

第一原理計算は、物質や材料のエネルギーや基礎的な物性、電子状態を得るための手段として不可欠であるだけでなく、より高度な計算である材料スクリーニング、結晶構造探索、未知物質予測、熱力学計算、分子動力学計算などを行う手段としても一般的になりつつある。しかし、このような高度な計算は、網羅的

な第一原理計算を必要とするため、実行可能な対象は限定的であるというのが現状である。本稿では、機械学習手法や効率的データ構造を導入することにより、第一原理計算に基づいた高度な計算を効率的に実施し、それらの応用範囲（原子数、時間範囲、自由度）を大幅に拡大するための手法について紹介する。

1. まえがき

第一原理計算は、量子力学に基づき、系の元素と結晶構造の情報のみからエネルギーや電子状態の高精度な計算が可能である。そのため、材料研究において、エネルギーや基礎的な物性、電子・原子レベルでの現象の解釈を得る手段として不可欠なものとなっている。また、近年の計算機や計算手法の進歩により、網羅的な第一原理計算に基づいた材料スクリーニング、結晶構造探索、熱力学計算、格子振動計算、分子動力学計算など多種多様な応用計算が一般的な手段になってきている。このような計算は、基礎的な第一原理計算と比べると、計算に必要となるコストは膨大で、実行可能な原子数、時間範囲、自由度は非常に限定的である。

そのような背景の中、第一原理的な高度な応用計算を高速化させる手段として、筆者は、機械学習やアルゴリズム論的手法を導入してきた。材料科学において、最小二乗法を含む回帰分析や行列分解など、機械学習は多岐にわたって古くから利用されており、すでになじみ深いものである。それと同時に、近年ではより高度な機械学習手法が数多く提案されており、これらも第一原理的な応用計算を高速化させる手段として有用である。そこで本稿では、第一原理計算と機械学習に基づいた高精度原子間ポテンシャル（機械学習ポテンシャル）の構築、機械学習ポテンシャルによる物性予測および大域的構造探索、構造列挙に基づいたベイズ最適化による多成分系構造探索、効率的データ構造を用いた構造列挙手法について紹介する。

2. 機械学習ポテンシャルとその応用

2.1 機械学習ポテンシャルの構築

機械学習ポテンシャルとは、与えられた結晶構造集合に対する網羅的な第一原理計算の結果に基づき、機械学習手法により原子間相互作用を推定するものである。機械学習ポテンシャルでは、隣接原子分布を表現する「構造特徴量」を定

義し、結晶構造のポテンシャルエネルギーと構造特徴量の関係をガウス過程モデル¹⁾、ニューラルネットワークモデル²⁾、多項式モデル^{3~6)}などの機械学習モデルにより記述する。このような機械学習ポテンシャルは、多数の構造特徴量と柔軟な機械学習モデルを用いるため、第一原理計算に近い精度での予測が、経験的ポテンシャルと同じような計算コストにより可能となる。これまで実行不可能であった探索範囲の広い結晶構造探索や高精度な大規模分子動力学計算などが可能となるため、機械学習ポテンシャルを構築するための手法開発や機械学習ポテンシャルの応用が、世界中で活発に行われている。

本稿で紹介する機械学習ポテンシャルでは、球面調和関数に基づいた多項式回転不変量を群論的手法により系統的に生成し、構造特徴量として用いている^{6,7)}。また、ポテンシャルエネルギーと構造特徴量の関係を多項式により記述している。それぞれの系について、ICSD (Inorganic Crystal Structure Database) のプロトタイプ構造から生成された1~3万程度の結晶構造に対して第一原理計算を行い、得られたエネルギーや原子に働く力に基づき、線形リッジ回帰を用いてポテンシャルエネルギーモデルを推定した。

それぞれの系において最適なポテンシャルを構築するため、カットオフ半径、構造特徴量の数、多項式回転不変量の次数などを変化させたさまざまなモデルを用いてポテンシャルを構築した。これらの予測精度と計算コストはトレードオフの関係にある。よって、最適なポテンシャルを探すことは、多目的最適化であり、最適なポテンシャルはパレート最適なポテンシャルの集合となる。

