

# 研 究 報 告 書

## 「機械学習と第一原理計算による新規スペクトル解析技術の確立」

研究タイプ: 通常型

研究期間: 2016 年 10 月～2020 年 3 月

研 究 者: 池野 豪一

### 1. 研究のねらい

現代の材料開発においては、物質の特性や化学反応を支配する特徴的な原子構造(格子欠陥・界面・表面など)を解明し、制御することが求められる。内殻電子分光法は物質中の電子が光子・電子と相互作用することにより、量子的なエネルギー準位間を遷移する過程を観測する手法であり、物質中の特定元素周辺の局所的な原子構造・電子状態を反映したスペクトルが得られる。最近では測定装置や技術の発達により、透過型電子顕微鏡を用いた 1 原子カラム毎のスペクトル測定や、化学反応の追跡するための時間分解分光測定の発展など、分光測定から得られる情報は飛躍的に増加している。

一方で、「得られたスペクトルを如何に解析するか」については現在の大きな課題となっている。最も簡便な方法は参照物質の実験スペクトルと比較する指紋照合法であるが、その比較方法は人間の「目視」による極めて原始的かつ定性的なものである。より信頼性の高い解析手法としては、第一原理計算を用いた理論指紋照合法がある。この方法では、候補となる原子配列に対して電子状態計算から理論スペクトルを求め、実験結果と照合することで局所原子配列・電子状態を明らかにする。高精度な解析が可能となる反面、個々のスペクトルの理論計算に必要なコストは高く、最新の分光実験で得られる大量のスペクトルデータ全てに対して第一原理計算を実施するには膨大な時間を要することになる。これらの分光実験で測定されるスペクトルデータが持つ情報を最大限活用し、材料開発に応用するためには、スペクトル形状と原子構造の関連を定量的かつ高速に予測することができる新しい手法の開発が求められる。

本研究のねらいは、第一原理計算と情報科学的学習を組み合わせ、大量のスペクトルデータを効率的かつ定量的に解析し、局所原子構造・電子状態に関する情報を抽出することができる新しい解析技術を確立することにある。電子状態計算を経由せずに、原子構造からスペクトル形状を直接予測できる回帰モデルを作成することにより数百～数千の構造に対するスペクトルを短時間で計算し、候補構造を絞り込むことができる。これにより、測定される大量のスペクトルデータが持つ情報を最大限活用することが可能となり、材料開発・設計につながる新しい知見が得られると期待される。

### 2. 研究成果

#### (1) 概要

本研究目標を達成するにあたり、以下の 3 つの研究テーマを設定した。

- (1) 遷移金属  $L_{2,3}$  端 X 線吸収スペクトルデータベースの作成
- (2) 圧縮センシングによるピーク分離法および実験・理論スペクトルの定量照合法の確立
- (3) 機械学習によるスペクトル予測モデルの構築

研究テーマ(1)について、遷移金属  $L_{2,3}$  端は、空間的に局在化した 2p,3d 軌道間の電子遷移を観測するものであり、応用上重要な遷移金属化合物の物性に関与する電子準位の情報が得られる。しかし、その計算にはこれらの軌道を占有する電子間に働く強い相互作用をあらわに取り扱う必要がある。計算には、研究者らが開発を進めてきた相対論多電子計算を用いる。高精度なスペクトルの予測モデルを作成するには、数万を越えるデータが必要であるため、計算の高速化を実現するため、幾つかのアルゴリズムの開発と実装を行った。これら新規に開発したプログラムを利用した自動計算システムを構築し、価数・配位子の種類・配位数・対称性が異なる遷移金属化合物のスペクトルデータの収集を進めている。研究期間内では、機械学習に用いるに十分なデータ数を収集することができなかったため、今後も継続してデータベース作成を進める。

研究テーマ(2)に関しては、圧縮センシングアルゴリズムの一つである ESPRIT 法を利用したピーク分解法を開発した。この方法では、スペクトルをローレンツ関数の和として分解し、スペクトルをピーク数( $J$ )、ピーク位置( $s_j$ )、ピーク幅( $\gamma_j$ )で表現し直すことにより、その「一致度」数値的に評価することを試みた。

