

一般セッション(ポスター講演) | 23 合同セッションN「インフォマティクス応用」：23 合同セッションN「インフォマティクス応用」(ポスター)

📅 2025年3月15日(土) 13:30 ~ 15:30 📍 P06 (森戸記念体育館)

## [15p-P06-1~15] 23 合同セッションN「インフォマティクス応用」(ポスター)

[15p-P06-1]

材料研究でのイジングマシン活用の試み

○尾崎 仁亮<sup>1</sup>、池田 潤<sup>1</sup> (1.村田製作所)

[15p-P06-2]

機械学習と量子アニーリングを活用した印刷型調光フィルムデバイス用インクの組成最適化

○田口 諒<sup>1</sup>、外岡 和彦<sup>1</sup>、渡邊 浩<sup>1</sup>、久保田 隆<sup>1</sup>、田嶋 一樹<sup>1</sup> (1.産総研)

[15p-P06-3]

MLIPsの特徴量を活用したNMRのケミカルシフト予測

○米澤 拓孝<sup>1</sup> (1.株式会社PFCC)

[15p-P06-4]

機械学習分子動力学シミュレーションによる高分子電解質膜内のプロトン輸送の解析

○吉本 勇太<sup>1</sup>、松村 直樹<sup>1</sup>、岩崎 有登<sup>1</sup>、中尾 宏<sup>1</sup>、坂井 靖文<sup>1</sup> (1.富士通)

[15p-P06-5]

分子計算と機械学習によるSi表面に対する低吸着エネルギー分子のデータ駆動型探索

○村島 広衛<sup>1</sup>、魚波 公希<sup>1</sup>、國枝 省吾<sup>1</sup>、塙 洋祐<sup>1</sup>、佐々木 悠太<sup>1</sup>、上島 仁<sup>2</sup>、新谷 俊了<sup>2</sup>、吉田 亮<sup>3</sup> (1.SCREEN ホールディングス、2.システム計画研究所、3.統計数理研究所)

[15p-P06-6]

イオン伝導体CIFファイルの違いがパーシステント図に与える影響

○吉武 道子<sup>1</sup>、長田 貴弘<sup>1</sup> (1.NIMS)

[15p-P06-7]

高密度シリカガラスの構造ユニットの定量的な形状解析

○嶋 佑泰<sup>1</sup>、志賀 元紀<sup>1,2,3</sup> (1.東北大院情報、2.NIMS、3.理研)

[15p-P06-8]

二段階LASSO回帰による多元素ドーブヘマタイト光電極の光電流予測

○西村 拓真<sup>1</sup>、隈部 佳孝<sup>1,2</sup>、原嶋 庸介<sup>3,4</sup>、藤井 幹也<sup>3,4</sup>、立川 貴士<sup>1,2</sup> (1.神戸大院理、2.神戸大分子フォト、3.奈良先端大物質、4.奈良先端大DSC)

[15p-P06-9]

内殻電子励起スペクトルを条件とした拡散モデルによる励起原子配位環境の予測

○大久保 怜<sup>1</sup>、高原 泉<sup>1</sup>、ワンルイス<sup>1</sup>、柴田 基洋<sup>1,2</sup>、溝口 照康<sup>1,2</sup> (1.東大院工、2.東大生研)

[15p-P06-10]

説明可能AIによる磁気構造とエネルギーの物理的情報変換モデルの開発

○(B)坪内 敢志<sup>1</sup>、長岡 竜之輔<sup>1</sup>、谷脇 三千輝<sup>1</sup>、町田 陽太郎<sup>1</sup>、Lira Foggiatto Alexandre<sup>1</sup>、小嗣 真人<sup>1</sup> (1.東理大先進工)

[15p-P06-11]

## XANESデータを用いる結晶構造予測モデル

○北井 孝紀<sup>1</sup>、志賀 元紀<sup>1,2,3</sup>、二宮 翔<sup>1</sup>、西堀 麻衣子<sup>1</sup> (1.東北大、2.NIMS、3.理研AIP)

---

[15p-P06-12]

セグメント回帰を用いた変化点抽出と線形化

○柳生 進二郎<sup>1</sup>、長田 貴弘<sup>1</sup> (1.NIMS)

---

[15p-P06-13]

2次元計測データよる誘電体薄膜材料探索効率化と既存データとの連携の検討

○長田 貴弘<sup>1</sup>、木野 日織<sup>1</sup>、河野 敬<sup>1</sup>、尾崎 仁亮<sup>2</sup>、池田 潤<sup>2</sup> (1.NIMS、2.村田製作所)

---

[15p-P06-14]

TADF材料の機械学習に向けた文献からのデータベース作成

○宮田 哲<sup>1,2</sup>、椿 真史<sup>2</sup>、中野谷 一<sup>3</sup>、畠山 琢次<sup>4</sup>、中山 泰生<sup>1</sup>、細貝 拓也<sup>1,2</sup> (1.東理大、2.産総研、3.九大、4.京大)

---

[15p-P06-15]

LLMを使った材料開発(II)：米国特許からのNdFeB磁石データの抽出

○岡 博之<sup>1</sup>、石井 真史<sup>1</sup> (1.NIMS)

---

## 材料研究でのイジングマシン活用の試み

### Trial for Utilizing Ising Machine in Materials Research

株式会社村田製作所<sup>1</sup> ◯尾崎 仁亮<sup>1</sup>, 池田 潤<sup>1</sup>

Murata Manufacturing, Co., Ltd.<sup>1</sup>, ◯Noriaki Ozaki<sup>1</sup>, Jun Ikeda<sup>1</sup>

E-mail: noriaki.ozaki@murata.com

電子部品が高電圧、高温などの過酷な環境下で使用される機会が増えており、部品を構成する材料への特性要求が高度化している。このため、要求を満たす材料の発見が難化し、開発における仮説立案—検証のサイクルの高速化が必要になっている。我々は、この高速化の鍵がデータ科学にあると考えている。中でも、メタヒューリスティクスの物理実装であるイジングマシンは、非制約二次二値最適化 (QUBO) に変換した問題の高速求解が可能であり、高効率なデータ解析手法として有望である。しかし、本研究に着手した当初、材料研究における課題を QUBO に変換した事例は少なく、実用可能性や汎用性は未知であった。そこで、我々は、どのような課題が、QUBO に変換可能か、検証に取り組んだ。この結果、下記のような活用方法があることを見出している。まず、実験計画段階では、球充填法を用いた非直方空間への空間充填実験計画を実装した。続くデータ解析段階では、線形回帰と組み合わせることで、特性発現の基盤となる要因構造を特定可能とした[1]の他、線形非ガウス因果モデルによる変数間の因果関係の探索を、高速化することに成功した。さらに、考察段階で有用となる、焼結や拡散のシミュレーション[2,3]が、実行可能であることも発見した。加えて、イジングマシンによる汎用的な最適化手法である FMA にグレイコードを適用し、連続変数の最適化を高精度化した。本研究により、イジングマシンが材料研究の様々な段階で有用なツールとなることを確認し、材料研究の効率化に寄与しうることを示した。

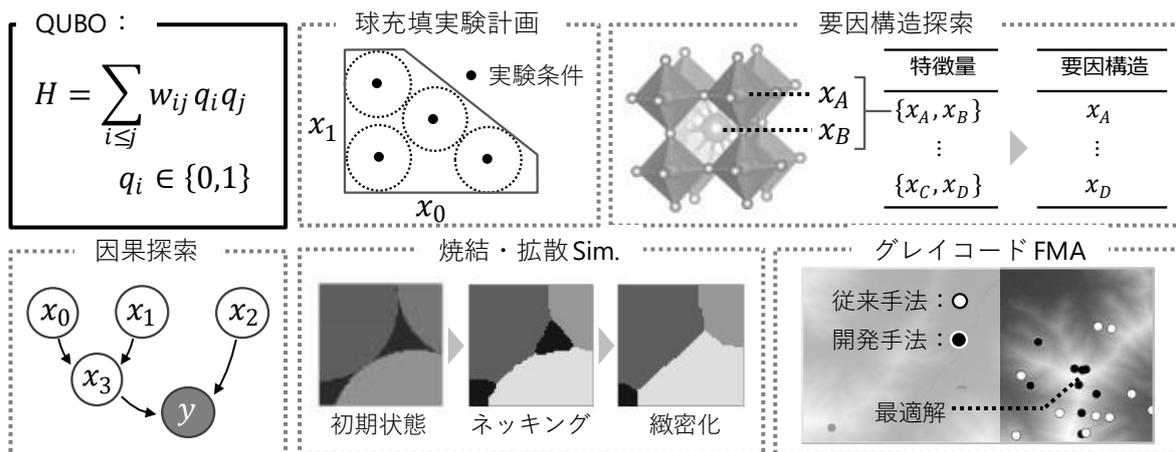


図 1. Tasks in materials research solved by Ising machine.