ポテンシャル構築に関する結果の一例として、図1にアルミニウム (Al) におけるさまざまなポテンシャルモデルの予測誤差と計算コストの関係を示す。図1より、パレート最適なポテンシャルよりも、予測精度と計算コストの両方が向上するポテンシャルは存在しないことがわかる。また、Alにおいては、予

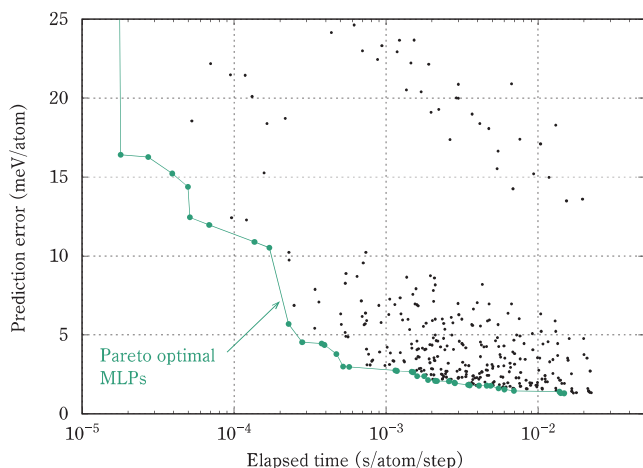


図1 アルミニウム (Al) における機械学習ポテンシャルの予測誤差と計算コストの関係。横軸は、エネルギーと原子に働く力の計算に必要な1原子1ステップ当たりの時間。緑丸はパレート最適なポテンシャル。

測誤差が非常に小さい3 (meV/atom) 以下となるポテンシャルが複数存在し、これらポテンシャルのエネルギーや力の計算に必要な時間は、 $10^{-4} \sim 10^{-2}$ (s/atom/step) 程度である。広く利用されている Embedded Atom Method (EAM) ポテンシャルと比べると計算コストは10~1000倍程度高くなるものの、高精度予測が可能な構造や物性は大幅に増加する。

このような方法で網羅的に構築されたパレート最適な機械学習ポテンシャルは、現在51種の単体や合金について、ウェブサイトにて公開中 (Machine Learning Potential Repository at Kyoto University)^{8,9)} であり、継続的にポテンシャルが追加される予定である。これらの機械学習ポテンシャルは、世界中で広く用いられている分子動力学計算ソフトウェアLAMMPSにて利用可能な形式であり、LAMMPSにて用いるためのパッケージも併せて公開している⁹⁾。また、それぞれのパレート最適なポテンシャルにより予測された多様な結晶構造に対する凝集エネルギー、エネルギー体積曲線、フォノン状態密度、弾性定数、格子定数などが併せて閲覧できるようになっており、パレート最適なポテンシャルの中から、目的に応じて適切なポテンシャルを選択することが可能である。

2.2 機械学習ポテンシャルを用いた大規模計算

機械学習ポテンシャルの最もわかりやすい応用は、第一原理計算が不可能であるような大規模構造での高精度計算、および分子動力学計算などの統計力学計算である。EAMなどの経験的ポテンシャルと比べると、計算可能な原子数やステップ数は少し減少するものの、一般的な第一原理計算では到底不可能である大規模な構造での計算が、第一原理計算と同じような精度で実行可能である。例えば、図2(a)は面心立方構造の金属において2112原子までの粒界構造モデルを考え、粒界構造および粒界エネルギーの予測を行ったものである¹⁰⁾。それぞれの粒界構造モデルについて、さまざまな剛体変位を与えることにより、粒界モデルの最適化を行った。機械学習ポテンシャルは粒界構造データを一切用いず構築した