研究テーマ(3)について、スペクトルの機械学習法の検討を行った。ここで機械学習の目的変数はスペクトルそのものであり、これは入射光のエネルギーに対する非負値の連続関数という特徴を持つ。このような目的変数に対する回帰モデル作成は殆ど行われておらず、回帰手法そのものの確立が必要となる。本研究では、既存の電子状態計算データベース(Materials Project)で提供されている X 線吸収スペクトルデータベースを利用して、様々な回帰手法の検討を行った。その結果、多層ニューラルネットワークモデルを用いて、8 割程度の高い確立でスペクトルの概形を、その原子配列から直接予測することが可能であることを示した。

## (2) 詳細

### 研究テーマ(1) 遷移金属 $L_{2,3}$ 端 X 線吸収スペクトルデータベースの作成

相対論多電子計算法の速度向上を目指して電子間相互作用を表す二電子積分の高速計算アルゴリズムの開発を行った。新たに開発した方法では、電子密度の多重極展開を行い、二電子積分をフーリエ空間で実行することにより、一次元振動の和として求める。また、動径関数をテーマ(2)で開発した ESPRIT 法を用いて指数関数の和としてフィッティングすることで、振動積分を解析的に、かつ高精度に計算できる手法を開発した。

また、異常高原子価イオンを含む化合物に対しては内殻電子の電子励起と同時に配位子から励起中心の遷移金属イオンへ電荷移動がスペクトル形状に大きく影響を与える。この配位子からの電荷移動は、配置間相互作用計算において新たな電子配置を付け加えることで計算が可能となるが、大規模なハミルトニアン行列の固有値問題を解く必要がある。本研究では、行列固有値の反復解法を用いて遷移の始状態の波動関数のみを求め、シフト線形方程式に基づいて系の Green 関数を高速に計算できるアルゴリズム(修正 RSCG 法)を開発した。これにより、多様な系に対して効率的かつ高精度なスペクトルを求めることが可能となった。

これらの手法を実装した上で、クラスター計算機上に自動計算システムを構築し、スペクトルデータベースの作成を試みた。しかし、研究期間内では、各遷移金属元素に対して得られ

ているスペクトルは千程度と、機械学習に用いるに十分なデータ数には至っていない。研究期間後もスペクトルデータの収集を継続的に実施し、機械学習モデルの作成を行う予定である。

### 研究テーマ(2) 圧縮センシングによるピーク分離法および実験・理論スペクトルの定量照合法の確立

圧縮センシングアルゴリズムの一つである ESPRIT 法を実装した。本アルゴリズムは、フィッティングパラメータの数、初期値に関する事前情報を必要とせず、等間隔メッシュ上で観測されるデータのみから指数関数の和として関数を復元できるという特徴がある。スペクトルデータをフーリエ変換し、指数関数の和でフィッティングした後、逆フーリエ変換を施すことでローレンツ関数の和としてピーク分離ができることを示した。当初の計画では、ピーク分離結果を元に、そこからの実験・理論スペクトルの変換関数を作成する計画であった。しかし、検討を進めた結果、分離後のピーク情報からスペクトルの一致度を判定することが困難であり、適切な変換関数の作成には至っていない。ただし、本手法はノイズのあるデータに対しても使用できること、ピーク位置に関する事前知識も不要という特徴から、実験スペクトルのピーク分離等、別の用途に活用できると考えられる。

### 研究テーマ(3) 機械学習によるスペクトル予測モデルの構築

Materials Project が提供する多重散乱理論(FEFF コード)に基づいて得られた O-K 端 X 線吸収スペクトルを収集し、結晶構造から理論スペクトルを直接予測する機械学習モデルの構築を試みた。まず、カーネル Ridge 回帰を使ってスペクトルの形状の直接予測を試みた。入力には物質の化学組成と、動径分布関数やクーロン行列といった原子構造から求められる記述子を用いた。回帰を行う前に、各エネルギーで離散化したスペクトルデータに対して非負値分解を行い、スペクトルを少数の独立成分の線形和として表現し、その重み係数を回帰の対象とすることで、計算量の削減を図った。約 2500 のデータに対してこの手法を適用したところ、テストデータに対する予測精度を表す決定係数( $R^2$ )の平均は 0.43 となり、十分な汎化性能のあるモデルの構築はできなかった。

