

博士論文

深層学習に基づく Skyrme エネルギー密度汎函数の
密度依存性及び原子核集団運動の解析

樋沢 規宏

京都大学大学院理学研究科
物理学・宇宙物理学専攻
物理学第二分野
原子核理論研究室

2023 年 12 月

要旨

量子多体系において、原理的に基底状態の厳密な adiabatic ダイナミクスを記述する密度汎関数理論 (Density Functional Theory, DFT) は、電子系のみならず原子核系においても多大な成功を収めてきた。DFT においてダイナミクスは、粒子数密度のみに依存したエネルギー密度汎関数 (Orbital-free Energy Density Functional, OF-EDF) により支配される。その一方で、実際の計算ではほとんどの場合、粒子数密度以外にも依存した Kohn-Sham 型の EDF (KS-EDF) が用いられてきた。現象論的に与えられた KS-EDF は様々な現象の再現に成功してきたものの、OF-EDF との等価性、及び粒子数が大きい場合における計算コストの問題を抱えている。

また KS-EDF では、系を外場で制御し、最終的な結果としてエネルギーや粒子数密度を得るという理論構造を持つ。そこで実験という観点で、電子系ほど自在に外場を扱えない原子核系では、外場は数学的な道具の側面が強い。したがって外場を通して仮想的に原子核を変形させることで、核分裂などの変形ダイナミクスが調べられてきた。しかし変形のさせ方には、外場の設計を通じて人間の強いバイアスが入り、多重極モーメントといった仮定された変形モードが現象の記述に用いられてきたが、核分裂の完全な記述はまだ成功していない。この“外場の呪い”は集団運動の記述において深刻な問題であり、人間のバイアスを排除した、ダイナミクスに基づく変形モードの導出法が求められる。

本論文では、KS-EDF に起因した 2 つの問題の解決に向けて、深層学習に基づいた data-driven アプローチによる解析を行なった。まず KS-EDF と OF-EDF の等価性と計算コストの問題に関して、電子系で開発されたランダムポテンシャルに基づくアプローチを原子核系に適用することで、Skyrme KS-EDF を再現するような OF-EDF の構築に取り組んだ。10 層の隠れ層を持つ neural network と教師データを用いて、 ^{24}Mg の核子数密度から対応する束縛エネルギーを推定するタスクを学習した。その結果、エネルギーの推定誤差は 43 keV であり、実験データとの比較において許容できるほどに小さい。つまり、エラーの範囲内で Skyrme KS-EDF は OF-EDF と等価であることが示唆された。また、与えられた密度や外場からエネルギーを推定する時間は、従来の KS-EDF に基づく手法に比べ、 10^5 倍から 10^6 倍短い。この圧倒的な計算速度は、研究プロセスの短縮を可能にし、原子核構造研究を加速させ得る。

次に、以上の ^{24}Mg の研究結果により示唆された Skyrme KS-EDF と等価な OF-EDF の存在に動機付けされ、DFT に基づく非経験的な集団座標の抽出を行なった。まず、低エネルギーの ^{236}U に関して、核子数密度と基底状態のエネルギーをランダムポテンシャルにより計算し、データセットを生成した。このデータを multi-task learning を用いて、様々な次元の潜在変数に圧縮し、もとのデータを復元する際のエラーを調べた。結果として、僅か 2 次元の潜在変数で低エネルギーの ^{236}U のダイナミクスの大部分が特徴づけられることが明らかになった。一方で、多重極モーメントによる密度やエネルギーの推定タスクを考え、そのエラーを調べたところ、多重極モーメントがダイナミクスの情報をほとんど持たないことが明らかになった。以上より本研究は、集団座標を抽出する全く新しい手法を提案しただけでなく、“外場の呪い”により仮定されてきた従来の変形パラメータの問題点を、定量的に示すことに成功した。

目次

第 1 章	密度汎関数理論と原子核物理	4
1.1	密度汎関数理論	4
1.2	Kohn-Sham scheme	6
1.3	Skyrme EDF	7
1.4	本論文の目的と構成	9
第 2 章	深層学習概論	10
2.1	ニューラルネットワーク	10
2.2	最適化手法	13
2.3	汎化とドメイン	14
第 3 章	Deep learning による Skyrme エネルギー密度汎関数の再構築	19
3.1	KS-EDF の問題点と機械学習	19
3.2	手法: ランダムポテンシャルと深層学習	21
3.3	結果: ^{24}Mg の OF-EDF	28
3.4	まとめ	41
第 4 章	Multi-task learning による ^{236}U の潜在変数の解析	43
4.1	外場の呪いと集団運動	43
4.2	データセット	45
4.3	Multi-task learning	48
4.4	評価基準	54
4.5	結果: ^{236}U の潜在変数	58
4.6	まとめ	64
第 5 章	結論	66
Appendix A	Artificial Neural Network	69
A.1	Convolutional neural network	69
A.2	Vision transformer	72

Appendix B	軸対称系の Skyrme-Hartree-Fock-BCS	76
B.1	Kohn-Sham 方程式	76
B.2	数値計算アルゴリズム	78
参考文献		82

第 1 章

密度汎関数理論と原子核物理

1.1 密度汎関数理論

量子力学の基本方程式であるシュレディンガー方程式が見出されて 100 年が経とうとしている。この 100 年間で計算機は著しい成長を見せた一方、同種粒子量子多体系のシュレディンガー方程式を厳密に解くことは、数値的にでさえ今だに難しい。単純な差分法や基底展開法、近年では Physics Informed Neural Network (PINN) [1] と呼ばれる機械学習を用いた方法など、様々な手法が提案されているが、どれも粒子数に対して計算コストが指数関数的に増大するボトルネックを持つため、問題解決の兆しは現時点では見えず、量子コンピュータの完成を待つばかりである。

一方で、量子多体系の厳密計算において我々の興味のあるエネルギー領域は限られている。現象として興味のあるエネルギースケールは多くの場合基底状態近傍であり、実際に我々の身の回りで観測可能な量子的現象もこのエネルギースケールに密集している。したがって、Hilbert space 全体を取り扱う必要性はない。特に電子系では、分子構造の解析という観点から、原子核が作るクーロンポテンシャル中の基底状態の性質は重要であり、精力的に研究されている。そのような限られた範囲では、多体の波動関数の大部分はもはや冗長な自由度であり、より本質的な自由度が創発される。その最たる例が密度汎関数理論 (Density Functional Theory, DFT) である。

1964 年、Hohenberg と Kohn は様々な外場のもとでの電子系の基底状態を考えた時、縮退がない場合、対応する電子数密度と外場の間に一対一対応が存在することを見出した [2]。この事実は次のように一般化することが可能である。ハミルトニアン \hat{H}_0 と粒子数密度演算子 $\hat{\rho}(\mathbf{r})$ で指定される空間 d 次元の同種粒子フェルミオン量子多体系に対して、外場付き変分問題

$$E^*[v] := \min_{|\Psi\rangle \in \mathcal{H}} \frac{\langle \Psi | \hat{H}_0 + \int d^d r v(\mathbf{r}) \hat{\rho}(\mathbf{r}) | \Psi \rangle}{\langle \Psi | \Psi \rangle}, \quad (1.1)$$

を考える。ただし、 \mathcal{H} は系の Hilbert 空間である。この最小化問題は、次のように分解することが可能である [3]:

$$E^*[v] = \min_{\rho} \min_{|\Psi\rangle \rightarrow \rho} \left(\frac{\langle \Psi | \hat{H}_0 | \Psi \rangle}{\langle \Psi | \Psi \rangle} + \int d^d r v(\mathbf{r}) \rho(\mathbf{r}) \right) =: \min_{\rho} \left(E[\rho] + \int d^d r v(\mathbf{r}) \rho(\mathbf{r}) \right). \quad (1.2)$$

ただし、左辺の二つ目の最小化は、ある密度 ρ を与える波動関数 $|\Psi\rangle$ の範囲内で、目的関数を最

小化するという操作を意味している。この時、 $E[\rho]$ はエネルギー密度汎関数 (Energy Density Functional, EDF) と呼ばれる。したがって系を固定した場合、基底状態に縮退がなければ、外場付き変分問題 (1.1) は密度に関する変分問題と等価である。

一方で系を固定した場合、変分問題 (1.1) は外場によって解が決まるため、基底状態の波動関数もまた外場 v の汎関数として与えられる:

$$|\Psi[v]\rangle := \operatorname{argmin}_{|\Psi\rangle \in \mathcal{H}} \frac{\langle \Psi | \hat{H}_0 + \int d^d r v(\mathbf{r}) \hat{\rho}(\mathbf{r}) | \Psi \rangle}{\langle \Psi | \Psi \rangle}. \quad (1.3)$$

したがって、波動関数が上記のように与えられる以上、全ての可観測量も外場の汎関数で表示される。粒子数密度に関しても同様に、

$$\rho[v](\mathbf{r}) := \frac{\langle \Psi[v] | \hat{\rho}(\mathbf{r}) | \Psi[v] \rangle}{\langle \Psi[v] | \Psi[v] \rangle} = \operatorname{argmin}_{\rho} \left(E[\rho] + \int d^d r v(\mathbf{r}) \rho(\mathbf{r}) \right), \quad (1.4)$$

という汎関数が存在し、同時に EDF に関する最小化問題の解として表示することも可能である*¹。したがって、形式的に

$$v[\rho](\mathbf{r}) = -\frac{\delta E[\rho]}{\delta \rho(\mathbf{r})}, \quad (1.5)$$

と表すことができる。

以上の議論を俯瞰すると、(1.2) 式が逆 Legendre 変換の形式になっていることが分かる。しかし $E[\rho]$ の凸性がわからないため、逆 Legendre 変換の Legendre 変換により元に戻せるかは非自明である。一方で外場の汎関数 $E^*[v]$ に関して、その定義 (1.1) から明らかのように

$$E^*[(1-t)v_1 + tv_2] \geq (1-t)E^*[v_1] + tE^*[v_2], \quad t \in (0, 1), \quad (1.6)$$

が成り立つため、 E^* が (広義) 凹関数であることがわかる。したがって、

$$E^{**}[\rho] = -\inf_v \left(\int d^d r v(\mathbf{r}) \rho(\mathbf{r}) - E^*[v] \right), \quad (1.7)$$

という Legendre 変換を考えれば、適切な密度と外場の定義域の設定により

$$E[\rho] = E^{**}[\rho], \quad (1.8)$$

が期待される。この時、密度と外場の間に対応が保証されるため、この範囲内で系の外場付き基底状態ダイナミクスは、相互作用のみによって決まる EDF により完全に支配されることが分かる。数値的に DFT を計算する場合、扱うべき量は粒子数密度 $\rho(\mathbf{r})$ のみである。したがって粒子数に対して計算コストが指数関数的に増大するボトルネックは存在せず、効率的な計算が可能になる [5]。

上記のように、DFT が成立するかどうかは逆変換により元に戻る Legendre 変換を実現するような、適切な外場と密度の定義域が与えられるかに依存する。Coulomb 相互作用に従う電子系では、外場の定義域は2種類の L^p 空間を用いて、

$$\mathcal{X}_v := L^{3/2}(\mathbb{R}^3) + L^\infty(\mathbb{R}^3), \quad (1.9)$$

*¹ 基底状態が縮退している場合は、 v から ρ への写像はユニークではないが、 ρ から v への写像はユニークになる [4]。

で与えられる Banach 空間で定義することが可能である [6]. この時, $E^*[v]$ の Legendre 変換をさらに逆 Legendre 変換することで元の汎関数 $E^*[v]$ に戻る. 実は, 対応する密度の定義域は $\mathcal{X}_\rho = L^3(\mathbb{R}^3) \cap L^1(\mathbb{R}^3)$ で与えられ, $\rho(\mathbf{r}) < 0$ となるものが含まれる. しかし, このような非物理的な密度は計算上問題にならないことが知られている [4].

原子核系では相互作用が完全にはわかっていない上, 電子系よりも複雑である. したがって原子核系において真に DFT が存在するかは非自明であり, 外場付きの基底状態ダイナミクスを完全に決める $E[\rho]$ の存在が証明されているわけではないことに注意が必要である.

1.2 Kohn-Sham scheme

DFT における最大の問題は, EDF を相互作用から解析的に計算することが極めて困難であるという点である. そこで実際の研究では, 現象論的に与えられた汎関数が用いられる場合がほとんどである. 特に運動項の密度依存性は非自明であり, Thomas-Fermi 近似 [7] と呼ばれる, 相互作用のない系から類推される形式の EDF など存在するが, 現象の再現性という観点から実用的ではない. そこで, 多くの計算では Kohn-Sham スキーム [8] と呼ばれる方法で DFT 計算が行われる:

$$\min_{\rho} \left(E[\rho] + \int d^d r v(\mathbf{r}) \rho(\mathbf{r}) \right) =: \min_{\rho} \min_{\sum_{i=1}^N |\varphi_i|^2 = \rho} \left(E_{\text{KS}}[\{\varphi_i, \varphi_i^*\}_{i=1}^N] + \int d^d r v(\mathbf{r}) \rho(\mathbf{r}) \right) \quad (1.10)$$

$$= \min_{\{\varphi_i^*\}_{i=1}^N} \left(E_{\text{KS}}[\{\varphi_i, \varphi_i^*\}_{i=1}^N] + \int d^d r v(\mathbf{r}) \sum_{i=1}^N |\varphi_i(\mathbf{r})|^2 \right). \quad (1.11)$$

ただし, N は系の粒子数である. また, $\{\varphi_i\}_{i=1}^N$ は L^2 内積の意味で正規直交化された N 個の関数であり, Kohn-Sham 軌道と呼ばれる. このとき, 以下の N 表現可能な密度,

$$\mathcal{N}_\rho := \left\{ \rho \mid \rho(\mathbf{r}) \leq 0, \int d^d r \rho(\mathbf{r}) = N, \int d^d r |\nabla \rho^{1/2}(\mathbf{r})|^2 < \infty \right\}, \quad (1.12)$$

を与えるような Kohn-Sham 軌道が存在することは, Harriman の構成法 [9] から保証することが可能である. つまり, ρ を与えるような補助的な自由度を導入することで, 密度に対して複雑な依存性を持った EDF を, 扱いやすい形式で記述するというのが, Kohn と Sham のアイディアの本質である*2. この時, Kohn-Sham 軌道はあくまで補助的, 数学的な自由度に過ぎず, 物理的な意味を持つかは不明であることに注意が必要である. 例えば, 形式的に Kohn-Sham 軌道を用いて構成された Slater determinant 型の多体波動関数を考えることが可能である. しかし, その波動関数を用いて計算される粒子数密度は基底状態の密度と等価であるものの, 他の用途で使用した場合, 結果が物理的に意味を持つかは注意深く検証する必要がある.

$E_{\text{KS}}[\{\varphi_i, \varphi_i^*\}_{i=1}^N]$ はもはや密度以外にも陽に依存した汎関数であり, 本論文では Kohn-Sham 型の EDF (KS-EDF) と呼ぶ. 一方で, 密度のみに陽に依存した EDF は Orbital-free EDF (OF-

*2 Kohn-Sham スキームは Hartree-Fock-Bogoliubov 理論の U, V 関数に対して, 直ちに拡張することが可能である.

EDF) と呼び、KS-EDF とは明確に区別する。ただし、両者には

$$E[\rho] = \min_{\sum_{i=1}^N |\varphi_i|^2 = \rho} E_{\text{KS}}[\{\varphi_i, \varphi_i^*\}_{i=1}^N], \quad (1.13)$$

という関係が存在する。したがって、DFT がよく定義される量子多体系に対して適切な KS-EDF を与れば、Kohn-Sham スキームにより基底状態の厳密なダイナミクスが得られる。しかし、KS-EDF から OF-EDF の解析的な表式を復元することは困難である。

前述の運動項部分に関しては、平均場理論から類推される KS-EDF が用いられることが多い:

$$T_{\text{KS}}[\tau] := \frac{\hbar^2}{2m} \int d^d r \tau(\mathbf{r}) =: \frac{\hbar^2}{2m} \int d^d r \sum_{i=1}^N |\nabla \varphi_i(\mathbf{r})|^2. \quad (1.14)$$

ただし、 m は系を構成するフェルミオンの質量である。この時、 $E_{\text{KS}}[\{\varphi_i, \varphi_i^*\}_{i=1}^N] = T_{\text{KS}}[\tau] + I_{\text{KS}}[\rho]$ と書かれる場合、最小化問題 (1.11) は以下と等価である:

$$\left(-\frac{\hbar^2}{2m} \nabla^2 + \frac{\delta I_{\text{KS}}[\rho]}{\delta \rho(\mathbf{r})} + v(\mathbf{r}) \right) \varphi_i(\mathbf{r}) = \epsilon_i \varphi_i(\mathbf{r}). \quad (1.15)$$

ただし、 ϵ_i は Kohn-Sham 軌道の正規直行化条件を課すため、Lagrange の未定乗数として導入されるパラメータである。この方程式は Kohn-Sham 方程式と呼ばれ、形式的に平均場理論における Hartree-Fock 方程式とほぼ等価であるため、同等のアルゴリズム、計算コストで解くことが可能である。ここでほぼ等価と言ったのは、1 粒子ポテンシャルに相当する部分に関して、Hartree-Fock 理論ではあり得ない項が、Kohn-Sham 方程式には登場し得るからである。

1.3 Skyrme EDF

原子核物理では、Skyrme 力 [10] の平均場近似から類推される以下の KS-EDF がしばしば用いられる:

$$E_{\text{Skyme}}[\rho_q, \tau_q, \mathbf{J}_q] = \int d^3 r H_{\text{Skyme}}(\mathbf{r}), \quad (1.16)$$

$$\begin{aligned} H_{\text{Skyme}}(\mathbf{r}) = & \frac{\hbar^2}{2m} \left(1 - \frac{1}{A} \right) \tau + \frac{b_0}{2} \rho^2 - \frac{b'_0}{2} \sum_q \rho_q^2 + \frac{b_3}{3} \rho^{\alpha+2} - \frac{b'_3}{3} \rho^\alpha \sum_q \rho_q^2 \\ & + b_1 \rho \tau - b'_1 \sum_q \rho_q \tau_q - \frac{b_2}{2} \rho \nabla^2 \rho + \frac{b'_2}{2} \sum_q \rho_q \nabla^2 \rho_q \\ & - b_4 \rho \nabla \cdot \mathbf{J} - b'_4 \sum_q \rho_q \nabla \cdot \mathbf{J}_q. \end{aligned} \quad (1.17)$$

ただし、各変数、パラメータの意味等は表 1.1 にまとめている。また、原子核ではスピンの自由度を考える必要があり、本論文では Kohn-Sham 軌道は

$$\varphi_{q,k} = \begin{pmatrix} \varphi_{q,k}^\uparrow \\ \varphi_{q,k}^\downarrow \end{pmatrix}, \quad (1.18)$$

という spinor 表記を取るものとする。

表 1.1 Skyrme EDF の各変数・パラメータの意味. $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ は Pauli 行列.

記号	意味	定義
m	nucleon mass	
A	mass number	
q	isospin	neutron (n) or proton (p)
$b_{0,1,2,3,4}$	model parameter	
$b'_{0,1,2,3,4}$	model parameter	
$\rho_q(\mathbf{r})$	number density	$\rho_q(\mathbf{r}) = \sum_{k=1}^N \varphi_{q,k}(\mathbf{r}) ^2$
$\tau_q(\mathbf{r})$	kinetic density	$\tau_q(\mathbf{r}) = \sum_{k=1}^N \nabla \varphi_{q,k}(\mathbf{r}) ^2$
$\mathbf{J}_q(\mathbf{r})$	spin density	$\mathbf{J}_q(\mathbf{r}) = \sum_{k=1}^N \varphi_{q,k}^*(\mathbf{r}) (-i\nabla \times \boldsymbol{\sigma}) \varphi_{q,k}(\mathbf{r})$
$\rho(\mathbf{r})$	total number density	$\rho_n(\mathbf{r}) + \rho_p(\mathbf{r})$
$\tau(\mathbf{r})$	total kinetic density	$\tau_n(\mathbf{r}) + \tau_p(\mathbf{r})$
$\mathbf{J}(\mathbf{r})$	total spin density	$\mathbf{J}_n(\mathbf{r}) + \mathbf{J}_p(\mathbf{r})$
$\varphi_{q,k}(\mathbf{r})$	Kohn-Sham orbital	
k	quantum number	

強相関量子多体系である原子核では, even-odd staggering [10] をはじめとした, Hartree-Fock 近似では十分に説明できない現象があり, 核子の超伝導性が重要であることが古くから知られている. この事実は Hartree-Fock 近似から類推された KS-EDF (1.16) 式 では, ansatz として不十分であることを示唆している [11]. そこで実際の計算では, 超伝導理論に動機付けされた汎関数が追加される. 以下では, 一貫して時間反転対称性が破られず, 陽子・中性子共に粒子数が偶数であるような状況に議論を限定する. 本論文でも一貫して用いる surface type の Density Dependent Delta Interaction (DDDI) [12] と呼ばれる KS-EDF は, 次のような形をしている:

$$E_{\text{DDDI}}[\rho, \tilde{\rho}_q] = \sum_{q=n,p} \frac{V_q}{4} \int d^3r \left\{ 1 - \left(\frac{\rho(\mathbf{r})}{\rho_0} \right)^\gamma \right\} \tilde{\rho}_q(\mathbf{r})^2. \quad (1.19)$$

ただし, γ, ρ_0 はパラメータであり, 本論文では一貫して $\gamma = 1, \rho_0 = 0.16 \text{ fm}^{-3}$ を用いる. また, V_q は pairing の強さを表すパラメータであり, 平均場理論における average pairing gap との関係式から類推することで定義される. 典型的には現象論的な関係式

$$\frac{12}{\sqrt{A}} \text{ MeV} = \frac{V_q}{2N_q} \int d^3r \left\{ 1 - \left(\frac{\rho(\mathbf{r})}{\rho_0} \right)^\gamma \right\} \tilde{\rho}_q(\mathbf{r}) \rho_q(\mathbf{r}), \quad (1.20)$$

を再現するようにして決められる [10, 13]. ただし N_q は中性子数 N_n または陽子数 N_p である. 本論文でも一貫してこの方法により V_q を定める.

