

2024年3月29日

東京工業大学

東北大学

理論計算と機械学習により無機材料表面の性質を 高精度かつ網羅的に予測

－光触媒材料などの探索や電子・光電子デバイスの設計を支援－

【要点】

- 最先端の理論計算と機械学習による新たな予測手法を開発
- 材料探索やデバイス設計において不可欠な指針
- マテリアルズインフォマティクスに立脚して材料開発を加速

【概要】

東京工業大学 科学技術創成研究院 フロンティア材料研究所の清原慎 JSPS 特別研究員（研究開始時。現：東北大学 助教）、大場史康教授は、産業技術総合研究所 エネルギー・環境領域 電池技術研究部門の日沼洋陽主任研究員と共同で、高精度と高速を両立させた最先端の**第一原理計算**（用語 1）により生成した大規模な理論計算データおよび機械学習を用いて、無機材料表面の基本的な電子構造を網羅的に予測することに成功した。

本研究により開発された手法は、光触媒や電子・光電子デバイスなどの設計において重要な指針を与える無機材料表面の**バンドアライメント**（用語 2）を多種多様な物質・表面を対象に予測することを可能にするだけでなく、表面以外の特性の予測にも使える汎用的なものであり、近年注目されている**マテリアルズインフォマティクス**（用語 3）に立脚した材料開発を加速することが期待される。

研究成果は 3 月 28 日（現地時間）に米国化学会誌「*Journal of the American Chemical Society*（ジャーナル オブ ジ アメリカン ケミカル ソサイエティー）」のオンライン速報版で公開された。

●背景

近年、大規模な材料データに対してデータ科学的手法を適用したアプローチであるマテリアルズインフォマティクスが注目を集めている。結晶構造などの比較的容易に入手できる情報から計測や理論計算に時間がかかる材料特性を予測することで、材料開発を加速することが可能である。本研究では、特に固体表面の性質の予測に着目している。

固体の表面では、内部とは異なる原子の配列や電子の状態により、**バルク**（用語 4）とは大きく異なる性質が発現する。特に**イオン化ポテンシャル**（Ionization Potential: IP）（用語 5）と**電子親和力**（Electron Affinity: EA）（用語 6）は半導体や絶縁体の電子状態に関する基本的な物理量であり、光触媒や電子・光電子デバイスなどを設計する際の重要な指針となる。しかしながら、固体表面には不純物が吸着しやすいため、実験により清浄な表面の IP・EA を精確に計測することは容易ではない。一方で、第一原理計算は、理想的な表面を考慮できることから、IP・EA を知るための強力なツールであるが、第一原理計算を用いて高精度に IP・EA を算出するには膨大な計算量を必要とする。さらに一つの物質においても表面はさまざまな面方位や原子配列を持ち、このような表面の多様性を考慮して網羅的に理論計算を実行することは現実的ではない。

以上のような背景をもとに、本研究では多様な固体表面を対象に IP・EA を高精度かつ高速に予測する手法の開発に取り組んだ。

●研究成果

本研究では、第一原理に基づいた最先端のハイスループット計算手法を用いて、約 3,000 種類の酸化物表面の原子位置の緩和構造と表面エネルギーおよび IP・EA を計算した。さらに、機械学習による予測モデルを用いて、表面の方位と終端面の位置の情報のみから IP・EA を高精度に予測することを可能にした。その概略を図 1 に示す。