そこで、学習に用いるデータ数を増やし、約 15 万のスペクトルデータに対する機械学習を試みた。データ数が 5 万を越えたあたりから非負値分解の計算コストが激増し、先の手法を適用することは困難となった。そこで、多層ニューラルネットワークを用いた予測モデルの作成を行った(図 1 参照)。この方法では、データベースからランダムに少数のスペクトルを選択し、ニューラルネットワークの重みを学習させる過程を繰り返すことで、予測精度を逐次更新することができる。ネットワークの層の数やバッチサイズ、エポック数などのハイパーパラメータについては、ベイズ最適化を用いて決定した。その結果、約 4 分の 3 のスペクトルに対して主なピークの位置や強度など、スペクトルの主な特徴を捉えた予測結果を得ることができた。一方、残りの約 4 分の 1 については、予測結果とオリジナルの理論スペクトルとの間に大きな誤差が見られた。データを検証したところ、実験的には不安定な仮想物質に対する予測精度が低いという傾向が見られたが、学習データに類似の構造からのスペクトルを含めることで、精度の改善が期待できる。

更に高精度なスペクトル予測モデルの構築を目指して、crystal graph convolution neural network (CGCNN) を用いた予測モデルの作成を行った。この手法では、結晶構造をグラフ構造として取り扱う。頂点に原子の特徴量、辺に原子間の結合を表す特徴量を保持している。さらに畳み込みニューラルネットワークを用いることで、目的変数に応じた特徴量ベクトルの最適化を行うことができる。前述のニューラルネットワークと同様のデータセットを用いて学習を行ったところ、同程度の精度のスペクトル予測モデルが得られた。

以上の結果から第一原理計算なしに、原子構造から直接スペクトル形状を捉える予測モデルを作成することが可能であるという確信を得た。

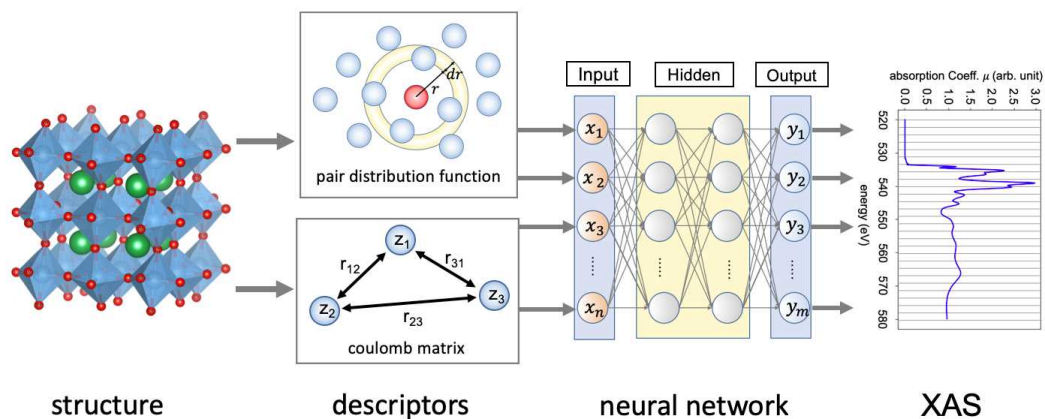


図 1 ニューラルネットワークを用いた XAS 予測モデルの概念図。電子状態計算なしに、原子構造から直接スペクトル形状を予測する。

### 3. 今後の展開

本研究を通して、原子配列から電子状態計算なしでスペクトル形状を予測することが可能であるという知見を得た。今回対象としたものでは O-K 端のみであるが、他の元素の任意の吸収端においても、十分な学習データを集めることができれば同様の手法ができると考えられる。したがって、今後の研究においては如何にして効率よくデータを収集するかが課題となる。本研究で実施していた遷移金属  $L_{2,3}$  端 X 線吸収スペクトルは複雑な物理現象を扱う必要がある。継続的な研究テーマとしては、このデータベース作成を継続し、機械学習モデルの作成を行うことが挙げられる。また、遷移金属  $L_{2,3}$  端の理論・実験スペクトルデータベースは存在しないことから、将来的にデータベースの公開も視野に入れる。

また、放射光分野において、実験スペクトルのデータベース作成に向けて議論が進められている。実験スペクトルの数は計算で得られるものよりもデータの数やバリエーションは下がるものと考えられる。これらのデータから如何に効率良く機械学習モデルを作成するかは、多岐に渡る科学分野において内殻電子分光法を活用する上で大事な研究テーマである。



#### 4. 自己評価

本研究の鍵となるのは、電子状態計算なしに、原子配列から直接 X 線吸収スペクトルを予測する機械学習モデルの構築である。本研究では、このようなモデルの作成が可能であることを実証できたことから、研究目的の根幹をなす部分については達成できたものとする。

