

jh190019

全電子混合基底第一原理計算法を活用したネットワーク型エネルギー絶対値算定マテリアルズ・インフォマティクス

川添 良幸 (東北大学)

概要

マテリアルズインフォマティクス (MI) を適用した材料研究が盛んになされる様になった。そこでは、計算量が比較的少なく多数のケースが算定出来る密度汎関数理論 (DFT) が適用されることが多い。しかし、本研究では、光発光素子用材料の理論的探索を目的とするため、バンドギャップ値 (イオン化ポテンシャル、電子親和力) の絶対値算定が必須である。そのため、計算量は大きくなるが理論的に正しくバンドギャップ値が算定可能な GW 計算をベースとし、その結果と DFT 計算結果を機械学習によって統合化することで、高精度・高効率 MI 技法の確立を図る。

1. 共同研究に関する情報

(1) 九州大学情報基盤研究開発センター

(2) 共同研究分野

超大規模数値計算系応用分野

(3) 参加研究者の役割分担

川添良幸	研究統括
水関博志	MI に依る新規材料探索
大野かおる	TOMBO プログラム開発
佐原亮二	TOMBO プログラム実行
本郷研太	MI プログラム作成・実行
南里豪志	システム高度化チューニング
Abhishek Kumar Singh	MI プログラム実行

2. 研究の目的と意義

機械学習技法を用いて、ゲーム、交通制御、市場予測、医療診断、材料設計等が人間以上の高速・高信頼性を持って実現される兆しが強くなっている。これらの中で、元々人間が作成・関与する事象は、データ収集や解析により極めて高効率に望ましい結果を得られる様になっている。しかし、自然を対象とした場合、人間の知り得る範囲は限られており、格段に困難となる。特に、最近喧伝されるマ

テリアルズ・インフォマティクス (MI) において、密度汎関数理論 (DFT) により多数のシミュレーション計算を行い、発光特性の予測に基づく新材料設計開発を効率化する、という方法論には大きな問題がある。DFT は基底状態の理論であり、励起状態の準位エネルギーは算定出来ない。しかし、発光波長を算定しているのは、最低伝導帯 (励起状態) と最高荷電子帯のエネルギー差であり、計算技法上発生する数値をパラメータを使って実験に合わせる (交換相関相互作用のパラメータ化) という現象論に基づくものである。多数のシミュレーション計算を実施していることを理由に正当性を議論すべきレベルの問題ではなく、理論的には明らかに予測能力のない方法である。更に、擬ポテンシャルを使うため、エネルギーの絶対値算定が出来ず、そのため、電子親和力やイオン化ポテンシャルの絶対値算定は不可能である。本研究では、我々独自開発の全電子混合基底第一原理シミュレーション計算法 (TOMBO、TOhoku Mixed-Basis Orbitals Ab initio Simulation Package) の GW 計算を MI 技法と結合し、これらの問題を抜本的に解決し、信頼性ある MI 技法の確立を目的とする。また、

副代表者所属の韓国科学技術研究院 (Korea Institute of Science and Technology, KIST) 計算科学研究センターで開発中の材料シミュレーション・プラットフォームを活用し、本研究成果を世界的に公開することも目的とする。

3. 当拠点公募型研究として実施した意義

本研究グループは、物性物理及び計算材料学の研究者の共同研究体制であり、独自の第一原理計算法の定式化や MI 技法の材料探索への適用を行ってはいるが、計算科学の専門家を必要としている。特に、TOMBO は平面波と原子軌道関数を含むオーバーコンプリートな基底系を採用しているため、スーパーコンピュータでの高効率稼働のためのチューニングに専門家の支援が必要である。TOMBO は従来 Power 系のシステム上でチューニングされており、広範な研究者の利用を促進するためには Intel 系マシンでのチューニングが必須である。平成 30 年度の JHPCN-Q プログラムとして、九州大学南里准教授の努力で、TOMBO の九州大学情報基盤センターの ITO スーパーコンピュータ・システムへの移植を完了し、36 並列および 144 並列での実行時間による実効並列化率を 99.8% (並列化効率 81.4%) まで向上させることが出来た。また、MI プログラムとの連携における高効率化技法や結果の可視化に関しても、ご支援をいただき、高度化・公開を試みる事が出来た。今回はネットワーク経由で、第一原理計算 (九大)、MI (北陸先端大)、公開プラットフォーム (KIST) を統合利用する計画を立て、この部分へのセンターご支援もお願いした。