ここで新たに導入した $\tilde{\rho}_q(\mathbf{r})$ は pairing density と呼ばれ, DFT の文脈では定性的には pairing の効果を表すと信じられている局所密度量である. そこで, $\tilde{\rho}_q(\mathbf{r})$ の定義も超伝導理論に基づいたものになり, Hartree-Fock-Bardeen-Cooper-Schrieffer (HF-BCS) 近似 [14] から動機付けられる各

Kohn-Sham 軌道の占有率 $v_i^2 \in [0, 1]$ を用いて,

$$\tilde{\rho}_q(\mathbf{r}) = -2 \sum_q \sum_{k>0} u_{q,k} v_{q,k} |\varphi_{q,k}(\mathbf{r})|^2, \quad (1.21)$$

として定義される。ただし, $u_{q,k} := \sqrt{1 - v_{q,k}^2}$ である。ここで, 本論文では時間反転対称性が破られていない範囲を考えているため, 各軌道は Kramers 縮退しており, 時間反転のペアに分けることができる。和の $k > 0$ とは, Kramers pair の片方に関してのみ和を取るという操作を意味している。この BCS-like な Kohn-Sham スキームにおいて, 他の local density についても HF-BCS 近似に動機付けされた変更がなされる:

$$\rho_q(\mathbf{r}) = 2 \sum_q \sum_{k>0} v_{q,k}^2 |\varphi_{q,k}(\mathbf{r})|^2, \quad (1.22)$$

$$\tau_q(\mathbf{r}) = 2 \sum_q \sum_{k>0} v_{q,k}^2 |\nabla \varphi_{q,k}(\mathbf{r})|^2, \quad (1.23)$$

$$\mathbf{J}_q(\mathbf{r}) = 2 \sum_q \sum_{k>0} v_{q,k}^2 \varphi_{q,k}^*(\mathbf{r}) (-i\nabla \times \boldsymbol{\sigma}) \varphi_{q,k}(\mathbf{r}). \quad (1.24)$$

以上の拡張のもとで, KS-EDF の最小化問題は形式的に,

$$\min_{\{v_{q,k}^2, \varphi_{q,k}^*\}_{q,k}} \left(E_{\text{Skyrme}}[\rho_q, \tau_p, \mathbf{J}_q] + E_{\text{DDDI}}[\rho, \tilde{\rho}_q] + \sum_{q=n,p} \int d^3r v_q(\mathbf{r}) \rho_q(\mathbf{r}) \right), \quad (1.25)$$

と定義され, 最小化は $v_{q,k}^2, \varphi_{q,k}$ の両方に対して実行される。

1.4 本論文の目的と構成

本論文の目的は Skyrme DFT と近年急速に発展している深層学習を組み合わせることで, 原子核物理における未解決問題に対して一石を投じることである。具体的には原子核における Skyrme EDF と等価な OF-EDF の実在性と計算コストの問題, 及び仮定された集団座標の問題に取り組む。

そこで第 2 章で本論文の理解に必要な深層学習の概略について述べる。次に第 3 章では ^{24}Mg に着目し, Skyrme EDF と等価な Orbital-free EDF の存在を深層学習により調べる。この時, Kohn-Sham EDF の数値計算上の問題点に触れ, 深層学習がこの問題を解決し得ることを同時に見る。第 4 章では第 3 章の手法を拡張し, ^{236}U の基底状態近傍の変形ダイナミクスに適用する。この時, 伝統的に用いられている原子核の変形パラメータが, Skyrme EDF に基づく外場付き基底状態のダイナミクスを記述する上で妥当かどうかを議論する。

第 2 章

深層学習概論

2.1 ニューラルネットワーク

2000 年代から始まったとされる第 3 次 AI ブームは、2023 年現在をもってしても衰えるどころか加速している。AlexNet [15] による、ILSVRC 2012 の圧倒的な成績での優勝、Transformer [16] による自然語処理における革命、ChatGPT [17] の衝撃など、枚挙にいとまがない。deep learning 技術は他分野にも大きく影響を及ぼし、医療、薬学、工学など、様々な分野が発展を遂げた。

deep learning の基本的なアイデアは、線形変換と非線形変換の繰り返しにより、任意関数がいくらかでも精度良く近似可能であるという、普遍近似定理 (Universal Approximation Theorem) [18, 19, 20] に根ざしている。歴史的には人間の神経細胞を模したモデルとして導入されたため、普遍近似定理に基づき与えられた近似器は、人工ニューラルネットワーク (Artificial Neural Network, ANN) と呼ばれる。現代的には Multi-layer perceptron (MLP), Convolutional Neural Network (CNN) [21], Transformer [16] など、様々なものが存在する*¹。

この章では第 3 章, 第 4 章を読む上で必要となる深層学習の基本的なアイデアを見るため、単純な 3 層の MLP を以下のように導入する:

$$\mathbf{y}(\mathbf{x}; V, W, \mathbf{a}, \mathbf{b}) := f(W\mathbf{z} + \mathbf{b}), \quad (2.1)$$

$$\mathbf{z} := g(V\mathbf{x} + \mathbf{a}), \quad (2.2)$$

ただし、各変数は以下のように定義される。

$$\mathbf{x} \in \mathbb{R}^{d_{\text{in}}}, \quad \mathbf{z} \in \mathbb{R}^{d_{\text{hid}}}, \quad \mathbf{y} \in \mathbb{R}^{d_{\text{out}}}, \quad (2.3)$$

$$\mathbf{a} \in \mathbb{R}^{d_{\text{hid}}}, \quad \mathbf{b} \in \mathbb{R}^{d_{\text{out}}}, \quad (2.4)$$

$$V \in \mathbb{R}^{d_{\text{hid}} \times d_{\text{in}}}, \quad W \in \mathbb{R}^{d_{\text{out}} \times d_{\text{hid}}}, \quad (2.5)$$

また f, g は何らかの非線形関数で、活性化関数 (activation function) と呼ばれる。本論文では活性化関数の入力が入力ベクトル量であった場合、表記上の問題として

$$\tilde{\mathbf{x}} = f(\mathbf{x}) \quad \Leftrightarrow \quad \tilde{x}_i = f(x_i), \quad (2.6)$$

*¹ 各々の詳細は Appendix A で個別に扱っている。

のようにベクトルの成分ごと非線形関数を作用させるものとする。つまり、入力ベクトル \mathbf{x} を重み行列 V とバイアス項 \mathbf{a} を用いて線形変換し、活性化関数 g を作用させる。得られた出力ベクトル \mathbf{z} に対して異なる重み行列 W 、バイアス項 \mathbf{b} 、活性化関数 f を同様に作用させることで、最終的に d_{out} 次元のベクトルを出力として得る。ここで、 $\{V, W, \mathbf{a}, \mathbf{b}\}$ はモデルパラメータであり、これらを目的のタスクに応じて最適化する。

この3層MLPを近似したい関数の近似器として用いるという状況を考える。そこでデータセット $D = \{(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})\}_i$ に対して損失関数 (loss function) $L(\mathbf{y}, \tilde{\mathbf{y}})$ を導入し、データセット D に対する損失を以下で導入する。

$$\mathcal{L}(D) = \frac{1}{|D|} \sum_{i=1}^{|D|} L(\mathbf{y}^{(i)}, \mathbf{y}(\mathbf{x}^{(i)}; V, W, \mathbf{a}, \mathbf{b})). \quad (2.7)$$

ただし、 $|D|$ はデータセット D の要素数を表すものとする。損失関数としては、Euclid 距離の自乗、cross entropy, Wasserstein 距離など、タスクに応じて様々なものが用いられる。deep learning の基本形では、この損失関数を最小化するようにモデルパラメータ $\{V, W, \mathbf{a}, \mathbf{b}\}$ を最適化することで、データセットを補間するような関数を得ることを目的とする。

損失の最小化は例えば最急降下法によって行われ、モデルパラメータをまとめて \mathbf{w} と書いた場合、

$$w_i \leftarrow w_i - \eta \frac{\partial \mathcal{L}(D)}{\partial w_i}, \quad (2.8)$$

により更新される。ここで $\eta > 0$ は学習率 (learning rate) と呼ばれ、一度の学習でどの程度パラメータを更新するか決めるハイパーパラメータである*2。MLP の最大の特徴は、この偏微分を連鎖率を用いることで効率よく計算できる点である：

$$\frac{\partial L}{\partial b_i} = \sum_j \frac{\partial L}{\partial y_j} \frac{\partial y_j}{\partial b_i} = \frac{\partial L}{\partial y_i} f'_i, \quad (2.9)$$

$$\frac{\partial L}{\partial W_{ij}} = \sum_k \frac{\partial L}{\partial y_k} \frac{\partial y_k}{\partial W_{ij}} = \frac{\partial L}{\partial y_i} f'_i z_j = \frac{\partial L}{\partial b_i} z_j, \quad (2.10)$$

$$\frac{\partial L}{\partial a_i} = \sum_{j,k} \frac{\partial L}{\partial y_j} \frac{\partial y_j}{\partial z_k} \frac{\partial z_k}{\partial a_i} = \sum_j \frac{\partial L}{\partial y_j} f'_j W_{ji} g'_i = \sum_j \frac{\partial L}{\partial b_j} W_{ji} g'_i, \quad (2.11)$$

$$\frac{\partial L}{\partial V_{ij}} = \sum_{k,l} \frac{\partial L}{\partial y_k} \frac{\partial y_k}{\partial z_l} \frac{\partial z_l}{\partial V_{ij}} = \sum_k \frac{\partial L}{\partial y_k} f'_k W_{ki} g'_i x_j = \frac{\partial L}{\partial a_i} x_j. \quad (2.12)$$

ただし、

$$f'_i = f'((W\mathbf{z})_i + b_i), \quad (2.13)$$

のように活性化関数の微分の引数を省略した。これら損失関数の微分に関し、データセットにわたる平均を取ることで、損失の偏微分を得ることができる。以上より、出力層に近い側から

$$\frac{\partial L}{\partial b_i} \rightarrow \frac{\partial L}{\partial W_{ij}} \rightarrow \frac{\partial L}{\partial a_i} \rightarrow \frac{\partial L}{\partial V_{ij}}, \quad (2.14)$$

*2 機械学習の分野において、learning rate のような人間が決める必要のあるパラメータはハイパーパラメータと呼ばれ、 \mathbf{w} のようなモデルパラメータとは区別される。

という順番で計算を行うことで、モデルパラメータの偏微分が効率的に計算できる。このように、入力から $\mathbf{x} \rightarrow \mathbf{z} \rightarrow \mathbf{y}$ と出力までのデータの流れを順方向と見た場合、逆方向に情報を伝播させて偏微分を計算することから、このアルゴリズムは backpropagation と呼ばれる。さらに、deep learning で用いられる損失関数や活性化関数は通常微分が簡単に実行できるものを用いる。例えば損失関数として最も標準的な Euclid 距離の自乗である平均自乗誤差 (Mean Square Error, MSE) を用いた場合、

$$\frac{\partial \mathcal{L}}{\partial y_i} = \frac{2}{|D|} \sum_{i=1}^{|D|} (y^{(i)} - y_i) = \frac{2}{|D|} \sum_{i=1}^{|D|} y^{(i)} - y_i, \quad (2.15)$$

である。一方で、活性化関数として最も簡単な Rectified Linear Unit (ReLU) を考えると、

$$\text{ReLU}(x) := \max\{0, x\}, \quad \frac{d\text{ReLU}(x)}{dx} = \begin{cases} 1 & (x > 0) \\ 0 & (x < 0) \end{cases}, \quad (2.16)$$

となり^{*3}、sigmoid 関数 $\sigma(x)$ では

$$\sigma(x) := \frac{1}{1 + e^x}, \quad \frac{d\sigma(x)}{dx} = (1 - \sigma(x))\sigma(x), \quad (2.17)$$

となる^{*4}。つまりこのような関数を用いる限り、損失関数や活性化関数の微分は各ニューロンの出力の簡単な四則演算のみで計算可能である。したがって、学習データから順方向に計算を進め出力を得る、つまり forward pass させる際に、各ニューロンの出力を保持しておけば、backpropagation により微分は高速に、そして浮動小数点の範囲内で厳密に実行できる。

より多層の MLP に対し、以上のアルゴリズムを拡張する方法は自明である。ここで MLP を多層化するとは入力と出力の間に、線形変換と非線形変換からなるモジュールを積み重ねることを指す。この時、線形変換と非線形変換の繰り返しという MLP の構造が backpropagation を可能にするため、多層化しても 3 層 MLP と同様の議論ができ、高速な最適化が可能となる。また、前述の CNN や Transformer といった MLP とは異なるモデルに関しても、backpropagation が使える形式になっている。さらに、MLP, CNN, Transformer といった異なるアーキテクチャーで定義されたブロックを柔軟に組み合わせることでモデルの設計を行うことも可能である。このフレキシブルさこそが、ANN の利点と言えよう。またオブジェクト指向プログラミングにより、扱いやすい deep learning フレームワークを作ることができるということも利点である。実際、PyTorch [22], JAX [23], TensorFlow [24] といった優れたフレームワークがオープンソースソフトウェア (OSS) ベースで開発されている。また、計算の並列化が容易なアルゴリズムになっているため、NVIDIA による Graphic Processor Unit (GPU) と、その Application Programming Interface (API) である CUDA により、高速な演算が可能である。以上のように、deep learning の優れたアルゴリズムを起点として、ソフトウェア面とハードウェア面での様々な発展に支えられ、deep learning は急速に普及した。

^{*3} $x = 0$ において微分は定義できないが、数値計算上 $x = 0$ 直上となる確率は非常に小さいため問題にはならない。

^{*4} sigmoid 関数は微分が限られた領域でのみ大きな値をとる。この性質に起因し、多層化を考えた際に、勾配消失問題と呼ばれる学習が進まなくなる問題が誘発される。したがって、現代的には隠れ層への使用は基本的に避けられている。

2.2 最適化手法

2.2.1 SGD

前節では簡単な 3 層の MLP を例に取り、最急降下法に基づく最適化に関して言及した。しかし最急降下法は、損失関数がモデルパラメータに対して多峰性を持つ関数であった場合、極小点からの脱出が困難であるという問題がある。したがって、実際の計算では確率的勾配降下法 (Stochastic Gradient Descent, SGD) と呼ばれる手法を基本とする、様々な最適化アルゴリズムが用いられる。

最急降下法と異なり、学習データ D に対する損失 $\mathcal{L}(D)$ を最小化するのではなく、 D をランダムに分割したのに対して backpropagation を逐次実行する。つまり、

$$D = \bigcup_i D_i, \quad i \neq j \Rightarrow D_i \cap D_j = \emptyset, \quad (2.18)$$

という分割に対して $\mathcal{L}(D_1) \rightarrow \text{backpropagation} \rightarrow \mathcal{L}(D_2) \rightarrow \text{backpropagation} \rightarrow \dots$ という操作の繰り返しを、全ての $\{D_i\}_i$ に渡って実行する。この時、各 D_i はミニバッチと呼ばれる。通常、端数を除き $|D_1| = |D_2| = \dots$ となるように等分割され、その際 $|D_i|$ の値はバッチサイズと呼ばれる。また、 $\{D_i\}_i$ 全てに渡って学習するという一連の操作は 1 epoch という単位で呼ばれ、通常一連の操作を繰り返す回数は epoch 数と呼称される。実際の計算では、 D は各 epoch で毎回ランダムに分割されることに注意が必要である。

SGD とはバッチサイズを B とした場合 $B = 1$ の場合を指し、 $1 < B < |D|$ の場合は mini-batch gradient descent と呼ばれる。しかし deep learning の文脈では SGD という用語は mini-batch gradient descent に対しても用いられるため [25]、本論文ではまとめて SGD と呼称する。

上記のミニバッチ学習は、データ全体ではなくミニバッチごとにパラメータを更新するため、最適化すべき損失は毎回異なる。したがって $\mathcal{L}(D_i)$ の勾配は、 $\mathcal{L}(D)$ の勾配に対してランダムノイズを加えたかのように解釈でき、勾配が更新される過程は Langevin 方程式によりモデル化される [26]。したがって、このノイズにより極小値から脱出することが可能になるため、より良い解に収束していくことが期待される。

2.2.2 Adam

十分な最適化に要する epoch 数は少なければ少ないほど良い。SGD は現在地の勾配の方向にパラメータを更新するため、細い谷のような場所に斜めから侵入すると、振動により学習は遅くなる。そこで、Momentum [27] や Root Mean Square Propagation (RMSProp) [28] という指数移動平均により振動を抑制する方法が開発された。Adaptive Moment Estimation (Adam) [29] はこれら 2 つの手法を組み合わせた最適化手法で、2023 年現在において最も主流な最適化手法の一つである*⁵。

*⁵ Adam の原論文 [29] は現時点で 16 万回引用されている。

Adam では 2 種類の指数移動平均を用いる:

$$m_{t,i} \leftarrow \beta_1 m_{t-1,i} + (1 - \beta_1) \frac{\partial \mathcal{L}}{\partial w_i}, \quad (2.19)$$

$$v_{t,i} \leftarrow \beta_2 v_{t-1,i} + (1 - \beta_2) \left(\frac{\partial \mathcal{L}}{\partial w_i} \right)^2, \quad (2.20)$$

$$m_{t=0,i} = v_{t=0,i} = 0. \quad (2.21)$$

ただし t は学習のタイムステップであり, $\beta_1, \beta_2 \in [0, 1)$ は Adam のハイパーパラメータである. これらを用いて Adam は次のような更新則として定義される:

$$w_{t,i} \leftarrow w_{t-1,i} - \eta \frac{\tilde{m}_{t,i}}{\sqrt{\tilde{v}_{t,i} + \epsilon}}. \quad (2.22)$$

ただし, $\epsilon > 0$ はゼロ除算回避のためパラメータである. また $\tilde{m}_{t,i}, \tilde{v}_{t,i}$ は,

$$\tilde{m}_{t,i} := \frac{m_{t,i}}{1 - \beta_1^t}, \quad \tilde{v}_{t,i} := \frac{v_{t,i}}{1 - \beta_2^t}, \quad (2.23)$$

で定義される. つまり, 勾配の振動を移動平均により抑制し, さらに振動する方向の学習率を移動平均で減衰させることで, 効率的な学習を実現可能にする. Adam には learning rate を含めた 4 つのハイパーパラメータが存在し, 原論文では $(\beta_1, \beta_2) = (0.9, 0.999)$, $\epsilon = 10^{-8}$ が推奨されている.

2.3 汎化とドメイン

2.3.1 Hold-out validation

deep learning ではデータセット D に対してモデルが最適化される. neural network の表現能力は非常に高いので, 何も対策せずに学習を行うと, 典型的には過学習 (overfitting) を引き起こす*6. つまり, 未知のデータに対してタスクの性能が悪くなる (図 2.1).

過学習を避けるため, hold-out validation という手法がしばしば用いられる. これはデータセットを訓練用のデータとテスト用のデータに分割し, 訓練データのみを用いて学習を行いつつ, テストデータに対するエラーを監視することで, 過学習を検知する方法である. hold-out validation により過学習の振る舞いを実際に確認するため, MNIST [31] と呼ばれるデータセットを用いた教師あり学習を行う. MNIST は 0 から 9 の手書き数字のグレースケール画像とその正解ラベルから構成されるデータセットであり, CC0 でライセンスされている. 与えられた画像からそこに書かれている数字を推定するモデルを作るため, 画像を入力, 各数字の確率を出力とする 4 層の MLP を用いた. そして cross entropy を損失関数, 最適化アルゴリズムを Adam とする教師あり学習によ

*6 モデルパラメータを徐々に増やしていくとはじめは表現能力が向上するため, テスト誤差も減少する. しかしある程度のところから過学習により, テスト誤差は増大し始める. 実は, さらにパラメータを増やしていくとテスト誤差は再度減少し始める. これは二重降下現象 [30] として知られており, 様々なモデルに対して確認されている. この節で示した MNIST の教師あり学習においても, さらに多層化していくことで, 二重降下現象が確認される. しかし以上の話は正則化を行わない場合の話であり, 現実的に使われているモデルとは状況が異なることに注意が必要である.

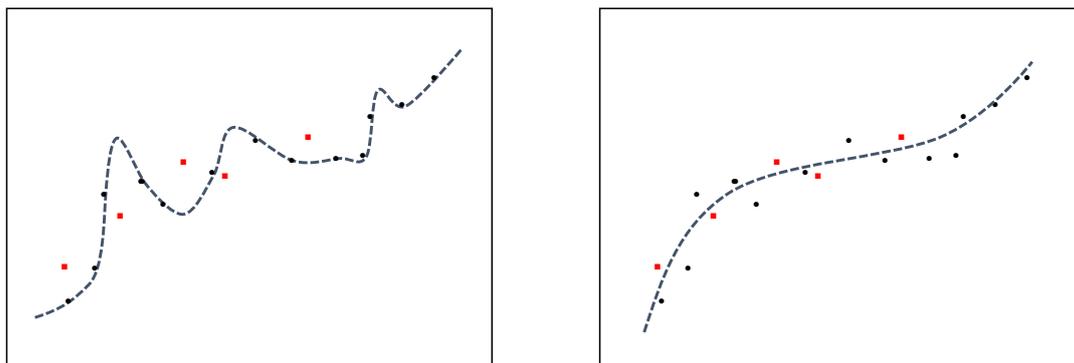


図 2.1 回帰分析における過学習の概念図. 訓練データを黒い点, テストデータを赤い点で示している. 左図は訓練データに対して過学習しておりテストデータに対して予測性能を持たない. 右図は訓練データに対して多少の誤差がある代わりに, テストデータに対する誤差が比較的小さい.

り, モデルの最適化を行った^{*7}. 図に学習データとテストデータそれぞれに対する epoch 毎の損失をプロットした. 図 2.2 の左は何も過学習対策をしていない場合であり, 学習データに過剰に適合した結果, 学習が進むにつれテスト誤差が増大していく様子が見て取れる. 過学習を避け, 良いモデルを得る最も簡単な方法は, early stopping である. これは, テスト誤差に改善が見られなくなったところで学習を打ち切るという手法である. 例えば, n 回連続でテスト誤差が最小値を更新しなかった場合, 学習を終了するといった基準に基づく early stopping 法などがある. この手法とは別に, learning rate decay, つまり学習が進むにつれ learning rate の値を減少させるという手法が存在する. learning rate の減衰方法には, 学習の振る舞いを見て人間が決めたり, テスト誤差の振る舞いから自動的に決定するといった方法などがある. この節では後者のアイディアに従い, テスト誤差が 5 回連続でこれまでの最小値を下回らなければ, learning rate を 0.1 倍するという手法で, learning rate decay を行った. その結果は図 2.2 の右に図示されており, 過学習が抑制されている様子が見て取れる. タスクの複雑さやモデルの深さ, そして他の過学習対策にも依存するが, シャープな learning rate decay をおこなうと, 次の瞬間急速に損失が低下する振る舞いを見せることがある. このように, learning rate decay の方が最終的なモデル性能が向上する確率が高くなる傾向にあるため, 単純な early stopping よりも, learning rate decay を用いたアプローチの方がよく見られる.

