

令和 2 年 5 月 28 日現在

機関番号：13102

研究種目：若手研究

研究期間：2018～2019

課題番号：18K13474

研究課題名(和文) 機械学習が拓く高効率結晶構造探索手法の開発

研究課題名(英文) Development of crystal structure prediction methods using machine learning

研究代表者

山下 智樹 (Yamashita, Tomoki)

長岡技術科学大学・産学融合トップランナー養成センター・特任准教授

研究者番号：60793099

交付決定額(研究期間全体)：(直接経費) 3,200,000円

研究成果の概要(和文)：結晶構造探索手法におけるアルゴリズムとして、新たに提案した選択型アルゴリズムであるベイズ最適化および近年広く利用されている進化的アルゴリズムのコード開発を行い、結晶構造探索ツール、CrySPYとして公開した(<https://github.com/Tomoki-YAMASHITA/CrySPY>)。広くユーザーに利用してもらえるようにCrySPYは無料で入手可能なPythonで設計し、オープンソースソフトウェアとした。また、ドキュメントおよびチュートリアルサイトも作成しウェブ上に公開した。

研究成果の学術的意義や社会的意義

本研究により、オープンソースの結晶構造探索ツールであるCrySPYを公開することができた。CrySPYは機械学習を用いた高効率な安定構造探索が可能であり、新材料設計の基盤となるツールとして誰もが利用可能である。半年に一度の頻度でCrySPYのチュートリアルセミナーも開催しており、大学に所属する研究者や学生および企業の研究者などの間で利用されるようになった。

研究成果の概要(英文)：We have developed crystal structure prediction methods. The code development of Bayesian optimization and evolutionary algorithm has done and implemented in the open source software, CrySPY(<https://github.com/Tomoki-YAMASHITA/CrySPY>). CrySPY is written in Python to be used widely. Moreover the document and web site are published.

研究分野：計算物理

キーワード：結晶構造探索 機械学習 第一原理計算

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

近年の計算機能力の飛躍的向上にともない、第一原理計算を軸にしたデータ駆動型材料デザインに関する研究が開拓され始めている。最近では、材料の組成を与えるだけで安定構造を予測することが可能な、結晶構造探索手法が注目を集めている。材料の最も基本的な情報である結晶構造が得られれば、種々の計算によりその性質も予測可能となることから、結晶構造探索手法は新材料デザインの根幹を担う役割を期待されている。ランダムに構造を生成して探索を行うランダムサーチを草分けに、最近では進化的アルゴリズムを用いた結晶構造探索手法が広く利用されている。どの手法も成功を収めてはいるものの、この分野の共通の課題は原子種および原子数が多い複雑な系では安定構造を見つけられないことである。探索できる構造の数は数百から千個程度までが現実的であり、安定構造を見つけるのにこれ以上の探索が必要となる複雑な系には従来の探索手法は適用不可能であった。

この問題を解決するため、データ科学と物性・材料科学の融合による新しい“選択型”アルゴリズムの開発に着手した。選択型アルゴリズムでは、多数の構造をあらかじめ生成しておき、その中から機械学習を用いて効率良く構造を選択して最適化していくことで、多数の候補の中から最も安定な構造を少ない試行回数で見つけ出すことが可能となる。機械学習の手法として、ベイズ最適化を用いることで、従来の手法と同程度の計算コストで数倍もの構造を扱うことができることを示した。しかしながら、作成したプログラムも機能ごとにバラバラで、誰もが扱うことができるようになっておらず、材料設計のためのツールとしてはまだまだ不便も多いのが現状である。

### 2. 研究の目的

#### (1) 高効率結晶構造探索ツール、CrySPYの開発および公開

開発したプログラムコード群を CrySPY に統合し、ランダムサーチおよびベイズ最適化が利用可能なオープンソースの高効率結晶構造探索ツールとして公開する。簡単な入力ファイルを準備するだけで構造の生成および最適化、ベイズ最適化の実施や計算結果の収集まで自動的に行えるようなシステムを構築する。プログラム言語には無料で入手出来る Python を用いる。また、広くユーザーに利用してもらえるように、ドキュメントの整備も行う。

