

インフォマティクスと融合させた高精度・高効率な

相対論的量子化学理論の開発

研究代表者 清野 淳司
(理工学術院総合研究所 次席研究員)

1. 研究課題

これまで量子化学の多くの場面で、データに潜む規則性・法則性を利用した高精度化・高速化が実現されてきた。例えば我々は、有機化合物に対して、非相対論的波動関数理論における高精度な手法および完全基底関数極限のエネルギーを推定する手法を開発した。また、電子密度の汎関数のみを用いて波動方程式と同等の形に変換して解く密度汎関数理論 (DFT) では、汎関数を構築する際に幾つかの化合物に対する分子物性を再現するようにパラメータ決定がされてきた。

本研究では相対論的量子化学とインフォマティクスを融合し、高精度化・高速化を達成することで、任意の元素を含んだ化合物のための、より実用性の高い手法とすることを目的とする。これにより実社会で求められるメソスケール領域まで踏み入れることができる相対論的量子化学計算を実現させる。さらに開発する手法が導入された、独自のプログラム **Relativistic And Quantum Electronic Theory (RAQET)** を用いて、本プロジェクト内での企業と共同研究を通して、材料設計・開発が加速されることを示す。最終的には産業界に根付かせることを目標とする。

2. 主な研究成果

2-1 機械学習を用いた半局所運動エネルギー汎関数のポテンシャルエネルギー曲線への適用

本研究では、インフォマティクスの手法を用いて、従来よりも圧倒的に高速な DFT 計算手法を開発することを目標とする。DFT ではホーヘンベルグ・コーン定理に基づき、原子・分子に対する電子のエネルギーが電子密度の情報のみで表現できることが保証されている。しかし、電子の運動エネルギーや量子的な交換・相関エネルギーに対する厳密な汎関数表現は不明である。これまでに本研究では、機械学習により電子密度の情報から半局所的な運動エネルギーを構築する手法を開発した。この手法は原子・分子のエネルギーについて、これまで開発されてきたすべての汎関数よりも、参照となる Kohn-Sham (KS) -DFT からの誤差が小さいことが確認された。2018 年度では、機械学習により得られた半局所的運動エネルギー汎関数をポテンシャルエネルギー曲線 (PEC) の計算へと適用することで化学結合の記述ができるかどうかを検証した。

本手法では、KS-DFT 計算で得られる記述子と目的変数の間の関係を、微小体積 (グリッド点) ごとに、機械学習における回帰により構築した。記述子として、各グリッド点における電子密度 ρ 、およびその 1 次、2 次、3 次勾配を用いた。また各グリッド点における最も近い n 個の原子の中心からの距離 ($|\mathbf{r}-\mathbf{R}_1|$ 、 $|\mathbf{r}-\mathbf{R}_2|$...、 $|\mathbf{r}-\mathbf{R}_n|$) の逆数も記述子として導入した。目的変数として、次のような KS 運動エネルギー密度 τ_{KS} に対する Enhancement factor F_{enh} を採用した。

$$F_{\text{enh}}(\mathbf{r}) = \tau_{\text{KS}}(\mathbf{r}) / \tau_{\text{TFvW}}(\mathbf{r}), \quad \tau_{\text{TFvW}}(\mathbf{r}) = \frac{3}{10} (3\pi^2)^{2/3} \rho^{5/3}(\mathbf{r}) + \frac{1}{8} \frac{|\nabla \rho(\mathbf{r})|^2}{\rho(\mathbf{r})} \quad (1)$$

F_2 分子の PEC に対して、本手法で構築した運動エネルギー汎関数の計算精度を検証した。結合距離 1.0 Å - 2.0 Å までの 0.05 Å 間隔の F_2 分子を KS-DFT により計算し、すべてのグリッド点のうち 10%を無作為に抽出して学習データとした。テストには、0.01 Å 間隔の F_2 を採用した。図 1 に記述子依存性を示す。この結果、電子密度とその m 次の勾配のみの汎関数(ML m)では正しく PEC を記述できないことがわかった。また n 原子の中心からの距離 (+D n)を加えることで、KS-DFT の PEC の形状およびエネルギー最小点をおおよそ再現できることが確認された。これにより結合領域を記述するためには、原子中心からの距離の情報が有効であることが示された。