テスト誤差を監視し, learning rate のようなハイパーパラメータのチューニングを行うことで, より良いモデルが得られる. しかし, 過度にチューニングを行うと, それはもはやテストデータを用いて学習しているようなものであり, 過学習が評価できているかは疑わしい. そこで, hold-out validation 法では教師データを訓練データとテストデータの 2 種類に分割するのではなく, 検証データを導入することで 3 種類に分割するという方法も存在し, 実際によく用いられている. この

^{*7} 隠れ層の次元は全て 100 で, 出力層を除き活性化函数には Sigmoid Linear Unit (SiLU) を用いた一方で, 出力層には softmax 関数を適用した. バッチサイズは 128 で, Adam のハイパーパラメータは learning rate を除き PyTorch のデフォルト値を採用し, learning rate の初期値として 10^{-3} を用いた.

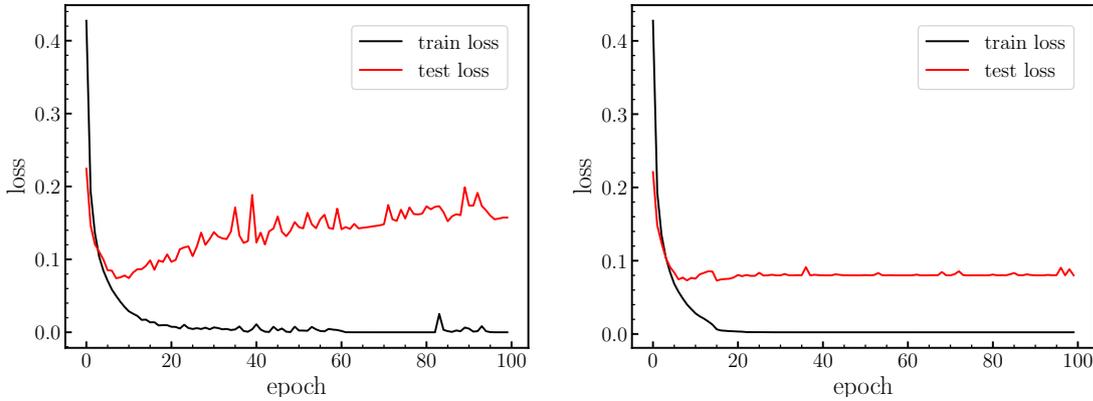


図 2.2 MNIST を用いた教師あり学習における損失の振る舞い. 横軸に epoch, 縦軸に cross entropy による損失の値をとり, 教師データ (黒線) とテストデータ (赤線) それぞれに関してプロットを行った. 図の左には何も過学習対策を行っていない場合の結果を示しており, 右にはテスト誤差を利用した learning rate decay の結果を図示している.

時, ハイパーパラメータのチューニングは検証データを用いて行い, テストデータは最終的なモデルの評価のみに用いられる.

2.3.2 ドメインシフト

前節では単一の学習データを分割し, モデルの訓練に用いないデータを作り出すことで, 過学習の評価を行った. しかしこのような手法では, 真に未知のデータセットに対してモデルの性能を保証することはできない. この議論を行うため, この節では転移学習 (Transfer Learning, TL) [32] におけるドメインの概念を導入する.

ドメインは feature (input) space \mathcal{X} と \mathcal{X} 上の確率分布関数 $P(X), X \in \mathcal{X}$ の対として定義される [32, 33, 34]:

$$\mathcal{D} := \{\mathcal{X}, P(\mathcal{X})\}. \quad (2.24)$$

またラベル付きデータの場合は, 各入力データ X に対して出力データ Y が割り当てられている. ここで output space を \mathcal{Y} とし, \mathcal{Y} 上の条件付き確率分布 $P(Y|X), Y \in \mathcal{Y}$ に対して,

$$\mathcal{T} := \{\mathcal{Y}, P(Y|X)\}, \quad (2.25)$$

で定義される \mathcal{T} はタスクと呼ばれる^{*8}. deep learning におけるデータセットは何らかのドメイン \mathcal{D} とタスク \mathcal{T} からサンプリングされたものと考えられる. 様々な問題に deep learning が適用されていることから分かるように, 人間から見たこの世界は多様なドメインとタスクで溢れている.

一般に学習に用いたデータセットのドメイン, タスクと, 実際のアプリケーションで使用されるデータセットのドメイン, タスクは異なる. 以下では簡単のためタスクが同じか, 存在しないよう

^{*8} タスクが決定論的ではなく条件付き確率で定義されているのは, 実際のデータはノイズやラベリングミスなどの様々な要因により, 決定論的なラベル付けが可能な場合は限られているからである.

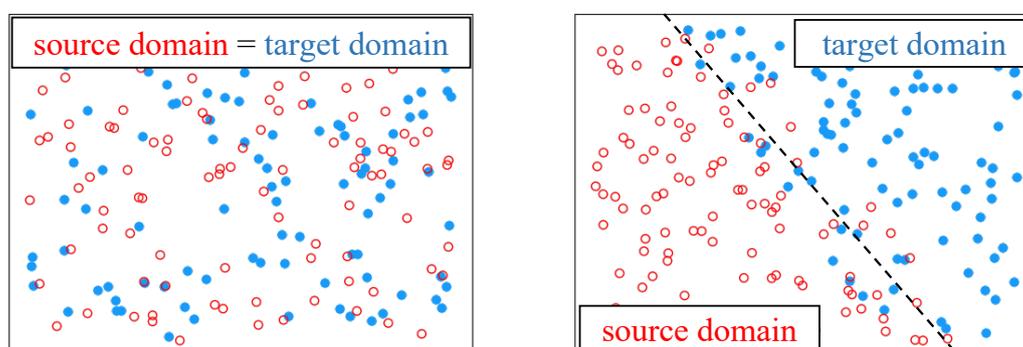


図 2.3 ドメインシフトの概念図. 赤い点, 青い点はそれぞれソースドメイン, ターゲットドメインからサンプリングされたデータ点である. 左図はターゲットドメインとソースドメインが等しい場合で, 右図は両者が異なる場合を表している.

な場合を考える. 学習に用いられるデータのドメインはソースドメイン D_s , 本番環境で用いられるデータのドメインはターゲットドメイン D_t と呼ばれる. 前節の Hold-out validation では単一のデータセットをランダムに分割するため, 単一のドメインのみで過学習を評価しており, ターゲット D_t に対する性能を保証できない. このような学習方法は転移学習の文脈では “traditional machine learning method” と呼ばれる [33, 34]. この traditional machine learning method が正当化されるのは $D_s = D_t$ の時であり, 現実社会の問題でこのような状況は稀であり, 時間や場所といった様々な要因で容易に等号は不等号に変わる. したがって, このドメインシフトは deep learning における中心的な話題の一つである.

この問題の深刻なところは, 現実のデータが定義されるような超多次元ベクトル空間において, 人間がドメインの違いを理解するのが困難な点である. つまり, 人間が理解できるような指標でソースドメインとターゲットドメインが近しかったとしても, neural network が両者は近いと判断するかは非自明である*⁹. したがってドメインシフトに対するモデルのロバストネスは注意深く調べる必要がある.

この情報化社会において, 様々なデータセットが配布, または販売されている. 教師データとは別にそのようなデータセットが利用できる場合には, 転移学習のテクニックを用いることができる. 転移学習は利用できるドメインとタスク, そしてそれらの違いなどによって分類することが可能である (図 2.4). 目的とするドメイン, タスクとは多少異なるデータであっても活用することができるため, 物理学においてもデータを集めて保管しておくことは, 極めて重要であることを強調しておく.

*⁹ 超大規模なデータセットが存在すれば, Variational Autoencoder (VAE) [35, 36] といった手法で潜在変数空間にマップするようなモデルを用意し, 手元のデータのドメインをより詳細に調べることができる.

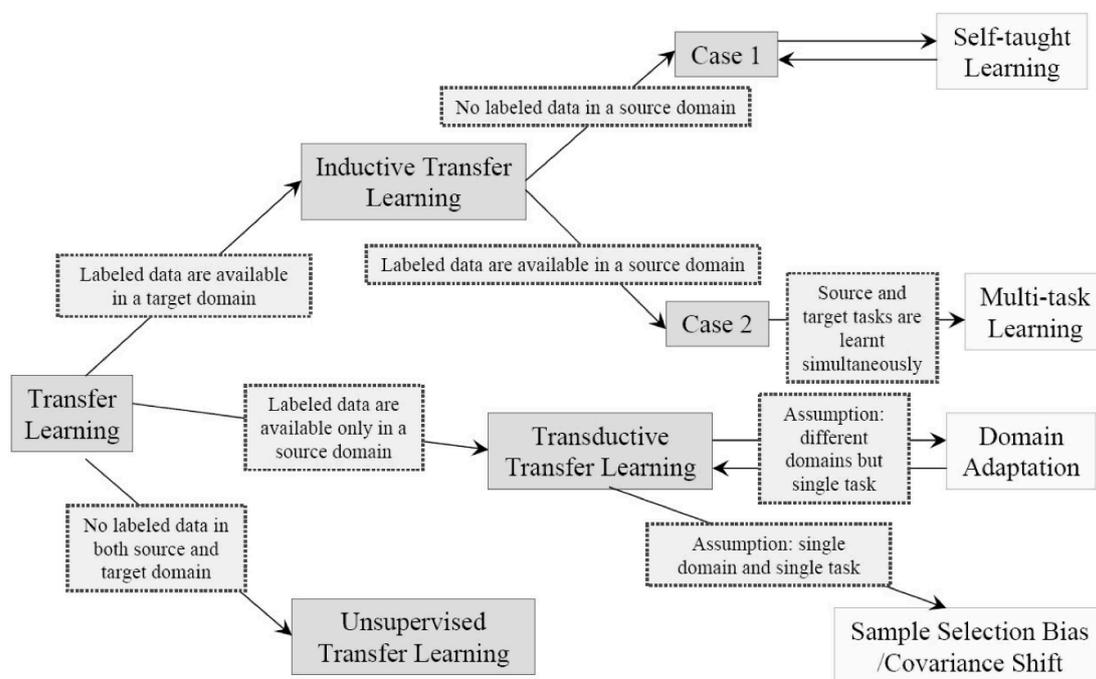


図 2.4 転移学習 (Transfer Learning, TL) の分類. ドメイン間の違いや, データラベルの有無, タスクの違いなどに応じて様々な手法が存在する. 図は文献 [32] より抜粋.

第 3 章

Deep learning による Skyrme エネルギー密度汎関数の再構築

3.1 KS-EDF の問題点と機械学習

Kohn-Sham スキームに基づく DFT 計算は, 電子系のみならず原子核系においても多大な成功を収めてきた [10]. 特に原子核系では核子間相互作用がまだ完全に解明されていないため, ほとんどの計算は厳密には一種のモデル計算である. そこで相互作用から出発して EDF を導出するという流れよりも, 直接 Kohn-Sham EDF からスタートするというアイデアは自然である. つまり KS-EDF に ansatz を導入し, それらのハイパーパラメータを実験値に合うように最適化する. Skyrme EDF (1.16) も, 上記の手法により得られた Kohn-Sham 型の EDF であり, 原子核の基底状態の解析に大きく貢献してきた. その一方で, 原子核理論における Kohn-Sham スキームの計算には, 2つの問題点が存在する.

3.1.1 計算コストの問題

Kohn-Sham スキームでは, 非線形 Schrödinger 方程式に似た形式を取る Kohn-Sham 方程式 (1.15) を解く必要がある. つまり, 1 粒子ハミルトニアン部分が解に依存しているため, 自己無撞着に解く必要があり, 計算コストが iteration の数に比例して増大する. 各 iteration においてボトルネックとなる部分は, EDF やアルゴリズムによって変わるため一概に言うことができない. したがって, 以下では代表的なものに関して個別に解説を行う.

有限差分法

まず, 本論文でも用いる Skyrme EDF, Coulomb direct term, Slater 近似した Coulomb exchange term [37] の和で構成される EDF に議論を限定する. 空間メッシュに区切って解くような場合には, 固有値問題を扱う際の対角化の操作がボトルネックになり, 計算コストはメッシュの総数を N_{mesh} とすると $O(N_{\text{mesh}}^3)$ でスケールする.

調和振動子基底展開

基底展開法の場合、ボトルネックはより複雑になる。簡単のため、3次元等方調和振動子基底を用いる場合に議論を限定する。数値計算に用いられる調和振動子基底は、各空間方向の固有値 (n_x, n_y, n_z) に対して

$$n_x + n_y + n_z \leq N_{\text{sh}}, \quad (3.1)$$

を満たすもののみが用いられることが多い。この時、基底の数は

$$N_{\text{basis}} = (N_{\text{sh}} + 1)(N_{\text{sh}} + 2)(N_{\text{sh}} + 3)/6, \quad (3.2)$$

で与えられる。したがって、1粒子ハミルトニアンを対角化する計算コストは $O(N_{\text{basis}}^3)$ でスケールする。その一方で、Coulomb direct term の平均場ポテンシャルの計算は、直接実行すると、 $O(N_{\text{basis}}^4)$ のコストがかかる。しかしこの部分は、調和振動子基底が空間各方向に関して separable であるという性質と、Coulomb 相互作用が Gaussian 型の相互作用の重ね合わせで効率よく近似できるという性質から、計算アルゴリズムの工夫により $O(N_{\text{Legendre}} N_{\text{sh}}^8)$ まで計算コストを削減することが可能である。ただし、 N_{Legendre} は Coulomb 相互作用の分解に利用する Gauss-Legendre quadrature の正の領域のノード数であり、典型的には 20 以上の数が取られる。調和振動子基底で類似のコードを書き、実際に計算を行った著者の経験 [38] に基づくと、 $O(N_{\text{Legendre}} N_{\text{sh}}^8)$ が支配的になる。

以上のボトルネックは粒子数の大きい計算を行う場合、深刻な問題になる。特に電子系では、分子量が 10^6 を超えるような高分子を取り扱うことがあり、 10^6 本の電子の Kohn-Sham 軌道を得るには、計算精度を度外視したとしても最低で $N_{\text{mesh}} = 10^6$ が必要になり、膨大な計算コストを要する。原子核系でも、 ^{236}U に代表されるような重い核を精度良く計算するには、多大な計算コストを要する。加えて近年の原子核研究では分野の進展に伴い、一度の研究で複数回の KS-EDF を用いた計算が求められることも多くなってきた。しかしながら、計算アルゴリズムの最適化はこれまでに十分行われてきており、オーダーが変わるほどの改良は、あまり見込めない。

3.1.2 OF-EDF の存在

Skyrme EDF は現象論的に与えられた ansatz から出発して得られた KS-EDF である。第1章で述べたように、DFT は外場付き変分問題により定義されるエネルギー汎関数の Legendre 変換により定式化される。DFT の数理においては、この Legendre 変換により得られた密度汎関数を逆 Legendre 変換した際に元に戻るような、外場の定義域が問われる。電子系よりも複雑な相互作用を持つ原子核物理において、この問題の解決は難しいと考えられる。したがって、厳密な意味で DFT の存在が証明されているわけではないため、現象論的に与えられた Skyrme DFT において、密度と外場の間の対応を調べることに意味がある。また実際の計算では、離散化や Appendix B で述べられているような、単純化や収束のためのテクニックを利用している。以上の数学的な根深い問題

と、テクニカルな要因が様々に入り乱れる実際の Skyrme DFT 計算において、エネルギーが密度から推定可能かどうかは、理論として興味深い。仮に Skyrme EDF を再現するような OF-EDF を構築できるのであれば、Hohenberg-Kohn の定理という DFT の原理に従う OF-EDF を得るための、重要な一歩となることが期待される。

3.1.3 本研究の目的

以上の問題は電子系原子核系問わず、DFT を用いた分野に共通して生じ得る。特に計算コストの問題は、DFT を用いた材料探索のような工業的な需要もあり、社会的な影響も考慮すると、極めて重要な問題であると言える。そこで近年電子系では、機械学習を用いた OF-EDF の構築が行われるようになってきた。特に文献 [39] では、深層学習を用いた 2 次元電子系において、様々な外場のもとで KS-EDF を用いて計算されたエネルギーと電子密度から、OF-EDF の再構築が行われた。本章では、電子系で発展した上記の手法を原子核系に応用し、Skyrme EDF の結果を再現するような OF-EDF の再構築を目指す。特に原子核系では核子の超伝導性が重要になるため、本研究では超伝導理論に動機付けられた Kohn-Sham スキーム (1.25) に基いた計算を行う。これは電子系における先行研究 [39] では考慮されていたかった効果であることは強調しておく。以上より、本研究では以下の 2 点に着目し議論を行う：

1. deep learning は Skyrme DFT の計算コストの問題を解決し得るか。
2. Skyrme DFT の結果を再現するような OF-EDF を構築できるか。

3.2 手法: ランダムポテンシャルと深層学習

3.2.1 Skyrme-Kohn-Sham DFT

Skyrme DFT に関しては第 1 章で既に解説しているため、この節では実際の計算に向けたより詳細な事項に関して述べる。原子核は 3 次元のオブジェクトである。しかしながら、対称性を一切要請することなく 3 次元密度を計算し neural network で学習するのは、現状においても不可能ではないが、現在著者が利用可能な計算資源では難しい。したがって軸対称性と時間反転対称性を要請し、円筒座標系 (r, z, ϕ) で議論を行う。この簡単化のもとでも、oblate や prolate などの重要な変形状態を扱うことができるため、核分裂といった現象への応用を考えれば我々の系は依然として興味深い。軸対称な系では、 z 方向の Kohn-Sham 軌道の全角運動量演算子 \hat{J}_z が 1 粒子 Hamiltonian と可換になるため、Kohn-Sham 方程式は演算子 \hat{J}_z の固有値 $\Omega\hbar$ を用いてブロック対角化される [40]。実際の計算ではこの固有値にカットオフを入れる必要があり、本研究では $\Omega \leq 9/2$ とした。3 次元のオブジェクトである Kohn-Sham 軌道は角度 ϕ 依存性を持つが、解として得られる粒子数密度は (r, z) のみに依存することに注意が必要である。

本研究では、2 次元空間格子を用いた有限差分法に基づき、KS-DFT の計算を行う。ただし、格子の刻み幅は $dr = dz = 0.8 \text{ fm}$ とし、空間格子点の数は r 方向に $N_r = 10$ 点、 z 方向に $N_z = 20$ 点とした。また、数値計算において $1/r$ という項が存在するため、計算の安定性という観点から r 方

向に関して原点を避ける. さらに, z 方向に関して同様に parity 変換に関して対称な格子点という観点から原点を避け, r, z 共に副格子上に点を取るものとする. したがって, 離散化された格子点 $\{r_i\}_{i=1}^{10}, \{z_i\}_{i=1}^{20}$ を明示すると

$$r_i \in \{0.4 \text{ fm}, 1.2 \text{ fm}, \dots, 7.6 \text{ fm}\}, \quad (3.3)$$

$$z_i \in \{-7.6 \text{ fm}, -6.8 \text{ fm}, \dots, 7.6 \text{ fm}\}, \quad (3.4)$$

となる. 離散化された Kohn-Sham 方程式の境界条件には, Box 境界条件を採用する. つまり, 格子点の外側で Kohn-Sham 軌道は 0 になることを要請する.

更なる簡単化のため, 原子核を ^{24}Mg に限定し, Coulomb 相互作用を無視する. この時, 系はアイソスピンに対して対称になるため, 陽子部分と中性子部分の波動関数として同じものを使うことで, 計算コストを半分にすることができる. ここで, Coulomb 相互作用を無視することの正当性に関して補足を行う. Skyrme DFT で用いられる Coulomb 相互作用は, 以下のような平均場理論に動機づけられた direct term と, Slater 近似 [37] に動機づけられた汎関数で表される exchange term の和で表示される:

$$E_{\text{Coulomb}}[\rho_p] = \frac{e^2}{2} \int d^3r \int d^3r' \frac{\rho_p(\mathbf{r})\rho_p(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} - \frac{3}{4} \left(\frac{3}{\pi}\right)^{1/3} e^2 \int d^3r \rho_p^{5/3}(\mathbf{r}). \quad (3.5)$$

ただし, e は電荷素量である. したがって, Coulomb term は陽子数密度のみに陽に依存した汎関数であり, OF-EDF の再構築という目的においては重要でない. また計算コストにおいて, 有限差分法では Coulomb direct term の計算はボトルネックにならない. これは Poisson 方程式を用いた, 効率の良い平均場ポテンシャルの計算アルゴリズムが存在するためである. したがって Coulomb term を無視することは, deep learning の結果と計算速度を比較する上でほとんど影響を与えない.

また, 対相関の効果を陽に取り入れた DFT 計算を行うため, surface type の DDDI pairing 汎関数 (1.19) を用いる. この時, BCS 波動関数に動機づけられた ansatz に基づく Kohn-Sham スキームにより計算を行う. ただし, pairing strength は現象論的な gap parameter を再現するように最適化し, 本研究では一貫して $V_0^{(n)} = V_0^{(p)} = -683.344 \text{ MeV fm}^3$ を用いる. また BCS スキームで計算を行う場合, 発散回避のため BCS 方程式において考慮する Kohn-Sham 軌道に制限をかける必要がある. 本研究では, Kohn-Sham 軌道の 1 粒子エネルギーが 60 MeV 以下のものに限るという, シャープなカットオフを採用した. 以上の設定のもとで, ^{24}Mg の基底状態のエネルギーを計算すると -222.59 MeV であり, 密度は球形であった*1.