#### (2) 進化的アルゴリズムとベイズ最適化のハイブリッドアルゴリズムの開発

選択型アルゴリズムの特徴は、これまで実現が難しかった千個を超えるような多数の構造の扱いのみではなく、拡張性が優れていることも挙げられる。現在最もよく利用されている進化的アルゴリズムを開発中のベイズ最適化に組み入れることが可能で、両手法の特徴を有した、現状で最も有望かつ効率的だと考えられるハイブリッドアルゴリズムの開発を行う。進化的アルゴリズムを用いて多数の構造を生成しておき、ベイズ最適化で効率的に選択することで、探索効率の飛躍的な改善を目指す。開発した手法は CrySPY で使用できるようプログラムの整備も行う。

#### (3) ランダム構造と最適化構造のエネルギー差予測

ランダムに生成される初期構造と最適化構造のエネルギー差を予測することができれば、最適化してもエネルギーが低くならない初期構造を候補からあらかじめ排除することが可能であり、探索効率改善の補助として機能する。本研究では、エネルギーと構造データをあらかじめ計算して用意しておき、ディープラーニングを利用した画像認識の技術を用いてエネルギー差の予測モデルを構築し、精度を検証する。

### 3. 研究の方法

#### (1) 進化的アルゴリズムとベイズ最適化のハイブリッドアルゴリズムの開発

Si<sub>16</sub> (ユニットセル内に Si 原子 16 個の系) および Si<sub>32</sub> の二つの系について、ランダムサーチ、進化的アルゴリズム、ベイズ最適化およびハイブリッドアルゴリズムを用いて安定構造の探索効率を検証する。構造探索シミュレーションには、本研究で開発する CrySPY を用いる。エネルギーの評価および構造最適化には、Si において高精度で計算が可能な Stillinger-Weber ポテンシャル<sup>1)</sup>を用いる。プログラムパッケージは soiap<sup>2)</sup>を CrySPY と連携させて利用する。進化的アルゴリズムにおける次世代構造生成のための進化的操作は crossover および strain を使用する。ベイズ最適化における結晶構造の記述子は Oganov と Valle の F-fingerprint<sup>3)</sup>を用いる。

#### (2) ランダム構造と最適化構造のエネルギー差予測

Si<sub>8</sub>、Si<sub>16</sub>、Ga<sub>4</sub>As<sub>4</sub> および Ga<sub>8</sub>As<sub>8</sub> の系を対象にエネルギー差をディープラーニングで学習する。構造とエネルギーのデータは第一原理計算プログラムパッケージの Quantum ESPRESSO<sup>4)</sup>を用いて事前に準備する。原子数が多い Si<sub>16</sub> と Ga<sub>8</sub>As<sub>8</sub> の系においては、学習データは原子数が少ない Si<sub>8</sub> および Ga<sub>4</sub>As<sub>4</sub> のものをそれぞれ用いて予測モデルを作成しておき、検証のテストデータは Si<sub>16</sub> と Ga<sub>8</sub>As<sub>8</sub> のものを用いる。結晶構造を三次元ボクセルデータに変換するため、Atomic Coordinate Voxel Descriptor (ACVD) と名付けた記述子生成手法を開発する。原子位置の座標を中心としたガウス分布に従って原子の存在確率をボクセルデータ化する。準備したデータセッ

トと記述子を用いて、畳み込みニューラルネットワークモデルを利用してエネルギー差予測モデルを構築する。

#### 4. 研究成果

##### (1) 高効率結晶構造探索ツール、CrySPY の開発および公開

探索アルゴリズムとして、ランダムサーチ、進化的アルゴリズム、ベイズ最適化および LAQA が使用可能な結晶構造探索ツールの CrySPY を公開し (<https://github.com/Tomoki-YAMASHITA/CrySPY>) ドキュメントの整備および複数回のバージョンアップを行なった。半年に 1 回ほどの頻度でチュートリアルセミナーも開催しており、学生や社会人が数十名参加している。当初の目的通り、広くユーザーに使用してもらえるような環境を整えることができた。

##### (2) 進化的アルゴリズムとベイズ最適化のハイブリッドアルゴリズムの開発

Si<sub>16</sub> では、各アルゴリズムにおいて 100 構造の探索を行い安定構造が得られたかどうか検証し、これを何度も繰り返すことで成功率を算出した。ベイズ最適化やハイブリッドアルゴリズムでは 300 構造から 100 構造を選択して探索した。Si<sub>32</sub> では探索が難しくなることも考慮して 300 構造の探索を行なった。ベイズ最適化とハイブリッドアルゴリズムに関しては 600 構造から 300 構造を選択して探索を行なった。