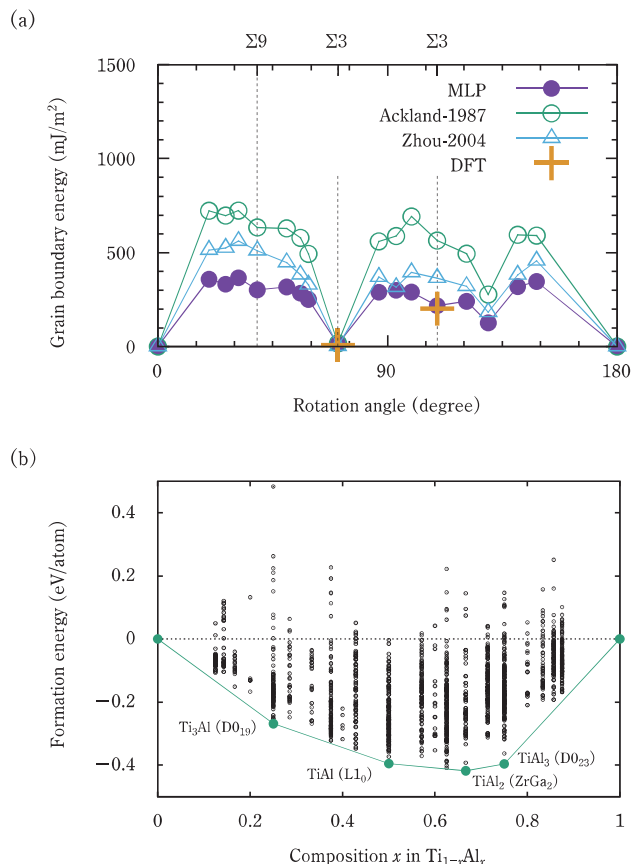


図2 (a) 金 (Au) における $\langle 110 \rangle$ 対称傾角粒界の過剰エネルギー (MLP: 機械学習ポテンシャル, Ackland-1987, Zhou-2004: EAM ポテンシャル^{11,12)}). (b) 機械学習ポテンシャルを用いた大域的最適化手法 (DIRECT法) により予測されたTi-Al系における形成エネルギーと安定構造。安定構造は、TiAl₃を除き、実験による構造と同じである。TiAl₃の実験構造はDO₂₂構造であるが、凸包上のDO₂₃構造の形成エネルギーとかなり近い値を示す (+6 meV/atom)。

ものであるが、粒界エネルギーの予測値は第一原理計算の値を再現しており、学習データに含まれない欠陥構造に対しても予測能力が高いことを示している。また、10,000原子、ステップ数1,000,000程度の摂動分子動力学計算やPhonon wave packet法の計算を通して、Si粒界における高精度な格子熱伝導計算を行った¹³⁾。そのほかにも、多くの応用が期待される。

2.3 機械学習ポテンシャルによる大域的構造探索

本節では、機械学習ポテンシャルの有効な応用の1つである大域的構造探索を紹介する。第一原理計算と遺伝的アルゴリズムなどの大域的最適化手法に基づいた構造探索は、近年におけるソースコードの公開により、開発者以外の研究者にも利用され始めており、今後さらに普及すると考えられる。しかし、安定構造が単純な系を除いては、何らかの構造が出力されるものの、真の安定構造は簡単に見つからないのが実情である。特に、多元系や高压相など構造が未知である場合や単位胞の原子数が多い構造を対象とする場合、大域的構造探索は非常に難しい。大域的構造探索の必要性や難しさは古くから認識されており¹⁴⁾、最適化の自由度が原子数に

対して指数関数的に増大することが本質的な問題である。

上述の方法で構築された機械学習ポテンシャルは、多くの結晶構造に対して高精度なエネルギー予測が可能である。その対象は、経験的ポテンシャルの予測精度が高い構造だけでなく、経験的ポテンシャルでは精度が足りない多くの構造を含む。この特徴を利用し、膨大な数の結晶構造について、機械学習ポテンシャルによるエネルギー計算をすることにより、化学組成とポテンシャルのみを入力とする探索範囲の広い効率的な大域的構造探索や準安定構造の列挙を行った。機械学習ポテンシャル自体が第一原理計算から構築されているため、機械学習ポテンシャルを用いた構造探索・列挙は第一原理的であるといえる。

