

令和 6 年 6 月 13 日現在

機関番号：13102

研究種目：基盤研究(C)（一般）

研究期間：2021～2023

課題番号：21K03386

研究課題名（和文）強化学習LAQAを用いた高効率結晶構造探索手法の開発およびソフトウェア化

研究課題名（英文）High-Efficiency Crystal Structure Prediction Method Using Reinforcement Learning
LAQA: Development and Software Implementation

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）：結晶構造探索手法のLAQAでは、第一原理計算や経験的原子間ポテンシャルを用いた計算から得られるエネルギーおよび原子に働く力を用いることにより、最終的にエネルギーがより低くなりそうな構造候補を選択して優先的に最適化することで、不要な最適化計算を行うことなく効率的に安定構造が探索可能となる。本研究課題では、応力テンソルを新たにスコアに取り入れることで計算コストを大幅に削減できる手法を開発した。開発したプログラムは結晶構造探索ソフトのCrySPYに組みこみ、GitHubおよびPyPIに公開した。

研究成果の学術的意義や社会的意義

本研究課題で改良されたソフトウェアであるCrySPYはPyPIにおいて現在まで8000以上のダウンロードがあり、月に600から700回ほどインストールされている計算になる。結晶構造探索ツールは、近年の計算機能力の飛躍的向上にともない、新材料設計においてなくてはならないツールになりつつある。また将来の発展に目を向けると、CrySPYは機械学習ポテンシャルと組み合わせ、アクティブラーニングを採用することで計算の高速化が期待できる。すでにその研究も開始しており、ここではLAQAの考え方を採用したスコア方式を採用している。本研究課題の成果がベースとなって、今後のこの分野の発展が見込まれる。

研究成果の概要（英文）：In the crystal structure method, LAQA, energy and atomic forces obtained from first-principles calculations or empirical interatomic potentials are used to select and prioritize the optimization of structure candidates that are likely to have lower final energy. This allows for efficient exploration of stable structures without performing unnecessary optimization calculations. In this research project, we developed a method that significantly reduces computational costs by incorporating stress tensors into the scoring. The developed program was integrated into the crystal structure search software CrySPY and made available on GitHub and PyPI.

研究分野：計算材料科学

キーワード：結晶構造探索 機械学習 第一原理計算 CrySPY LAQA

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マテリアルズインフォマティクスに関する研究が開拓され始めている。従来のシミュレーションでは、結晶構造データを元に計算を行うことで物性等を評価するが、構造が未知の新材料に対しては適用することができない。所望する機能・性質を有する新材料のデザインを可能とするためには、目的の組成や物性を持つ未知材料の結晶構造を予測する手法が不可欠である。このような流れを受け、新材料デザインを行うための基盤技術として、組成(原子数)を指定するだけで安定な構造を求めることが可能な、結晶構造探索手法が注目を集めている。これまでは、第一原理計算の構造最適化と種々の構造生成アルゴリズムを組み合わせた手法が主流であり、中でも、進化的アルゴリズムが広く利用されている。しかしながら、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、構造候補の数が爆発的に増加し、安定構造を見つけられないことであった。

この問題を解決するため応募者を中心とする研究グループは、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を行った。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、第一原理計算を用いて最適化することで、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。選択型アルゴリズムの利点は、従来の進化的アルゴリズムなどの構造生成手法と相反することなく、併用することができる点である。これまでは安定ではない構造の最適化に不必要に時間を費やしていたが、選択型アルゴリズムの登場により効率よくシミュレーションが進められるようになった。開発したプログラムはオープンソースの結晶構造探索ツール CrySPY (<https://github.com/Tomoki-YAMASHITA/CrySPY>) として公開している。

我々が開発した選択型アルゴリズムの LAQA は第一原理計算から得られるエネルギーや原子に働く力を用いて強化学習を行うので、結晶構造の記述子は必要ではなく、シンプルに扱うことが可能であり、幅広いユーザー向けの手法であると言える。探索効率も非常に良く、従来の手法と比較して計算コストを約 50%から 90%ほど削減可能であることが示されている。しかしながら、LAQA のアルゴリズムは発展途上であり、構造データセットによっては効率よく探索を実行できない場合があることがわかっている。これは構造候補の選択に用いるスコア関数の評価に応力テンソルの項を導入できていないためである。また、CrySPY ではエネルギーの評価や構造最適化には外部の第一原理計算ソフトを利用しており、有料ソフトウェアの VASP やオープンソースの Quantum ESPRESSO が利用できる。しかしながら、現状では LAQA と連携して利用できるのはすでにインターフェースを開発した VASP のみであり、Quantum ESPRESSO には対応していない。VASP は有料ソフトなので、CrySPY の利用者(特に企業の研究者)からは LAQA と Quantum ESPRESSO の連携への要望も多く、インターフェースのプログラム開発が強く望まれている。