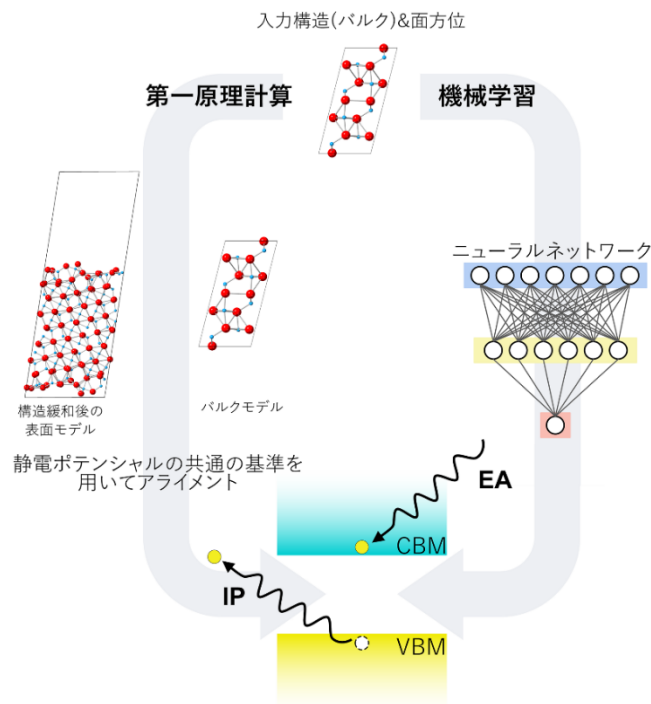


図 1 第一原理計算および機械学習を用いた IP・EA の予測の概念図

第一原理計算を用いて IP・EA を算出する際の標準的な流れは次の通りである。まずバルクの結晶構造と面方位・終端面位置の情報から表面のモデルを作成し、表面エネルギーが低下するように原子位置を緩和させる。一方で、バルクのモデルから静電ポテンシャルの基準に対する**価電子帯上端** (Valence Band Maximum: VBM) と**伝導帯下端** (Conduction Band Minimum: CBM) (用語 7) のエネルギーを計算し、表面モデルの結果と組み合わせることで IP・EA を求める。このような電子のエネルギーを高精度に算出するには、**ハイブリッド汎関数** (用語 8) などの高いレベルの近似を用いる必要があり、また表面モデルが多く原子を含むため、両面から非常に大きな計算コストを要する。

本研究では高精度と高速を両立した第一原理計算手法を用いて、まず約 2,200 種類の二元系酸化物無極性表面のデータベースを構築した (図 2 上)。IP・EA の実験値が報告されている酸化物表面を対象に理論計算値と実験値 (文献値) を比較したところ、よく一致していることを確認した (図 2 下)。これほどの高精度かつ大規模な表面特性の第一原理計算データベースの構築は他に類を見ない。

