

第一原理計算を用いた非調和フォノン特性データベースの構築

大西 正人

統計数理研究所 先端データサイエンス研究系

1. はじめに

フォノンは原子の集団振動であり熱輸送特性に大きな影響を及ぼす。一方、フォノンは電子、マグノン、フォトンなどと相互作用することにより、熱物性以外にも電気、磁気、光学、機械特性などに影響を及ぼす。近年、マテリアルズ・インフォマティクス (MI) 技術が急速に発達しており、情報科学と材料科学を融合した様々な材料開発手法により、電池、触媒、磁性材料などに関する物性において成果が挙がっている。また無機材料のデータベースとして Materials Project、AFLOW、AtomWork などの従来のデータベースに加え、2023 年末頃から DeepMind[1]、META[2]、Microsoft[3] などが超大規模データベースを発表するなどデータ量の増強も加速している。しかし、特に従来のデータベースでは、主に電子物性（バンドギャップ、電子バンド構造等）をまとめたものであり、熱機能材料に関しては熱伝導率などのフォノン非調和物性のデータが不足している。

非調和フォノン物性の計算には多数の超格子構造の力計算が必要であるため計算コストが高くなるが、近年の計算能力の発達と HPCI を含めた我が国の優れた計算環境により、非調和フォノン特性の大規模計算も可能になってきた。そこで申請者らは世界中の共同研究者と協力して、非調和フォノン特性の解析プロセスの標準化とそれを組み込んだ自動化計算ソフトウェアの開発、さらにそれを利用した非調和フォノン特性データベース構築を進めている。本年度は、自動計算によるデータベースの構築を進めながら、構築したデータベースと機械学習を用いたスペクトル熱伝導率などの非調和フォノン特性の予測モデル構築を進めた。過去の自動計算も合わせ、約 6,000 材料の非調和フォノン特性の計算に成功し、この

データとグラフニューラルネットワークを用いた非調和フォノン特性の予測を行った。また、構築した予測モデルを用い、DeepMind が発見した約 38 万の新材料データベースから低熱・高熱伝導材料のスクリーニングを行うことで、高性能な熱機能材料の候補材料を特定した。

このように、本研究で構築した非調和フォノン特性データベースを用いることで、高精度な非調和フォノン特性の予測が可能であることを示した。今後は、電子・フォノン相互作用や高次の非調和フォノン相互作用などを考慮したデータを追加することで、より広範囲な材料開発への発展が期待できる。

2. 第一原理非調和フォノン特性解析

本研究では図 1 に示すような、第一原理非調和フォノン特性解析プロセスを自動化した[4]。まず、与えられた任意の結晶構造の基本単位胞 (primitive cell)、慣用単位胞 (conventional cell)、超格子構造 (原子数 150 個以下) を決定する。次に、フォノン計算では力の計算精度が重要なため、複数ステップの構造緩和計算を行う。このように用意した構造を用いてフォノン特性を計算する。複素周波数が発生した場合は、さらに厳しい構造緩和計算、大きな超格子構造の利用、4 次の非調和効果によるフォノン繰り込みの考慮などを行う。これまでの自動計算ではフォノン繰り込みは考慮しておらず、最も基本的な計算方法である、