本研究では Kohn-Sham 方程式を様々な外場のもとで解き, deep learning で学習するための教師データを生成する. この時, 外場の定義域を工夫するなどしない限り, 一般には原子核密度の重心位置は原点に一致しない. しかし, 原子核構造研究において原子核の重心位置は物理的に興味のある対象ではない. また, 本研究で使用する格子空間のサイズは, 原子核が原点近傍にある場合十分な計算精度を与えるような大きさである. したがって, 本研究では一貫して 2 次の拘束を用いた重心

*1 pairing strength を弱めると, prolate に変形することは確認している.

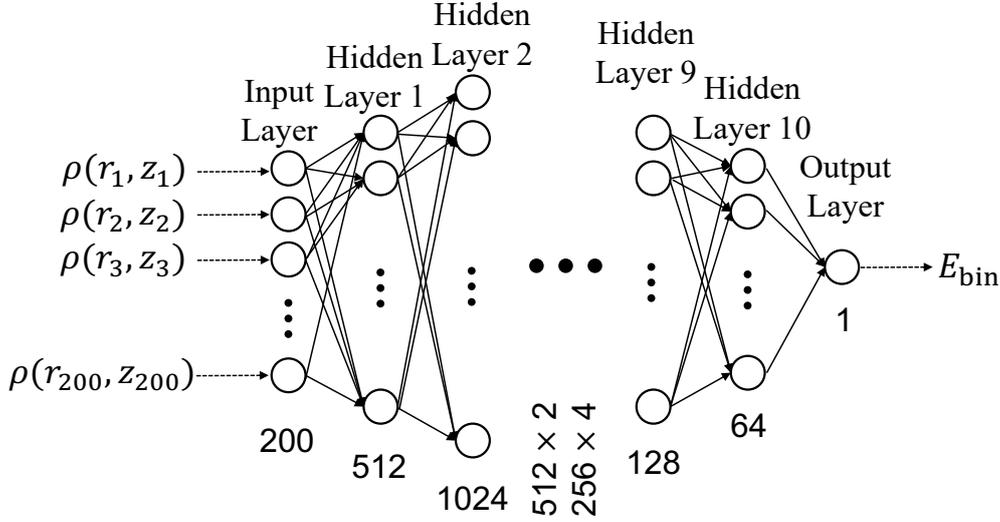


図 3.1 OF-EDF の学習に使用したニューラルネットワーク. 10 層の隠れ層を持ち, それら全ての活性化関数には ReLU を用いた. また, 出力層に関しては sigmoid を用いている. 図中の数字は, 各隠れ層のニューロンの数を表している.

固定を行う. つまり, 以下の汎関数を Skyrme EDF に加える:

$$E_{\text{CoM}}[\rho] = \frac{C}{2} \left(\int d^3r z \rho(\mathbf{r}) \right)^2. \quad (3.6)$$

ただし, C は拘束の強度を決めるハイパーパラメータであり, 本研究では一貫して $0.625 \text{ MeV}/\text{fm}^2$ を用いる.

3.2.2 Neural network

モデル

本研究では, KS-DFT により得られた様々なエネルギーと密度の対からなるデータセット $D = \{E^{(i)}, \rho^{(i)}\}_i$ に対して, 回帰分析を用いることで OF-EDF, $E = E[\rho]$ の再現を試みる. この OF-EDF の近似器として, fully-connected layer のみからなる multi-layer perceptron (MLP) *2 を採用する. fully-connected layer は, 入力されたベクトルに行列を作用させ, バイアス項を足すことで出力ベクトルを用いる. この行列とバイアス項がモデルパラメータであるため, パラメータの数はおおよそ入力ベクトルの二乗で増大する. したがって, Full HD の画像といった, ベクトル換算した場合に 10^6 次元を超えるような大きいデータを取り扱う場合には, Convolutional neural network (CNN) [15, 21] や Vision transformer (ViT) [41] といった異なるアーキテクチャーを使用する必要がある. しかし, 本研究における入力データは核子数密度 ρ であり, 離散化の結果, $1 \times 10 \times 20$ という, 非常に小さなサイズのグレースケール画像データとみなせる. したがって, 本研究では fully-connected layer を採用し, 図 3.1 で表示される 10 層の隠れ層を持つ MLP を用い,

*2 第 2 章を参照.

OF-EDF の学習を行う。ただし、隠れ層の活性化関数は全て Rectified Linear Unit (ReLU) を用い、output layer の最後には sigmoid 関数を作用させる。各隠れ層のニューロン数に関しては図 3.1 中に記載してある。

最適化手法

次に、学習方法に関して述べる。教師あり学習では、教師データの正解ラベル $E_{\text{ans}}^{(i)}$ とモデルの予測値 $E_{\text{pred}}^{(i)}$ の間の誤差が最小になるように、モデルパラメータが最適化される。本研究では、損失関数として一般的に用いられている平均自乗誤差 (Mean Square Error, MSE) を作用する:

$$\mathcal{L}_{\text{MSE}}(D) = \frac{1}{|D|} \sum_{i=1}^{|D|} \left| E_{\text{ans}}^{(i)} - E_{\text{pred}}^{(i)} \right|^2. \quad (3.7)$$

ただし、 $|D|$ はデータセット D の要素数であり、括弧付き上付き添えはデータの要素のラベルであり、本論文では一貫してこのノテーションを採用している。また最適化ではバッチサイズを 128 としたミニバッチ学習を用い、500 epoch まで学習を行った。

deep learning では勾配法に基づき、損失関数から計算された微分係数を用いてパラメータは更新される。本研究ではパラメータの更新アルゴリズムとして Adaptive Moment Estimation (Adam) [29] を用いた。ベーシックな Adam のハイパーパラメータには、移動平均に関するパラメータが 2 つ、正則化のパラメータが 1 つ、learning rate というモデルパラメータ更新の大きさを制御するパラメータが 1 つ存在する。このうち最も重要なのは、learning rate で、より良い学習結果を得るためには、学習の過程で変化させることが望ましい。本研究では epoch に応じて段階的に learning rate を減衰させるというアプローチを採用した。初期値を 10^{-4} とし、 10^{-5} (epoch = 101), 10^{-6} (epoch = 201), 5.0×10^{-7} (epoch = 301), and 10^{-7} (epoch = 401) と、減少させた。実際の計算は Keras API [42] を用いて実装し、Adam の learning rate 以外のハイパーパラメータに関しては、Keras API のデフォルト値を用いた*3。

3.2.3 データセット

一度 KS-EDF が与えられれば、基底状態のエネルギー、粒子数密度、外場の対応が得られる。この事実を利用し、様々な外場のもとで KS-DFT を実行することで、OF-EDF を学習するためのデータセットを得ることができる。この節では、比較のため 2 種類の多様性の異なるデータセットを導入する。これらの基本的なアイディアは 2 次元電子系を扱った先行研究 [44, 39] に基づいている。しかし、我々の系は 3 次元軸対称原子核系であるため、多少の変更を加える。

*3 Adam の改良版である、Rectified Adam optimizer (RAdam) [43] でも計算を行ったが、結果に有意な改善は見られなかった。

Simple Harmonic Oscillators (SHO)

一つ目の外場として, Simple Harmonic Oscillator (SHO) を導入する. これは, 名前が表す通り以下の式で表現される, z 方向の中心をシフトした変形調和振動子型のポテンシャルである:

$$v_{\text{SHO}}^{(i)}(r, z) = \frac{1}{2}k_r^{(i)}r^2 + \frac{1}{2}k_z^{(i)}(z - z_0^{(i)})^2. \quad (3.8)$$

ただし, 3 種類のパラメータは, それぞれ $0 \leq k_r, k_z \leq 1.1 \text{ MeV/fm}^2$, $-1.6 \text{ fm} \leq z_0 \leq 1.6 \text{ fm}$ の範囲から一様乱数としてサンプリングした*4.

SHO は想定される外場の定義域*5 のうち, 非常に限られた領域しか被覆しない. しかし, 実際の原子核構造計算では, 多重極モーメントのような非常に限られた領域の外場だけに焦点を当てた計算が行われている. したがって, SHO のような単純な構造のデータセットに関して, その学習プロセスを調べることは, 今後の OF-EDF の研究において重要である.

2次元電子系の先行研究 [44, 39] では, 2次元座標 (x, y) に対して, 中心位置 (x_0, y_0) をランダムにした変形調和振動子型のポテンシャルが用いられていた. したがって, 相違点は座標系の変更に伴う変数の取り替えと, r 方向に関して中心位置を 0 に固定した点のみである.

Random Potentials (RND)

2つ目の方法として, Random Potential (RND) を導入する. 本研究の目的は Skyrme EDF を再現するような OF-EDF の構築であり, この目的のため, 様々な範囲の粒子数密度とエネルギーの対を生成する必要がある. しかしながら, どのような確率密度で生成された外場が, 効率よく多様な密度を与えるかは非自明な問題である. 格子空間においてランダムな外場を生成する方法は, 基底展開, Gaussian Processes (GPs) [45] など, 様々考えられるが, 本研究では先行研究 [44, 39] と同様, グラフィカルなアプローチを採用する:

$$v_{\text{RND}}^{(i)}(r, z) = m(r, z) \times \text{sr}^{(i)}(r, z). \quad (3.9)$$

ただし, $m(r, z)$ と $\text{sr}^{(i)}(r, z)$ は次のように定義される.

$$m(r, z) = e^{-4.0 \max\{0, \sqrt{r^2+z^2}-r_0\}^2/r_0^2}, \quad (3.10)$$

$$\text{sr}^{(i)}(r, z) = \sum_{r', z'} s^{(i)}(r, z; r', z') \text{rnd}^{(i)}(r', z'). \quad (3.11)$$

ここで, $s^{(i)}$ は Gaussian filter であり,

$$s^{(i)}(r, z; r', z') = e^{-\{(r-r')^2+(z-z')^2\}/\mu_2^{(i)}(r', z')}, \quad (3.12)$$

で定義される smoothing 関数である.

つまり, ランダムポテンシャルは以下の流れに沿って生成される. まず各格子点 (r, z) 上において, $[v_{\min}, v_{\max}]$ の範囲から一様乱数を生成し割り当てる. このようにして生成されたポテンシャル

*4 本研究では, 特に言及のない限り乱数は全て一様乱数である.

*5 原子核における DFT の数理はまだ完成しておらず, 外場の定義域の同定は未解決問題である (第 1 章を参照).

は、非常に高周波の成分が多く含まれたものである確率が高い。そこで $s^{(i)}$ で表される Gaussian filter (3.11) 式を作用し、畳み込みを行うことで、滑らかな低周波成分が支配的な関数に変換する。この時、(3.12) 式中の Gaussian width $\mu_2^{(i)}(r, z)$ に関して、外場が特定の長さスケールを学習することを避けるため、 $[\mu_{2\min}, \mu_{2\max}]$ の範囲でランダムに生成する。最後に、格子の空間の境界付近での外場の振る舞いに起因した、数値的な不安定性を抑制するため、(3.10) 式で定義されるマスクを (3.9) 式のように作用させる。

RND にはハイパーパラメータが5つ存在する。例えば v_{\max} は大きく取るほど多様な外場が実現可能になるが、我々の興味の対象は基本的に低エネルギーの原子核であるため、それほど大きな値は必要ない。文献 [46] によると、 ^{24}Mg の回転バンドに関して、基底エネルギーから 4_1^+ までの励起エネルギーの実験値は、およそ 4 MeV である。そこで、本研究では RND のハイパーパラメータとして、 $r_0 = 1.4 \times 1.2A^{1/3} \text{ fm}$, $v_{\min} = -1.1 \text{ MeV}$, $v_{\max} = 1.1 \text{ MeV}$, $\mu_{2\min} = 0.8 \text{ fm}^2$, and $\mu_{2\max} = 1.2 \text{ fm}^2$ を採用した。これらは binding energy のみに着目した場合、励起状態が 5 MeV 以下の ^{24}Mg のデータが多く得られるようにチューニングされている。

マスクによる処方箋によって、外場のコンスタントシフトを抑制し、適切な領域から外場をサンプリングすることも可能になる。第1章で述べたように、DFT は外場と密度の間の Legendre 変換により定式化される。したがって、この Legendre 変換が成立するような外場の定義域を適切に与えることが、DFT の数理において重要になる。特に外場のコンスタントシフトは pairing のない場合の Kohn-Sham スキームにおいて、解として同じ密度を与える。例えば電子系の DFT では、外場の定義域 X_v は (1.9) 式で与えられ、外場のコンスタントシフトは禁止される。原子核系の X_v は未解決問題であるが、電子系と同様にコンスタントシフトは自明に問題となるため、外場のコンスタントシフトを禁止するマスク (3.10) の存在は重要となる。ただし、BCS 近似に動機付けされた Kohn-Sham スキームでは、外場のコンスタントシフトは粒子数の拘束条件に吸収されるため、自動的に調整される。しかし、外場から密度の情報を推定する上では、コンスタントシフトの抑制は重要となる。

最後に、2次元電子系の先行研究 [44, 39] との違いに関して述べる。2次元電子系では、 $\text{rnd}^{(i)}(r, z)$ としてランダムな binary データを使用していた。つまり、空間各点で 0 または 1 をランダムに割り当てた配列の生成を行なった。この操作は、原子核に起因した Coulomb ポテンシャルが外場として扱われることの多い電子系においては、物理的に興味のあるドメインを被覆するという観点からもっともらしい。しかし、原子核系の性質を議論する上で適切なポテンシャルが何かという問いは、極めて非自明で難しい問題である。伝統的に多重極モーメントを外場に用いた計算が多く行われてきたが、その選択が最適である保証はない。したがって、本研究では binary データより複雑な、実乱数を用いて $\text{rnd}^{(i)}(r, z)$ を生成した。また、本研究では KS-EDF の定義に重心位置の拘束を用いたことに伴い、マスク m に関しても、異なるものを用いている。

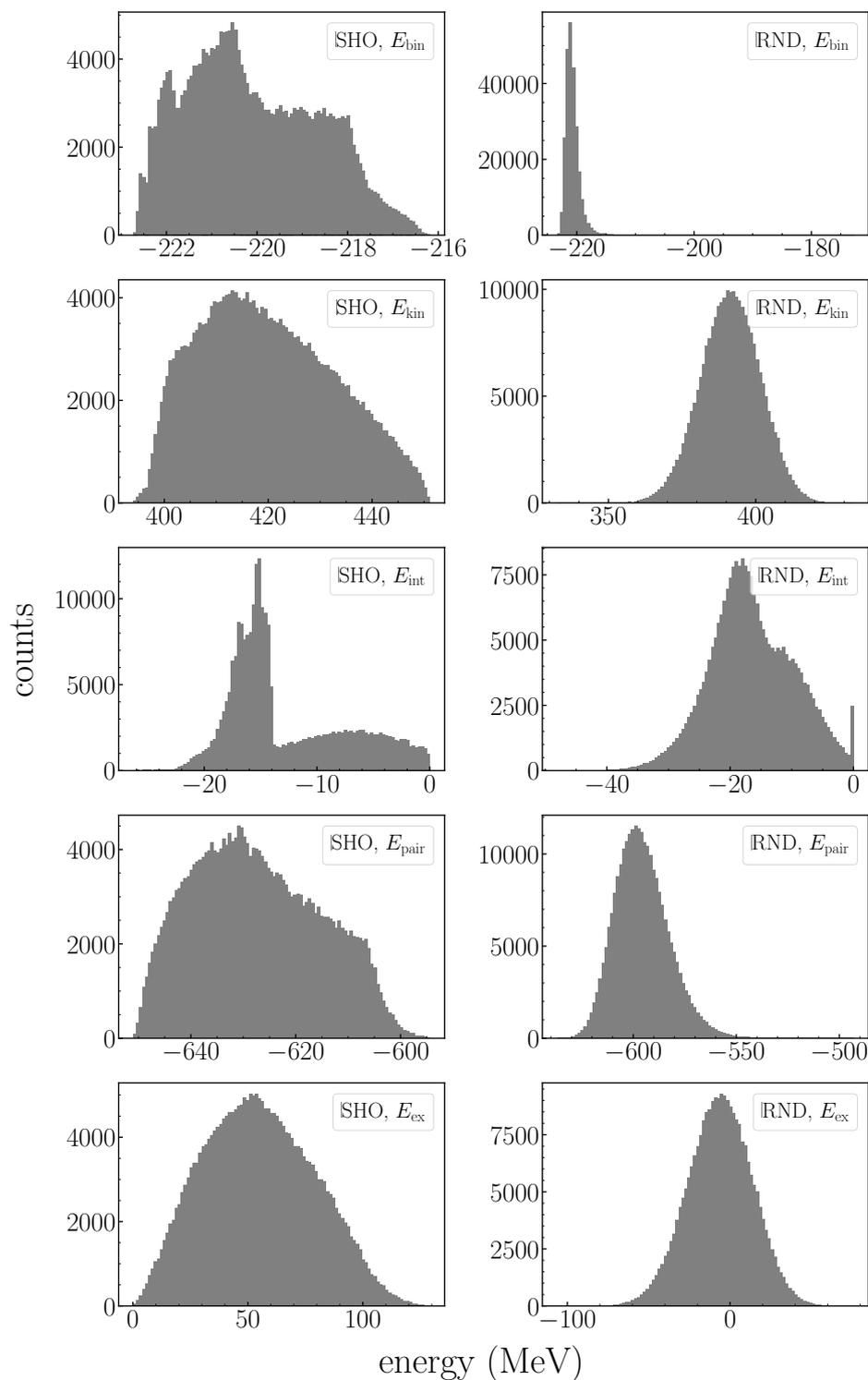


図 3.2 異なる 250 000 サンプルの外場に対する Skyrme EDF 計算の結果のヒストグラム. 左側の列は SHO, 右側の列は RND に対する結果である. また, 最上段から順に binding energy, kinetic energy, interaction energy, pairing energy external field energy の結果を表示している. 全てのパネルに関して横軸の単位は MeV である.

3.3 結果： ^{24}Mg の OF-EDF

3.3.1 データセットと前処理

前節で示した学習プロトコルを KS-EDF に適用した結果に関して紹介する前に、データセットの前処理に関して説明を行う。本研究では、SHO, RND それぞれに関して 250 000 個のデータを用意した。各データには、核子数密度 $\rho^{(i)}$ と原子核の binding energy $E_{\text{bin}}^{(i)}$ だけでなく、以下のエネルギーの情報も含まれている：

$$E_{\text{kin}}[\tau] := \frac{\hbar^2}{2m} \left(1 - \frac{1}{A}\right) \int d^3r \tau(\mathbf{r}), \quad (3.13)$$

$$E_{\text{int}}[\rho, \tau, \mathbf{J}] := E_{\text{Skyrme}}[\rho, \tau, \mathbf{J}] - E_{\text{kin}}[\tau], \quad (3.14)$$

$$E_{\text{pair}}[\rho, \tilde{\rho}] := E_{\text{DDDI}}[\rho, \tilde{\rho}], \quad (3.15)$$

$$E_{\text{ex}}[\rho] := \int d^3r v(\mathbf{r})\rho(\mathbf{r}). \quad (3.16)$$

上から順に kinetic energy, interaction energy, pairing energy, external field energy であり、右辺の定義は 1.3 節で既に述べている。ただし本研究では、簡単化のために陽子数の local density と中性子の local density が一致するため、アイソスピンの足は全て省略した。 $E_{\text{bin}} := E_{\text{kin}} + E_{\text{int}} + E_{\text{pair}}$ の関係があることに注意が必要である。これら全てのエネルギーは KS-DFT と OF-EDF の等価性 (1.13) 式から、全て核子数密度 ρ のみに陽に依存した EDF として表現可能であると期待される。

ここで念の為、用語に関する注意事項を述べる。例えばここでは E_{kin} を kinetic energy と呼称するが、これはハミルトニアン の運動項の期待値という意味での kinetic energy とは別物であり、その物理的な意味は不明である。定性的には運動項に近いと考えられるが、あくまで定性的であって、Skyrme EDF の範囲内においてこれが運動項と近いことを示す定量的な根拠はない。 $E_{\text{int}}, E_{\text{pair}}$ に関しても同様である。しかし定性的な議論として、kinetic energy, interaction energy, pairing energy が密度のみの汎函数として表現可能か調べることは可能である。一方、 E_{ex} に関しては外場に起因したエネルギーを厳密に指し示すが、原子核において外場そのものが数学的な道具の側面が強いため、物理的な量であるとは言い難い点に注意が必要である。

図 3.2 に SHO, RND それぞれに関して、各エネルギーの分布を示した。 deep learning の学習データとして用いる場合、例外値とみなせるデータは除外することが望ましい。そこで本研究では、表 3.1 に示した範囲外のデータを除去した。残ったデータに関して、ランダムに 200 000 データをサンプリングすることで、SHO, RND, 各種エネルギーそれぞれに関して 200 000 データからなるデータセットを作成した。200 000 という数は、先行研究 [39] と同じである。hold-out validation 法に基づいて学習結果を評価するため、これら 200 000 のデータのうち、ランダムに選ばれた 90 % のデータを教師データ、残りの 10 % をテストデータとした。つまり、10 % のテストデータは訓練には使用しないため、学習済みモデルの汎化性能評価に使うことができる。