[1] 尾崎仁亮, 池田潤, 特徴量の組合せ最適化による材料特性決定要因の解明, 応物秋季学術講演会(2024). [2] 尾崎仁亮, 池田潤, イジングマシンによる固体材料の焼結シミュレーション, 応物秋季学術講演会(2023). [3] 尾崎仁亮, 池田潤, イジングマシンによる拡散のシミュレーション, 応物春季学術講演会(2024).

## 機械学習と量子アニーリングを活用した 印刷型調光フィルムデバイス用インクの組成最適化

Machine learning and quantum annealing optimize the composition of ink for fabricating  
printable smart windows

産総研<sup>1</sup>, °田口諒<sup>1</sup>, 外岡和彦<sup>1</sup>, 渡邊浩<sup>1</sup>, 久保田隆<sup>1</sup>, 田嶋一樹<sup>1</sup>

AIST<sup>1</sup>, °Ryo Taguchi<sup>1</sup>, Kazuhiko Tonoka<sup>1</sup>, Hiroshi Watanabe<sup>1</sup>, Takashi Kubota<sup>1</sup>, Kazuki Tajima<sup>1</sup>

E-mail: r-taguchi@aist.go.jp, k-tajima@aist.go.jp

【緒言】環境面・安全面から車載応用に向けたスマートウィンドウの開発が活発である。これまで我々は、無機化合物に電圧を印加することで、光学特性が変化するエレクトロクロミズム (EC) を利用した調光フィルムデバイスの開発に取り組んでいる<sup>1-3)</sup>。本デバイスは、EC 材料を水溶性溶媒に分散したインクを塗布することで簡便に作製でき、印刷機により大型デバイスの作製にも成功している<sup>3)</sup>。一方、EC インクの組成がデバイスの調光機能に与える影響については未だに不明な点が多く、インク組成を最適化するには至っていない。

そこで本研究では、インク組成の異なる EC インクを調製し、インク組成が調光性能に与える影響を調べた。また、得られたデータを学習データとして、インク組成から調光機能を予測する機械学習モデルを構築した。さらに、組み合わせ最適化を得意とする量子アニーリングを用いて、構築した機械学習モデルから EC インク組成を最適化した。

【実験・結果】三酸化タングステン ( $\text{WO}_3$ ) を EC 材料、ポリビニールアルコール (PVA) バインダーとして、それぞれの混合比を変えた EC インクを調製した。スピncerをういて EC 層が約  $1 \mu\text{m}$  の膜厚になるようにスピncer条件を最適化し、EC インクを ITO (インジウムスズ酸化物) 基板上に成膜し、無色透明な薄膜デバイスを作製した。作製した薄膜デバイスに電圧印加すると、青色に着色し、透過率が変化することを確認した。また、この透過率の変化幅は  $\text{WO}_3$  と PVA の混合比率で異なることが分かった。得られたデータをトレーニングデータとテストデータに分割し、インク組成から透過率変化幅の予測する機械学習モデルの構築を検討した。その結果、決定木モデルと量子アニーリングを組み合わせた QBoost を利用することで、テストデータの決定係数 0.87 と、精度よく透過率変化幅を予測できることが明らかになった。さらに、インク組成から着色効率や着色速度の予測する機械学習モデルの構築にも成功した。最終的に、構築した複数の機械学習モデルと量子アニーリングを用いることで、インク組成の組み合わせ最適化を行い、EC インク組成を決定した。

- 1) C. Y. Jeong, T. Kubota, K. Tajima, M. Kitamura and H. Imai, *Materials Chemistry and Physics* **277** (2022).
- 2) K. Tajima, H. Watanabe, M. Nishino and T. Kawamoto, *RSC Adv* **10**, 2562-2565 (2020).
- 3) K. Tajima, C. Y. Jeong, T. Kubota, T. Ito, K. Araki, T. Kamei and M. Fukui, *Solar Energy Materials and Solar Cells* **232** (2021).

# MLIPs の特徴量を活用した NMR のケミカルシフト予測

## NMR chemical shift prediction using machine learning interatomic potentials

(株)Preferred Computational Chemistry<sup>1</sup>

○米澤 拓孝

Preferred Computational Chemistry, Inc.<sup>1</sup>

○Hirotaka Yonezawa<sup>1</sup>

E-mail: hirotakayonezawa@pfcc.co.jp

核磁気共鳴装置 (NMR) は有機分子の構造決定には欠かせない測定装置である。複雑な分子の構造決定の際には、シミュレーションで算出したスペクトルと実測データを比較することでピークを帰属することもあり、高精度なピーク予測手法の開発が長らく研究されてきた。<sup>[1]</sup>

ケミカルシフト値の予測は *ab initio* な手法とデータ駆動型の手法の双方から検討されてきた。データ駆動型の手法については、さらに原子団寄与法<sup>[2]</sup>、HOSECode<sup>[3]</sup>、機械学習的な手法<sup>[4]</sup>の3タイプに分類される。現在のところ、HOSECode と機械学習的な手法が同等の精度を達成しているが、いずれの手法においても距離が近い分子内の原子の影響しか考慮できず、対象原子周辺の局所環境の表現が十分ではないことが課題となっている。

精度向上のためには、より繊細に原子周辺の局所環境を表現する必要がある。多様なデータセットを学習した汎用的な機械学習ポテンシャル(MLIPs)は、原子周辺の局所環境を適切に表現可能と期待される。そこで、本研究では汎用的な MLIPs を特徴量の抽出器として活用することを考え、MLIPs 用いて抽出した特徴量を活用して NMR のケミカルシフト予測モデルの構築を検討した。

NMR のケミカルシフトデータは nmrshiftdb2<sup>[5]</sup> から取得し、各原子の特徴量は MLIPs の各層から抽出してデータセットを構築した。構築したデータセットと一般的な機械学習ライブラリを利用してケミカルシフト予測モデルを構築した。当日はデータセット構築・検証、特徴量の構築・分析、および機械学習モデルの比較と展望について議論する予定である。

### [参考文献]

- [1] E. Jonas, S. Kuhn, N. Schlörer, *Magn. Reson. Chem.* 2022, 60(11), 1021.
- [2] E. G. Paul, D. M. Grant, *J. Am. Chem. Soc.* 1963, 85(11), 1701.
- [3] W. Bremser, *Anal. Chim. Acta* 1978, 103(4), 355.
- [4] Y. Kwon, D. Lee, Y.-S. Choi, M. Kang, S. Kang, *J. Chem. Inf. Model.* 2020, 60(4), 2024.
- [5] S. Kuhn, N. Schlörer, *Magn. Reson. Chem.* 2015, 53(8), 582

# 機械学習分子動力学シミュレーションによる 高分子電解質膜内のプロトン輸送の解析

## Analysis of Proton Transport in Polymer Electrolyte Membranes

### Using Machine-Learning Molecular Dynamics Simulations

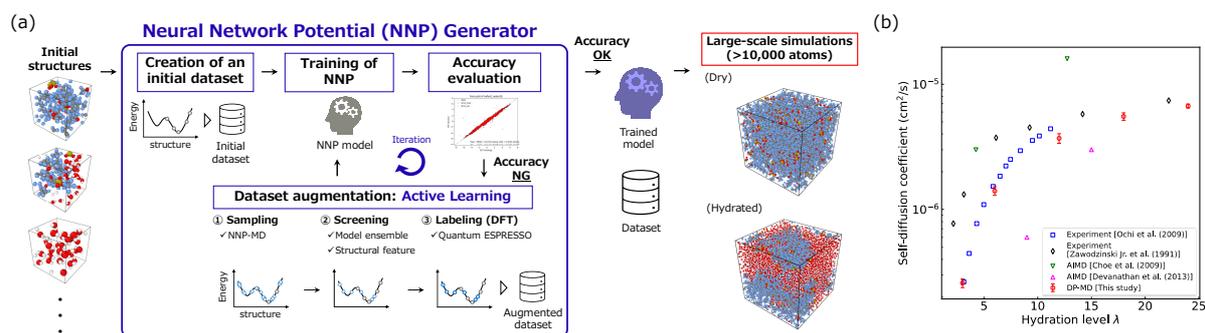
富士通株式会社<sup>1</sup> °吉本 勇太<sup>1</sup>, 松村 直樹<sup>1</sup>, 岩崎 有登<sup>1</sup>, 中尾 宏<sup>1</sup>, 坂井 靖文<sup>1</sup>