図 2(b) は、決定論的大域的最適化手法の 1 つである DIRECT 法により、Ti-Al 系の安定構造および準安定構造列挙を行ったものである¹⁵⁾。DIRECT 法では、解の存在する可能性の高い探索領域を分割し、分割した部分探索領域の中央に該当する構造のエネルギーを計算する。解の存在する可能性の高い探索領域の判定、探索領域の分割およびエネルギー計算を繰り返すことで、大域的に最安定な構造を探索する。また、その過程で得られるエネルギー計算を行った構造集合には低いエネルギーをもつ準安定な構造が含まれている。図 2(b) のデータ点は、12 原子以下の組成を対象に DIRECT 法を行い、エネルギー計算を行った構造集合から取り出した代表構造について局所最適化を行ったものである。機械学習ポテンシャルによる合計 10^9 回程度のエネルギー評価の結果、六方最密充填構造置換の Ni_3Sn 型 (D0_{19}) 構造や面心立方構造置換の AuCu 型 (L1_0) 構造などが、凸包上に安定構造として現れている。また、それと同時に、 AlB_2 型 (C32)、 Cu_3Au 型 (L1_2)、 AuCd 型 (B19)、 Cu_2Mg -Laves 型 (C15)、 AuTe_2 型 (C34)、 Ca_7Ge 型構造など、さまざまな準安定構造が DIRECT 法の結果として求まった。

3. 結晶構造集合に基づいた効率的な結晶構造探索

3.1 機械学習を利用した結晶構造探索

構造探索において、結晶構造に関する事前知識や仮定を利用できる場合、それらに基づいた離散的な構造候補集合を列挙し、構造候補集合の中から探索することが常套手段である。多成分系における置換型の結晶構造探索などが一例であり、格子モデルを仮定して、格子における原子配置を列挙し、列挙された構造集合の中から解を探索する。このような結晶構造に関する制限により、連続的な探索空間を用いた構造探索よりも広範囲な構造探索が可能となる。

候補となる構造集合が与えられたとき、最も単純な探索方法は、全ての候補構造について第一原理計算を実行することであるが、実行できる構造集合の大きさはかなり限定的である。より大きい構造集合を用いる場合には、以下の機械学習を利用した効率的な構造探索が必要である。①構造集合に含まれる構造を表現する構造特徴量を導入し、構造のエネルギー

を構造特徴量の関数としてモデリングする。②構造集合の一部について第一原理計算を行い、機械学習手法によりエネルギーモデルを推定する。③計算コストの低いエネルギーモデルを用いて、構造集合を探索する。エネルギーモデルの誤差が小さい場合には、エネルギーモデルの予測値をそのまま利用することが可能であり、エネルギーモデルの誤差が大きい場合には、ベイズ最適化のような機械学習手法を用い、逐次的構造探索を行うことが有効である。

このような機械学習を利用した効率的な構造探索の例として、 ACuO_{3-x} ($\text{A}=\text{Sr}, \text{Ca}, \text{Ba}, \text{La}, \text{Sc}, \text{Y}$) 酸素欠損ペロブスカイトの構造予測¹⁶⁾を紹介する。まず、ペロブスカイト単位胞の 8 倍までの拡大で表される 80 万程度の非等価な酸素欠損配置を列挙した。それぞれの系について、構造集合から 400~500 個程度の構造をサンプリングし、第一原理計算を行った。次に、酸素サイトの対の相関関数を構造特徴量とし、第一原理計算結果に基づきガウス過程回帰によるエネルギーモデルを構築した。酸素欠損ペロブスカイトの場合、ガウス過程回帰モデル予測値と第一原理計算値との間には強い相関があるものの予測誤差は大きく、例えば、 $\text{A}=\text{Sr}$ の場合、予測誤差は 0.13 (eV/atom) 程度であった。これは、Cu 原子の価数変化や、酸素欠損が存在することによる立方晶ペロブスカイト格子からの大きい緩和が理由である。