研究成果の欄に挙げた 3 つの研究テーマそれぞれに関しての達成度合いを見ると、まずテーマ(1)のデータベース作成については時間的要因により達成には更なる時間を要する。テーマ(2) に関しては、圧縮センシングを用いた事前知識を必要としないピークの分離方法については開発できた。しかし、この結果を元にスペクトルの照合を行うという目的に対しては、未達成であり、今後の研究課題とする。テーマ(3) については、前述の通りおおむね達成できたものとする。

研究期間全体に渡り、学生または研究補助者 1 名と協力する体制で進めてきた。研究費は前半(1、2 年目)にはデータベース作成のための計算機クラスターの導入、後半(3、4 年目)には機械学習に利用する GPU 搭載の計算機の導入に利用された。研究の進捗に応じて適切な、経費の執行を行ったものと判断する。

現代の自然科学研究において、X 線吸収分光法は物質の微視的状态を知るための重要な実験手法として認識され、多岐に渡る分野で活用されている。本研究で得た機械学習モデル作成に関する知見を用いて様々な元素・吸収端に対して信頼できる予測モデルを作成することができれば、スペクトルから局所原子配列に関する情報を迅速に抽出することが可能となり、物質評価のプロセスを効率化することができる。近年では、学術用途のみならず、二次電池や触媒などの、民間企業で開発されている先端材料の物質評価に X 線吸収分光法が利用されるようになってきている。本研究の成果に基づいて、これらのデータ解析が機械的に実行できるようになれば、新たな材料開発指針を得ることができ、開発サイクルの改善につながるものと期待される。

#### 5. 主な研究成果リスト

##### (1) 論文(原著論文)発表

1. Hidekazu Ikeno and Teruyasu Mizoguchi, “Basics and applications of ELNES calculations”, *Microscopy* **66** (2017) 305–327.
2. Masato Urasaki, Hidekazu Ikeno, “*Ab-Initio* Multiplet Calculations Using Iterative Algorithms for X-ray Absorption Spectra at Transition Metal  $L_{2,3}$ -edges”, *Materials Transactions* **59** (2018) 1057–1061.
3. Hidekazu Ikeno, “Spherical Bessel transform via exponential sum approximation of spherical Bessel function” *Journal of Computational Physics* **355** (2018) 426–435.
4. Hidekazu Ikeno, “mxpfit: A library for finding optimal multi-exponential approximations”, *Computer Physics Communications* **230** (2018) 135–144.
5. Ikuya Yamada, Akihiko Takamatsu, Kaisei Asai, Takuto Shirakawa, Hideo Ohzuku, Akihiro Seno, Tasuku Uchimura, Hiroshi Fujii, Shogo Kawaguchi, Kouhei Wada, Hidekazu Ikeno and Shunsuke Yagi, “Systematic Study of Descriptors for Oxygen Evolution Reaction Catalysis in Perovskite Oxides” *The Journal of Physical Chemistry C* **122** (2018) 27885–27892.

(2)特許出願

研究期間累積件数:0 件

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

1. 池野豪一、「遷移金属  $L_{2,3}$  端 X 吸収分光と多電子系第一原理計算による正極材料の化学状態解析」、立命館 SR 公開シンポジウム「軟 X 線分光を用いた二次電池研究の最前線」、2016 年 11 月 11 日、(立命館大学)。[招待講演]
2. Hidekazu Ikeno, “Ab-initio Multiplet Calculations for X-ray Absorption Spectroscopy: Application to Cathode Materials”, Frontiers in Materials Science (FMS2017), Sep. 4–6, 2017 (Greifswald, Germany). [招待講演]
3. H. Ikeno and N. Miyamoto, “Machine Learning Approach for the Analysis of X-ray Absorption Spectra”, PRESTO International Symposium on Materials Informatics, Feb. 9–11, 2019 (University of Tokyo, Japan).
4. 池野豪一、宮本直幸、「X 線吸収スペクトルの機械学習」、第 22 回 XAFS 討論会、2019 年 9 月 2–4 日、(京都大学)。[招待講演]
5. Hidekazu Ikeno, DFT-CI calculations for RIXS and RIXS-MCD”, Frontier of theoretical approach in x-ray spectroscopy (PF 研究会「X 線分光理論の新展開: 構造・電子状態解析から磁性研究まで」), Oct. 3–4, (Photon Factory, Tsukuba, Japan). [招待講演]