Fujitsu Ltd.<sup>1</sup>, °Yuta Yoshimoto<sup>1</sup>, Naoki Matsumura<sup>1</sup>, Yuto Iwasaki<sup>1</sup>, Hiroshi Nakao<sup>1</sup>, Yasufumi Sakai<sup>1</sup>

E-mail: yoshimoto.yuta@fujitsu.com

密度汎関数理論 (DFT) 計算に近い精度を保ちつつ、より大規模な系の動力学解析が可能な方法として、機械学習原子間ポテンシャルを用いた分子動力学 (MD) シミュレーションが注目されている。特に、近年はニューラルネットワークポテンシャル (NNP) が活発に開発されているが、大規模・長時間 MD シミュレーションを安定に実行可能な NNP の構築方法は未だ発展途上であり、長時間スケールの動的特性等の評価が困難である。

我々は、NNP 生成に係る煩雑な手続きを自動化し、NNP を簡便に作成するためのツールを開発している [Matsumura et al., arXiv:2411.17191 (2024)]。本研究では、開発ツールを用いて、高分子電解質膜の一つであるナフィオン膜を対象とした NNP を構築し、大規模・長時間 NNP-MD シミュレーションを実施した。NNP 訓練用の小規模構造として、含水率の異なる複数のナフィオン系及び水系を使用し、能動学習を用いたデータセット拡充により NNP の高精度化を図った (図 1(a))。なお、NNP モデルには Deep Potential [Zeng et al., *J. Chem. Phys.* 159, 054801 (2023)] を使用した。最終的に得られた NNP を用いることで、10,000 原子以上から成る大規模ナフィオン系の NNP-MD シミュレーションを 30 ns 以上の長時間にわたって安定に実行することに成功した。また、我々のシミュレーションでは、幅広い含水率にわたって、水素原子の自己拡散係数を従来の第一原理 MD シミュレーションよりも高精度に再現できることを確認した (図 1(b))。



**Figure 1.** (a) A workflow for constructing an NNP for Nafion systems. Small systems of Nafion with varied hydration levels and pure water (<300 atoms) are used as training structures. The NNP is automatically improved via active learning implemented in our NNP generator. The resultant NNP is used for large-scale (>10,000 atoms) and long-time (>30 ns) MD simulations of Nafion systems. (b) The self-diffusion coefficients of hydrogen atoms in the hydrated Nafion systems.

## 分子計算と機械学習による Si 表面に対する低吸着エネルギー分子のデータ駆動型探索

## Data-Driven Method for Discovering Low Adsorption Energy Molecules on Si Surfaces

SCREEN ホールディングス<sup>1</sup>, システム計画研究所<sup>2</sup>, 統計数理研究所<sup>3</sup>○村島 広衛<sup>1</sup>, 魚波 公希<sup>1</sup>, 國枝 省吾<sup>1</sup>, 埴 洋祐<sup>1</sup>, 佐々木 悠太<sup>1</sup>, 上島 仁<sup>2</sup>, 新谷 俊了<sup>2</sup>, 吉田 亮<sup>3</sup>SCREEN Holdings Co., Ltd.<sup>1</sup>, Research Institute of Systems Planning, Inc.<sup>2</sup>, The Institute of Statistical Mathematics<sup>3</sup>, ○Hiromori Murashima<sup>1</sup>, Koki Uonami<sup>1</sup>, Shogo Kunieda<sup>1</sup>, Yosuke Hanawa<sup>1</sup>, Yuta Sasaki<sup>1</sup>, Hitoshi Kamijima<sup>2</sup>, Toshiaki Shintani<sup>2</sup>, Ryo Yoshida<sup>3</sup>

E-mail: h.murashima@screen.co.jp

近年、半導体デバイスの微細化や三次元化が進み、半導体洗浄プロセスの乾燥工程にて微細パターンが倒壊する問題が深刻化している [1]。パターン倒壊はパターン間に充填された液体を固化・昇華させる昇華乾燥法により抑制できることが知られている [2]。本手法の課題の一つとして有機残渣があり、膨大な数に上る候補材料から効率的な低残渣材料探索手法の開発が求められている。広大な材料空間の探索手法として、ベイズ最適化 (BO) による逐次実験計画法とシミュレーションでの物性計算の融合による、材料科学の事前知識を活用しないデータ駆動型探索手法が提案されている [3]。本研究では、データ駆動型探索の有効性確認と低残渣昇華剤候補の選定を目的とし、吸着エネルギー値を予測するサロゲートモデルとシミュレーションでの物性計算を用いた自動探索手法から Si 基板表面に対する低吸着エネルギー分子の探索を行った。

化学分子データベースから収集した 7 万種の候補材料群から、疎水性表面の Si 平板上の吸着エネルギー値をシミュレーションにより計算しデータセットを構築した。データセットから 2 値分類モデルとサロゲートモデルを作成し、分類モデルにて候補材料を絞り込んだ後に、吸着エネルギー値の回帰と次の計算分子の選定、シミュレーションによる物性計算のサイクルを繰り返し探索した。サロゲートモデルとしてガウス過程回帰 (GPR) と Random Forest (RF) を使用し、GPR では獲得関数 (Expected Improvement) の最大値、RF では候補空間内の最小予測値を基に、次の計算分子を選定した。

GPR および RF を用いることで、探索サイクルの初期から無作為抽出を大幅に上回る効率で候補材を探索出来ることが確認された (Fig)。RF が GPR よりも効率的に探索が行えたのは初期学習データにベンチマーク値を超える材料が十分に存在するためであると考えられる。一方、ベンチマーク値をより厳しい条件に設定するにつれて GPR の方が RF よりも効果的に探索でき、初期学習データを超える候補材の探索には GPR が有効であることが確認された。

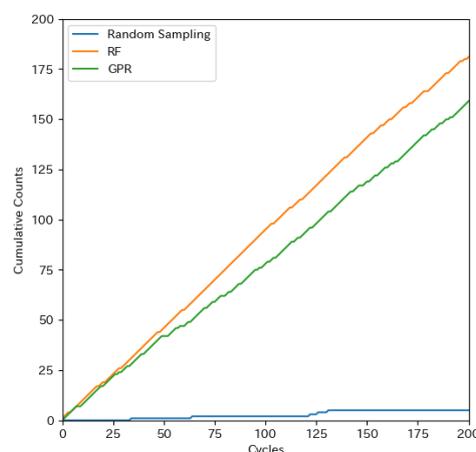
[1] D. Chandra, *Acc. Chem. Res.* 2010, **43**, 8, 1080–1091[2] Y. Sasaki, *et al.*, *ACS Appl. Nano Mater.* 2022, **5**, 7, 9495–9502[3] S. Nanjo, *et al.*, arXiv:2408.05135

Fig. Cumulative Counts versus the number of search cycles. Cumulative Counts represents the number of candidate materials exceeding the benchmark value.