そのため、ベイズ最適化を採用し、誤差の大きい構造特徴量を利用したガウス過程回帰モデルによる逐次的な構造探索を行った。具体的には、構造集合に含まれる全ての構造に対して、ガウス過程回帰モデルを用いた予測を行い、安定構造である確率が高い構造を選択した。その構造に対して第一原理計算を行い、モデルを更新し、逐次的に構造探索を行った。その結果、すでに報告されている酸素欠損ペロブスカイト構造に加え、図 3 に示すような酸素欠損が規則化した新たな構造が予測された。

本稿で紹介した第一原理計算、非等価な結晶構造の列挙、回帰分析、構造特徴量によるモデリング、ベイズ最適化を組み合わせた方法は、他の多成分系においても効率的な構造探索が期待される。また、置換型の構造探索例を紹介したが、機械学習に基づいた構造探索手法は、構造集合さえ与えることができれば、置換型に限らず応用可能である。

3.2 効率的なデータ構造を用いた結晶構造列挙およびスクリーニング

前節で述べた構造集合を作る際、等価な構造を含めた全構造を列挙した後、非等価な構造を列挙することが一般的である。小さい単位胞のみを対象とする場合は問題とならないが、大きな単位胞までの広い探索範囲とする場合には、構造を列挙する箇所が構造探索のボトルネックとなる。一般的な方法では、 10^7 程度の非等価な構造列挙が限界であり、構造列挙に基づいた構造探索も同様に限界を迎える。より大きな単位胞での構造探索を行いたい場合や、多成分系や最小の原子数が大きくなる複雑な化学組成を対象とする場合など、

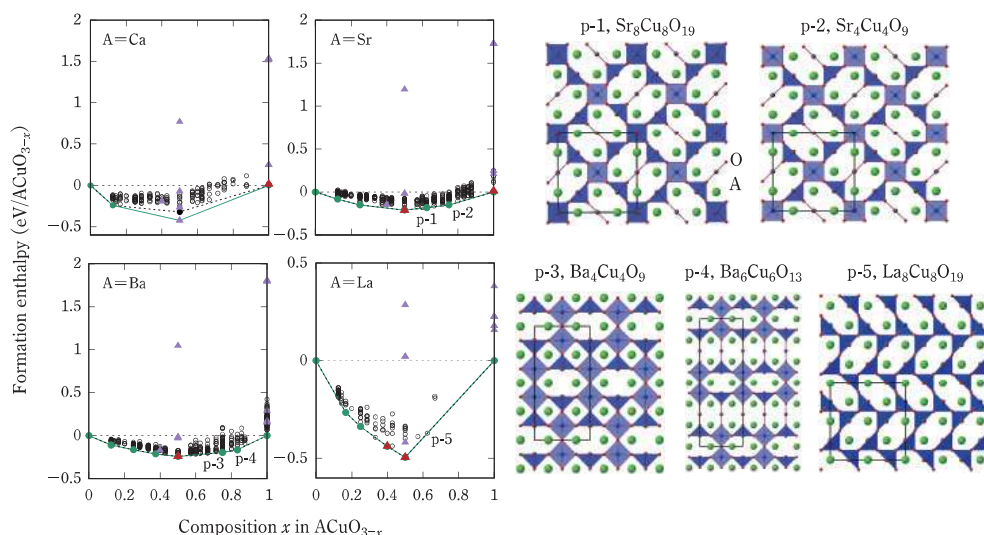


図3 ACuO_{3-x} (A = Sr, Ca, Ba, La) 酸素欠損ペロブスカイトの形成エネルギーと予測安定構造 (at 10 GPa).