4. 前年度までに得られた研究成果の概要

全電子混合基底法第一原理シミュレーションプログラム TOMBO は、本研究グループが 25 年以上の歳月を掛け、定式化から独自開発している我が国では希有の独自定式化・開発

の材料設計用ソフトウェアである。特に、通常の擬ポテンシャルを使う計算法とは異なり、絶対値でエネルギー準位の算定が可能である。絶対値でエネルギー算定の可能な GW 近似プログラムは世界的に見ても TOMBO のみである。TOMBO は発光波長=バンドギャップ値の実験と一致させるために現象論的なパラメータを用いている他の方法とは一線を画し、パラメータを一切使わずに絶対値算定が可能であり、本研究の基盤として活用する。信頼性ある材料データベース構築と発信に関しては、研究代表者が非平衡系材料データベース構築に携わって 25 年以上になり、ドイツ国 Springer 社の Landolt-Boernstein (LB) シリーズとして出版を継続している。(現在までに 6 冊の刊行を済ませ、3 冊が完了間近な状態にある。(オンラインでのデータ提供は Springer 社から実施中である。) 我が国では LB シリーズを使う研究者は多いが、Chief-Editor は川添を含め極めて限られている。Springer 社の編集担当者との連携でデータの信頼性に関する多くの知見を得ており、その方策を本研究へ適用する。我々は 15 年前に既に多数の原子の組合せに対する第一原理シミュレーション計算 (DFT) を実施し、その結果を踏まえて実験グループが新規有用材料を開発するという体制的研究を実施していた (その成果は既に実用化されている)。世界的に見ても、現在の MI 研究者は、この当時の第一原理計算レベル (DFT) を基礎とした研究に MI 技法を適用するレベルに終始しており、本研究では、我々は独自に開発している信頼性の高い研究方策としての GW 近似を採用することにより、他研究グループとの抜本的な差別化・高度化が実現できる。しかし、如何にスーパーコンピュータが高速化したと言っても GW 計算には $O(N^5)$ 以上、(N=電子数) の計算量がかかるため、電子数が 100 程度の小規模系に限られる。この問題の抜本的改善のため、限られた数の

GW 計算に多数実行する DFT 計算結果を機械学習を活用して統合する。本研究成果の公開は、副代表者の水関が所属する KIST のウェブベースの材料シミュレーション・プラットフォームでの実績がある。

5. 今年度の研究成果の詳細

5.1. 第一原理計算を用いた結晶構造と機械学習用データの作成

III-V 族化合物半導体の組み合わせとして、III 族元素は B, Al, Ga を、V 族元素は N, P, As を計算対象にした。結晶構造は全組成で Zinblende 構造および Wurtzite 構造の単相であると仮定して、Zinblende 構造は $2 \times 2 \times 2$ 、Wurtzite 構造は $2 \times 2 \times 1$ のスーパーセルを計算に用いた。両構造ともに III 族、V 族は 8 個ずつの原子を含むため、各元素の組成を 12.5% 刻みで表現可能である。V 族サイトに 1 種類の元素を置いた組成、例えば、N 原子のみの nitride に限定すると III 族側の組成の組み合わせは Zinblende 構造、Wurtzite 構造ともに 45 通りである。2 元系、すなわち、III 族、V 族ともに 1 種類の元素の場合には、スーパーセル内の原子配置は 1 通りしか存在しないが、3 元系以上の場合には、同じ組成でも複数の原子配置が可能である。計算量の都合上、Energy convex hull により、エネルギーの最も低い構造を求めることはできない。本研究では、V 族サイトに N のみ、P または As を 12.5% 刻みの組成の原子配置の各 45 通りの組成の構造を、イオン半径の分布が最も小さくなるように決定した。これらの組成の原子構造に対して、VASP コードを用いて PBE 汎関数で構造最適化を行い、その構造に対して mBJ 汎関数によるバンドギャップを求め、機械学習用データを作成した。