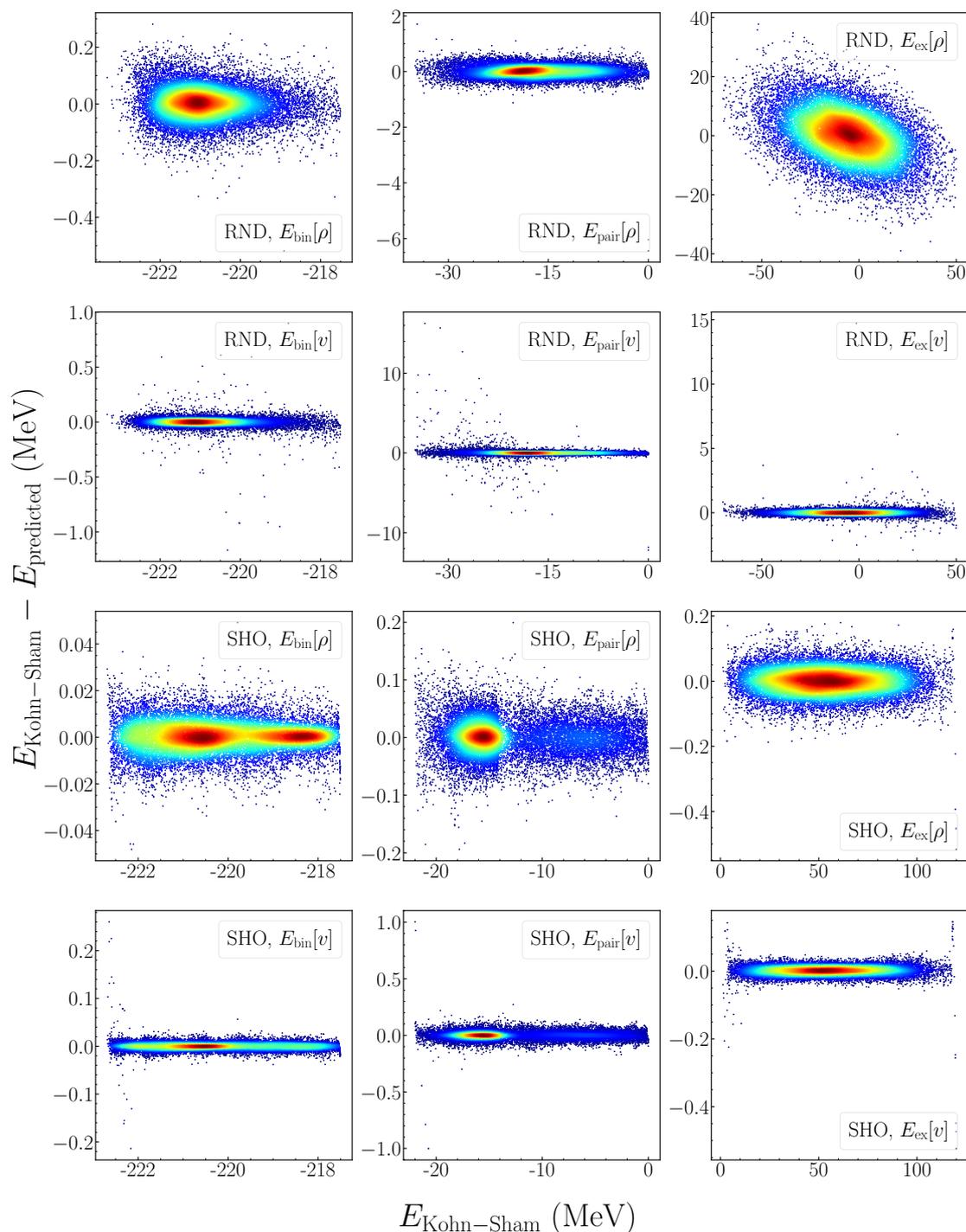


図 3.3 Kohn-Sham の結果と $E[\rho]$, $E[v]$ による推定値の差の散布図。横軸は Kohn-Sham のエネルギーの値であり、各パネルにテストデータ (20000 点) に対する結果を示している。RND の結果を第 1, 2 行, SHO の結果を 3, 4 行目に表示している。また、左から順に binding energy, pairing energy, external field energy を表している。各点の色はカーネル密度推定によるもので、赤い部分ほど点が密集、青いほど疎であることを表している。また、軸の単位は全ての MeV である。

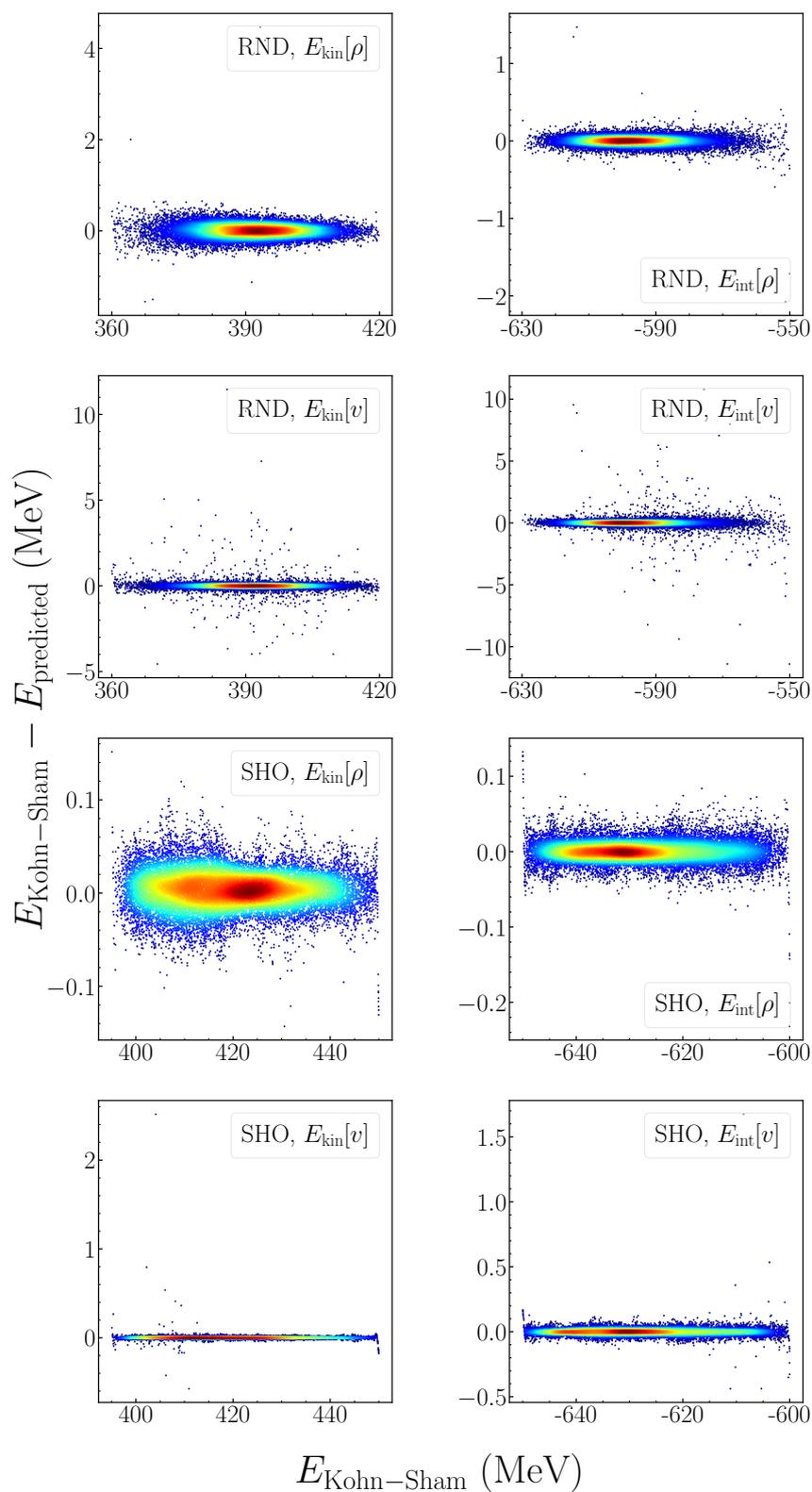


図 3.4 図 3.3 と同じ Kohn-Sham の結果と $E[\rho]$, $E[v]$ による推定値の差の散布図. ただし, kinetic energy と interaction energy の結果を示している.

表 3.1 SHO, RND, 各種エネルギーに対するカットオフの下限と上限. 全ての単位は MeV である. 括弧内の数字はオリジナルの 250 000 個のデータに対し, カットオフを適用して残ったデータの数を表している. カットオフの値は, 図 3.2 から, およそ全てのデータが含まれるように選んだ.

type	SHO			RND		
	lower	upper		lower	upper	
E_{bin}	$-\infty$	-217.5	(239 827)	$-\infty$	-217.5	(245 740)
E_{kin}	395.0	450.0	(249 491)	360.0	420.0	(249 310)
E_{int}	-650.0	-600.0	(248 782)	-630.0	-550.0	(249 398)
E_{pair}	-22.0	$+\infty$	(248 911)	-35.0	$+\infty$	(249 112)
E_{ex}	$-\infty$	120.0	(249 704)	-70.0	50.0	(249 468)

3.3.2 $\rho \rightarrow E[\rho]$

この節では核子数密度を入力変数, 各種エネルギーを出力変数とした教師あり学習の結果について議論する. 言い換えると, OF-EDF, つまり密度のみに陽に依存した EDF の再構築を行う. この時, 電子系の先行研究では考慮されていなかった, 超伝導性の効果を定性的に表す E_{pair} を再現できるかは興味深い問題である.

図 3.3 の最上段は, 横軸に Kohn-Sham スキームに基づく E_{bin} の結果を, 縦軸に neural network の推定値との差を取り, RND のテストデータ (10%) に対して結果を示している. 各点の色は, カーネル密度推定に基づいたデータ点の密度を表しており, 密集している領域ではより赤く, 疎な領域ではより青く表示されている. SHO の E_{bin} の結果に関しては, 上から 3 番目の段に描かれており, RND に比べより正確に推定できている様子が見て取れる. SHO と RND の乱数の数の差からくる, データセットとしての多様性を考慮すれば, この結果はもっともらしい. 図 3.3 には E_{pair} と E_{ex} の結果も示しており, E_{kin} と E_{int} の結果に関しては, 図 3.4 に図示されている. 横軸のスケールと縦軸のスケールの比較から明らかなように, RND の E_{ex} を除き, 本研究で用いた neural network は Skyrme DFT の結果の推定に成功していると言える.

興味深いことに, RND の E_{ex} の大きい誤差に関しては, 学習法やモデルの変更を行っても改善が見られなかった*⁶. これは, 原子核における引力と飽和性のため, 低エネルギーの原子核が形状の意味で同じような粒子数密度を好むという性質に起因していると考えられる. DFT の原理に基づけば, 適切に定義された密度と外場の定義域の間には, 全単写が存在すると期待される. しかし実際の計算では外場の詳細な構造は, 丸め誤差, 差分誤差, iteration の収束誤差といった様々な数値誤差により, 埋もれてしまう. 特に原子核密度は同じような形状を好むため, 密度間の微細な違いに含まれると期待される外場の詳細な構造の情報は, 数値誤差により保持されないと考えられる. 仮に外場の詳細な情報が含まれていたとしても, そのような微細な特徴量をニューラルネットワークに認

*⁶ CNN のモデルを用いた教師あり学習の結果に関しては後述する.

識させるには、学習テクニックが必要となる。したがって、密度から外場を復元するタスクが暗に含まれる E_{ex} の推定に関して、誤差が大きいのはもっともらしい。external field energy の予測誤差が大きいという問題は、電子系の先行研究 [39] でも報告されているが、引力系である原子核系ではより顕著になっている。次節でみるように、この問題は外場を入力変数とすることで解消される。

定量的な誤差評価のため、テストデータに関して以下で定義される平均絶対誤差 (Mean Absolute Error, MAE) の計算を行った:

$$\mathcal{M}_{\text{MAE}} = \frac{1}{|D|} \sum_{i=1}^{|D|} |E_{\text{ans}}^{(i)} - E_{\text{pred}}^{(i)}|. \quad (3.17)$$

結果は表 3.2 にまとめている。ここで、binding energy の MAE が SHO で 0.0051 MeV, RND で 0.0433 MeV と、理論研究において要求される精度に比べ、十分高いことは特筆すべきである。例えば、r-process の核分裂の研究では、およそ 100 keV 程度の精度が要求される [47]。また、pairing energy に関しても SHO で 0.0233 MeV, RND で 0.1567 MeV の誤差で予測に成功している。このことは、BCS に動機づけられた Kohn-Sham スキームにおいて、対相関の効果を粒子数密度のみから推定可能であるということを示唆している。

最後に、計算時間に関して言及する。本研究で用いた簡単化のもとで、Kohn-Sham スキームにより ^{24}Mg の学習データを生成すると、典型的にはシングルコアで数分要する。その一方で、学習済みのニューラルネットワークを用いて、与えられた密度や外場からエネルギーを推定する時間はシングルコアを用いて 0.1 ms 程度であり、Kohn-Sham スキームの 10^5 から 10^6 倍程度高速である。この圧倒的な計算速度の違いによる恩恵は、特に重い原子核において顕著になることが期待される。例えば、重い核の核分裂の研究において、多次元ポテンシャルエネルギーを高速で描くことができれば、様々な変形度に基づく核分裂の記述を効率よく検証することが可能になる。この意味で、neural network を用いた OF-EDF は、核分裂をはじめとした原子核理論研究を加速させるポテンシャルを秘めている。

表 3.2 SHO, RND, 各種エネルギーに対する平均絶対誤差 (Mean Absolute Error, MAE). $E[\rho]$ と $E[v]$ に関して全ての単位は MeV である一方, $\rho[v]$ に関しては無次元量である.

type	SHO		RND	
	$E[\rho]$	$E[v]$	$E[\rho]$	$E[v]$
E_{bin}	0.0051	0.0054	0.0433	0.0237
E_{kin}	0.0165	0.0071	0.1131	0.0900
E_{int}	0.0105	0.0182	0.0431	0.1499
E_{pair}	0.0233	0.0261	0.1567	0.1411
E_{ex}	0.0318	0.0105	6.6973	0.1338
	$\rho[v]$		$\rho[v]$	
	0.1107		0.4101	

3.3.3 $v \rightarrow E[v]$

KS-EDF から OF-EDF を再構築するという目的からは逸れるが、与えられた外場から基底状態のエネルギーを推定するような neural network を考えることも可能である。このようなモデルは、外場が物理的に意味のあることの多い電子系において、特に大きな需要が見込める。そこで電子系の先行研究 [39] に従い、本研究でも OF-EDF の再構築と全く同じ手順で、エネルギー汎関数 $E[v]$ の構築に臨む。ただし原子核 DFT における外場は、多くの場合数学的な道具に過ぎず、OF-EDF の再構築が可能になった現状を鑑みるに、外場を起点とするような $E[v]$ に基づく研究にどの程度の将来性があるかは非自明であることに言及しておく。しかし、先行研究 [39] との比較という観点から、本研究において $E[v]$ の学習を行うことには大いに意味がある。 $E[v]$ の学習には、前節で用いたものと全く同じデータセットと neural network (図 3.1) を用いる。ただし入力変数が密度から外場に変わる点にだけ、注意が必要である。

$E[v]$ の MAE に関しては表 3.2、テストデータ各々に関する Kohn-Sham の結果と予測値の差は、図 3.3 と図 3.4 に示されている。MAE を用いた定量的な評価において、密度を入力変数とした場合に比べ、外場を入力とした方が誤差が小さくなる傾向が見て取れる。RND の external field energy の推定において、 $E[\rho]$ に比べ 1/50 という著しい誤差の改善が得られたことは、特筆すべきである。以上のことから、外場による MAE の改善は前節で議論した通り、外場が密度に比べ詳細な情報を持っているためであると考えられる。

他方で SHO の binding energy の推定において、外場を入力とした場合の方が MAE が僅かに増大した。この原因を調べるため、図 3.3 の 4 段目、一番左のパネルを見ると、基底状態近傍に大きなエラーを持つデータ点がいくつか存在する様子が見て取れる。そこでこのデータ点のポテンシャルを調べたところ、ほとんどフラットな調和振動子ポテンシャルであることがわかった。したがって、neural network が曲率の比較的大きな外場のデータを基準に学習を行ったため、比較的小さな曲率のデータへの最適化が疎かになったためであると考えられる。一方密度に関して、フェルミオン系ではパウリ原理による飽和性があるため、入力変数のスケールによる問題は比較的穏やかになると考えられる。

3.3.4 $v \rightarrow \rho[v]$

DFT におけるダイナミクスは粒子数密度のみで決まるため、全ての可観測量もまた密度の汎関数として与えられる。特にモーメントのような external field energy と同じ形式で与えられる可観測量は、汎関数を厳密に与えることができる。ただし 2 体以上の演算子や、座標表示した one-body density matrix の非対角成分に依存するような可観測量に関しては、汎関数の形状が非自明であるため、別途導出しなければならないことに注意が必要である。いずれにせよ、密度の情報は DFT において不可欠な要素であるため、その推定は重要な問題である。この節では先行研究 [39] と同様に、与えられた外場から原子核密度を推定するモデルの学習に取り組む。

ここでは汎関数 $\rho[v]$ を再現するような、外場を入力、核子密度を出力とする教師あり学習を行う

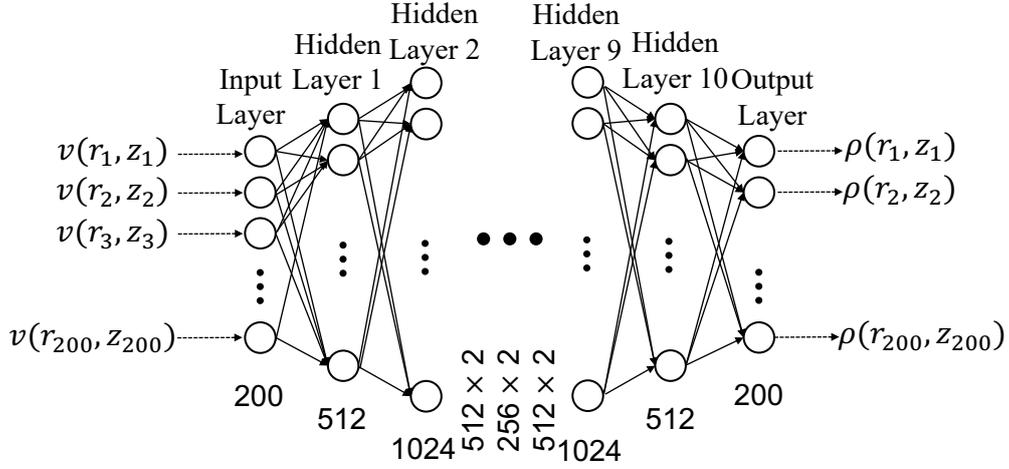


図 3.5 v と ρ の間のマップを得るために使用した encoder-decoder 構造を持つ neural network. 10 層の隠れ層を持つ MLP で, 活性化関数には ReLU を用いている. ただし, 出力層には密度の規格化のため, softmax 関数を用いた.

ため, OF-EDFE[ρ] や外場の汎関数 $E[v]$ の教師あり学習からモデルアーキテクチャーを変更する必要がある. まず, 出力変数となる原子核密度には,

$$\int d^3r \rho(\mathbf{r}) = A = 24, \quad \rho(\mathbf{r}) \geq 0, \quad (3.18)$$

という条件が要請される. 本研究で用いられている離散化された軸対称系では, 粒子数に関する条件は次のような和で表現される:

$$\sum_{i,j} 2\pi r_i \rho(r_i, z_j) dr dz = A. \quad (3.19)$$

ただし $dr = dz = 0.8 \text{ fm}$ であり, 有限差分計算では (3.19) 式を満たすように規格化が行われるため, データセットの点は全てこの規格化条件を倍精度実数の範囲内で厳密に満たす. この特殊な性質を自動的に満たす出力を得るために, 多クラス分類問題や Attention 機構 [16] で用いられる softmax 関数を用いることができる:

$$\text{softmax}(\mathbf{x})_i = \frac{e^{x_i}}{\sum_i e^{x_i}}. \quad (3.20)$$

定義から明らかなように, $\sum_i \text{softmax}(\mathbf{x})_i = 1$, $\text{softmax}(\mathbf{x})_i \geq 0$ である. つまり, 出力変数を密度から,

$$\rho^{\text{out}}(r_i, z_j) := \frac{2\pi r_i dr dz}{A} \rho(r_i, z_j), \quad (3.21)$$

に変更することにより, 出力が自動的に規格化条件と $\rho \geq 0$ を満たすような neural network を構成することができる*7.

*7 出力層として,

$$\rho(r_i, z_j) \rightarrow A \frac{\rho(r_i, z_j)}{\sum_{i,j} 2\pi r_i \rho(r_i, z_j) dr dz} \quad (3.22)$$

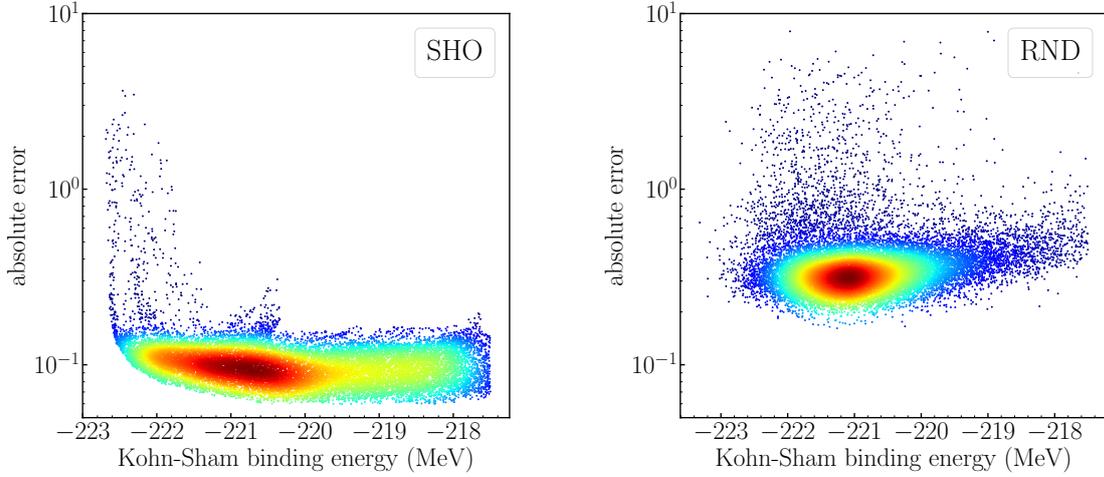


図 3.6 外場から推定された密度と Kohn-Sham の密度の間の絶対誤差。横軸には各点に対応する Kohn-Sham の binding energy を用いている。左右のパネルはそれぞれ SHO と RND の結果を表している。各点の色はカーネル密度推定によるもので、赤いほど点が密集、青いほど粗であることを意味している。

図 3.5 に、外場から密度を推定するために使用する neural network を図示した。基本的な構造として、encoder-decoder の形式を持った MLP を用いている。また、出力層として softmax 関数を使用した規格化層を適用している。また、最適化のための損失関数として、 ρ^{out} に関する以下で定義される MSE を用いた：

$$\mathcal{L}_{\text{MSE}}^{\rho}(D) := \frac{1}{|D|} \sum_{i,j} \sum_{k=1}^{|D|} \left| \rho_{\text{ans}}^{\text{out}(k)}(r_i, z_j) - \rho_{\text{pred}}^{\text{out}(k)}(r_i, z_j) \right|^2. \quad (3.23)$$

ただし下付き添え字の ans, pred はそれぞれ、教師データの正解ラベル、モデルの予測値を意味している。また、評価関数として以下で定義される ρ に関する MAE [48] を用いた。

$$\mathcal{M}_{\text{MAE}}^{\rho}(D) := \frac{1}{|D|} \sum_{k=1}^{|D|} 2\pi \iint r dr dz \left| \rho_{\text{pred}}^{(i)}(r, z) - \rho_{\text{ans}}^{(i)}(r, z) \right|. \quad (3.24)$$

この積分に対して三角不等式を用いることにより、

$$0 \leq \mathcal{M}_{\text{MAE}}^{\rho}(D) \leq 2A, \quad (3.25)$$

を得る。したがって、 $2A = 48$ がエラーの大きさに関する指標となる。また教師データとテストデータには、binding energy E_{bin} の時と全く同じカットオフを適用した (表 3.1 を参照)。

図 3.6 はテストデータの各点に対する密度の絶対誤差と、対応する Kohn-Sham の binding energy の関係を示している。各点の密度の絶対誤差の平均が MAE $\mathcal{M}_{\text{MAE}}^{\rho}$ に一致し、SHO では 0.1107, RND では 0.4101 である。したがって基準値である 48 に比べ、SHO, RND 共に MAE は

という層を導入することで、直接規格化条件を要請することも可能である。例えば PyTorch [22] という deep learning のフレームワークでは、このような層は容易に導入できる。

十分小さい. また図 3.7 に RND のいくつかのサンプルに関して, 上から順に外場, Kohn-Sham で計算された密度, 推定された密度, 2つの密度の差を示した. また差の図には密度の絶対誤差を表示しており, $M_{\text{MAE}}^{\rho} = 0.4101$ に比べ特別誤差が小さいサンプルを恣意的に選んでいないことがわかる. これらの例から, 定性的にも neural network が Kohn-Sham の密度の再現に成功していると言える.

