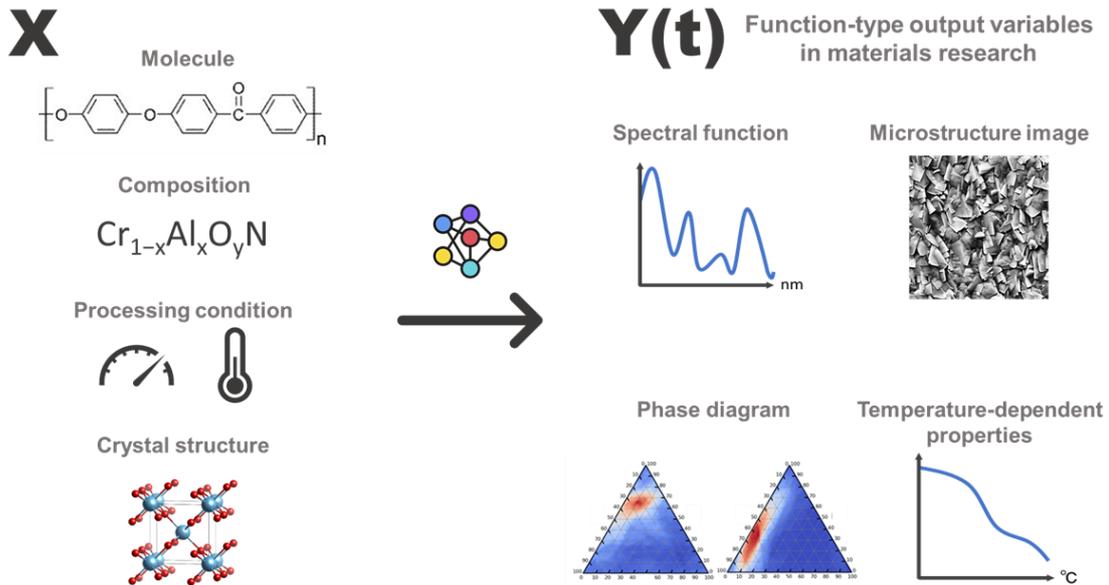


図 6. 材料科学における関数出力回帰の汎用的方法論の構築

(4) 「プロジェクトの総合的推進」

2021年度は、統計数理研究所に加えて3大学14企業に所属する計62名が本プロジェクトに参画した。参画者の一部は、RadonPyの開発に加わり、ガラス転移温度、溶媒和自由エネルギー、レオロジー特性、熱硬化性樹脂の自動計算手法の開発に取り組んだ。また、多数の参画者が富岳を利用してデータ生産を実施した。参画者の多くは、MDシミュレーションやスパコンの専門知識を持たないため、毎月定例で開催されるミーティングやSlackを活用して技術指導や情報交換を行った。定例ミーティングの開催履歴は以下の通りである。

テクニカルミーティング：プロジェクトの進捗報告・情報交換に関するミーティング

- 10月5日 (40名)
- 12月7日 (41名)
- 2月8日 (42名)

運営会議：プロジェクトの企画・運営に関するミーティング

- 9月21日 (56名)
- 11月9日 (53名)
- 1月11日 (37名)
- 3月8日 (55名)

また、2022年3月31日に成果報告会を開催し、プロジェクトの研究紹介、RadonPy ファーストリリースのアナウンス、関連コミュニティからの意見収集を行った。参加者数は219名であった。

「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」公開セミナー（成果報告会）

【日時】2022年3月31日（木）15:00-17:00

【開催方法】ウェブ開催（Zoom）

【プログラム】

15:00-15:45 プロジェクト概要：データ駆動型高分子材料研究の諸問題

吉田 亮（統計数理研究所 ものづくりデータ科学研究センター センター長）

15:45-16:45 高分子物性自動計算システム RadonPy の開発と産学連携によるデータプラットフォームの共創

林 慶浩（統計数理研究所 ものづくりデータ科学研究センター 特任助教）

16:45-17:00 質疑応答・クロージング

2-3. 活動（研究会の活動等）

- 2022年3月31日「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」公開セミナー（成果報告会）（参加人数219名）
- RadonPy データベース共同開発事業コンソーシアム 運営会議4回 テクニカルミーティング3回（参加人数のべ324名）

2-4. 実施体制

業務項目	担当機関	担当責任者
(1) 高分子物性データベースの開発	統計数理研究所	吉田 亮
(2) 高分子物性自動計算アルゴリズム・ソフトウェアの開発	東京工業大学 統計数理研究所	古屋 秀峰 吉田 亮
(3) マテリアルズインフォマティクスの基盤技術の研究	統計数理研究所	吉田 亮
(4) プロジェクトの総合的推進	統計数理研究所	吉田 亮