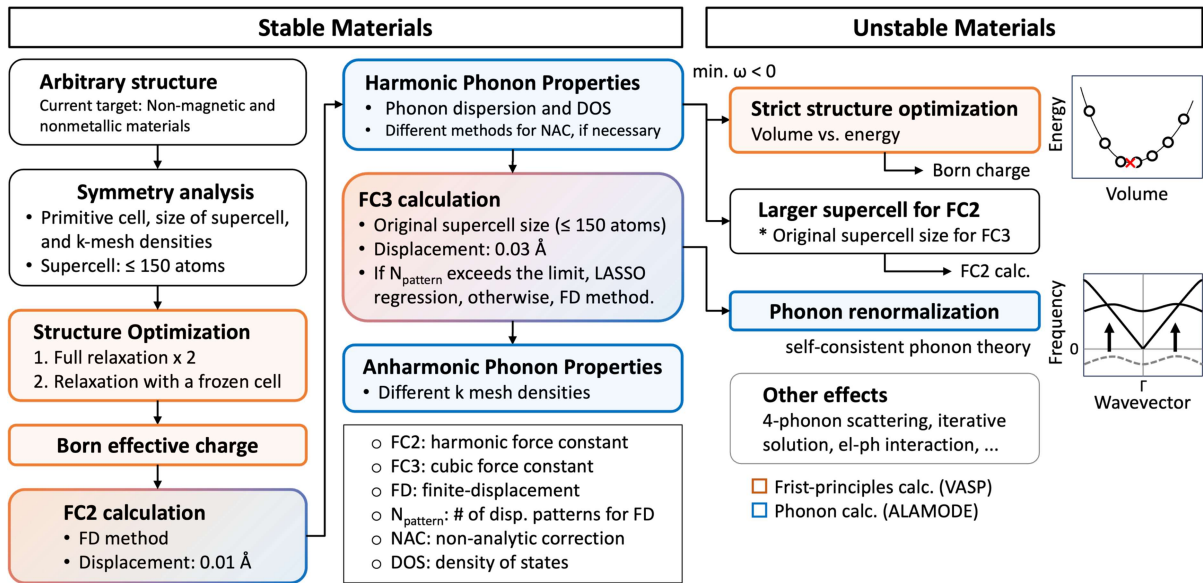


図 1. 自動化した第一原理フォノン計算プロセス

緩和時間近似を用いてボルツマン輸送方程式を解くことで 3 フォノン散乱の散乱頻度を計算している。対象とした材料は、Phonondb に含まれる全て材料と Materials Project に含まれる材料のうち非金属・非磁性材料で合わせて 21,452 材料であり、そのうち約 6,000 材料の非調和フォノン特性の計算が終了している。

上記の第一原理フォノン特性解析には、第一原理計算に VASP [5]、フォノン計算には ALAMODE[6] を用いており、Python を用いた自動化や結晶構造の操作には、ASE[7]、Pymatgen[8]、Spglib[9]、Pymatgen[8]、Phonopy[10, 11]、Custodian[8]、SeeK-path library[9, 12] などの Python ライブラリを利用している。

3. 計算結果と考察

3-1. 深層学習スケーリング則

自動計算により得られた約 6,000 材料の非調和フォノン特性のデータセットと機械学習を用いて格子熱伝導率 (κ_p : p は粒子性を表す添字) の予測をした。機械学習には CGCNN (crystal graph convolution neural network) [13] を用いた。学習データに用いるデータ数を変化させた時に、データ

サイズと誤差 (mean absolute error) をプロットしたところ図 2(a)左図のように、対数プロットで直線となるようなスケーリング則 [14] が確認された。スケーリング則は $(MAE) = (A/N_{\text{train}})^{\alpha}$, ($A, \alpha > 0$) と表されるが、今回観察されたスケーリング則の強さ (係数 α) は 0.125 であり、大規模言語モデル (0.095) [14] や結晶材料の力の予測 (0.21) [1] で得られたスケーリング係数と同程度の値となった。

それぞれのデータ数に対する第一原理計算と予測値を比較したところ、図 2(a)右図に示すように、予測値と計算値がよく一致していることが分かる。データ数が多いほど計算値と予測値が一致する線 (黒線) 上にデータが集まっており、データ数の増加に伴う予測精度の向上が確認できる。

スケーリング則が観察されたことは、データ数を増加させると誤差が減少を続けることを意味しており、今後のさらなるデータ数増強による予測精度向上が期待される。一方で、例えば、MAE を 0.2 まで減少させるためには、200,000 材料以上の非調和フォノン特性のデータが必要になる。この計算コストは現在利用している、全て第一原理計算で力の計算をする手法では現実的ではなく、今後は機械学習ポテンシャルを利用したデー