5.2. 機械学習モデル構築と適用

電界効果トランジスタ、太陽電池材料、発光素子等の分野において幅広く利用されている

III-V 族化合物半導体は III 族元素と V 族元素の組み合わせにより、その物性値が連続的に変化することが知られている。発光素子の場合、発光波長の制御には、特定のバンドギャップと格子定数を実現できる、最適な組成を見出すことが求められている。しかし、バンドギャップの値を正確に求めるためには計算コストの高い GW 計算を必要とすることから、III-V 族化合物半導体を網羅的に探索することは難しい。それを解決するために本研究では、ニューラルネットワークを用いた物性予測コードを作成している。現在までに十分な数の GW 計算のデータを得られていないため、mBJ 汎関数によるバンドギャップ値を教師学習用のデータとして用いた。図 1 は組成に相当するスーパーセル内の元素毎の原子数をもとにバンドギャップを予測するモデルを適用した結果である。

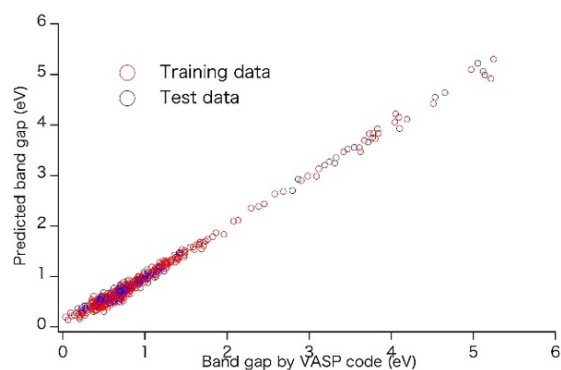


図 1 ニューラルネットワーク(NN)を用いた Wurtzite 構造 III-V 族化合物半導体バンドギャップ予測 横軸は第一原理計算で得られた値、縦軸は予測値 赤丸は Training data (400 個)、青丸は Test data (50 個) を示す。

Wurtzite 構造の $B_xGa_{1-x}N$ は B が低濃度の組成のみ単相の実現が容易であることが知られている。本研究では最大 $4 \times 4 \times 4$ のスーパーセル ($B_1Ga_{127}N_{128}$) を用いて B 濃度 0.78% までのバンドギャップを求めた。実験グループから報告されている $B_xGa_{1-x}N$ が GaN よりもバンドギャップが小さくなる振る舞いは見られず、