## イオン伝導体 CIF ファイルの違いがパーシステント図に与える影響

### The Influence of Difference in CIF Files of Ion Conductors on the Persistent Diagram

NIMS °吉武 道子, 長田 貴弘

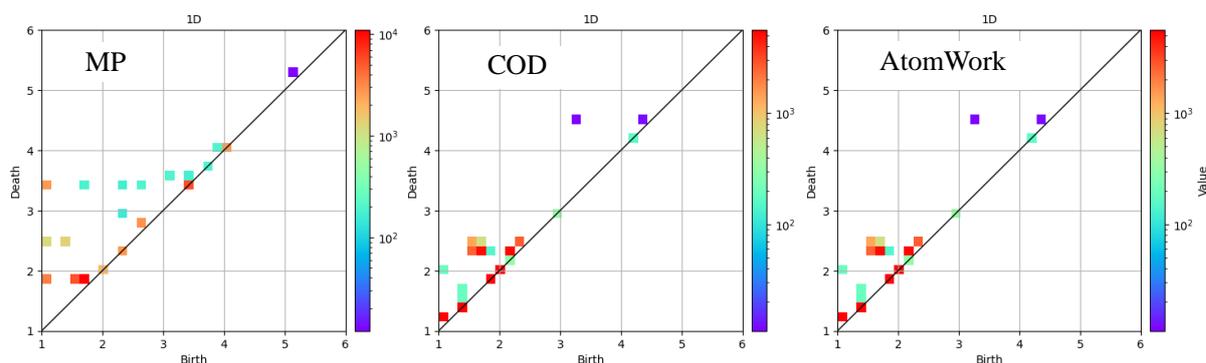
NIMS °Michiko Yoshitake, Takahiro Nagata

E-mail: yoshitake.michiko@nims.go.jp

全固体電池用の電解質として、高いリチウムイオン伝導率を示す材料の探索が行われてきた。多くの研究により、リチウムイオン拡散の活性化エネルギーが低くなる拡散経路を形成することが重要であることが知られている。先に、このような拡散経路を持つかどうかを第一原理計算や MD シミュレーションなどを使わずに簡便に知る方法として、パーシステントホモロジーの適用を検討し、マテリアルプロジェクト (MP) の CIF ファイルから作成したパーシステント図 (PD) を用いて、多形のうちの空間群の結晶がイオン伝導体として良さそうかが予測できた[1]。

本来の目的では、第一原理計算が行われていない系について PD を作成する。そこで今回、実験的に得られている CIF ファイルから作成した PD と MP の CIF ファイルから作成した PD を比較検討した。実験的に得られている CIF ファイルが無料でダウンロードできる代表的サイトとして Crystallography Open Database (COD) [2]と AtomWork [3] があり、先に検討した結晶構造の MP の CIF ファイルとこれらの実験 CIF ファイルによる PD を比較した。なお、PD 作成は HomCloud [4] を用い、原子半径の違いは考慮していない。

Li10Ge(PS6)2 では、実験で得られている CIF は P42/nmc, 137 (COD と AtomWork)なのに対し MP は P4\_2mc, 105 であるがその PD は非常に似ていた。一方 Li5La3Ta2O12 では、実験も MP も Ia-3d, 230 であるにも関わらず、下図のように MP の PD と実験 (COD&AtomWork) の PD はかなり異なっていた。実験 CIF ファイル間においても CIF ファイルに違いがあり、異なる PD 図が得られる場合もある。CIF ファイルの内容に注意が必要である。



[1] 吉武道子, 長田 貴弘 : 2024 年第 85 回応用物理学会秋季学術講演会 17p-P04-8.

[2] <https://www.crystallography.net/cod/browse.html>

[3] <https://crystdb.nims.go.jp/crystdb/search-materials>

[4] Ippei Obayashi, Takenobu Nakamura, Yasuaki Hiraoka, J. Phys. Soc. Jpn. 91, 091013 (2022): <https://homcloud.dev/index.html>

## 高密度シリカガラスの構造ユニットの定量的な形状解析

## Quantitative Order Analysis Based on Structural Units in Densified Silica Glass

東北大院情報<sup>1</sup>, NIMS<sup>2</sup>, 理研<sup>3</sup> ◦嶋 佑泰<sup>1</sup>, 志賀 元紀<sup>1, 2, 3</sup>Tohoku Univ.<sup>1</sup>, NIMS.<sup>2</sup>, RIKEN<sup>3</sup> ◦Yudai Shima<sup>1</sup>, Motoki Shiga<sup>2, 1, 3</sup>

E-mail: shima.yudai.q7@dc.tohoku.ac.jp

ガラス材料の構造は結晶とは異なり、並進・回転操作に不変でなく対称性をもたないものの、完全にランダムな原子配置にならない。しかしながら、シリカガラス等では化学結合長を超えた距離での構造秩序（中距離構造秩序）の存在が回折実験で観測され、その秩序の起源となる構造ユニットに関して多くの議論がなされている。この課題に対して、原子間の空隙や穴を特徴付けるパーシステントホモロジーによる解析[1]や化学結合からなるネットワークに内在するリング（閉ループ）の形状を解析するリング解析[2]などによって、構造秩序を解析する研究が取り組まれてきた。さらに、そのために sovary [3]などの網羅解析可能なソフトウェアが開発されてきた。一方、極限環境下でのガラス構造の秩序を評価する研究もなされている。例えば、高圧化・高温化で高密度化したシリカガラスに対して、回折実験で計測される構造因子において FSDP (First Sharp Diffraction Peak)が通常よりも高く、構造秩序のあるものが発見されている[4]。この結果から、合成条件ごとに構造秩序が変動する様子を網羅的に解析することは、中距離秩序の理解のために重要な課題と考えられる。

本研究では、まず、様々な圧力下 (0~175GPa) で合成されたシリカガラスの構造モデル[5]に対して、化学結合からなるリングを網羅的に列挙した。そして、得られたリングのペアの中で、共有ノード（原子）を持たずに絡み合っているものを自動抽出し、その形状を定量評価する試みを行った。この解析の結果、およそ 20~40GPa の間で多くの絡み合ったリングが存在することを確認した。また、含まれるリングの多くはリング形状指標である丸みが大きく、同時に、リングの形状指標の粗さが小さくなっていた。したがって、絡まっているリングは、真円に近くかつ平らな形状のものが含まれていることがわかった。また、シリカの結晶に関しても同様の解析を行なったところ、Coosite のみが絡み合ったリングを有することがわかった。得られたリングも同様に平らで真円に近いものが多かった。さらに、比較的小きなリングである Guttman リング[6]を列挙した場合でも、結晶の中で唯一 Coosite において絡み合ったリングを抽出することができた。得られたリングは Coosite で確認可能な他のリングに比べ、より真円に近くより平らであるものであることがわかった。

(1) Y. Hiraoka, et al., *Proc. Natl. Acad. Sci. USA*, 113 7035–7040 (2016).(2) M. Shiga, et al., *Commun. Mater.*, 4, 91 (2023).(3) <https://github.com/MotokiShiga/sova-cui>(4) Y. Onodera, et al., *NPG Asia Mater.*, 12, 85 (2020).(5) L.C. Erhard, et al., *Nat. Commun.*, 15, 1927 (2024).(6) Guttman. L, et al., *J. Non-Cryst. Solids*, 116, 145–147 (1990).

## 二段階 LASSO 回帰による多元素ドーピングヘマタイト光電極の光電流予測 Photocurrent Prediction of Multi-Element-Doped Hematite Photoelectrodes by Two-Step LASSO Regression

神戸大院理<sup>1</sup>, 神戸大分子フォト<sup>2</sup>, 奈良先端大物質<sup>3</sup>, 奈良先端大 DSC<sup>4</sup>

○(D)西村 拓真<sup>1</sup>, 隈部 佳孝<sup>1,2</sup>, 原嶋 庸介<sup>3,4</sup>, 藤井 幹也<sup>3,4</sup>, 立川 貴士<sup>1,2</sup>

Grad. Sch. of Sci., Kobe Univ.<sup>1</sup>, Mol. Photosci. Res. Center, Kobe Univ.<sup>2</sup>,

MS, Nara Inst. Sci. Tech.<sup>3</sup>, DSC, Nara Inst. Sci. Tech.<sup>4</sup>

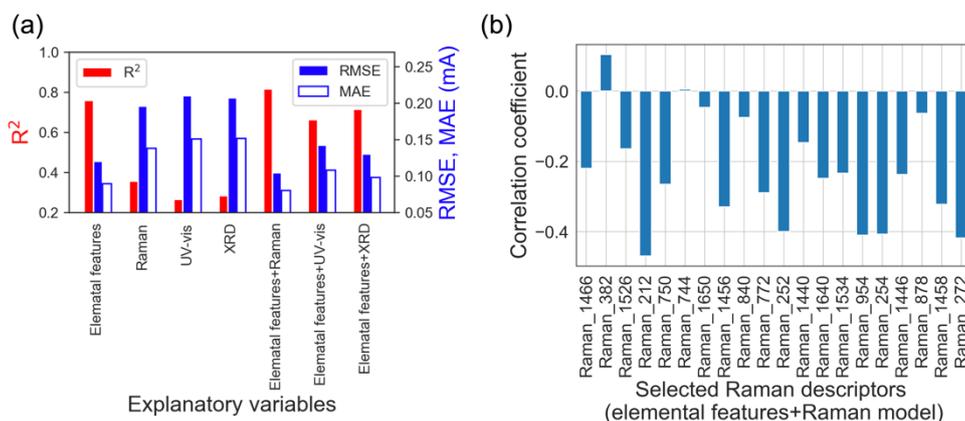
○Takuma Nishimura<sup>1</sup>, Yoshitaka Kumabe<sup>2</sup>, Yosuke Harashima<sup>3,4</sup>,