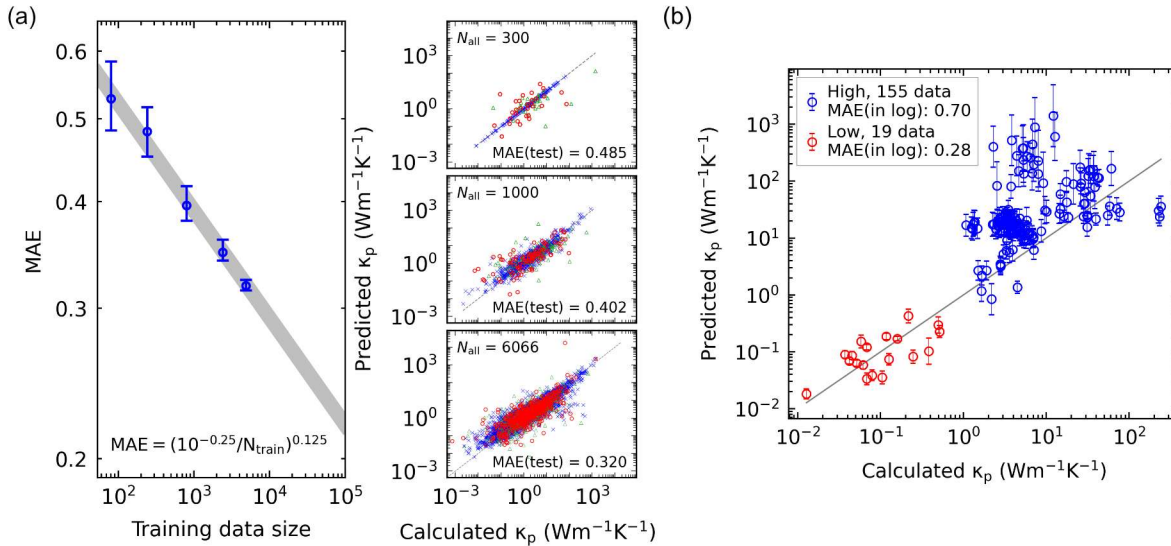


図 2. 構築した非調和フォノンデータベースを用いた格子熱伝導率の予測。(a) スケーリング則の実証。学習データサイズを変化させた時の平均誤差 (mean absolute error, MAE) の変化(左)。MAE は $\log \kappa_p$ に対して計算した値であり、エラーバーは 20 回のアンサンブル平均の 90%信頼度を表す。灰色の線は図中に記載したフィッティング曲線である。異なるデータ数における第一原理計算値と予測値のパリティプロット (右)。データ数がデータ数が多い場合に、計算値と予測値が一致する直線周りにデータが集中していることが分かる。青、赤、緑のマーカーは、それぞれ学習 (training)、検証 (validation)、テストデータの結果である。(b) 予測モデルと DeepMind の GNoME データベースを用いたスクリーニングの結果。

タ数増強も検討する必要がある。

3-2. 低熱・高熱伝導材料の探索

上記のように GNN を用いて構築した予測モデルを適用して、DeepMind の 38 万の新材料結晶構造データベース[1]から低熱・高熱伝導材料をスクリーニングした。その際の予測結果と予測後に第一原理計算をした値を比較すると、図 2(b) に示すように、高熱・低熱伝導材料の分類は可能であることが確認できた。また、低熱伝導材料(赤いプロット)では MAE (mean absolute error) が 0.28 となり、6,000 材料全て使った場合 (MAE ≈ 0.32) よりも予測精度が高かった。一方で、高熱伝導材料(青いプロット)では MAE が 0.70 となり予測精度が低くなった。どちらの場合も、予測がより低いあるいは高い材料に着目していることから、平均の予測精度よりは悪くなることは問題ない。その上で、高熱伝導に関しては、今回

の計算では 4 フォノン散乱の影響を考慮していないことから、学習データの理論値に誤差が乗りやすいことが原因で、予測値が実際の値を過大評価しやすいと予測される。

いずれにしても、本課題では、6,000 材料といった大規模データを用いることで、熱機能性材料の予測が可能になることを実証に成功した。今後は、熱伝導率などのスカラー値だけでなく、モード依存のフォノン緩和時間など複雑なフォノン特性の予測も進める予定である。

4. おわりに

本プロジェクトは 2022 年頃から始めたものであり、ようやくある程度の量のデータが集まり機械学習を行うことが可能になった。本報告書には熱伝導率の値のみを記載したが、構築したデータベースを用いてスペクトル熱伝導率やフォノン分散の予測も可能である。これらをまとめた成果

は近日中に学術論文として投稿予定であり、構築したデータベースも併せて公開予定である。今後は、高次の非調和ポテンシャルの考慮や電子・フォノン相互作用などのデータも追加する予定である。データ数の増強には機械学習ポテンシャルの利用も検討しているが、これらの特性を精度良く解析するには現状では第一原理計算が必要となる。構築したデータベースは熱物性のみではなく、電子・磁気・光学・機械物性など様々な分野の材料開発に貢献すると期待される。

参考文献

- (1) A. Merchant et al., *Nature*. **624**, 80–85 (2023).
- (2) L. Barroso-Luque et al., *arXiv* (2024), doi:10.48550/arxiv.2410.12771.
- (3) J. Li et al., *arXiv* (2025), doi:10.48550/arxiv.2503.11568.
- (4) M. Ohnishi et al., (2025), doi:10.48550/arxiv.2504.21245.
- (5) G. Kresse and J. Furthmüller, *Phys. Rev. B*. **54**, 11169–11186 (1996).
- (6) T. Tadano et al., *J. Phys.: Condens. Matter*. **26**, 225402 (2014).
- (7) A. H. Larsen et al., *J. Phys.: Condens. Matter*. **29**, 273002 (2017).
- (8) S. P. Ong et al., *Comput. Mater. Sci.* **68**, 314–319 (2013).
- (9) A. Togo et al., *Sci. Technol. Adv. Mater.: Methods*. **4**, 2384822 (2024).
- (10) Togo et al., *J. Phys.: Condens. Matter*. **35**, 353001 (2023).
- (11) Togo, *J. Phys. Soc. Jpn.* **92**, 012001 (2023).
- (12) Y. Hinuma et al., *Comput. Mater. Sci.* **128**, 140–184 (2017).
- (13) T. Xie and J. C. Grossman, *Phys. Rev. Lett.* **120**, 145301 (2018).
- (14) J. Kaplan et al., *arXiv* (2020), doi:10.48550/arxiv.2001.08361.