2. 研究の目的

結晶構造探索手法において、最も時間を要するのは第一原理計算による構造最適化の部分である。従来の手法では、生成した構造に対して順番に最適化が実行されるが、最適化してもエネルギーが高く安定構造ではない場合が圧倒的に多いことが問題である。このように不要な最適化計算を如何に省くかを動機付けに開発された手法が LAQA である。LAQA では強化学習により最終的にエネルギーがより低くなりそうな候補を選択して優先的に最適化することで、不要な最適化計算を行うことなく効率的に安定構造が探索可能となる。

何を基準に候補を選択するかがここでの焦点となるが、LAQA では第一原理計算から得られる情報を用いたスコア関数を評価することで優先度を決める。構造最適化によりエネルギーが大きく下がるならば、必然的に原子に働く力や応力テンソルの値が大きくなる。従って、エネルギーが低く、力や応力テンソルが大きい候補ほどスコアが高くなるような関数を導入すればよい。現状のスコア関数は、現時点でのエネルギーおよび原子に働く力から計算されており、応力テンソルの項を導入できていない。原子に働く力が小さく応力テンソルが大きい候補がある場合は、うまく学習が働かず効率の良い探索ができない。応力テンソルは原子に働く力に比べて振る舞いが複雑であり、スコア関数に導入するには、様々な材料に関して最適化ステップ毎の応力テンソルの値をデータ化し、その変化の特性を解析することで汎用的なスコア関数を開発する必要がある。

本研究では、選択型アルゴリズム LAQA において、応力テンソルの項を新たに導入したスコア関数の開発を行う。また、オープンソースの Quantum ESPRESSO とのインターフェースを開発し、誰でも容易に LAQA を利用できるように CrySPY に組み込み、ソフトウェアとして公開する。

3. 研究の方法

(1) LAQA の改良

LAQA のスコア関数として、ストレス項を付け加えた以下の式を用いた。

$$L = -E + w_F F^2 / (2\Delta F) + w_S S$$

ここで、 E はエネルギー、 F は原子に働く力（の平均）、 S はストレステンソルの要素の絶対値の平均であり、これらは第一原理計算から求められる。 w_F および w_S はそれぞれ力とストレス項の重みである。この重みをいくつか変えながら、安定構造が既知の系に対してLAQAによる構造探索を実施し、ランダムサーチの結果と比較した。

(2) 結晶構造生成プログラムの並列化

LAQAの開発が順調に進んだこともあり、LAQAやランダムサーチで使用している初期構造生成プログラムの効率化にも取り組んだ。MPIを用いた並列化を行うため、mpi4pyを用いてPythonコードを改良した。

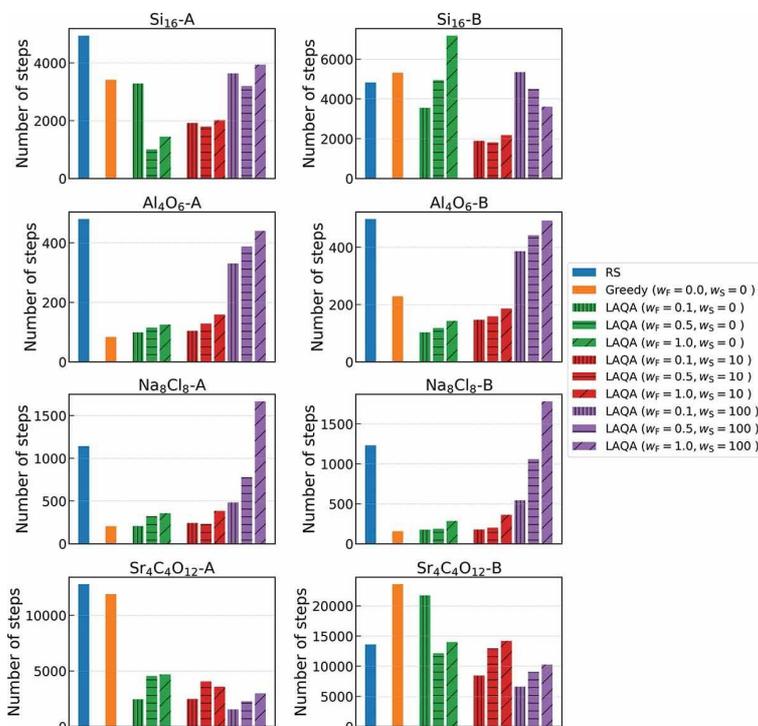
(3) ソフトウェアの公開およびドキュメント作成

作成したプログラムは、オープンソースソフトウェアのCrySPYに統合し、GitHubに公開した。同様にドキュメントもGitHubに公開している。また、ユーザーが容易にインストールできるように、Python Package Index (PyPI)に登録し、pipコマンドでインストールできるようにした。