Mikiya Fujii<sup>3,4</sup>, Takashi Tachikawa<sup>1,2</sup>

E-mail: tachikawa@port.kobe-u.ac.jp

光触媒を用いた光分解反応は太陽光エネルギーからクリーン水素を生成する極めて重要なエネルギー変換反応である。赤錆として知られるヘマタイト ( $\alpha\text{-Fe}_2\text{O}_3$ ) は安定・安価、無毒な光触媒であり光電極として用いることで光電気化学的に水分解反応を進行させることができる。本研究では、多元素ドーピングヘマタイト光電極を 97 サンプル作製し、蛍光 X 線測定 (XRF) による組成式の決定を行い、さらに実験データを取得した (Raman スペクトル、紫外可視透過率測定 (UV-vis)、X 線回折測定 (XRD))。これらの実験データから水素性性能の指標である光電流を予測する機械学習モデルの作成に取り組んだ。

XRF で決定したサンプルの組成情報から作成した XenonPy 記述子と、ドーパントの結晶半径、価数、金属-酸素結合生成エネルギーを用いて作成したオリジナル記述子を合わせて元素特徴量とした。元素特徴量と実験データを組み合わせて説明変数とし、二段階 LASSO 回帰によって光電流 (@1.6 V vs. reverse hydrogen electrode (RHE)) を予測した。図 1a に様々な説明変数を用いたときの回帰モデルの予測精度を示す。単独でデータを用いたときには元素特徴量を用いたモデルが最も予測精度が高かった。これは、サンプルに含まれる元素物性に関わる記述子を用いることで光電極性能を説明できることを示唆している。また、元素特徴量と Raman スペクトルを組み合わせることで予測精度が向上したことから、サンプルの振動情報が光電極性能に影響を与えると考えられる。回帰係数が大きい波数における信号強度は目的変数に対して負に相関しているものが多いことが明らかとなった (図 1b)。この結果から、金属-酸素の分極率の変化が小さくなり、反応中心金属上にキャリアが密集することで光分解反応が進行しやすくなると考えられる。



**Figure 1 a.** Results of prediction by two-step LASSO regression using various types of descriptors as explanatory variables. **b.** A correlation coefficient of selected Raman descriptors and the target variable.

## 内殻電子励起スペクトルを条件とした拡散モデルによる 励起原子配位環境の予測

Generation of coordination environment around excitation atom by diffusion model  
conditioned on core electron excitation spectra

東大院工<sup>1</sup>, 東大生研<sup>2</sup> ◯(M1)大久保 怜<sup>1</sup>, (D1)高原 泉<sup>1</sup>, (D1)ワン ルイス<sup>1</sup>, 柴田 基洋<sup>1,2</sup>,  
溝口 照康<sup>1,2</sup>

Grad. Sch. of Eng., UTokyo<sup>1</sup>, IIS, UTokyo<sup>2</sup>, ◯Ren Okubo<sup>1</sup>, Izumi Takahara<sup>1</sup>, Louis Wong<sup>1</sup>,

Kiyou Shibata<sup>1,2</sup>, Teruyasu Mizoguchi<sup>1,2</sup>

E-mail: utokyo-okubo@g.ecc.u-tokyo.ac.jp

物質開発ではイメージングや回折、分光など実験的に得られる計測データをもとに、結晶構造や結晶欠陥、局所的な組成や電子構造などに関する分析が行われている。特に、透過型電子顕微鏡やシンクロトロンで計測される電子線エネルギー損失分光と X 線吸収分光は構造に関する情報を豊富に含む。その中でも特に、吸収端近傍微細構造(ELNES/XANES)は内殻準位の電子の非占有軌道への励起に由来し、励起原子の配位環境や化学結合に関する情報を含む。ELNES/XANES からの構造情報抽出には、参照物質のスペクトル形状との定性的な比較を通して、構造情報を抽出する指紋照合法が用いられている。しかし、近年では分析手法の発展に伴い膨大なスペクトルデータを得られるようになり、指紋照合法による個別の情報抽出は現実的ではなく、スペクトル解析の自動化・効率化が求められている。近年では Lee ら[1]が機械学習を用いて粉末 X 線回折パターンから多相無機化合物の相同定を行うなど、情報科学的手法によって実験データを自動・自律的に解析し、物質開発に有益な情報を得る手法の開発が進められている。特にコンピュータビジョンや自然言語処理など広い分野で応用が進んでいる拡散モデルは生成モデルの一種であり、変分自己符合化器や敵対的生成ネットワークといった他の手法と比較して、拡散モデルは高い学習安定性や高精度な条件付け生成が可能であるといった優位性を持つため、分子構造の生成にも利用されている。実際に、Hoogeboom ら[2]及び Xu ら[3]は O(3)同変なグラフニューラルネットワークを用いた拡散モデルによる三次元分子座標および元素種の生成や有機分子のコンフォメーションの生成、Kwon ら[4]は XANES を条件とする拡散モデルを用いてアモルファスカーボンの原子構造予測を報告している。しかし、ELNES/XANES の由来する局所性に注目してサイトごとのスペクトルから配位構造などを拡散モデルで生成する報告はなされていない。このような背景から、本研究では ELNES/XANES を拡散モデルの条件付けにもちいて、固体における励起原子近傍の配位環境の 3 次元座標を生成する構造情報抽出手法の開発を目的とした。

Materials Project[5]に登録されている Si-O 系の各物質の非等価な O サイトを対象として平面波基底擬ポテンシャル法による第一原理計算を行って、O-K 端 ELNES/XANES スペクトルデータベースを作成した。拡散モデルを用いてスペクトルを条件とした励起 O サイト近傍の配位構造の予測を行った。拡散モデルのアーキテクチャは先行研究[2,3]を参考に 3 次元構造の生成を行う上で帰納バイアスとなる同変性を考慮したグラフニューラルネットワーク EGNN[6]を使用して作成した。励起 O サイトと二つの Si からなる配位環境を持つ Si-O-Si のグラフデータに限定した。評価には Si-O 間の結合距離および Si-O-Si からなる結合角を指標とした。

拡散モデルのアーキテクチャの検証のために行った条件なし生成においては、結合距離、および結合角が訓練データと類似した確率分布で生成された。さらに ELNES/XANES を条件に行った生成については条件として渡すスペクトルごとに正規分布から異なる 5 つの初期座標をサンプリングしてそれぞれから生成を行い、結合距離、結合角について平均をとることで生成された座標の妥当性や予測の不確かさを検討した。

当日は手法と結果の詳細について発表する。

### 参考文献

- [1] J. W. Lee, W. B. Park, J. H. Lee, S. P. Singh, and K. S. Shon Nat Commun 11, 86 (2020).
- [2] E. Hoogeboom, *et al.* PMLR, 8867-8887 (2023).
- [3] M. Xu, *et al.* arXiv preprint arXiv:2203.02923, (2022).
- [4] H. Kwon, *et al.* arXiv preprint arXiv:2312.05472, (2023).
- [5] A. Jain, *et al.* APL materials, 1.1 (2013).
- [6] V. G. Satorras, E. Hoogeboom, and M. Welling arXiv preprint arXiv:2102.09844, (2021).