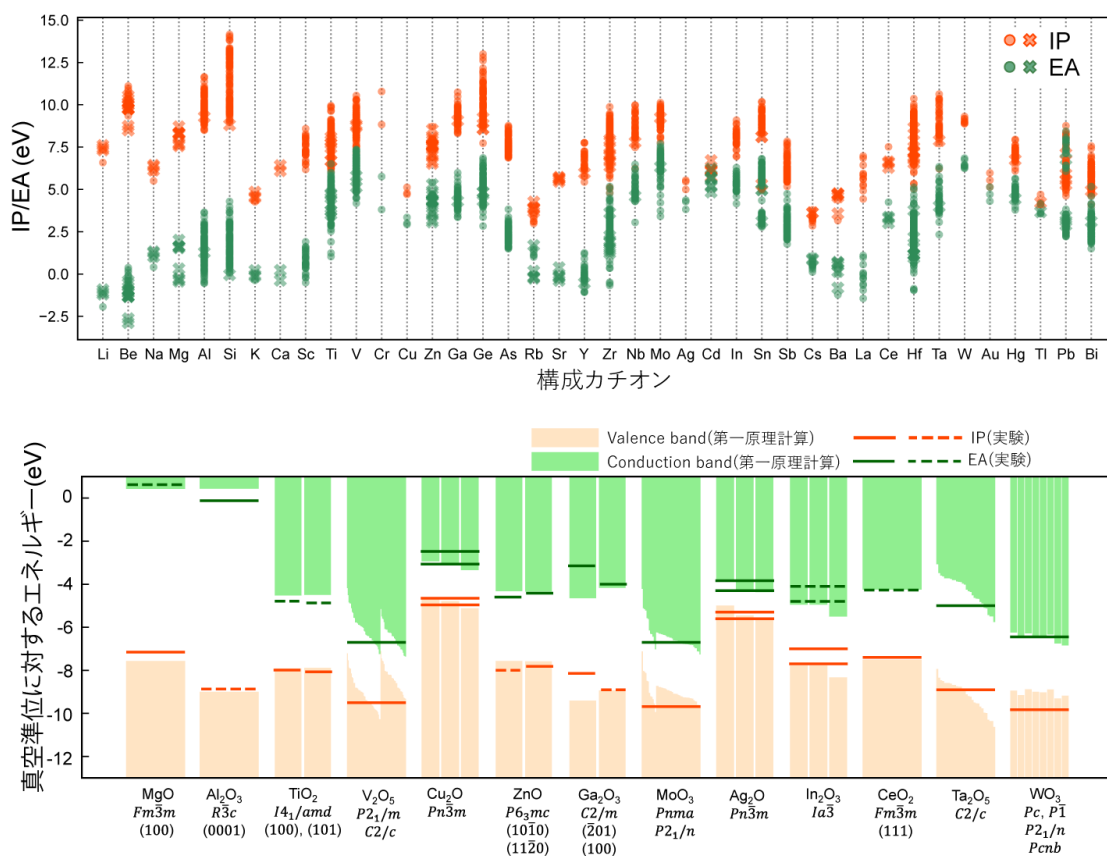


図 2 ハイスループット第一原理計算により得られた約 2,200 種類の二元系酸化物表面の IP・EA (上) と実験値 (文献値) との比較 (下)。薄オレンジ色のバーの上端・薄緑色のバーの下端がそれぞれ真空準位に対する VBM・CBM の第一原理計算値であり、IP・EA に対応する。

次に、上述の二元系酸化物表面データベースを用いて、構造緩和前の表面原子配列から構造緩和後の IP・EA を予測するニューラルネットワークを構築した。原子配列の記述子として、ニューラルネットワークに接続することで構成元素の数に対してスケールブルになるように拡張した **Smooth Overlap Atomic Positions (SOAP)** (用語 9) を開発し (Learnable SOAP: L-SOAP)、原子配列をベクトル化した。図 3 左に示すように、第一原理計算値とニューラルネットワークによる予測値を比較した際、**訓練データ・テストデータ** (用語 10) の多くの点に対角線に近い位置にあり、IP・EA の第一原理計算値を高精度に予測可能なことが分かる。さらに本手法は**アテンション層** (用語 11) を導入しているため、どの原子が IP・EA に大きな影響を与えるのかを推定可能である。例として Sb_2O_3 の(001)表面における影響の大きさを図 3 右に示す。各原子の IP と EA への影響は大きく異なることが分かる。