他の理論計算と同様に B 濃度がゼロに向かってバンドギャップの値が GaN に近づく様子が見られた。

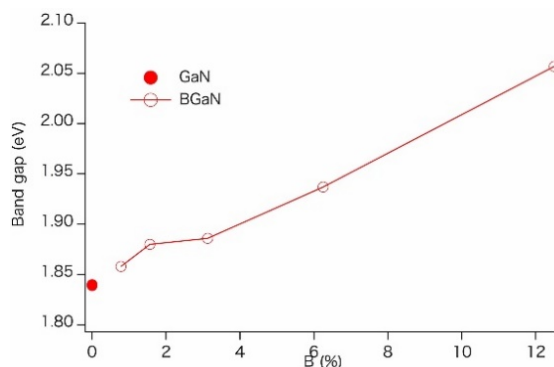


図 2 B_xGa_{1-x}N のバンドギャップの組成変化

5.3. 種々の機械学習モデル構築と検証

前節に示したニューラルネットワーク (NN) では、物性の予測精度の信頼性は担保されているが、複雑な非線形変換のために、物性チューニングに寄与する記述子を同定することは困難である。目的の物性に対して主要な寄与を与える記述子を同定できれば、物質設計の指針となり得る。本研究では、そのような記述子解析の観点から、NN とは別に、解釈可能性の高い機械学習モデルの構築にも取り組んでいる。本研究では、そのような機械学習モデルとして、ランダムフォレスト (RF) モデルによるベンチマークを実施した。図 3 と図 4 は、それぞれ、RF による Wurtzite のバンドギャップの予測性能と記述子の重要度を示している。図 3 から、RF の予測性は NN に比べて若干劣るものの、その相関係数は 0.98 程度と十分な精度で予測ができており、物性予測モデルとして十分利用可能であると考えられる。更に、NN と RF の大きな違いは、記述子の解釈可能性にある。すなわち、RF では、図 4 に示すように、各種記述子の重要度を含めて、学習が可能である。今後の展開として、各種記述子セットを考慮した RF モデルを利用して、物質設計に利用可能な記述子セットを同定する予定である。

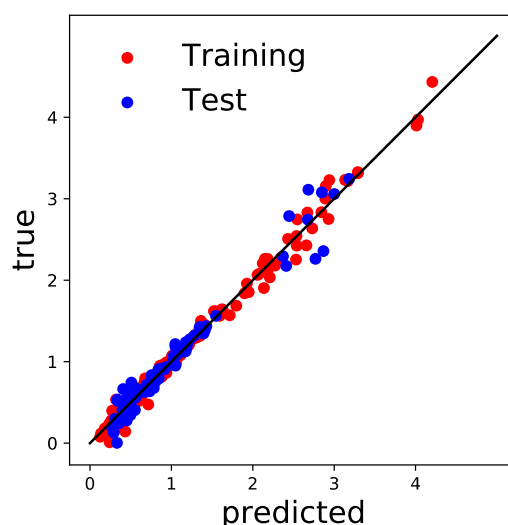
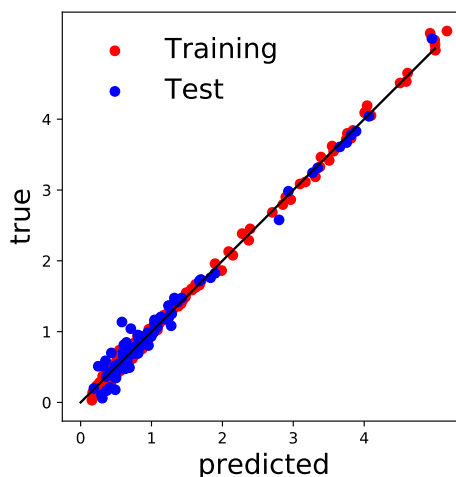
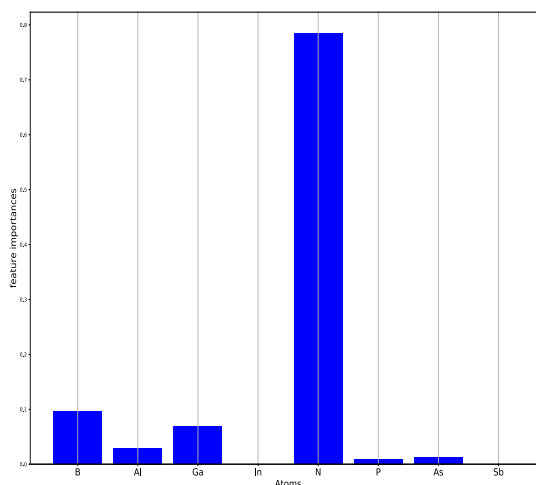


図 3. (上) Wurtzite、及び(下) Zincblende のバンドギャップに対するランダムフォレストの予測性能 (Wurtzite では $R^2 = 0.98$ 及び Zincblende で $R^2 = 0.96$) .