# 説明可能 AI による磁気構造とエネルギーの物理的情報変換モデルの開発

## Development of a physics-informed conversion model of magnetic structure and energy using explainable AI

東京理科大学<sup>1</sup>

○(B)坪内 敢志<sup>1</sup>, 長岡 竜之輔<sup>1</sup>, 谷脇 三千輝<sup>1</sup>, 町田 陽太郎<sup>1</sup>, Lira Foggianto Alexandre<sup>1</sup>, 小嗣 真人<sup>1</sup>  
 Tokyo Univ. of Sci.<sup>1</sup>, ○(B)Kanji Tsubouchi<sup>1</sup>, Ryunosuke Nagaoka<sup>1</sup>, Michiki Taniwaki<sup>1</sup>, Yotaro Machida<sup>1</sup>,  
 Alexandre Lira Foggianto<sup>1</sup>, Masato Kotsugi<sup>1</sup>  
 Email: 8221073@ed.tus.ac.jp

ミクロな磁気構造はマクロな磁気特性を司る重要な情報源である。磁気デバイスの高度化には磁気構造の物理的メカニズムの理解が必要であり、因果関係の解析は基礎と応用の両面で重要である。しかし、磁気構造は複雑な形状を示すため、定量化は困難なのが現状である。そこで、本研究では説明可能 AI を活用し、定量的かつ解釈性の高いモデルの構築を目指す。特に機能の背後にあるエネルギーに着眼し、AutoEncoder を用いて学習モデルを構築する。

本モデルで用いたデータセットは、mumax3 による Landau-Lifshitz-Gilbert 方程式に基づくシミュレーションを用いて生成した。物理パラメータは、交換スティフネス定数、一軸磁気異方性定数、界面 DMI 定数をそれぞれ変化させ、ゼロ磁場における 30000 枚の磁気画像を取得した。さらに、学習のターゲットとして磁気画像と対応する交換エネルギーを計算した。得られた磁気画像を入力とし、出力としてエネルギー密度画像が得られるように Attention 層を組み込んだ AE モデルを構築した(Fig.1)。学習の際は、出力画像とターゲットの情報損失が最小となるように学習させた。また、特徴量で構成される潜在空間を用いて、交換エネルギーの予測を行った。

潜在空間を記述する特徴量(潜在変数)と交換エネルギーの相関を解析した結果(Fig.2a), 潜在変数が磁気構造のエネルギー傾向を捉えていることが確認された。これにより磁化情報をエネルギーに情報変換可能なことが示された。また、Attention 層における活性領域を可視化した結果、モデルは磁気構造の微細な変化を捉えることができ、磁区のエッジやカーブなどの特徴がエネルギーの高い領域として強調されていることが分かった(Fig.2b)。

したがって、エネルギーと構造の関係を定量的に説明可能な潜在空間が生成でき、その過程で重要な特徴領域を Attention 層が捉えていることが明らかとなった。

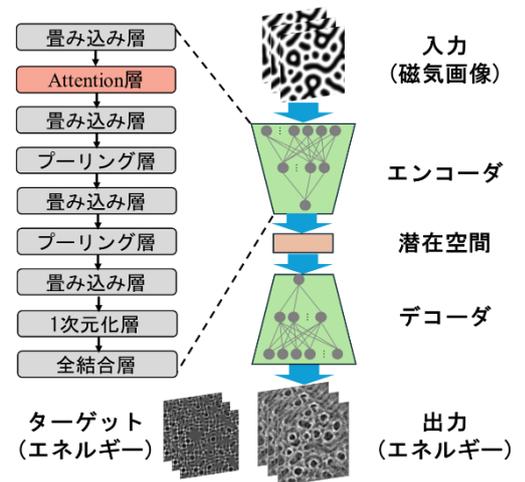


Fig.1 オートエンコーダモデル構造

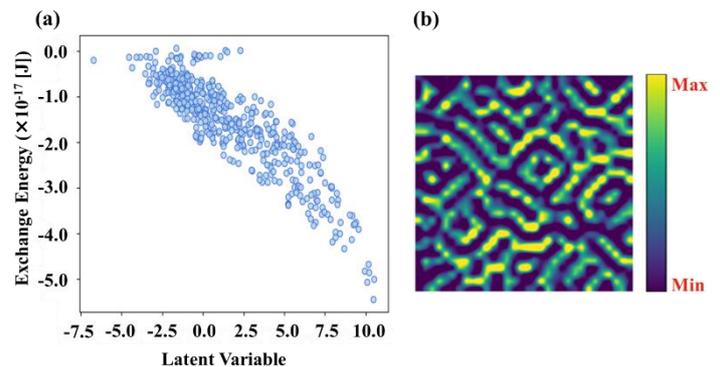


Fig.2 (a) 潜在変数と交換エネルギーの相関解析  
 (b) Attention マップ

[1] Ruoyan Feng *et al*, *J. Phys.: Condens. Matter*. 36, 1-3 (2024) [2] Kawaguchi, M. *et al*, *npj Comput. Mater.* 7, 1-6 (2021)  
 [3] A Oluwasanmi *et al*, *Sensors*, 22, 4-12 (2021)

## XANES データを用いる結晶構造予測モデル

### Crystal Structure Prediction Model from XANES spectra

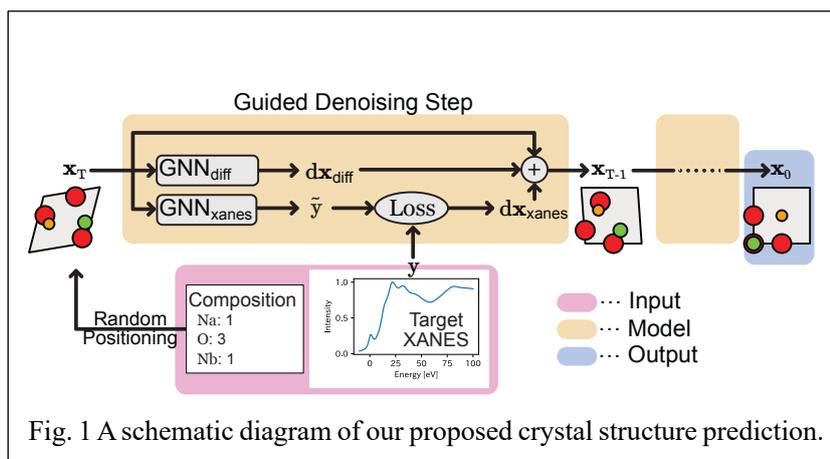
東北大<sup>1</sup>, NIMS<sup>2</sup>, 理研 AIP<sup>3</sup> 北井孝紀<sup>○1</sup>, 志賀元紀<sup>1,2,3</sup>, 二宮翔<sup>1</sup>, 西堀麻衣子<sup>1</sup>

Tohoku Univ.<sup>1</sup>, NIMS<sup>2</sup>, Riken<sup>3</sup> <sup>○</sup>Koki Kitai<sup>1</sup>, Motoki Shiga<sup>2</sup>, Kakeru Ninomiya<sup>1</sup>, Maiko Nishibori<sup>1</sup>

E-mail: k-kitai@tohoku.ac.jp

結晶構造予測 (Crystal Structure Prediction; CSP) は与えられた組成の物質が取りうる安定な結晶構造を主として計算機を用いて求める課題である。現在は進化的戦略、粒子群最適化、ベイズ最適化等のアルゴリズムによって構成原子の配置の提案を繰り返し、その中で生成エネルギーを最小化するような構造を発見する手法が一般的である。そこに近年では新たに、結晶拡散モデルと呼ばれる深層学習技術を応用する手法が提案された<sup>1</sup>。結晶拡散モデルは訓練データとして与えられた安定な結晶構造のデータ群を参考にして、新規の結晶構造をランダムに生成することができる統計モデルである。通常のモデルでは結晶構造の記述子 (格子ベクトル、組成、原子配置) もランダムに生成されるが、構造を予測したい化学組成を指定する条件付きの生成も可能である。結晶拡散モデルへは、このように補助的な情報を与えることが可能であり、例えば隣接原子間の微小な位置変化に敏感な X 線吸収端近傍構造 (X-ray Absorption Near Edge Structure; XANES) のデータを用いてアモルファス構造を生成する手法が報告されている<sup>2</sup>。

本研究では結晶の組成および XANES スペクトルを入力とし、対応する結晶構造を予測する結晶拡散モデルを提案する。訓練されたモデルで結晶構造を予測する手順を Fig. 1 に示す。モデルの訓練および検証のためのデータを、Materials Project



データベースに登録されている中の Nb 元素を含む結晶構造と、その X 線吸収スペクトルの K 端付近を理論計算することで用意した。スペクトルの計算には、第一原理計算に基づくシミュレーションソフト FDMNES を用いた。このデータセットを用いて、XANES スペクトルの情報を加えることが構造推定の精度向上に寄与することの検証を行った。また、通常 (構造情報のみに関する) の結晶拡散モデルに相当する Fig. 1 中の  $GNN_{diff}$  と表記された部分のモデルパラメータに関しては、Materials Project 由来のデータセット MP-20 を使用して事前学習を行った。これにより XANES スペクトルが得られている結晶構造のデータのみでモデルを学習する場合に比べて、過学習を抑えつつ高い予測精度が得られることを確認した。

[1] L. Jiao, et al., *Advances in Neural Information Processing Systems*, **36**, 17464–97 (2023).