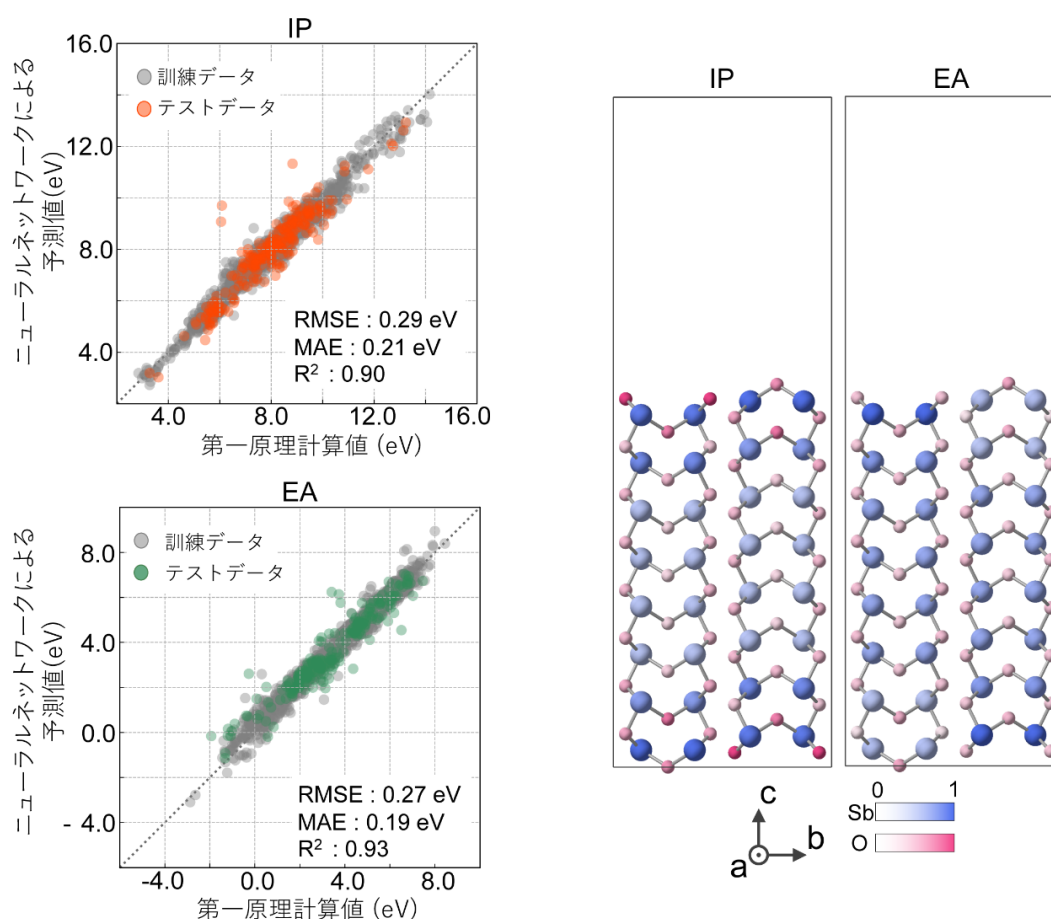


図 3 ニューラルネットワークを用いた二元系酸化物の IP・EA の予測 (左)。図中に**二乗平均平方根誤差 (Root-Mean-Square Error: RMSE)**、**平均絶対誤差 (Mean Absolute Error: MAE)**、**決定係数 (R^2)** (用語 12) の値を示す。表面モデル内の各原子の IP・EA への影響の大きさ (右)。 Sb_2O_3 の(001)表面の例であり、色が濃いほど影響が大きいことを表している。

さらに、三元系酸化物表面への展開を行った。三元系酸化物は多くの場合、二元系酸化物より複雑な結晶構造を持つことから、その表面について大規模な理論計算データを生成することは困難である。このため、本研究では約 700 種類の三元系酸化物無極性表面の理論計算データを用意し、二元系酸化物表面について構築したニューラルネットワークをベースとして**転移学習**（用語 13）を行った。その結果の一例として、EA の予測精度を図 4 に示す。訓練データの割合が増えるにつれて誤差が小さくなり、決定係数が 1 に近づいている。すなわち、予測精度が向上していくことが分かり、今後、三元系酸化物表面の理論計算データが増えることで、さらなる精度改善が期待できる。また、通常の SOAP を用いた場合と比べると、L-SOAP を用いて転移学習を行う方がはるかに予測精度が高く、本研究で開発した L-SOAP は転移学習に向いていると言える。

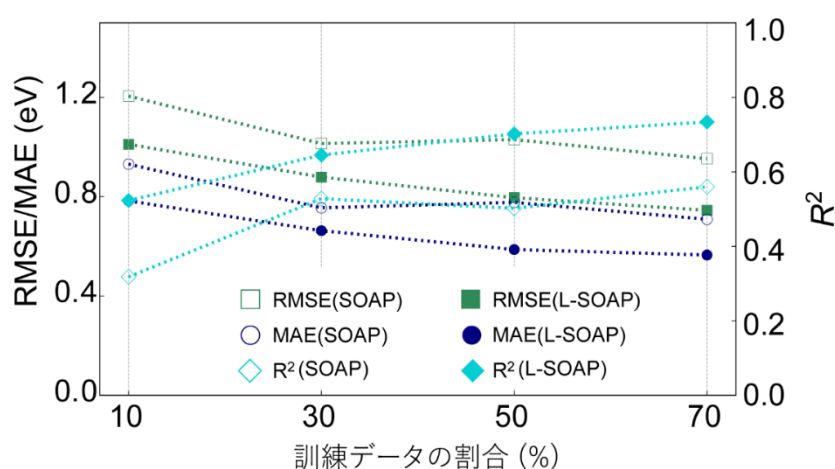


図 4 EA に関する三元系酸化物表面への転移学習の精度。二乗平均平方根誤差 (Root-Mean-Square Error: RMSE) と平均絶対誤差 (Mean Absolute Error: MAE) は左の軸、決定係数 (R^2) は右の軸を参照している。

●社会的インパクト

無機材料のバンドアライメントは、光触媒や電子・光電子デバイスなどの設計において不可欠な指針を与えるが、多種多様な物質の表面を対象に系統的に評価することはこれまで困難であった。本研究は、近年注目されているマテリアルズインフォマティクス的アプローチにより、この状況を打開するものであり、今後の材料探索やデバイス設計を加速することが期待できる。