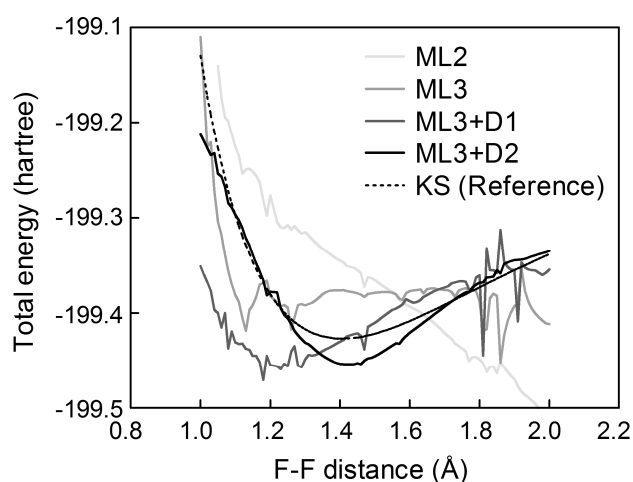


図 1. F_2 分子のポテンシャルエネルギー曲線

2-2 半局所運動エネルギー汎関数を用いた軌道非依存 DFT 計算手法の開発

これまで機械学習を用いた運動エネルギー汎関数の数値検証において、KS-DFT により算出された電子密度を用いてきた。しかし、効率的な計算を実現するためには、KS-DFT の軌道情報を用いない、軌道非依存(OF-)DFT 計算手法を確立することが不可欠である。そこで 2018 年度は OF-DFT 計算手法を実現するために、二乗が電子密度と一致する関数を導入し、原子中心に置いた幾つかのガウス基底関数の線形結合を用いてエネルギーを表現する手法を開発・実装した。機械学習により構築された運動エネルギー、およびその電子密度微分を組み込み、それぞれの原子・分子について最も低い全エネルギーを探索することで、最適な電子密度が得られる。図 2 に H_2O 分子に対する OF-DFT による電子密度の KS-DFT からの絶対誤差率 (APE) を示す。既存の TF9vW と MGE2 汎関数の結果も併記する。本手法 (ML) はすべての領域で既存の汎関数より KS-DFT に近い電子密度を与えた。全エネルギーについても 7 つの原子・分子 (Be, Ne, H_2 , HF, H_2O , N_2 , CO) の平均絶対誤差が 0.451 hartree となり、同様に従来法よりも小さい誤差となることが確認された。