以上の学習により得られた $\rho[v]$ と, 2.3.1 節で得られた OF-EDF $E[\rho]$ を組み合わせることで, $E[\rho[v]]$ という汎関数を得られる. したがって, 密度に関するエラーの指標の一つとして, $E[\rho[v]]$ の MAE を用いることも可能である. SHO, RND のテストデータに関して計算したところ, それぞれ 0.0847 MeV, 0.3059 MeV であり, 表 3.2 の値に比べおよそ一桁ほど大きい. この大きいエラーは density-drive error (DDE) [49, 39] と呼ばれ, $\rho[v]$ の学習の損失関数に, エラーの OF-EDF への影響が考慮されていないことが原因であると考えられる. 従って損失関数に関して, 学習済みの $E[\rho]$ を用いた以下のような工夫が求められる:

$$\tilde{\mathcal{L}}_{\text{MSE}}^{\rho}(D) := \mathcal{L}_{\text{MSE}}^{\rho}(D) + \frac{1}{|D|} \sum_{i=1}^{|D|} |E_{\text{ans}}^{(i)} - E[\rho_{\text{pred}}^{(i)}]|. \quad (3.26)$$

このような損失関数を用いる場合には, 信頼できる学習のため, $E[\rho]$ として高い汎化性能を持つモデルが必要となる. そのような信頼度の高いモデルを作るには, 多量の学習データや高度な学習テクニックが必要になり, 現段階では難しいためこの論文では計算は行わない. したがって現在のところ, 外場からエネルギーを推定する場合は $E[v]$ を用いた方が適切である.

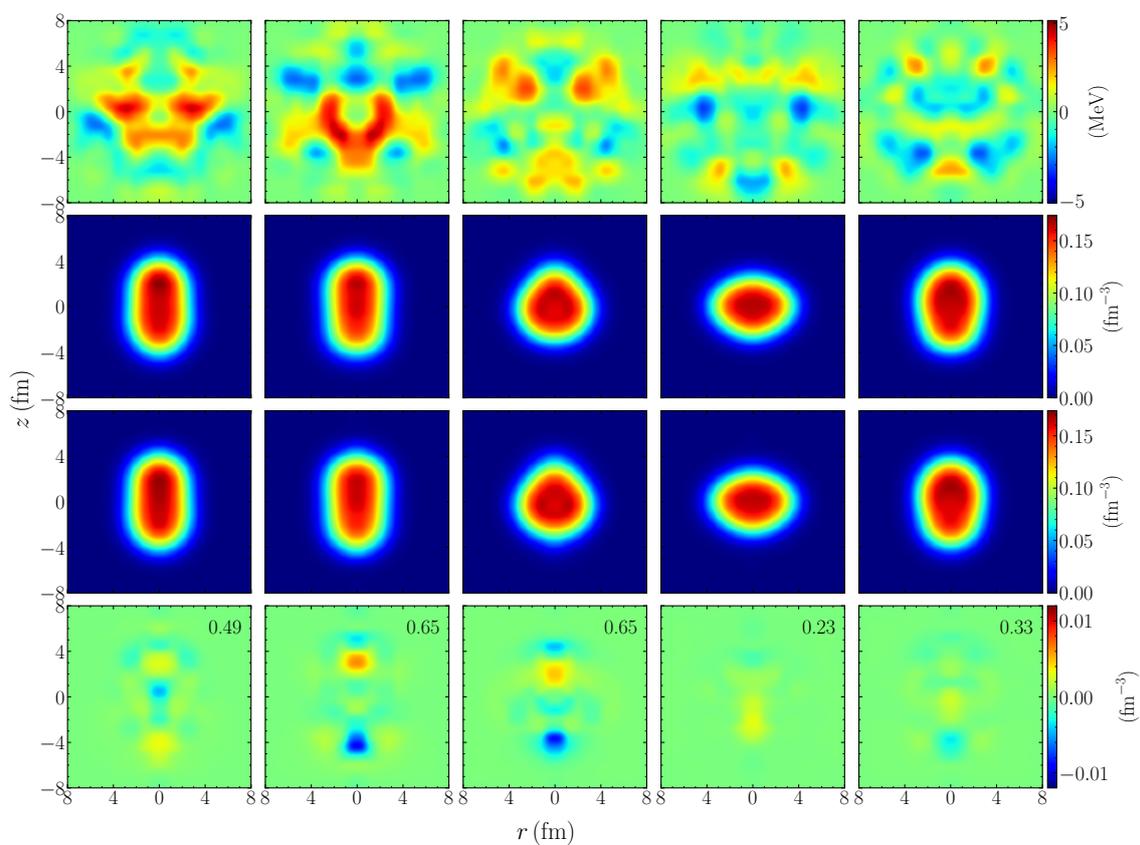


図 3.7 RND の外場 (第 1 列) から推定された密度のサンプル (第 3 行目). 縦方向に同一データ点に対するものを示しており, 比較として Kohn-Sham による密度の結果を第 2 列目に示している. 第 4 列目に密度の推定値と Kohn-Sham の密度の差を示しており, パネルの数字は各データの絶対誤差を意味している. 色の軸の単位は外場に対しては MeV であり, それ以外では fm^{-3} である.

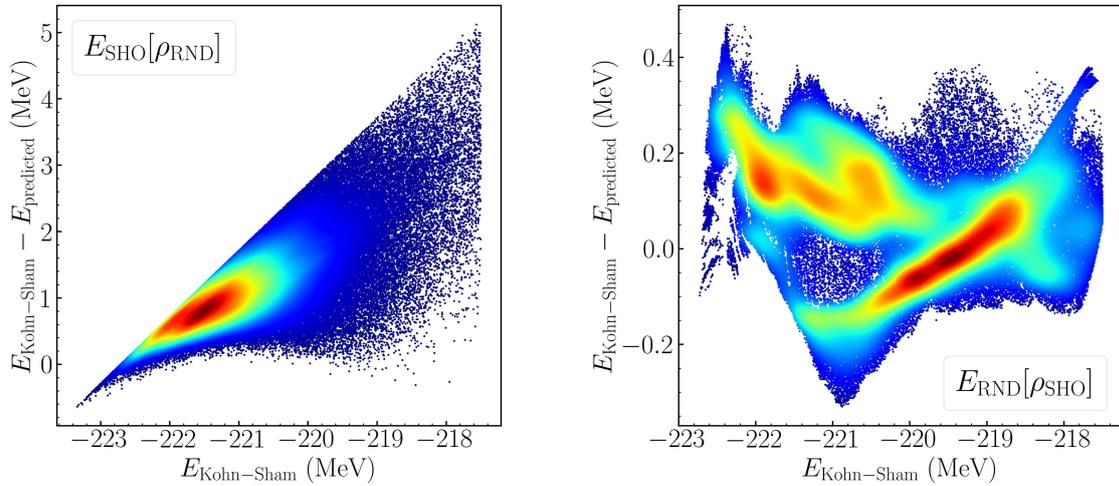


図 3.8 学習済みモデルの異なるドメインに対する汎化性能の評価. 左右のパネルはそれぞれ RND, SHO のデータに対する推定誤差を示している. ただし推定に使う学習済みモデルは, 左図が SHO で学習された OF-EDF $E_{\text{SHO}}[\rho_{\text{RND}}]$, 右図は RND で学習された OF-EDF $E_{\text{SHO}}[\rho_{\text{RND}}]$ である. 横軸には Kohn-Sham の binding energy を載せており, 全ての軸の単位は MeV である.

3.3.5 汎化性能

本研究では, SHO と RND という, 2 種類の異なる性質をもつデータセットを導入し, 教師あり学習を行った. その結果として, 各々のデータセットにおいて, 訓練データで学習したモデルはテストデータに対して汎化性能を獲得することに成功した. しかしこの事実は, 未知のデータセットに対する汎化性能を保証しない. 例えば, RND で訓練されたモデルは, SHO に対する予測性能を持つかは非自明である. このような異なるドメインに対する汎化性能は, 機械学習の実社会へのアプリケーションにおいて極めて重要な問題である.

Skyrme EDF の再構築の文脈で, この汎化性能を調べるため, $E_{\text{SHO}}[\rho_{\text{RND}}]$ と $E_{\text{RND}}[\rho_{\text{SHO}}]$ を考える. ただし, ρ_{SHO} と ρ_{RND} はそれぞれ, SHO と RND の核子数密度であり, E_{SHO} と E_{RND} は, SHO と RND で学習した binding energy の EDF である. 2.3.1 節で調べたのは $E_{\text{SHO}}[\rho_{\text{SHO}}]$ と $E_{\text{RND}}[\rho_{\text{RND}}]$ の性能であり, この節で調べるのは異なるドメインに対する性能であることに注意が必要である. 図 3.8 の左は, Kohn-Sham で得られた binding energy と $E_{\text{SHO}}[\rho_{\text{RND}}]$ による推定値の差を示しており, 右図はその逆 $E_{\text{RND}}[\rho_{\text{SHO}}]$ の結果を図示している. 図から明らかのように, SHO で学習した OF-EDF は RND のデータに対して予言能力をほとんど持たない. 一方で E_{RND} は, 図 3.3 に示した $E_{\text{RND}}[\rho_{\text{RND}}]$ のエラーに比べ大きいものの, SHO に対して一定の汎化性能を持つ. Kohn-Sham の結果と推定値の MAE は, $E_{\text{SHO}}[\rho_{\text{RND}}]$ に対して 1.1523 MeV, $E_{\text{RND}}[\rho_{\text{SHO}}]$ に対して 0.122 MeV である. 同様の振る舞いは定量的にはないものの, 先行研究 [39] においても言及されている. したがって, RND は deep learning において異なるドメインに対して一定の汎化性能を獲得可能なほどランダムであると結論づけられる.

表 3.3 CNN を用いた SHO と RND に対する学習結果の平均絶対誤差 (Mean Absolute Error, MAE). $E[\rho]$ と $E[v]$ の単位は MeV である一方, 密度の MAE は無次元量であることに注意が必要である.

type	SHO		RND	
	$E[\rho]$	$E[v]$	$E[\rho]$	$E[v]$
E_{bin}	0.0049	0.0055	0.0336	0.0245
E_{kin}	0.0151	0.0079	0.1010	0.1134
E_{int}	0.0119	0.0191	0.0484	0.1619
E_{pair}	0.0179	0.0250	0.1505	0.1852
E_{ex}	0.0220	0.0108	4.7059	0.0889
	$\rho[v]$		$\rho[v]$	
	0.1092		0.3484	

3.3.6 Convolutional Neural Network の結果

先行研究 [39] では, OF-EDF の計算に CNN が用いられていた. MLP と比較した場合, CNN の画像を認識するために効率的なモデルパラメータの制限を行うことで, 大きなサイズの画像に対しても少ない計算コストで学習が可能になる.

本研究の学習データは非常に小さい画像であり, fully-connected layer のみで十分よく学習ができる. しかし, 比較のため CNN でも同様の計算を行ったところ, 結果に有意な改善は見られなかった (表 3.3 を参照).

各タスクで学習に用いたニューラルネットワークは表 3.4, 3.5 に述べられている. 学習方法は MLP の場合と全く同じである. ここで, 同程度のモデルパラメータ数に対しては CNN の方が計算が重いことを考慮すれば, 少なくとも本研究と同じようなセットアップにおいては, CNN を使う優位性はない. ただし, deep learning にはハイパーパラメータが多くあり, それら最適化の度合いによって性能は変化する. また, データの事前処理を工夫することで性能を向上させることも可能である. しかし, これらには技術や経験が必要である. 幸いなことに, Kaggle [50] といった機械学習のコンペティションを行うサイトが存在し学習データを投稿することで, 費用はかかるものの, 多くの専門家に様々なアイデアを出してもらうことが可能である. 例えば, IceCube プロジェクトは実際に Kaggle にてコンペティションを行なっている [51].

表 3.4 $E[\rho]$ と $E[v]$ の学習で用いた CNN モデル. 各層の表現は, Keras API [42] の言葉で書かれているため, 再現が容易に可能である. 各層で言及されていないパラメータに関しては, 全て Keras API のデフォルト値を用いている.

layer	type
input	Input(shape=(10, 20, 1))
1	Conv2D(filters=32, kernel_size=3, activation='relu')
2	Conv2D(filters=64, kernel_size=3, activation='relu')
3-8	Conv2D(filters=64, kernel_size=4, padding='same', activation='relu')
9	Conv2D(filters=128, kernel_size=3, activation='relu')
10-13	Conv2D(filters=128, kernel_size=3, padding='same', activation='relu')
14	Flatten()
15	Dense(units=128, activation='relu')
output	Dense(units=1, activation='sigmoid')

表 3.5 $\rho[v]$ の学習に用いた CNN モデル. encoder-decoder 構造を持つ neural network を用いており, 1 から 4 層に画像サイズを下げる convolutional layer, 5 から 7 層に画像サイズを変えない convolutional layer を用いている. これらの操作により, 画像サイズは (2, 12) まで圧縮される. 一方で 8 から 11 層には deconvolution layer 層を用い, 画像サイズは (10, 20) に拡大される. 各層で言及されていないパラメータに関しては Keras API のデフォルト値を用いている.

layer	type
input	Input(shape=(10, 20, 1))
1	Conv2D(filters=32, kernel_size=3, activation='relu')
2	Conv2D(filters=64, kernel_size=3, activation='relu')
3	Conv2D(filters=128, kernel_size=3, activation='relu')
4	Conv2D(filters=256, kernel_size=3, activation='relu')
5-7	Conv2D(filters=256, kernel_size=4, padding='same', activation='relu')
8	Conv2DTranspose(filters=256, kernel_size=3, activation='relu')
9	Conv2DTranspose(filters=128, kernel_size=3, activation='relu')
10	Conv2DTranspose(filters=64, kernel_size=3, activation='relu')
11	Conv2DTranspose(filters=32, kernel_size=3, activation='relu')
12	Flatten()
13	Dense(units=200, activation='softmax')
output	Reshape(target_shape=(10, 20, 1))

3.4 まとめ

3.4.1 結論

本研究では、Kohn-Sham 型のエネルギー密度汎関数 (Kohn-Sham Energy Density Functional, KS-EDF) である Skyrme EDF から始め、密度のみに依存した EDF の再構築を行なった。得られた EDF は Kohn-Sham 軌道を必要としないため、Orbital-free EDF (OF-EDF) と呼ばれる。OF-EDF の再構築に向けて、2種類の異なるランダムな外場 (SHO, RND) に対し、Kohn-Sham スキームを用いて得られた核子数密度とエネルギーを計算し、学習データを生成した。これらの外場は、電子系の先行研究 [44, 39] をベースに、原子核系に合わせて改良されたものである。deep learning による教師あり学習の結果得られた様々な OF-EDF は、RND の external field energy を除いて、Kohn-Sham DFT の結果を再現することに成功した。この結果は、Skyrme EDF を再現するような OF-EDF の存在を示唆している。RND の external field energy は、密度を入力変数とした場合には再現することができないが、外場を入力変数とすることで十分良い精度での推定が可能である。この性質は電子系に比べ、引力系である原子核においてより顕著に見られる特徴である。また、SHO と RND の予測性能を比べると、ランダム性の小さい SHOの方がエラーが小さいという結果が得られた。一方で、与えられた密度、または外場からエネルギーを推定する時間は、従来の Kohn-Sham スキームに比べ 10^5 から 10^6 倍高速である。

3.4.2 展望

本研究では軸対称性を要請したが、原子核物理では非軸対称変形は、特に核分裂において、重要な役割を果たすと考えられている。非軸対称変形を考慮するには対称性を一切要請しない DFT 計算を行い、3次元の密度を取り扱う必要がある。時間反転対称性も課さず、陽子と中性子で異なる外場を扱う場合には、スピンの自由度とアイソスピンの自由度も考慮しなければならない。この粒子数密度データを空間方向に関して離散化すれば、“4色”の立体データと見なすことができる。このようなデータは1次元ベクトル化した場合、高次元になると予想されるため、fully-connected layer は適さない。一方、CNN には3次元データを扱うものも存在し、Keras API [42] や PyTorch では、2次元の CNN と同様簡単に実装することができる。また近年では、Vision Transformer (ViT) [41] と呼称される、CNN とは全く異なる画像認識モデルも登場し、CNN 以上の性能を見せている。オリジナルの ViT は2次元データ用の neural network であるが、そのアーキテクチャーを3次元データ用に拡張する方法は自明である。したがって、データを用意するコストは大きく増えるものの、本研究を対称性を課さない原子核系に拡張する方法は自明である。

また、deep learning を用いた手法の最大の利点は、一度大規模なデータを用意すれば、エネルギーを高速で推定可能な点である。このような低コストの計算手法は、数値実験を加速させる。近年では研究が進展し高度化するに伴い、一本の学術論文を執筆するのに必要な DFT の計算回数が徐々に増加している。典型的な例として、多次元ポテンシャルエネルギー曲面中の核分裂の計算が挙げられる。計算機の性能は向上し続けているものの、研究に必要な計算コストも同時に増大し、数

値実験でアイデアを検証するプロセスはより煩雑になっている。本研究で開発されたような高速な計算モデルは、特に Python ライブラリのような便利な形式で提供されることで、アイデアのテストや検証に要する時間を大幅に短縮することができる。計算精度が問題になる場合は、従来の Kohn-Sham スキームを用いて、deep learning でテストされたアイデアを再検証することができる。これは materials informatics (MI) [52] の考え方を、理論研究に応用したものと解釈できる。この意味で、deep learning により再構築された EDF は、原子核理論研究を加速させ得る。

data-driven アプローチにおける deep learning の潜在的な問題は、データを収集するためのコストである。本研究では比較的軽い原子核である ^{24}Mg を採用し、軸対称性を課したので、データの収集コストも比較的低かった。しかしウラン同位体のような重い原子核や超重元素は、核分裂経路の探索といった原子核物理における重大な理論的課題を内包し、非常に興味深い研究対象であるものの、本研究を適用するには多大な計算コストがかかる。ただし、データ収集は個人で行う必要はなく、多くの研究チームで協力して行うことも可能である。さらに、deep learning に利用できるような質の高いデータが、研究者たちのパソコンの中に、既に数多く存在している可能性もある。したがって、原子核理論コミュニティにおいて数値データを収集する枠組みを作ったり、格子間隔といったハイパーパラメータを統一して数値計算を行う枠組みを整備することが望ましい。このような協力的なアプローチは、研究を効率的かつ効果的に進めるのに役立ち、核物理コミュニティ全体に利益をもたらすと期待される。

第 4 章

Multi-task learning による ^{236}U の潜在変数の解析

4.1 外場の呪いと集団運動

4.1.1 集団座標

DFT において、系のダイナミクスは外場により制御される。電子系の場合、外場は原子核が作り出す Coulomb ポテンシャルやレーザー等が該当し、物理的に意味のある量である。一方で、原子核系では電子系ほど多様な外場を実現することは困難であるため、その点で原子核 DFT と電子系の DFT は明確に異なる。したがって、原子核 DFT では変形した原子核の性質を調べるという目的で、伝統的に多重極モーメント演算子

$$\hat{Q}_{\lambda k} = \int d^3r r^\lambda Y_{\lambda k}(\Omega) \hat{\rho}(\mathbf{r}), \quad (4.1)$$

などが用いられる^{*1}。ただし、 $Y_{\lambda k}(\Omega)$ は球面調和関数であり、 Ω は立体角である。仮定された外場の使用は平均場理論や、生成座標法 (Generator Coordinate Method, GCM) といった平均場を超えた量子相関を扱う理論でも同様に現れる [10]。

原子核系における興味深い物理現象として、原子核集団運動が存在する。原子核集団運動とは、量子多体系の複雑な量子相関により創発される複数の核子の協調的な振る舞いであり、少数の集団的な自由度 (集団座標) で記述可能であると信じられている。核分裂はその代表例であり、核子自由度に基づく記述を目指して、精力的に研究が行われている。しかしながら、核分裂の研究のほぼ全てにおいて、多重極モーメントなどの仮定された集団座標が用いられてきた。理想的には系のダイナミクスから得られた集団座標を用いるべきであり、多重極モーメントなどが核分裂を記述できる保証できるはない。実際、多重極モーメントを用いた GCM に基づく解析では、核分裂収率 (Fission Product Yield, FPY) の再現は比較的成功的なものであるものの、半減期の再現は未だ多くの課題があ

^{*1} 多重極モーメント演算子は形式的に DFT における外場 (1.1) と同じであり、Lagrange の未定乗数によりこの演算子の期待値に対して拘束条件をかけるという操作は、外場付き変分問題を解くことと、形式的に等価である。したがって、原子核を変形させるために多重極モーメントに対して拘束条件をかけるという操作は、原子核に外場をかける操作と同じとみなせる。