[2] H. Kwon, et al., *Machine Learning: Science and Technology*, **5**, 045037 (2024).

## セグメント回帰を用いた変化点抽出と線形化

## Change point detection and linearization using segment regression

NIMS<sup>1</sup> ○柳生 進二郎<sup>1</sup>, 長田 貴弘<sup>1</sup>NIMS<sup>1</sup>, °Shinjiro Yagyu<sup>1</sup>, Takahiro Nagata<sup>1</sup>

E-mail: YAGYU.Shinjiro@nims.go.jp

近年、実験の自動化により大量の測定スペクトルが生成され、それに伴い解析の効率化が望まれている。これまで我々は、セグメント回帰（今後 PW 法と呼ぶ）を利用することで、スペクトルの特徴（変化点とその前後の傾き）を数値微分とは異なる方法で自動で捉えることができることを報告してきた[1,2]。この手法は、一連の自動自律化実験におけるスペクトル解析の自動化に利用できるだけでなく、スペクトル解析によって生じた二次加工データの解析においても有用である。PW 法はデータに対して変化点の前後を直線で近似することから、PW 法を適用する前にオリジナルデータを線形変換処理により効率的に特徴を捉えることができる。本報告では、指数関数、べき乗関数、フィードバック項を含む指数関数やべき乗関数について、代表的な線形変換方法である片対数、両対数を適用したものに対して PW 法を適用した例について報告する。また、実際に測定した Schottky 効果を持つ I-V スペクトルについて適用結果について報告する。

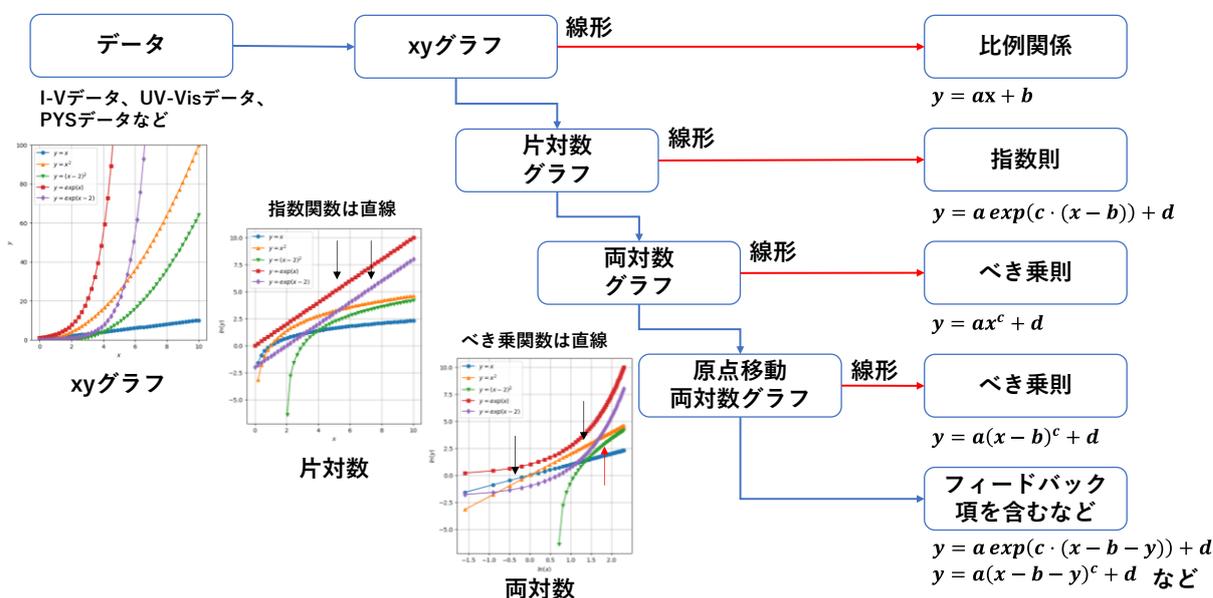


図 指数則・べき乗則と片対数と両対数による線形化

[1] 柳生進二郎、吉武道子、長田貴弘、”セグメント回帰を用いたスペクトル前処理法” 第84 回応用物理学会秋季学術講演会 20a-P04-7(2023).

[2] S. Yagyu, M. Yoshitake, T. Nagata; J. Surface Analysis, 31, 1, 2, (2024), <https://doi.org/10.1384/jsa.31.2>

## 2次元計測データによる誘電体薄膜材料探索効率化と 既存データとの連携の検討

### Improvement of Efficiency of Dielectric Thin Film Material Development by 2D Measurement Data and Study of Linkage with Existing Data

NIMS<sup>1</sup>, 村田製作所<sup>2</sup> °長田貴弘<sup>1</sup>, 木野日織<sup>1</sup>, 河野敬<sup>1</sup>, 尾崎仁亮<sup>2</sup>, 池田潤<sup>2</sup>

NIMS<sup>1</sup>, Murata co. ltd.<sup>2</sup>, °Takahiro Nagata<sup>1</sup>, Hiori Kino<sup>1</sup>, Takashi Kono<sup>1</sup>, Noriaki Ozaki<sup>2</sup>, Jun Ikeda<sup>2</sup>

E-mail: NAGATA.Takahiro@nims.go.jp

機械学習などを活用したデータ駆動による材料探索は、従来手法に比べて大幅な時間短縮が期待できる。しかし、機械学習のための材料データ集めと整理は非常に手間がかかる。これを自動化する手法の研究は進んでいるが、多様な構造情報を持ち、薄膜材料開発において物性理解に重要な手法である2次元 X 線回折手法 (2D-XRD) などの2次元計測データの自動的な解析手法の研究は途上である。これに対し我々は、2D-XRD データセットを非負値行列因子分解(Non-negative matrix factorization: NMF)で分類し、変分オートエンコーダ(Variational AutoEncoder: VAE)を用いて分類結果を視覚化する方法を提案している[1]。例えば XRD のデータセットに対し、それぞれの XRD データが持つ回折ピークを特定し、各データを分類するのは非常に手間がかかる。NMF を用いることで、XRD データを自動的に分類でき、各分類で特徴的にみられる回折ピークを知るには、分類の基準となった少数の XRD パターンを分析するだけで十分であると報告されている[2]。これらに2次元データを複数手法組み合わせた解析は薄膜成長の理解と物性制御が期待できる。本報告では高誘電体材料に注目し、多様な構造を有するペロブスカイト、積層ペロブスカイト構造の強誘電・誘電体薄膜材料を用いて2D-XRD、原子間力顕微鏡(AFM)の表面形状データとその誘電率について NMF と VAE を用いた可視化を行った[Fig. 1]。また、これらの構造と誘電率の関係性については NIMS が提供しているデータベース AtomWork-Adv.[3]のデータ群を基に作成した誘電率と結晶構造の関係性を示すマテリアルマップと比較を行った結果、基本的な構造と誘電率の変化の傾向が一致する領域もあるが、薄膜構造によって実現する組成に対する準安定構造は異なる傾向を示すことがわかった。データ数に対する信頼性の評価は必要であるが、既存データと実験データ群の相関性を用いた材料開発に展開できる可能性が示唆される結果が得られた。

[1] Akihiro Yamashita *et al.*, Science and Technology of Advanced Materials: Methods 2, 23 (2022).

[2] Christian J. Long, *et al.*, Review of Scientific Instruments 80, 103902 (2009).