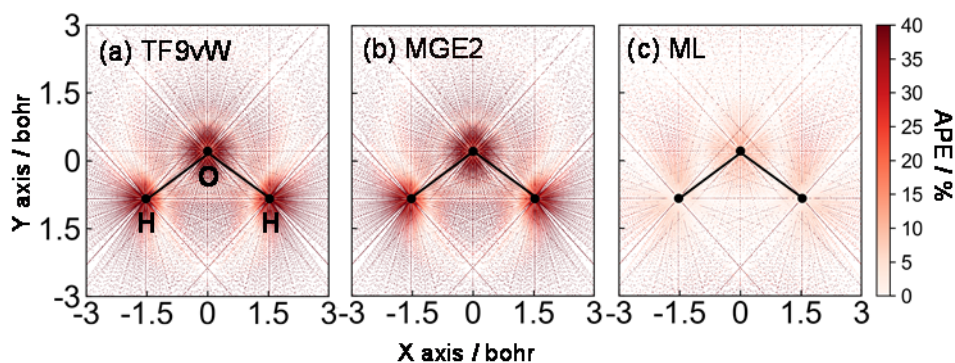


図 2. H_2O 分子に対する OF-DFT の KS-DFT からの APE

2-3 インフォマティクスを用いた結合エネルギー密度解析手法の開発

化学結合の強弱の情報は化学反応や分子の性質の解析において重要な役割を果たす。量子化学計算から結合エネルギーを見積もる方法として、これまでに全エネルギーを構成原子および結合の寄与へと分割する様々な手法が開発されてきた。しかし、従来の手法では、多重結合のような強い結合とイオン結合や非結合部位のような弱い結合を同時かつ定性的に正しく記述することが困難であった。本研究ではインフォマティクス手法を用いた新たな方法を開発し、RAQETへと連結した。

エネルギー分割法は全エネルギーを量子化学計算の定義に基づき原子成分 E^{AA} と結合成分 E^{AB} の総和として分割する。本研究では各成分を、非経験的パラメータ α を用いて以下のように定義した。

$$E^{AA} = T_S^{AA} + E_{Ne}^{AA} + E_{ee}^{AA} + E_{HFx}^{AA} + E_{DFTxc}^{AA} + \sum_{A>B} (1 - \alpha^{AB}) (T_S^{AB} + E_{HFx}^{AB} + E_{DFTxc}^{AB}) \quad (2)$$

$$E^{AB} = E_{Ne}^{AB} + E_{ee}^{AB} + E_{NN}^{AB} + \alpha^{AB} (T_S^{AB} + E_{HFx}^{AB} + E_{DFTxc}^{AB}) \quad (3)$$

α は原子成分と結合成分の分割が曖昧となる部分を補正する。すべての α を同時に決定するための最適化条件として、原子成分は結合 R^{CD} に対して不変である、という条件を用いた。この条件により、 α に基づく連立方程式が構築される。この連立方程式において、エネルギー成分の大きな項と小さな項が共存する場合、最小二乗法により解くと過適合が生じてしまう。そこで LASSO 回帰により過適合の回避を狙った。

共有結合やイオン結合を含んだ 32 分子・42 結合に対して、本手法に基づき、連立方程式を LASSO 回帰で解くことで得られた結合解離エネルギー (BDE) の結果を図 3 に示す。ここで BDE は該当する結合の関与するすべての結合エネルギーの総和で評価した。最安定構造から解離極限のエネルギーを差し引いて得た BDE (QC) との決定係数 R^2 も示す。比較として従来法の M03、NK05、NOIK11 の結果も載せた。本手法は QC との R^2 が 0.91 と良い相関を表していることから、すべての結合における強弱を再現していることが確認された。

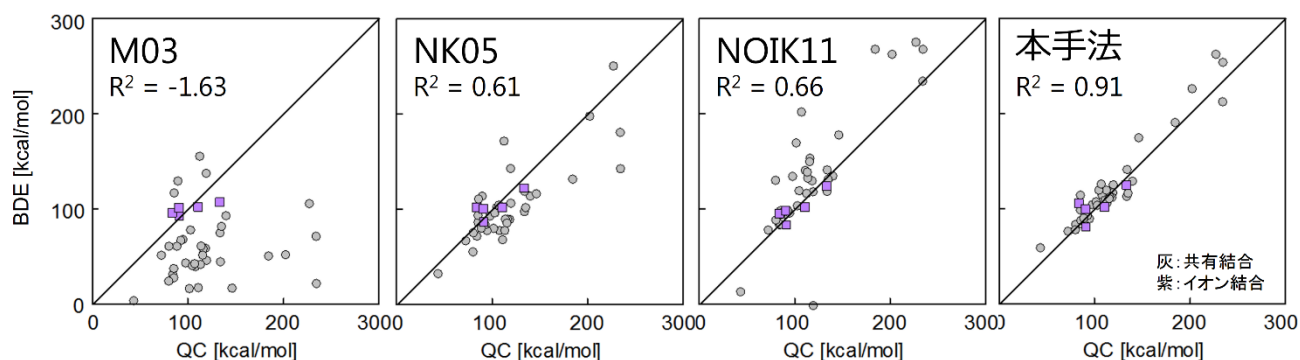


図 3. 42 結合に対する BDE の QC と各分割手法の比較

3. 共同研究者

中井 浩巳 (先進理工学部 化学・生命化学科・教授)