各アルゴリズムにおける安定構造探索の成功率を表 1 に示す。Si<sub>16</sub> の系では、進化的アルゴリズムの優位性が見られなかった一方で、ベイズ最適化およびハイブリッドアルゴリズムの成功率が高い結果が得られた。原子数が比較的少ない系であったため、ランダムな構造生成で十分に安定構造を探索できたと考えられる。進化的アルゴリズムに優位性が見られなかったのは、探索構造数が 100 構造とそれほど多くなく、進化的操作による構造生成数が少なかったからだと考えられる。ベイズ最適化を用いた選択型のアルゴリズムはより多くの候補の数を扱うことが可能なので、探索効率の向上に貢献している。ハイブリッドアルゴリズムの成功率が最も高かったのは、進化的操作による構造生成数が多く、進化的アルゴリズムとベイズ最適化の両方がうまく機能した結果である。

Si<sub>32</sub> の系では、ランダムサーチは成功率が悪く、相対的に進化的アルゴリズムの成功率が上がった。原子数が多く系が複雑になると、もはやランダムな構造生成では安定構造を見つけることが難しくなる。この系では探索数が 300 構造に増加されており、進化的操作における構造生成数も十分多くなったことから、進化的アルゴリズムの成功率がランダムサーチに比べて高くなった。この系においてもハイブリッドアルゴリズムの成功率は最も高く、現状で最も効率の良い探索アルゴリズムであると言える。

表 1 各アルゴリズムにおける安定構造探索の成功率。

アルゴリズム	ランダムサーチ	進化的アルゴリズム	ベイズ最適化	ハイブリッドアルゴリズム
成功率 (Si <sub>16</sub> )	64%	58%	70%	82%
成功率 (Si <sub>32</sub> )	46%	60%	54%	64%

##### (3) ランダム構造と最適化構造のエネルギー差予測

結晶構造の記述子として ACVD を用い、畳み込みニューラルネットワークを利用してエネルギー差の予測モデルを構築した。Si<sub>8</sub> および Si<sub>16</sub> の系ではエネルギー差の平均絶対誤差はそれぞれ 0.20 および 0.17 eV/atom という結果が得られた。Ga<sub>4</sub>As<sub>4</sub> および Ga<sub>8</sub>As<sub>8</sub> の系においてはそれぞれ 0.16 および 0.17 eV/atom であった。いずれも少ない原子数の系で学習して構築した予測モデルを使用した結果であり、原子数が多くなっても構築した予測モデルが使用できることがわかる。単純な系でデータを用意して学習し、予測モデルを構築することで、複雑な系においてランダムに生成した初期構造が最適化によってどれくらいエネルギーが変化するかを大雑把ではあるが予測することができた。これらの結果より、結晶構造探索の初期構造のスクリーニングに活用できる可能性を見出すことができた。

#### < 引用文献 >

- F. H. Stillinger and T. A. Weber, Phys. Rev. B **31**, 5262 (1985).  
<https://github.com/nbsato/soiap>
- A. R. Oganov and M. Valle, J. Chem. Phys. **130**, 104504 (2009).
- P. Giannozzi, et al., J. Phys.: Condens. Matter **21**, 395502 (2009).

## 5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 Yamashita Tomoki, Sato Nobuya, Kino Hiori, Miyake Takashi, Tsuda Koji, Oguchi Tamio	4. 巻 2
2. 論文標題 Crystal structure prediction accelerated by Bayesian optimization	5. 発行年 2018年
3. 雑誌名 Physical Review Materials	6. 最初と最後の頁 13803
掲載論文のDOI（デジタルオブジェクト識別子） 10.1103/PhysRevMaterials.2.013803	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Terayama Kei, Yamashita Tomoki, Oguchi Tamio, Tsuda Koji	4. 巻 4
2. 論文標題 Fine-grained optimization method for crystal structure prediction	5. 発行年 2018年
3. 雑誌名 npj Computational Materials	6. 最初と最後の頁 32
掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s41524-018-0090-y	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Sato Nobuya, Yamashita Tomoki, Oguchi Tamio, Hukushima Koji, Miyake Takashi	4. 巻 4
2. 論文標題 Adjusting the descriptor for a crystal structure search using Bayesian optimization	5. 発行年 2020年
3. 雑誌名 Physical Review Materials	6. 最初と最後の頁 33801
掲載論文のDOI（デジタルオブジェクト識別子） 10.1103/PhysRevMaterials.4.033801	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計10件（うち招待講演 1件/うち国際学会 7件）