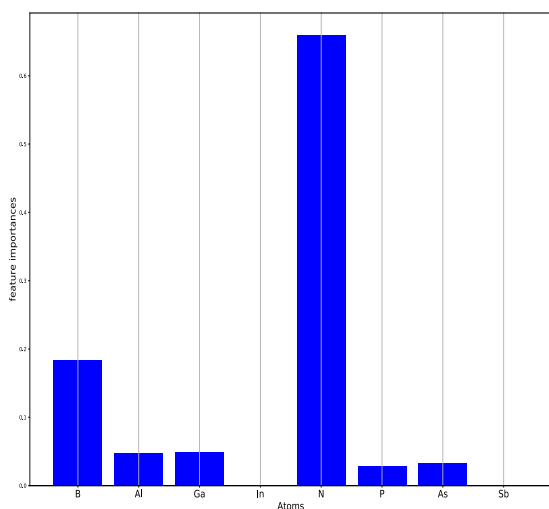


図 4. RF モデルにおける (上) Wurtzite、及び (下) Zincblende の各記述子 (元素) の重要度

図 4 の横軸は今回の記述子の元素であるが、興味深いことは、Wurtzite と Zincblende で、バンドギャップに対するそれらの寄与度が異なる。窒素は一定にしているので除外すると、Zinc Blende において明らかにホウ素の寄与が目立つ。ガリウムの寄与に大きな違いは見られない。現在、元素以外の記述子として、電気陰性度等の基本的な化学情報を試し、評価する予定である。

6. 今年度の進捗状況と今後の展望

平成 31 年度 (令和元年度) は、スタート時点で外国居住日本人研究者への連絡が遅れ、さらに追加インド人研究者の登録に時間が掛かり、実際の研究は 9 月になってやっと本格化出来た。さらに 1 月末までで終了であったため、実質半年以下での研究となり、特に繁忙期に重ったため予定していたシミュレーション計算が進まなかった。特に、最初に予定した GW 計算は極めて限られた件数の計算にしか実施出来ず、MI への適用は見送らざるを得なかった。大量の計算時価を要する GW 計算は、システムのキュー設定上、処理までの待ち時間が長く、ほぼ進まなかった。ただ、

計算の進展、特に共同研究者の大阪大学実験グループとの打合せによる内容的充実には大きなものがあつた。現在、まとめと論文化のための議論を継続している (外国との連絡、さらにコロナウィルスの問題で TV 会議を活用)。MI 研究は現在継続中であり、本共同研究で得られた成果を順次公表予定である。

7. 研究業績一覧 (発表予定も含む)

(1) 学術論文 (査読あり)

1. Y. Kawazoe, H. Mizuseki, K. Hongo, and N. Sarukura, “MI applied to understand deeply the properties in $Ga_xB_{1-x}N$ ”, in preparation.

(2) 国際会議プロシーディングス (査読あり)

1. 2019 年 6 月 28 日、Y. Kawazoe, ICMAT - 2019, Singapore, “Materials Informatics based on Reliable Database” (Invited Talk)
2. 2019 年 7 月 10 日、Y. Kawazoe, MRS-Thailand, Pattya, “Materials Informatics based on Reliable Database”, (Keynote Talk).
3. 2020 年 2 月 5 日、Hiroshi Mizuseki, Babu Ram, Ryoji Sahara, Yoshiyuki Kawazoe, and Kenta Hongo, “Machine Learning on Band Gap Prediction of III-V Compound Semiconductors” ACCMS - International Conference on Materials Genome, ACCMS-ICMG 2020, SRM University, AP - Amaravati, India, (Invited Talk).
4. 2019 年 7 月 26 日、Kenta Hongo, “Computational Materials Design from *Ab Initio* Simulations to *Ab Initio* Materials Informatics” ACCMS10, Hong Kong, China, (Invited Talk).
5. 2019 年 9 月 2 日、Kenta Hongo, “Data-Driven Approach to Computational

Materials Design” IUMRS2019, Perth,
Australia, (Invited Talk).

(3) 国際会議発表 (査読なし)

なし

(4) 国内会議発表 (査読なし)

1. 2020年3月12日、本郷研太、「計算科学とデータ科学の協働による物質探索の進展」、凝縮系の理論化学、沖縄、(招待講演)。

(5) その他 (特許, プレスリリース, 著書等)

なし