五十幡 康弘 (理工学術院総合研究所・次席研究員)

吉川 武司 (先進理工学部 化学・生命化学科・講師)

4. 研究業績

4.1 学術論文

- [1] M. Hayami, J. Seino, Y. Nakajima, M. Nakano, Y. Ikabata, T. Yoshikawa, T. Oyama, K. Hiraga, S. Hirata, and H. Nakai, “RAQET: Large-scale two-component relativistic quantum chemistry program package” *J. Comput. Chem.* **39**, 233 (2018), Front Cover.
- [2] Y. Ikabata, T. Oyama, M. Hayami, J. Seino, and H. Nakai, “Extension and acceleration of relativistic density functional theory based on transformed density operator”, *J. Chem. Phys.* **150**, 164104 (2019).
- [3] M. Fujinami, J. Seino, T. Nukazawa, S. Ishida, T. Iwamoto, and H. Nakai, “Virtual Reaction Condition Optimization based on Machine Learning for a Small Number of Experiments in High-dimensional Continuous and Discrete Variables” *Chem. Lett.* in press, Editor’s Choice.

4.2 総説・著書

- [1] 藤波 美起登、清野 淳司、中井 浩巳、「理論化学とインフォマティクスの融合による反応設計」、*化学と工業*、**71**, 653 (2018).
- [2] 藤波 美起登、清野 淳司、中井 浩巳、「人工知能を用いた化学反応の予測と反応条件の最適化」、*マテリアルズ・インフォマティクスによる材料開発と活用集*、技術情報協会、379 (2019).

4.3 招待講演

- [1] 清野 淳司、「理論・計算・実験化学とインフォマティクスの融合研究」、分子研 理論・計算分子科学領域セミナーシリーズ、分子科学研究所、2018/06/21.
- [2] 清野 淳司、「人工知能を用いた化学反応の予測と反応条件の最適化」、技術情報協会・講習会 触媒開発における人工知能、計算科学の活用—新しい触媒の探索、設計の具体的手法—、技術情報協会セミナールーム、2018/08/27.
- [3] 清野 淳司、「相対論的量子化学における理論・計算手法の基礎」、第8回量子化学スクール、岡崎コンファレンスセンター、2018/12/19.
- [4] 清野 淳司、「人工知能(AI)技術と電子状態情報を用いた化学反応予測および反応条件最適化」、R&D支援センター講習会、商工情報センター（カメラプラザ）9F会議室、2019/01/22.

4.4 受賞・表彰

- [1] 第41回ケモインフォマティクス討論会ポスター賞、藤波 美起登 (D2)：共同発表者

4.5 学会および社会的活動

該当なし

4.6 外部資金

- [1] 科学技術振興機構 (JST) 戦略的創造研究推進事業 (さきがけ) 『理論・実験・計算科学とデータ科学が連携・融合した先進的マテリアルズインフォマティクスのための基盤技術の構築』 “量子化学と情報学との融合による次世代密度汎関数理論と均一系触媒における反応予測システムの開発”, (研究代表, 平成 29-32 年度).

5. 研究活動の課題と展望

本研究では、相対論的量子化学における最新の理論・手法が実装された RAQET プログラムを用いて、大学/企業問わず様々な研究者が研究を進めることができるように、今後も引き続きプログラムの機能拡充を行う。また機械学習を用いた運動エネルギー汎関数において、これまでは数値検証のための必要最低限のデータセットを用いて学習を行ったため、構築された運動エネルギー汎関数の汎用性は低い。そのため今後は、膨大なデータを用いたオンライン学習による逐次学習システムを構築し、任意の化合物に適用できるように拡張する研究を進める。