[3] <https://atomwork-adv.nims.go.jp>

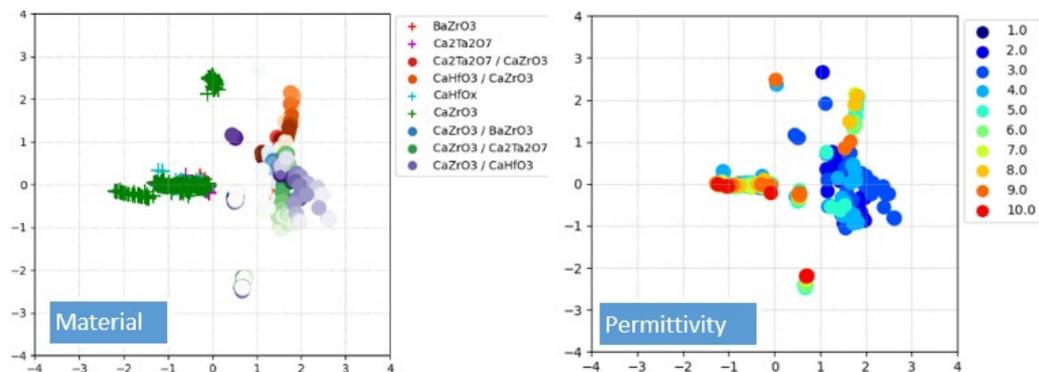


Fig. 1. The distribution of the 2D-XRD data in the latent space labeled by (a) composition and (b) permittivity of films

## TADF 材料の機械学習に向けた文献からのデータベース作成 Development of Database from Literatures for Machine Learning of TADF Materials

東理大<sup>1</sup>, 産総研<sup>2</sup>, 九大<sup>3</sup>, 京大<sup>4</sup>

(M2)°宮田 哲<sup>1,2</sup>, 椿 真史<sup>2</sup>, 中野谷 一<sup>3</sup>, 畠山 琢次<sup>4</sup>, 中山 泰生<sup>1</sup> 細貝 拓也<sup>1,2</sup>

Tokyo Univ. Sci.<sup>1</sup>, AIST<sup>2</sup>, Kyushu Univ.<sup>3</sup>, Kyoto Univ.<sup>4</sup>

°Satoshi Miyata<sup>1,2</sup>, Masashi Tsubaki<sup>2</sup>, Hajime Nakanotani<sup>3</sup>, Takuji Hatakeyama<sup>4</sup>, Yasuo Nakayama<sup>1</sup>,  
Takuya Hosokai<sup>1,2</sup>

E-mail: 7223570@ed.tus.ac.jp

TADF 材料は希少金属を使わない次世代の有機 EL 用発光材料として注目を集めている。2016 年以降、機械学習を応用して新規 TADF 材料を開発する研究事例が増えてきている。Gomez らは量子化学計算と機械学習を組み合わせ、仮想空間上に生成した 160 万個の分子の光物性スクリーニングを行い、高効率な TADF 材料の探索に成功した[1]。我々も機械学習による TADF 材料の探索を目指しており、特に分子構造データから TADF 材料の発光ダイナミクスのパラメーター(瞬時蛍光と遅延蛍光の発光寿命、およびその二つの発光の強度比)を予測する学習モデルの開発を進めている。昨年度の春季応用物理学会において、実験データの収集基盤の構築に成功したことを報告した[2]。収集した 207 種類の有機分子の発光データを使用して機械学習モデルを作成したところ、予測精度が低く満足いく性能を示さなかった。この要因の一つとしてデータ数の不足が影響していると考えている。

今回このデータ数の不足を解消する方法として、論文等の文献からのデータ収集を行った。調べる文献の対象は、2012 年に九州大学の安達らが報告した高効率 TADF 材料に関する論文の引用文献(数は 7320 件以上)に設定した[3]。

現在、我々は該当する文献のうち古いものから順に発光データの収集を進めている。この過程において、文献リストの作成や発光データの取り出し、化学構造の描画、収集したデータの整理といった工程が文献からのデータベース作成を律速させることが分かった。

本発表では、我々が行っている文献からの TADF 材料のデータ収集の取り組みについて報告する。また当日には、実験データ 207 種類に加え、文献 1000 件から取得した 900 種類以上の TADF 材料のデータを用いた機械学習の結果についても紹介する。

### 参考文献

[1] R. G. Bombarelli, et al., *Nature Materials*, 15, 1120-1127 (2016).

[2] 宮田哲、椿真史、中野谷一、畠山琢次、中山泰生、細貝拓也、「TADF 材料の機械学習に向けた実験データ収集基盤の構築」第 71 回応用物理学会春季学術講演会, 24p-P07-10 (2024).

[3] H. Uoyama, et al., *Nature*, 492, 234-238 (2012).

## LLM を使った材料開発(II)：米国特許からの NdFeB 磁石データの抽出 Materials Development with LLM (II): Extraction of NdFeB Magnet Data from US Patents

NIMS ○岡 博之, 石井 真史  
NIMS ○Hiroyuki Oka, Masashi Ishii

E-mail: OKA.Hiroyuki@nims.go.jp

### 1. 序

磁石は現代社会において必要不可欠な材料であり、特に NdFeB 磁石は高い磁気特性から部品の小型化・省エネに貢献しており、様々な分野での利用が広がっている。一方で、高温での磁気特性維持などの要求が増えており、その改良には今までの蓄積データを活用するマテリアルズ・インフォマティクス (MI) が欠かせない。MI 用のデータは大量の既報文書から集めるのが有効で、我々はその自動化を検討しており、磁気特性は米国特許の表から、磁石試料製造条件はその本文から抽出するようにしてきた<sup>[1]</sup>。自動抽出では、数値、単位、元素記号、特定の文の認識・分類が必要であり、これらの機械学習を行うようにしてきた。しかし、最近では、生成 AI と呼ばれる大規模言語モデル (LLM) が文書を読解でき、適した質問を与えれば、上記のような機械学習が無くてもデータを認識して抽出してくれる。これは新たに抽出対象を追加したいときに有効で、また、機械学習を行うよりも高精度にデータを抽出してくれる可能性もある。そこで本研究では、LLM を用いて、米国特許の本文から NdFeB 磁石の製造条件の抽出を検討した。その結果を報告する。

### 2. 抽出手順および結果

LLM には Gemma 2 を使い、使用する GPU (NVIDIA RTX A6000 (48GB)×2) を考慮して、gemma-2-9b-it のモデルを用いた。Gemma 2 の最大コンテキストサイズが 8,192 トークンであるため、入力する米国特許の内容は試料製造を記述している部分に限定した。また、試料製造は、実験を変えて、複数セクションに分けての記述が多いため、セクション単位で入力するようにした。特許内容のテキスト抽出とセクション単位への分割は[1]で報告した方法 (機械学習による文書レイアウト解析と手修正) で行った。この後、各セクションに対して、次の3つを抽出する質問を LLM に与えた：①試料製造条件、②試料化学組成、③試料製造における前述参考実験名。①では、表 1 に示す条件を抽出対象とし、左列の文字列を質問に含めておいて、該当する条件の記述があれば、その文字列をキーとして、数値、単位を JSON

形式で抽出するようにした。この後、②では、化学組成に関する記述があれば、それを抽出するようにした。③では、試料製造について、前述の実験 (セクション) と同様の方法で行った旨の記述があれば、その名前 (セクションタイトル) を抽出するようにした。抽出されれば、そこで抽出されたデータをコピーした。

表 1 抽出対象の試料製造条件

製造条件	内容
particle-size	微粉碎後の粒径
compact-press	磁界中成型での成型圧力
compact-mag-field	磁界中成型での印加磁場
sinter-temp	焼結温度
sinter-time	焼結時間
aging-1-temp	1回目の熱エージング温度
aging-1-time	1回目の熱エージング時間
aging-2-temp	2回目の熱エージング温度
aging-2-time	2回目の熱エージング時間

上記のようにして 11 特許からデータを抽出したところ、精度は f1 で 0.72 であった。[1]で報告した方法 (ルールベース手法と機械学習) では 0.53 であり、より高い精度で抽出できることが分かった。この要因の1つとして、[1]では製造条件や化学組成に関する文の分類が必要であったが、その精度は低く、一方、LLM はいくつかの文があっても、該当記述箇所を的確に認識できるため、その差が出たと考えられた。

### 3. まとめ

LLM の回答は質問表現や投入する文書の量などによって変わる。そのため、これらの検討は日々行っており、発表当日は新たに得られた結果も含めて報告する予定である。

### 謝辞

本研究は、データ創出・活用型マテリアル研究開発プロジェクト事業 (Grant Number : JPMXP1122715503) の助成によって成し得ましたので、ここに感謝の意を表します。

[1] 岡博之他, 本文からの磁石製造条件の自動抽出, 応物 85 秋, 朱鷺メッセほか 2 会場 (新潟市), 2024.9.16-20.