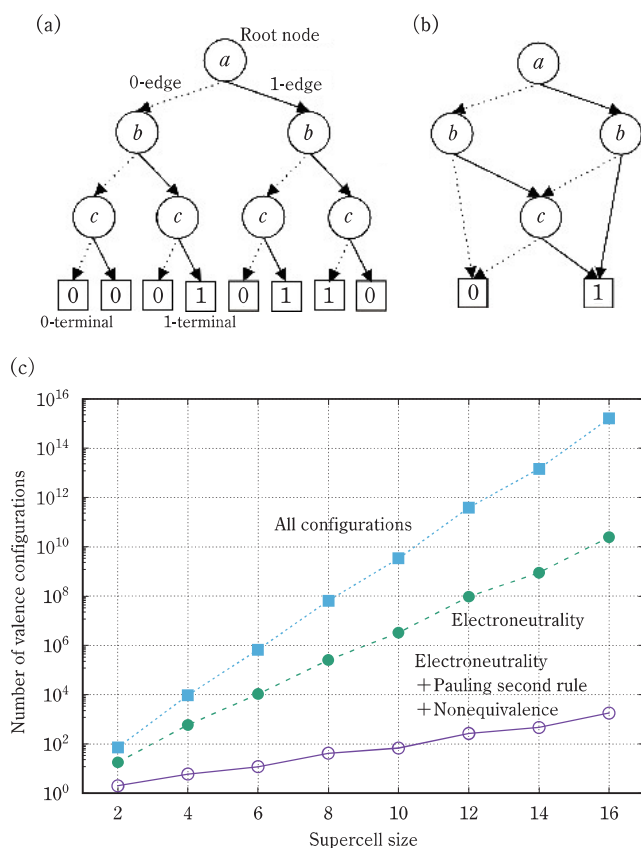


図4 組み合わせ集合 $S = \{\{a,b\}, \{a,c\}, \{b,c\}\}$ の (a) 2 分木表現と (b) ZDD による表現¹⁹⁾. (c) 電気的中性条件とポーリング第 2 則によりスクリーニングされた酸素欠損ペロブスカイト A₂B₂O₅ の非等価なイオン配置数.

大きな数の構造列挙が必要な場合には、効率的なデータ構造の活用が有効である。

その 1 つが、Zero-suppressed Binary Decision Diagram (ZDD) と呼ばれるデータ構造^{17,18)} である。ZDD では、図 4 のように、冗長な表現を削除し、等価な表現を共有することで、組み合わせ (構造) 集合を圧縮する。ZDD を用いて非等価な構造のみを列挙することは実装も含め容易ではないも

の、現実的な計算資源において、 $10^{15} \sim 10^{16}$ 程度の非等価な構造が列挙可能である¹⁹⁾。

しかし、ZDD の構造集合を用いた構造探索を行う場合、構造集合に含まれる 10^{15} 程度の構造全てを利用することはほぼ不可能である。よって、多くの非等価な構造を含む構造集合から、解となりうる可能性がある構造のみに絞り込むことが望ましい。その選択枝の 1 つが、事前知識や制約による構造集合のスクリーニングである。ZDD は、ある制約を満たす一部だけを取り出すことができるデータ構造となっており、構造集合の ZDD 形式を保持したままスクリーニングすることができるため、非常に効率的である。

図 4(c) は、酸素欠損ペロブスカイト A₂B₂O₅ のイオン配置について、固体化学的な制約である電気的中性条件、ポーリング第 2 則、対称性に関する非等価性に関する制約を順に適用し、構造集合をスクリーニングしたものである²⁰⁾。A サイトは 2 価の陽イオン、B サイトは 2 価あるいは 3 価の陽イオン、O サイトは 2 価の陰イオンあるいは酸素欠損の配置を考慮している。その結果、制約を導入することにより、構造数 10^{15} の集合から数千程度の構造を取り出すことができている。大幅に構造探索の候補を削減できており、その効果は単位胞に含まれる原子数が多いほど顕著である。このように、ZDD 形式の構造集合をスクリーニングすることで、広い範囲における構造探索が可能となる。

本稿では、電気的中性条件やポーリング第 2 則を厳密に満たすことを条件としたが、条件を緩め、閾値を設定したうえで構造列挙することも選択枝の 1 つである。また、電気的中性条件やポーリング第 2 則以外にも、配位数や局所環境、クラスタ数なども制約として扱うことが可能である。例えば、Special quasirandom structure と呼ばれる多成分系のランダム状態を模倣した規則構造を取り出す ZDD もその 1 つである²¹⁾。これは、局所環境の平均がランダムな状態と同じになるという制約の下で、大きな構造集合を表現する ZDD から少