1. 発表者名 Tomoki Yamashita, Shinichi Kanehira, N. Sato, Hiori Kino, Koji Tsuda, Takashi Miyake, and Tamio Oguchi
2. 発表標題 Crystal Structure Prediction by Bayesian Optimization and Evolutionary Algorithm
3. 学会等名 APS March Meeting 2019（国際学会）
4. 発表年 2019年

1. 発表者名 佐藤暢哉, 山下智樹, 小口多美夫, 福島孝治, 三宅隆
2. 発表標題 Bayes最適化を用いた結晶構造探索の効率と記述子のパラメータ
3. 学会等名 日本物理学会第74回年次大会
4. 発表年 2019年

1. 発表者名 Tomoki Yamashita, Kei Terayama, Shinichi Kanehira, N. Sato, Hiori Kino, Koji Tsuda, Takashi Miyake, and Tamio Oguchi
2. 発表標題 Development of crystal structure prediction tool
3. 学会等名 PRESTO International Symposium on Materials Informatics (国際学会)
4. 発表年 2019年

1. 発表者名 Tomoki Yamashita
2. 発表標題 Crystal structure prediction by machine learning
3. 学会等名 CSRN-OSAKA Annual Workshop 2018 (招待講演) (国際学会)
4. 発表年 2018年

1. 発表者名 佐藤暢哉, 山下智樹, 小口多美夫, 福島孝治, 三宅隆
2. 発表標題 Bayes最適化を用いた結晶構造探索における記述子
3. 学会等名 日本物理学会2018年秋季大会
4. 発表年 2018年

1. 発表者名 Tomoki Yamashita, Shinichi Kanehira, Nobuya Sato, Hiori Kino, Koji Tsuda, Takashi Miyake, and Tamio Oguchi
2. 発表標題 Hybrid Algorithm of Bayesian Optimization and Evolutionary Algorithm in Crystal Structure Prediction
3. 学会等名 MATERIALS RESEARCH MEETING 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 Nobuya Sato, Tomoki Yamashita, Tamio Oguchi, Koji Hukushima, and Takashi Miyake
2. 発表標題 Descriptor for Efficient Crystal Prediction Using the Bayesian Optimization
3. 学会等名 MATERIALS RESEARCH MEETING 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 Nobuya Sato, Tomoki Yamashita, Tamio Oguchi, Koji Hukushima, and Takashi Miyake
2. 発表標題 Parameter in a Descriptor for Efficient Crystal Structure Search Using the Bayesian Optimization
3. 学会等名 The 22nd Asian Workshop on First-Principles Electronic Structure Calculations (国際学会)
4. 発表年 2019年

1. 発表者名 Tomoki Yamashita, Shinichi Kanehira, Nobuya Sato, Hiori Kino, Koji Tsuda, Takashi Miyake, and Tamio Oguchi
2. 発表標題 Searching Efficiency of Bayesian Optimization and Evolutionary Algorithm in Crystal Structure Prediction
3. 学会等名 10th International Conference on Materials for Advanced Technologies (国際学会)
4. 発表年 2019年

1. 発表者名 山下智樹, 兼平慎一, 佐藤暢哉, 木野日織, 津田宏治, 三宅隆, 小口多美夫
2. 発表標題 結晶構造探索におけるベイズ最適化と進化的アルゴリズムのハイブリッドアルゴリズム
3. 学会等名 日本物理学会2019年秋季大会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

CrySPY <a href="https://github.com/Tomoki-YAMASHITA/CrySPY">https://github.com/Tomoki-YAMASHITA/CrySPY</a> チュートリアルと解説 <a href="https://tomoki-yamashita.github.io/cryspy/tutorial/outline.html">https://tomoki-yamashita.github.io/cryspy/tutorial/outline.html</a>
--

6. 研究組織		
	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)
		備考