4. 研究成果

(1) LAQAの改良

LAQAの探索効率を比較するため、 Si_{16} （ユニットセル内にSiが16原子）、 Al_4O_6 、 Na_8Cl_8 および $Sr_4C_4O_{12}$ の安定構造が既知の系について、構造探索シミュレーションを実施した。それぞれの系において初期構造のデータセットを二つ用意してシミュレーションを実施した。下図に安定構造を見つけるまでに要した第一原理計算の構造最適化ステップ数を示す。



図：LAQAの探索効率。縦軸は第一原理計算による計算量を示している。

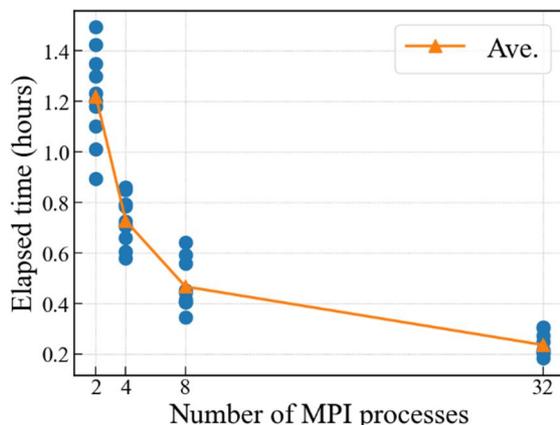
図は Sci. Technol. Adv. Mater. Meth. 2, 84 (2022). から引用した。

水色で示された棒グラフはランダムサーチの場合の計算量、橙色が上記スコアでエネルギー項だけを用いた場合、緑色がエネルギーと力の項を用いた場合、そして赤色および紫色がストレス項も含めて用いた場合の計算量を示している。図を見てわかるように、LAQAを用いることで基本的にはランダムサーチよりも大幅に計算コストを削減できている。しかしながら、橙色や緑はデータセットによっては非効率な場合があることもわかる。ストレス項を導入することで、このような不安定さがなくなり、安定して効率の良い探索ができるようになった。これらの成果は Sci. Technol. Adv. Mater. Meth. 2, 84 (2022) に掲載された。

(2) 結晶構造生成プログラムの並列化

Pythonライブラリのmpi4pyを用いて、CrySPYの結晶構造生成の並列化を行なった。下図は Li_3PS_4 のランダム構造を100構造生成したときのMPIプロセス数と計算時間の関係を示したものである。各プロセス数において10回試行（図の青丸）し、平均をとったものが図の実践で示されている。MPIプロセス数の増加に伴い、計算時間が効率よく減少していることがわかる。

LAQA やランダムサーチなどの結晶構造探索では、はじめに多数の初期構造を生成する必要があり、その大幅な時間短縮が可能になった。



図：結晶構造生成の並列効率。

(3) ソフトウェアの公開およびドキュメント作成

LAQA や MPI 並列のコード開発を取り入れた CrySPY を GitHub および PyPI に公開した。オープンソースソフトで利用しやすい Quantum ESPRESSO のインターフェース開発も行い、CrySPY に統合している。本研究課題開始時の CrySPY のバージョンは 0.9.2 であったが、研究実施期間の 2024 年 3 月までの間に 11 回のバージョンアップを行い、CrySPY 1.2.3 まで開発が進んだ（2024 年 6 月現在の最新版は CrySPY 1.3.0）。PyPI での現在までのダウンロード数は 8,116 であり、月に 600 から 700 回ほどインストールされている。本研究課題で進めたソフトウェア開発が大いに役に立ち、CrySPY の普及を促進していると言える。また、ドキュメントの作成も行い、例えば LAQA を用いた探索シミュレーションのチュートリアルページも公開している。

(4) 今後の展望

近年の機械学習ポテンシャル構築手法の発展により、第一原理計算にかかる時間が大幅に削減できるようになりつつある。CrySPY も機械学習ポテンシャルと組み合わせて、アクティブラーニングを採用することで計算の高速化が期待できる。海外のグループとの共同研究もすでに開始しており、そこでは LAQA の考え方を用いたスコア方式を採用している。また、計算の高速化に伴い、組成可変型の構造探索の機能も取り入れているところであり、今後も精力的にソフトウェア開発を行う予定である。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1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、Sekine Hirotaka	4. 巻 2
2. 論文標題 Improvement of look ahead based on quadratic approximation for crystal structure prediction	5. 発行年 2022年
3. 雑誌名 Science and Technology of Advanced Materials: Methods	6. 最初と最後の頁 84 ~ 90
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/27660400.2022.2059335	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 -

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 Tomoki Yamashita and Hirotaka Sekine
2. 発表標題 Improvement of Searching Efficiency in Crystal Structure Prediction by LAQA
3. 学会等名 IUMRS-ICYRAM 2022（国際学会）
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

CrySPY https://github.com/Tomoki-YAMASHITA/CrySPY CrySPY document https://tomoki-yamashita.github.io/CrySPY_doc/

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
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------