別添 1 学会等発表実績

学術論文

1. Hayashi, Y., Shiomi, J., Morikawa, J., Yoshida, R., RadonPy: automated physical property calculation using all-atom classical molecular dynamics simulations for polymer informatics. arXiv preprint. arXiv:2203.14090 (2022). DOI: <https://doi.org/10.48550/arXiv.2203.14090>
2. Zhang Q, Liu C, Wu S, Yoshida R. Bayesian sequential stacking algorithm for concurrently designing molecules and synthetic reaction networks. ChemRxiv. (2022). DOI: 10.26434/chemrxiv-2022-5qp9
3. Iwayama M, Wu S, Liu C, Yoshida R. Functional output regression for machine learning in materials science. ChemRxiv. (2022). DOI: 10.26434/chemrxiv-2022-vbnnh
4. Kusaba, M., Liu, C., Yoshida, R. Crystal structure prediction with machine learning-based element substitution. Computational Materials Science. 11:111496 (2022). DOI: <https://doi.org/10.1016/j.commatsci.2022.111496>
5. Torres P., Wu S., Ju S., Liu C., Tadano T., Yoshida R., Shiomi J. Descriptors of intrinsic hydrodynamic thermal transport: screening a phonon database in a machine learning approach. Journal of Physics: Condensed Matter. 34(13):135702 (2022). DOI: <https://doi.org/10.1088/1361-648X/ac49c9>
6. Ma, R., Zhang, H., Xu, J., Hayashi, Y., Yoshida, R., Shiomi, J., Luo, T., Machine learning-assisted exploration of thermally conductive polymers based on high-throughput molecular dynamics simulations, arXiv preprint. arXiv:2109.02794v1 (2021). DOI: <https://doi.org/10.48550/arXiv.2109.02794>

ソフトウェア等

1. 高分子物性自動計算ライブラリ RadonPy: <https://github.com/RadonPy/RadonPy>
2. 分子・合成経路自動生成アルゴリズム Seq-Stack-Reaction: <https://github.com/qi-zh/Seq-Stack-Reaction>
3. 関数出力変数カーネル回帰: https://github.com/yoshida-lab/XenonPy/blob/master/samples/kernel_neural_network.ipynb
4. 結晶構造予測プログラム CSPML: <https://github.com/Minoru938/CSPML>

口頭発表

1. 【国際会議・招待講演】 2022.1 PiAI Seminar Series: Physics informed AI in Plasma Science, Scientific Understanding from Machine Learning in Materials Science (R. Yoshida)
2. 【国際会議・招待講演】 2021.12. Materials Research Meeting 2021, online, “Machine Learning Phase Prediction of Quasicrystals”, (R. Yoshida)
3. 【国際会議・招待講演】 2021.9 The 11th International Conference on Flexible and Printed Electronics, online, “Machine Learning for Inverse Materials Design” (R. Yoshida)
4. 2022.3 第5回マテリアルズインフォマティクスセミナー（「富岳」成果創出加速プログラム成果報告会），ウェブ開催，“データ駆動型高分子材料研究の諸問題”（吉田 亮）
5. 【基調講演】 2022.3 日本化学会 第102春季年会（2022），ウェブ開催，“材料研究における逆問題と統計的機械学習”（吉田 亮）
6. 2022.3 日本化学会 第102春季年会（2022），ウェブ開催，“Data-driven crystal structure prediction using

structure similarity" (草場 穂*, Chang Liu, 吉田 亮)

7. 2022.3「富岳」成果創出加速プログラム研究交流会, ウェブ開催, “高分子物性自動計算ソフトウェア RadonPy の開発とデータ駆動型材料研究に資するデータベースの創出” (吉田 亮)
8. 2022.2 第2回プラズマインフォマティクス研究会, ウェブ開催, “少数データに基づく多次元出力変数の教師あり学習” (岩山 めぐみ*, Stephen Wu, 吉田 亮)
9. 【招待講演】2022.2 名古屋大学生物統計セミナー, ウェブ開催, “統計的機械学習の応用と実際” (吉田 亮)
10. 【招待講演】2022.2 CMC Research ウェビナー, ウェブ開催, “マテリアルズインフォマティクス概論” (吉田 亮)
11. 【招待講演】2021.12 国立精神・神経医療研究センター 第2回脳病態数理・データ科学セミナー, ウェブ開催, 統計的機械学習による予測・発見・理解 (吉田亮)
12. 【優秀発表賞】2021.12. 第44回ケモインフォマティクス検討会, ウェブ開催, “ベイズ推論を用いた高潜熱蓄冷材向け分子探索” (外海駿輔, 榊原功次, 板倉智也, Stephen Wu, 吉田亮)
13. 2021.12 「富岳」成果創出加速プログラム 物質・材料系課題 合同研究会, ウェブ開催, “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” (吉田亮)
14. 【特別講演】2021.11 第4回 ROIS 産学連携・知的財産セミナー データサイエンスにおける産学連携シーズ～ ROIS・統数研 産連知財セミナー ～, ウェブ開催, “統計的機械学習による新材料創製：産学連携の現状と可能性” (吉田亮)
15. 2022.10 第42回日本熱物性シンポジウム, ウェブ開催, “分子動力学計算による高分子熱物性データベースの構築とデータ科学的手法による物性支配因子の解析” (林 慶浩*, ウ ステファン, 野口 瑠, 塩見 淳一郎, 森川 淳子, 吉田 亮)
16. 【招待講演】2022.9 CMC Research ウェビナー, ウェブ開催, “マテリアルズインフォマティクス概論” (吉田 亮)
17. 2021.9 2021 年度統計関連学会連合大会, ウェブ開催, “高分子のバーチャルスクリーニング” (ウ ステファン*, 野口 瑠, 林 慶浩, 難波江 裕太, 早川 晃鏡, 森川 淳子, 吉田 亮)
18. 2021.9 第70回高分子検討会, ウェブ開催, “分子動力学計算による高分子熱物性データベースの構築と機械学習的手法による物性支配因子の解析” (林 慶浩*, ウ ステファン, 野口 瑠, 塩見 淳一郎, 森川 淳子, 吉田 亮)