数の該当構造のみを取り出すものである。ほかにも、結晶構造データベースや第一原理計算データを解析し、データの傾向から結晶構造に関する制約を導入することも可能である²⁰⁾。

4. むすび

本稿では、機械学習手法や効率的なデータ構造、結晶構造を表現する構造特徴量、構造特徴量を用いた予測モデルを導入することにより、第一原理計算に基づいた応用計算を効率的に実施する手法について紹介した。本稿で紹介した手法を用いることにより、第一原理計算に基づいた結晶構造探索、熱力学計算、分子動力学計算などの実行可能な原子数、時間範囲、自由度を大幅に拡大することができる。これまででは不可能であった多くの興味ある系での第一原理的な結晶構造予測や分子動力学計算により、新物質探索など幅広い応用が期待される。

謝 辞

本研究は科研費基盤研究 (B) (19H02419) および新学術領域研究 (19H05787) の助成を受けたものである。

文 献

- 1) A.P. Bartók, M.C. Payne, R. Kondor, and G. Csányi: Phys. Rev. Lett. **104**, 136403 (2010).
- 2) J. Behler and M. Parrinello: Phys. Rev. Lett. **98**, 146401 (2007).
- 3) A. Seko, A. Takahashi, and I. Tanaka: Phys. Rev. B **90**, 024101 (2014).
- 4) A. Thompson, L. Swiler, C. Trott, S. Foiles, and G. Tucker: J. Comput. Phys. **285**, 316 (2015).
- 5) A.V. Shapeev: Multiscale Model. Simul. **14**, 1153 (2016).
- 6) A. Seko, A. Togo, and I. Tanaka: Phys. Rev. B **99**, 214108 (2019).
- 7) A. Seko: Phys. Rev. B **102**, 174104 (2020).
- 8) A. Seko: arXiv:2007.14206.

- 9) A. Seko: <https://sekocho.github.io>
- 10) T. Nishiyama, A. Seko, and I. Tanaka: Phys. Rev. Mater. **4**, 123607 (2020).
- 11) G.J. Ackland, G. Tichy, V. Vitek, and M.W. Finnis: Philos. Mag. A **56**, 735 (1987).
- 12) X.W. Zhou, R.A. Johnson, and H.N.G. Wadley: Phys. Rev. B **69**, 144113 (2004).
- 13) S. Fujii and A. Seko: arXiv:2105.11959, Comput. Mater. Sci., in press.
- 14) J. Maddox: Nature **335**, 201 (1988).
- 15) A. Seko, K. Toyoura, and J. Honda: in preparation.
- 16) A. Seko and S. Ishiwata: Phys. Rev. B **101**, 134101 (2020).
- 17) S. Minato: Proc. 30th Int. Design Automation Conf. (DAC '93), p. 272 (1993).
- 18) 湊真一 (編): 超高速グラフ列挙アルゴリズム (森北出版, 2015).
- 19) K. Shinohara, A. Seko, T. Horiyama, M. Ishihata, J. Honda, and I. Tanaka: J. Chem. Phys. **153**, 104109 (2020).
- 20) A. Seko and K. Shinohara: in preparation.
- 21) K. Shinohara, A. Seko, T. Horiyama, and I. Tanaka: Phys. Rev. Mater. **5**, 113803 (2021).

(2021 年 8 月 20 日 受理)

用語解説

多項式回転不変量

ある原子を中心とした結晶構造の回転に対して不変である構造を表現する多項式。

プロトタイプ構造

ICSD にて分類されている結晶構造型。

Profile



世 古 敦 人 (せこ あつと)

京都大学大学院工学研究科准教授 (材料工学専攻)。博士 (工学)。2007 年京都大学大学院工学研究科博士課程修了。専門分野は計算材料科学。