る。特に FPY は、定性的には分裂片のシェル効果に強く支配されるため、多重極モーメントでなくとも、シェル効果を取り込むことが可能であり、原子核を2つに分裂させるような外場を用いた計算であれば、どのようなものでもある程度 FPY の再現に成功する可能性を否定できない。

一方で系の厳密な時間発展ダイナミクスを記述する理論として、時間依存密度汎関数 (Time-Dependent Density Functional Theory, TDDFT) が存在する。これはいくつかの仮定のもとで、与えられた初期状態に対する系の時間発展が、時間依存した密度の汎関数のみで支配されることを利用した理論である。TDDFT は ^{240}Pu の induced fission に適用され、大きく変形した初期条件から時間発展により核分裂を引き起こす様子が再現された [53]。しかしこの研究は核分裂バリアの外からの時間発展であり、核分裂のトンネリングダイナミクスが TDDFT により記述可能であることを示唆する結果ではないことに注意が必要である。TDDFT を用いた核分裂の記述には、3つの大きな問題点がある。まず1つめは、核分裂の記述にはトンネリング現象を扱う必要があるため、長時間の時間発展が必要になり、膨大な計算コストを要するという問題である。次に、TDDFT で用いるべき汎関数が非自明であるという点である。TDDFT においても Kohn-Sham スキームにより計算が行われ、TDHF 理論に動機付けされた形式で Kohn-Sham 軌道が導入されるが、この時 static な場合の KS-EDF がしばしば流用される。しかし、核分裂に対してその ansatz が妥当である保証はない。最後の問題点は、TDDFT がそもそも核分裂を記述できる理論であるかわからないという点である。TDDFT の仮定には、外場の時間に対する周期性や Taylor 展開可能性が含まれるが、核分裂においてその仮定が成り立つかは非自明である。したがって、集団座標を用いたアプローチは今後とも必要とされる。

しかしながら、ダイナミクスから多体系の集団座標を抽出する操作は非常に困難である。Self-consistent Collective Coordinate Method (SCC) [54, 55] では、Time-dependent Hartree-Fock (-Bogoliubov) (TDHF(B)) 近似により得られる変分多様体上の軌跡に対し、それを含むような symplectic 多様体を考えることで、集団的な情報を引き出す。しかし、あくまで TDHF(B) という近似的な手法に基づく理論である上、計算コストの問題から近似的な手法である Adiabatic SCC (ASCC) [56] が実際の計算では用いられている。したがって、(A)SCC による核分裂の記述に関しては、今後の発展を期待するばかりである。

このような現状から、現象論的に仮定された変形モードに基づく集団運動の解析は、依然として盛んに行われている。変形モードの仮定は Kohn-Sham DFT や平均場理論の文脈においては外場の選択と等価であり、外場の選択を通じて理論に人間の強いバイアスが入り込む。したがってこの“外場の呪い”とも言うべきバイアスを回避し、ダイナミクスに根差したより適切な変形パラメータの新しい導出法が求められている。

4.1.2 多様体仮説

我々の身の周りはデータで溢れている。写真、動画、音声、文章。様々な形式で保存されるそれらは、すべて高次元 Euclid 空間上のベクトルとして表現することが可能である。例えばフルカラーの写真であれば、ピクセル数を3倍した次元数を持つ Euclid 空間上のベクトルとみなせ、Full HD の写真であればその次元はおおよそ 622 万である。しかしながら、この広大な空間からランダムにサン

プリンクされた点に対応する画像は、高確率でただのノイズであり、人間が意味を見出せるようなものではない。逆に、我々が意味のある画像として認識できるデータの分布は、超高次元 Euclid 空間上の非常に限られた領域に密集していることが期待される。このようなアイデアは多様体仮説 [57] と呼ばれ、deep learning のモデル設計といった根底の部分に自然に現れる [35, 58].

物理学と多様体仮説は親和性が高い。物理学では、エネルギースケールに応じて様々な自由度や理論が存在する。一般に高エネルギーの膨大な自由度の大部分は、低エネルギーといった注目するスケールでは抑制され、新奇な自由度が創発される。有効場理論 (Effective Field Theory, EFT) [59] はまさに上記の哲学に基づいた理論であり、集団運動が少数のパラメータで記述できるという仮定も同様の哲学を共有する。この哲学は、超高次元空間上の理論は、興味のあるスケールにのみ着目すればより低次元の理論として記述できる、と言い換えることができる。これは多様体仮説の主張と非常に近い。したがって、deep learning の手法により集団運動の記述可能であるという期待が生まれる。

本研究の目的は、多様体仮説に動機付けられた deep learning の手法により、 ^{236}U の低エネルギー変形ダイナミクスに関する集団的な情報を抽出することである。ここで低エネルギー原子核の自由度は多体の波動関数であるが、これは数値的に取り扱うことさえ難しい。そこで本研究では、DFT に基づき基底状態の adiabatic なダイナミクスから、系の潜在的な変数の抽出を目指す。

類似の研究は文献 [60] にて近年行われ、四重極、八重極変形した原子核の HFB 波動関数の情報が、autoencoder (AE) [36] と呼ばれる教師なし学習の一種を用いることで、10 次元または 20 次元のベクトルに圧縮され、そして復元された。彼らの動機は GCM 等への応用に向けた波動関数の低次元表現の獲得にあり、そのアイデアは非常に興味深い。特に、彼らの論文では言及はされていないものの、AE という手法の背景には多様体仮説があるため、その点で本研究と類似している。ただし、彼らの学習データは四重極、八重極モーメントという 2 つのパラメータで指定できることが事前にわかっているため、多様体仮説は自明であることに注意が必要である。本研究では、前章でも用いたランダムポテンシャルアプローチ [44, 39] を採用することで、非自明な多様性を持ったデータセットを生成する。このデータを、OF-EDF の教師あり学習と autoencoder を組み合わせた multi-task learning (MTL) [61, 62, 63] により解析し、潜在変数の抽出に取り組む。

4.2 データセット

4.2.1 Skyrme Kohn-Sham DFT

我々の目的は DFT と MTL により、 ^{236}U の低エネルギー変形状態の潜在的な情報を解析することである。DFT では与えられた外場のもとでの基底状態のダイナミクス、つまり adiabatic なダイナミクスは OF-EDF のみで与えられる。しかしながら、原子核物理で用いられているのは KS-EDF であり、与えられた外場からエネルギー E と密度 ρ のペアを得るためには、Kohn-Sham 方程式を解く必要があり、 E と ρ の対応関係は極めてわかりにくい。したがって本研究では、前節と同様にランダムポテンシャルアプローチ [44, 39] を用いることで、一度データセットの表現に還元した DFT の情報から、潜在変数を抽出する。ただし、前章の研究からデータセット表現の DFT は、OF-EDF

を復元可能であることが期待される。この意味で、データセット表現に還元された DFT の解析は、OF-EDF の解析とエラーの範囲内で等価である。

本研究では引き続き、Skyrme EDF [11, 64, 10] を用いる。ただし、計算コストの観点から軸対称性と時間反転対称性も同様に要請する。また、本研究では対象の原子核が質量数 236 のウランという重い原子核であるため、計算精度や速度の観点から有限差分法ではなく、3次元軸対称調和振動子基底による基底展開アルゴリズムによって Kohn-Sham 方程式を解く。実際の計算では HFBTHO (v3.00) [65] と呼称される、DFT のオープンソースソルバーを用いた*2。HFBTHO では、以下の調和振動子基底により Kohn-Sham 軌道は展開される:

$$\varphi_{n_z, n_r, \Lambda}(\mathbf{r}; b_z, b_\perp) = \varphi_{n_z}(\xi; b_z) \varphi_{n_r; b_\perp}^{|\Lambda|}(\eta) \frac{e^{i\Lambda\phi}}{\sqrt{2\pi}}, \quad (4.2)$$

$$\varphi_{n_z}(\xi; b_z) = \frac{1}{\sqrt{b_z 2^{n_z} n_z!} \sqrt{\pi}} H_{n_z}(\xi) e^{-\xi^2/2}, \quad (4.3)$$

$$\varphi_{n_r}^{|\Lambda|}(\eta; b_\perp) = \sqrt{\frac{2n_r!}{b_\perp^2 (n_r + |\Lambda|)!}} \eta^{|\Lambda|/2} L_{n_r}^{|\Lambda|}(\eta) e^{-\eta/2}, \quad (4.4)$$

$$\xi = \frac{z}{b_z}, \quad \eta = \left(\frac{r_\rho}{b_\perp} \right)^2. \quad (4.5)$$

ただし、 $\mathbf{r} = (r, z, \phi)$ は3次元円筒座標であり、 b_\perp, b_z はそれぞれ r 方向、 z 方向の調和振動子の長さパラメータである。また、 H_{n_z} と $L_{n_r}^{|\Lambda|}$ はそれぞれ Hermite 多項式と、Laguerre 陪多項式である。ここで、前章の研究に引き続き、本研究でも z 方向の parity 対称性は要請していないため、八重極変形のような変形も可能であることを強調しておく。つまり、本研究は非対称核分裂を扱える枠組みになっている。

前章の研究と同様に Skyrme EDF のパラメータとして SLy4 [66] を使い、surface type の対相関相互作用にモチベートされた KS-EDF [11] を用いた。さらに、本研究ではより現実的なデータセットを作るため、Coulomb 相互作用を取り入れた。ただし、Coulomb direct term と、Slater 近似をした Coulomb exchange term の和の形のものを用いている。また、HFB 近似 [10] から類推される Kohn-Sham スキームに基づき、DFT 計算を行う。pairing に関する KS-EDF の pairing strength parameter V_q に関しては、陽子中性子それぞれに関して average pairing gap が現象論的な値 $12/\sqrt{A}$ MeV を再現するように決めた。質量数 $A = 236$ に対する最適化の結果、陽子に対して $V_p = -650.98$ MeV/fm³、中性子に対して $V_n = -526.53$ MeV/fm³ という結果が得られた。また、調和振動子基底の長さパラメータ b_\perp, b_z に関しては、外場のない場合の基底状態のエネルギーが最小になるように最適化を行ない、 $b_z = 2.08$ fm, $b_\perp = 1.92$ fm を得た。ここで、 b_\perp, b_z, V_n, V_p という4つのモデルパラメータの最適化は同時に行われていることに注意が必要である。以上のパラメータのもとで、外場のない ^{236}U の基底状態のエネルギーは -1780.87 MeV である。最後に、後述するようにランダムポテンシャルのもとで最適化を行うため、重心位置の拘束が必要になる。そこで本研究では2次の拘束条件 (3.6) を採用し、一貫して $C = 1.0$ MeV fm⁻² を用いている。

*2 本研究で行うようなランダムポテンシャルに基づく計算は、HFBTHO 自体に実装されているわけではない。したがって、ソースコードを追加する必要がある。

4.2.2 ランダムポテンシャル

前章の ^{24}Mg に対する OF-EDF の再構築の研究で示されたように、ランダム性の低いデータセットで学習したモデルは、より多様なデータに対して全く汎化しない。したがって、系の非自明なダイナミクスを含むようなデータセットは、非常に多様な粒子数密度を包含しなければならない。以上の観点から、本研究では3次元軸対称調和振動子基底 (4.2) に適した形状の、以下のようなランダムポテンシャルを導入する:

$$v^{(i)}(r, z) = \sum_{n_z/q+2n_r+|\Lambda| \leq N_{\text{shell}}} v_{n_z, n_r, |\Lambda|}^{(i)} \varphi_{n_z}(\xi; b_z^{(i)}) \varphi_{n_r}^{|\Lambda|}(\eta; b_{\perp}^{(i)}). \quad (4.6)$$

ただし、 $q = (b_z/b_{\perp})^2 = 1.1736$ は調和振動子の非等方性を表すパラメータである。また、 i はデータのインデックスを表し、 $v_{n_z, n_r, |\Lambda|}^{(i)}$, $b_z^{(i)}$, $b_{\perp}^{(i)}$ のような括弧付き上付き添え字の付いた変数は、一貫して各データ毎にランダムに生成されるパラメータを意味する。ここで、基底展開の重み $v_{n_z, n_r, |\Lambda|}$ だけでなく、長さパラメータ b_z , b_{\perp} もランダムに生成していることに注意が必要である。調和振動子基底の完全性から、固定した b_z , b_{\perp} のみであっても任意関数は実現可能である。しかし、本研究では有限の $N_{\text{shell}} = 26$ を用いているため、データセットが特定のスケールを持つ可能性がある。このようなスケールをモデルが学習してしまった場合、異なるドメインのデータに対して汎化性能が著しく低下する恐れがある。したがって、本研究ではランダムな b_z , b_{\perp} を採用する。

^{236}U の低エネルギー変形ダイナミクスを調べるという目的のため、外場のない状態の binding energy を基準とし、そこからの励起エネルギーが 15 MeV 以下の状態が多く含まれるという戦略のもと、ランダムポテンシャルのハイパーパラメータを決定した。15 MeV は ^{236}U の核分裂バリアの高さよりも大きい値であることに注意が必要である*³。まず、最適化された調和振動子の長さパラメータを用いることで、 $b_{z, \perp}^{(i)} = (1 + t_{z, \perp}^{(i)})b_{z, \perp}$ によりランダムな長さパラメータを生成した。ただし、 $t_{z, \perp}^{(i)}$ は $[0, 0.5]$ の範囲から一様ランダムにサンプリングされた数である。他方 $v_{n_z, n_r, |\Lambda|}^{(i)}$ は各量子数 $(n_r, n_z, |\Lambda|)$ に関して、それぞれ $[-v_0, v_0]$ の範囲からサンプリングされた一様乱数である。ここで、単一のカットオフパラメータ v_0 を用いると binding energy の分布に大きなバイアスが生じることに注意が必要である。この問題を回避するため、本研究では4種類の v_0 を用いた。その具体的な値は $0.6 \text{ MeV fm}^{3/2}$, $1.2 \text{ MeV fm}^{3/2}$, $1.8 \text{ MeV fm}^{3/2}$, $2.4 \text{ MeV fm}^{3/2}$ である。複数の v_0 を用いたとしてもなお、binding energy の分布にはバイアスが生じるため、量の多いデータに関して除去する必要がある。そこで本研究では、binding energy に関して 100 本のビンで構成されるヒストグラムを考え、各ビンの中で 2250 個を超えた部分に関しては、ランダムにデータを除去した。さらに、binding energy が $-1780.87 \text{ MeV} \leq E \leq -1765.00 \text{ MeV}$ の範囲外のデータをすべて取り除いた。最後に大きく変形した状態を取り除くため、四重極モーメントにカットオフを課し、 $[3.0 \text{ b}, 16.0 \text{ b}]$ の範囲外のデータを除去した。ただし、本研究では多重極モーメントは一貫して以下の演算子の期待値で定義される:

$$\hat{Q}_{\lambda k} = \hat{R}^{\lambda} Y_{\lambda k}(\hat{\theta}, \hat{\phi}). \quad (4.7)$$

*³ 例えば文献 [67] によれば、HFB 計算に基づく Q_{20} 方向のバリアの高さは約 9MeV である。

ここで $Y_{\lambda k}$ は球面調和関数であり, $\mathbf{r} = (R, \theta, \phi)$ は 3 次元極座標である. 四重極モーメントは $\lambda = 2, k = 0$ の場合に相当する.

以上の戦略に基づき, 298 982 個のランダムポテンシャルによるデータを生成し, 除去の結果 181 134 個のデータが残った. 得られたデータセットに関して, 20 % が $v_0 = 0.6 \text{ MeV fm}^{3/2}$, 23 % が $v_0 = 1.2 \text{ MeV fm}^{3/2}$, 49 % が $v_0 = 1.8 \text{ MeV fm}^{3/2}$, 8 % が $v_0 = 2.4 \text{ MeV fm}^{3/2}$ 由来である. 図 4.1, 図 4.2 にはそれぞれ, 181 134 のデータに対する binding energy のヒストグラム, 四重極モーメントと八重極モーメントの 2 次元散布図が図示されている. 後者に関して, 色軸は各データの binding energy を表している. 基底展開法では差分法と異なり, 粒子数密度は空間全ての点で定義されている. そこで deep learning に用いるため, r, z 方向共に $dr = dz = 0.4 \text{ fm}$ の幅で離散化した. この格子幅は, Skyrme DFT の 3 次元格子計算で用いられる幅のおよそ半分程度である. また点の数は, r 方向に 48 点, z 方向に 128 点取り, 格子は前章の ^{24}Mg に対する研究同様, 副格子上に配置した. したがって核子数密度は, サイズが $1 \times 48 \times 128$ のグレースケール画像と等価である. ここで画像のサイズに関して, deep learning におけるテクニカルな問題に関して言及する. deep learning において画像を取り扱うようなモデルは, 画像サイズを縦横それぞれに関して半分に縮小したり, 反対に 2 倍に増やすようなモジュールにより構成されることが多い. したがって縦横サイズの因数に, 2 ができる限り多く含まれているような画像データの方が既存のモデルを流用しやすいため便利である. 本研究の画像サイズである $48 = 2^4 \times 3$ と $128 = 2^7$ はこの条件を満たしており, deep learning における取り扱いが容易である.

4.3 Multi-task learning

多様体仮説と集団座標の仮定によれば, 前節のようにして生成されたデータセットは少数のパラメータでよく記述できると期待される. このような潜在変数を抽出するための基本となるテクニックとして, autoencoder [36] が古くから知られている. この手法を ^{236}U のデータセットに適用することで, 潜在変数の解析を行うことが可能である. つまり, 入力された核子数密度 ρ は潜在変数にエンコードされ, デコーダーにより潜在変数からオリジナルの密度が復元される. このスキームは以下のような式で表現される:

$$\rho \rightarrow \mathbf{z} = \mathcal{E}[\rho] \rightarrow \rho = \mathcal{D}(\mathbf{z}). \quad (4.8)$$

ただし, $\mathbf{z} = \{z_1, z_2, \dots, z_{d_z}\}$ は抽出された d_z 次元潜在変数である. また, 関数 \mathcal{E}, \mathcal{D} はそれぞれエンコーダー, デコーダーを表し, neural network でモデリングされる. この時デコーダーに, 精度よく密度の復元をするよう要請すると, データセットに多様なデータが含まれているほど潜在次元 d_z は大きくなる傾向があり, 人間の理解が難しくなる. したがって集団的な情報の抽出という観点では, 潜在次元を下げることで, ある程度精度に関して妥協する必要がある. 加えて, そのような密度に大きな復元エラーが生じる場合, density driven error (DDE) [49] により, 学習済み OF-EDF に復元した密度を通して正しいエネルギーが得られる保証はない. つまり, 潜在変数がエネルギーの情報を保持しているかは非自明である.

他方, encoder-decoder 構造を OF-EDF の教師あり学習そのものに組み込むことで, 潜在変数の抽出をするという操作も考えられる. つまり, 入力された密度は潜在変数に圧縮され, decoder は潜

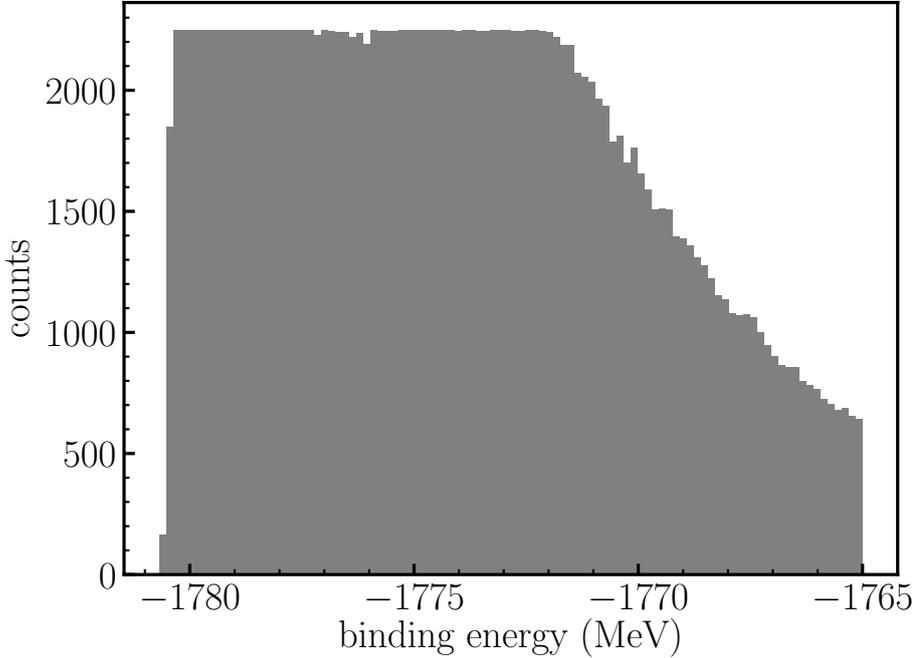


図 4.1 本研究で生成したデータセットの binding energy に対するヒストグラム. 過剰なデータはランダムに取り除いたため, データの分布に大きなバイアスは見られない. 基底状態のエネルギーは -1780.87 MeV であり, -1772 MeV, つまり励起エネルギーがおおよそ 8 MeV のところまでの間で, 分布は均一になっている.

在変数からエネルギー密度を再現するように学習される. 数式で表すと, 以下のようになる:

$$\rho \rightarrow \mathbf{z} = \mathcal{E}[\rho] \rightarrow E = \mathcal{D}(\mathbf{z}). \quad (4.9)$$

しかし, このような教師あり学習は意味のない自明な潜在変数を導く危険性がある. 例えば, 潜在変数の各成分すべてがエネルギーの変換で書けるような形式で学習が進む可能性があり, このような場合, 潜在変数のすべての成分は実質的に等価な情報しか見えていない.

以上の問題を回避するため, 本研究では 2 つのタスクを組み合わせる:

$$\rho \rightarrow \mathbf{z} = \mathcal{E}[\rho] \rightarrow E = \mathcal{D}_E(\mathbf{z}) \text{ and } \rho = \mathcal{D}_\rho(\mathbf{z}). \quad (4.10)$$

ここで \mathcal{D}_E と \mathcal{D}_ρ はそれぞれ, binding energy と密度に関するデコーダーである. エンコーダーとデコーダーはニューラルネットワークでモデリングされ, 学習データにより最適化される. 図 4.3 はこのモデルのアーキテクチャーの概念図を表している. この時抽出された common feature representation \mathbf{z} は, エネルギーと密度両方の情報を豊富に含んでいることが期待される. つまり, ^{236}U の基底状態近傍の adiabatic な変形ダイナミクスに対する重要な情報である.