●今後の展開

本研究において開発した手法をバンドアライメントの観点での酸化物材料のスクリーニングに応用していくとともに、硫化物・窒化物などの他の物質系の表面特性の予測へと展開する。さらに本手法の適用範囲は表面に限られないため、表面以外の特性の予測に応用する。

●付記

本研究は日本学術振興会 科学研究費助成事業 (20J00773、20H00302、23K13811)、神奈川県立産業技術総合研究所 脱炭素化対策事業、文部科学省 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 (JPMXP1122683430)、科学技術振興機構 戦略的創造研究推進事業 CREST (JPMJCR17J2) の助成を受けて行われた。また、京都大学 学術情報メディアセンターおよび九州大学 情報基盤研究開発センターの計算資源を使用した。

【用語説明】

- (1) **第一原理計算**：量子力学の基本原則に基づいた理論計算。物質の性質を支配する電子の状態だけでなく、安定性や構造を決定する際の指標となる全エネルギーが得られ、結晶・分子、表面・界面などの構造を予測できる。
- (2) **バンドアライメント**：複数の物質の電子のエネルギーバンドをある基準でそろえること。光触媒や電子・光電子デバイスなどの設計において重要な指標となる。
- (3) **マテリアルズインフォマティクス**：実験および理論計算の結果に対してデータ科学的手法を適用することで、膨大な種類の材料やその性質の予測・解析を行い、材料の最適化や新材料・新機能の開拓を行うアプローチ。
- (4) **バルク**：ここでは格子欠陥が入っていない理想的な結晶のこと。
- (5) **イオン化ポテンシャル**：半導体・絶縁体（および原子・分子）から電子を1個取り去るのに要するエネルギー。電子が固体から出る際には必ず表面を通るため、その値は表面の原子配列や電子状態に依存する。
- (6) **電子親和力**：半導体・絶縁体（および原子・分子）に電子を1個与える際のエネルギーの利得。イオン化ポテンシャルと同様に固体表面の原子配列や電子状態に依存する。
- (7) **価電子帯・伝導帯**：半導体・絶縁体において、価電子が詰まっている・空いているエネルギーバンド。VBM・CBM はその最大・最小のエネルギーにそれぞれ対応する。
- (8) **ハイブリッド汎関数**：密度汎関数理論をベースとした第一原理計算における交換相関項の近似の一つ。交換項をハートリー=フォック理論の厳密なものに部分的に置き換えることで計算精度を向上させる。
- (9) **Smooth Overlap Atomic Positions (SOAP)**：固体や分子の原子配列をベクトル化して表現する手法の一つ。
- (10) **訓練データ・テストデータ**：訓練データは、機械学習モデルの係数を調整するためのデータ。構築したモデルの精度は、学習に使用しなかったテストデータにより評価する。
- (11) **アテンション層**：ニューラルネットワークに関する技術の一つ。目的の値に対して、入力情報の重要度を自動的に決定してくれる。
- (12) **二乗平均平方根誤差 (Root-Mean-Square Error: RMSE) ・ 平均絶対誤差 (Mean Absolute Error: MAE) ・ 決定係数 (R^2)**：予測精度を測る代表的な指標。

(13) **転移学習**：機械学習手法の一つ。あるデータで一度予測モデルを構築し、他のデータを用いて予測モデルの再学習を行うこと。

【論文情報】

掲載誌：*Journal of the American Chemical Society*

論文タイトル：Band Alignment of Oxides by Learnable Structural-Descriptor-Aided Neural Network and Transfer Learning (学習可能な構造記述子を用いたニューラルネットワークと転移学習による酸化物のバンドアライメント)

著者：Shin Kiyohara, Yoyo Hinuma, and Fumiyasu Oba (清原慎、日沼洋陽、大場史康)

DOI：10.1021/jacs.3c13574

【問い合わせ先】

〈理論計算に関すること〉

東京工業大学 科学技術創成研究院 フロンティア材料研究所 教授

大場史康

Email: oba@msl.titech.ac.jp

〈機械学習に関すること〉

東北大学 金属材料研究所 助教

清原慎

Email: sin@tohoku.ac.jp

【取材申し込み先】

東京工業大学 総務部 広報課

Email: media@jim.titech.ac.jp

TEL: 03-5734-2975 FAX: 03-5734-3661

東北大学 金属材料研究所 情報企画室広報班

Email: press.imr@grp.tohoku.ac.jp

TEL: 022-215-2144