複数のタスクを同時に扱う機械学習の手法は multi-task learning (MTL) [61] と呼ばれ, 通常転移学習 (Transfer Learning, TL) [32, 33, 34] の文脈で定義されることが多い. TL では, MTL は複数のタスクの情報を学習に利用することで, 個別のタスクにおける汎化性能向上を行う手法であ

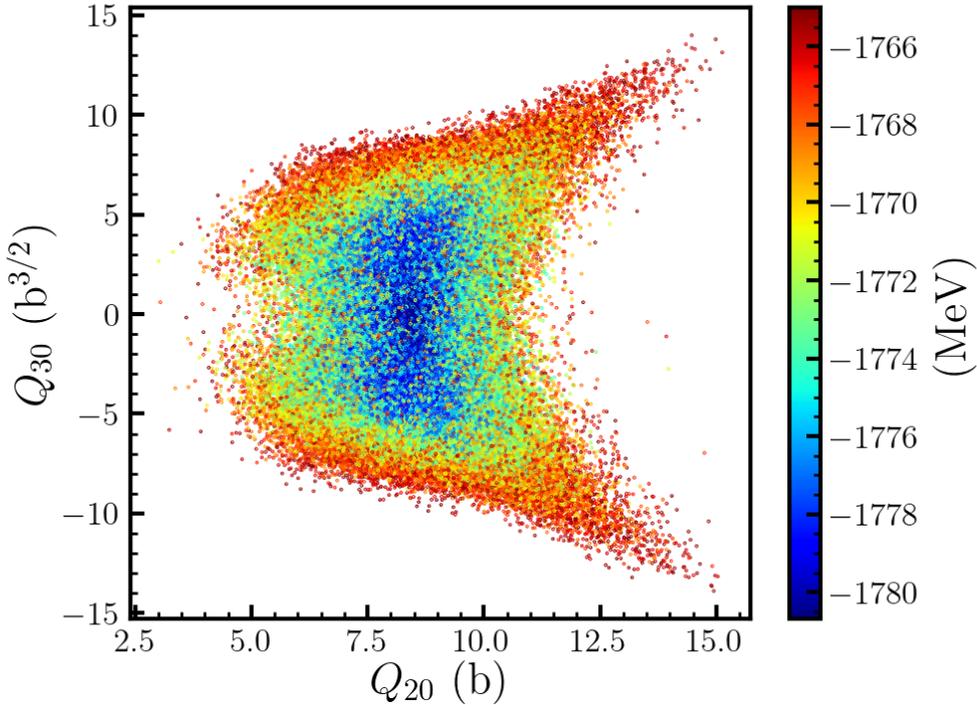


図 4.2 データセットにおける四重極モーメントと八重極モーメントの散布図. 各点の色は対応する束縛するエネルギーを表している. Skyrme EDF が z 軸方向に対する parity 変換に対して普遍であることを反映して, $Q_{30} = 0$ 軸に沿って点の色は反転対称である.

る. その一方で本研究における MTL は, common feature representation の獲得のみに焦点を当てた用途であるため, 標準的な目的と異なることに注意が必要である.

4.3.1 アーキテクチャー

MTL を実行するため, エンコーダー \mathcal{E} とデコーダー $\mathcal{D}_E, \mathcal{D}_\rho$ に対して具体的なモデルを定義する必要がある. この研究では入力変数は核子数密度 $\rho(r, z)$ であり, 離散化の結果, サイズが $1 \times 48 \times 128$ の白黒画像と等価である. この画像サイズは MLP のみで扱うには大きいため, 画像認識分野で多大な成功を収めてきた畳み込みニューラルネットワーク (Convolutional Neural Network, CNN) を主に用いる. 本研究では ResNet [68] と呼称される, CNN 系のモデルの中で最も有名なものの一つを用いる. ResNet の基本的なアイディアは, skip (residual) connection と呼ばれる層を跨ぐデータのフローによるもので, 典型的には一つ前の層からの出力が, 現在の層の出力に足される形で実装される. residual connection はニューラルネットワークの各層が恒等写像に近い関数を学習する手助けをすることで, 多層化に対してモデルの性能が向上しないという degradation problem を解決する手法として提案された. オリジナルの ResNet の論文 [68] では, 18, 34, 50, 101, 152 層という 5 種類のモデルが提案されており, それぞれ ResNet18, ResNet34,

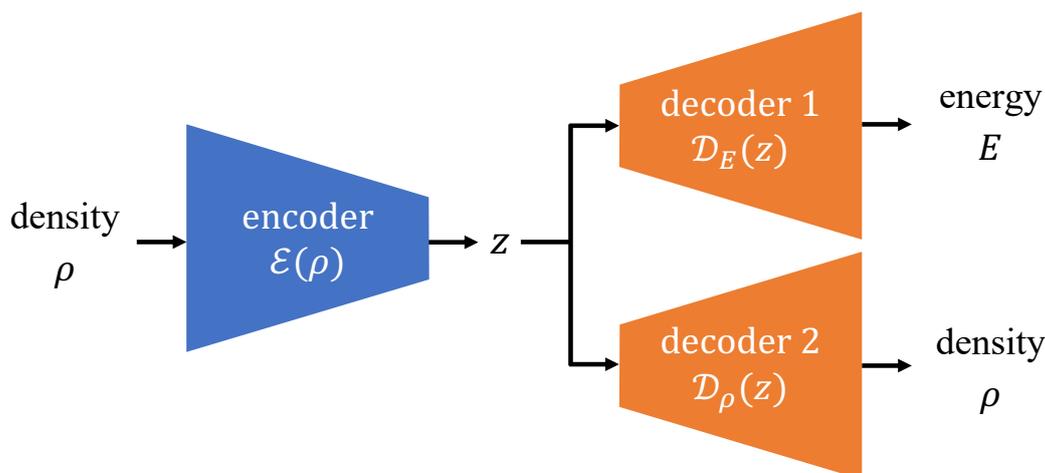


図 4.3 本研究で使用する MTL モデルの概念図. エンコーダー ε が密度 ρ の情報を潜在変数という形に圧縮し, 密度のデコーダー \mathcal{D}_ρ が密度自身を復元する. 同時にエネルギーのデコーダー \mathcal{D}_E により, 潜在変数からエネルギーの推定が行われる.

ResNet50, ResNet101, ResNet152 と呼称されている. residual connection という優れたアイデアにより ResNet は 2015 の著名な画像認識コンペティション [69] で優勝し脚光を浴びた. 2023 年 12 月現在 ResNet の論文は約 19 万回引用されており, 物理学だけでなく医療, 工学といった様々な分野の deep learning においてそのアイデアが用いられているため, 非常に信用度の高いモデルである.

したがって本研究でも ResNet18 をエンコーダー ε に用いる. ただし, ResNet が扱っていた ImageNet [70] というデータセットでは, 画像サイズが $3 \times 256 \times 256$ であり, ^{236}U の密度データより大きい. そこで本研究で用いる全ての ResNet では, kernel size が 3, stride が 2, padding が 1 の 2D max pooling layer を取り除いている. また出力層のニューロン数は潜在次元 d_z と一致するように調整し, この値はハイパーパラメータとして扱い, 様々のものを用い数値実験を行う.

核子数密度のデコーダ \mathcal{D}_ρ に関して, ResNet を直接用いることはできない. そこで, 図 4.4 に示される, ResNet18-like なモデルを採用した^{*4}. ここで前章の研究でデコーダーに用いていた transposed convolutional layer を, 本研究では用いなかった理由について言及する. transposed convolutional layer では画像の拡大を行ったのち, 通常の畳み込み演算を行う. この時適切なハイパーパラメータを選択しない場合, checkerboard artifacts [72] と呼ばれる, transposed convolutional layer 特有の模様が出力に現れやすいことが知られている. 一方, 本研究で用いられている upsampling layer は, 画像を整数倍にしかできず, ジャギーに注意する必要があるものの, checkerboard artifacts ほど深刻な問題は生じない. 加えて upsampling は画像に特定の方向を陽に発生させるような構造を持たず, z 軸方向に関しても同様であるため, 物理的に好ましいモデルの構築に役立つ. 本研究では画像サイズの適切な選択により, upsampling layer が扱いやすくなっている. 一方でエネルギー側のデコーダー \mathcal{D}_E として, 6 層の fully-connected layer からなる MLP を

^{*4} 活性化関数として Parametric Rectified Linear Unit (PReLU) [71] を用いた計算も行ったが, 性能に優位な改善は見られなかった.

採用した. ここで潜在変数に対して過剰に複雑な依存性を持ったエネルギーを避けるため, ResNet に比べて比較的浅いモデルを選択していることに注意が必要である.

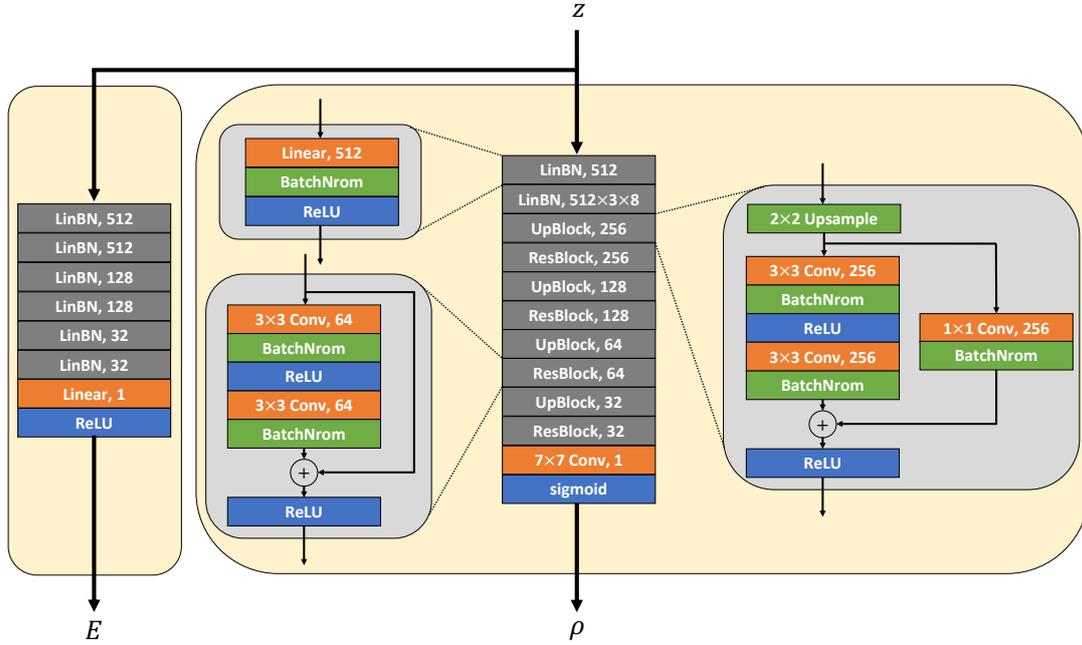


図 4.4 MTL モデルのデコーダー側のモデルアーキテクチャー. 左右のブロックはそれぞれ, \mathcal{D}_E と \mathcal{D}_ρ に対応する. 畳み込み層において, 左側の数字は kernel size を意味しており, 右の数字は出力チャンネルを表している. また全ての padding は same padding となるように取られている. また, upsampling layer は入力画像を縦横それぞれ 2 倍にする層であり, 本研究では内挿は用いていない.

4.3.2 損失関数

multi-task learning では複数のタスクを含むため, 最適化手法は極めて重要な問題である. 基本的なアプローチは, 個別のタスクに対する損失関数を適切な非で足し合わせ, バランスさせる方法である [61]. MTL の分野では, 異なる損失関数の組み合わせ方として uncertainty weighting (UW) [73] や dynamic weight average (DWA) [74] といった手法が知られている. 本研究の MTL では, エネルギーの推定タスクと密度の復元タスクは, 特に潜在次元 d_z が小さい時に, 互いに競合する. また, 2つのタスクの性能を同時に評価可能な単一の評価関数を定義することが望ましいが, MTL を世界ではじめて集団運動に適用する本研究では, 事前にそのような関数を決めることは難しい. したがって, 本研究では単純にエネルギー損失関数 \mathcal{L}_E と密度の損失関数 \mathcal{L}_ρ を定数比 α で組み合わせる:

$$\mathcal{L}_{\text{MTL}}(D; \alpha) = \mathcal{L}_E(D) + \alpha \mathcal{L}_\rho(D). \quad (4.11)$$

ただし,

$$\mathcal{L}_E(D) := \frac{1}{|D|} \sum_{i=1}^{|D|} \left(E_{\text{pred}}^{(i)} - E_{\text{true}}^{(i)} \right)^2, \quad (4.12)$$

$$\mathcal{L}_\rho(D) := \frac{1}{|D|} \sum_{i=1}^{|D|} \int d^3r \left(\rho_{\text{pred}}^{(i)}(\mathbf{r}) - \rho_{\text{true}}^{(i)}(\mathbf{r}) \right)^2, \quad (4.13)$$

である. ここで, $|D|$ はデータセット D の要素数を表している. ハイパーパラメータ α はエネルギーの推定と密度の復元タスクのどちらを優先するか決定するパラメータである. 次節ではいくつか異なる α で計算を行い, MTL のパフォーマンスを評価する. neural network のすべてのモデルは, 訓練データに対してこの損失関数 \mathcal{L}_{MTL} を最小化するように最適化される. ここで計算時に, すべてのデータは min-max normalization を取ることに注意が必要である. 一方, 評価関数などで性能を議論するときは, その逆操作を行うことで, 物理的な次元を元に戻す. 損失関数中の, エネルギー E と密度 ρ は normalization により無次元量であるため, 損失関数 $\mathcal{L}_E(D)$ は無次元, $\mathcal{L}_\rho(D)$ は (4.13) 式の積分に起因して fm^3 の次元を持つ. したがって, α の物理的な次元は fm^{-3} である.

4.3.3 optimization

本研究では neural network を最適化するため, Adam optimizer [29] を用い, ハイパーパラメータに関しては learning rate を除き PyTorch [22] のデフォルト値を用いる. learning rate の初期値は 10^{-4} とし, 文献 [60] と同様に動的に減衰させていった. つまり訓練データとは別に用意された検証データに対する誤差を epoch 毎に計算し, 15 回連続で改善が見られなかった場合, learning rate を 0.5 倍する. これは PyTorch では “torch.optim.lr_scheduler.ReduceLROnPlateau” を用いることで容易に実装可能である. このような動的な learning rate decay アプローチを用いる場合, 検証用データを学習に利用しているため, 公正な評価のため最終的なモデルの性能評価に別のデータが必要になる. したがって, 181 134 個のデータをランダムに訓練データ (90%), 検証データ (5%), テストデータ (5%) に分割した. この時, テストデータは最終的な性能評価のみに用いているという点を強調しておく.

本研究では最適化はミニバッチ学習を用いて行われ, バッチサイズは 128 を選び, epoch の数は 1000 とした. 本研究で用いる MTL モデルでは, 特に潜在次元が小さい場合に, 2 つのタスクは互いに敵対的に振る舞う. そのために, 検証誤差は訓練誤差に比べて不安定になり, その収束性は悪い. そこで, 1000 epoch の学習の間で最も小さい検証誤差 \mathcal{L}_{MTL} を与えるモデルを最終的な結果として選んだ. 次節以降で見ると, 本研究では異なるタイプの deep learning の計算も行う. それら全てに対して, 特に言及のない部分に関しては, すべて上記の学習法に従い最適化を行なっている.

最後に計算コストに関して言及する. 本研究では単一のハイパーパラメータに対する MTL モデルを訓練するのに, NVIDIA RTX 3090TI を用いて 50 GPU-hours 要した. つまり, n 個の NVIDIA RTX 3090TI を連続して $50/n$ 時間稼働させる必要がある.

表 4.1 OF-EDF $E[\rho]$ の教師あり学習に対する平均絶対誤差 (MAE) と平均自乗誤差 (MSE). 計算には 4 種類の ResNet と ViT を用いた. 結果からエネルギー推定タスクのエラーの基準を 64 keV と定義することができる. 比較のため最下段に, 陽子数密度と中性子数密度を同時に入力とした OF-EDF の学習結果を載せた.

model	input(s)	validation data (5%)		test data (5%)	
		MAE / MeV	MSE / MeV ²	MAE / MeV	MSE / MeV ²
ResNet18	ρ	0.0633	0.00965	0.0638	0.00963
ResNet34	ρ	0.0638	0.00965	0.0646	0.00949
ResNet50	ρ	0.0649	0.01084	0.0649	0.01023
ResNet09	ρ	0.0724	0.01197	0.0720	0.01161
ViT [41]	ρ	0.2613	0.15161	0.2704	0.16302
ResNet18	ρ_n, ρ_p	0.0641	0.00973	0.0635	0.00951

4.4 評価基準

4.4.1 エネルギー推定タスク

MTL の結果を評価する前に, 性能評価のための基準値に関して議論を行う. 本研究で用いるデータの数は有限であるため, ^{236}U の低エネルギーダイナミクスを完全に復元することは不可能である. むしろ今のような少ないデータセットに対して*5 極限まで小さい誤差を目指すような訓練を行った場合, ソースドメインに過剰に適合したモデルが得られるのみで, ドメインシフトに対して著しく脆弱となる可能性があるため, 注意が必要である. そこで学習データの有限性から来るもっともらしい誤差を見積もるため, MTL における 2 つの異なるタスクに相当する学習を別途行う. ここではまず, 教師あり学習により OF-EDF $E[\rho]$ の再構築に臨み, そのエラーをエネルギー推定タスクの下限值とする.

表 4.1 にさまざまなモデルを用いて計算された, エネルギー推定タスクの平均自乗誤差 (Mean Square Error, MSE) と平均絶対誤差 (Mean Absolute Error, MAE) を示した. ResNet は出力層を除き, 3.3.1 節で説明したものと同じである. ただし, binding energy というスカラー量を出力するために, 出力層のニューロン数を 1 とし, また最後に ReLU を演算させた. ResNet09 は原論文 [68] には含まれないモデルであり, ResNet18 の Residual ブロックをちょうど半分にしたモデルに相当する. 全てのモデルは (4.12) 式で定義される MSE 型の損失関数を用いて訓練されている. ここで, ResNet34 と ResNet50 のエラーが ResNet18 と同等であることが見て取れる. このことから ResNet18 は我々の ^{236}U のデータセットの特徴を捉えるのに, 十分な表現能力を持っていると期待される. 本研究で ResNet18 を主に使うのは, このためである. 以上の結果をもって, エネルギー推定タスクに対するエラーの基準を 64 keV と定義する.

*5 画像生成の分野では, 例えば LAION-5B [75] と呼ばれる, 58.5 億枚の画像とテキストのペアからなるデータセットが CC-BY 4.0 ライセンスで公開されている. それと比べた場合, 本研究の 18 万というデータ数は小規模である.

本研究では Vision Transformer (ViT) [41] を用いた OF-EDF の再構築も行なった。ViT は画像認識タスクにおいて CNN 系のモデルを打ち破り, state-of-the-art (SOTA) を達成したモデルである。その基本的なアーキテクチャーは大規模言語モデル (Large Language Model, LLM) で著しい成功を収めた Transformer [16] を, 画像処理タスクに応用したというものである。本研究では画像サイズの違いに起因して, オリジナルの ViT モデルにモチベートされたモデルを使用した。本研究で用いた ViT 型のモデルは 16 個の multi-head self attention からなる transformer ブロックにより構成される。各 attention における head の数は 16 で, transformer block の feedforward 部分のニューロンの次元は 2048 である。またパッチサイズ, embedding dimension, dropout rate はそれぞれ 8×8 , 128, and 0.3 である。表 4.1 によると, この ViT 型モデルのエラーが ResNet のものに比べて大きい様子が見て取れる。しかしながら原論文でも述べられているように, ViT の性能を引き出すには大規模なデータが必要である [41]。その一方で, 将来的にデータの数が増えれば ViT は重要な役割を果たすと期待される。したがって, 2023 年現在の段階で ViT の計算を行い, 性能を示すことには意味がある。

本研究では陽子と中性子に同じ外場を与えるという限られたセクターに限定してデータセットを生成したため, この範囲内ではエネルギーは核子数密度 ρ のみの汎関数で与えられる。したがって neural network への入力として ρ を用いている。しかしデータとして陽子数密度 ρ_p と中性子数密度 ρ_n の値も保持しているため, 検証のため ρ_p と ρ_n を入力変数とした OF-EDF $E[\rho_n, \rho_p]$ の学習も同様に行なった。図 4.1 の最下段に示されているように, 結果は核子数密度 ρ のみを入力変数とした場合とほとんど変わらない。したがって理論的に想定されるように, 本研究のデータセットにおけるダイナミクスの自由度は, 核子数密度のみである。

4.4.2 密度復元タスク

エネルギー推定タスクの場合と同様に, 密度に対しても誤差の基準を定義する。ここで密度は DFT の文脈において入力変数であり, 教師あり学習は使えない。したがって, 教師なし学習の一種である autoencoder (AE) [36] により誤差を評価する。図 4.3 に示される本研究の MTL モデルに対して, 仮に \mathcal{D}_E を取り除き, 損失関数からエネルギーの項を取り除けば, 密度に関する AE と等価になる ((4.8) 式を参照)。そこで, さまざまな潜在次元 d_z に対して AE を計算し, テストデータに対して誤差を評価した。結果は図 4.5 にまとめている。ただし, 密度の誤差に関する評価関数は, 前節の ^{24}Mg に対する研究と同様, 平均絶対誤差 (MAE) を採用する:

$$\mathcal{M}_{\text{MAE}}^\rho := \frac{1}{|D|} \sum_{i=1}^{|D|} \int d^3r \left| \rho_{\text{pred}}^{(i)}(\mathbf{r}) - \rho_{\text{ans}}^{(i)}(\mathbf{r}) \right|. \quad (4.14)$$

ただし, MAE は物理的な次元を持った密度を用いて計算されるため, 無次元量である。ここで小さな潜在次元に対しては, 質量数 $A = 236$ の数パーセント程度の MAE で密度を復元できていることがわかる。この大きい誤差は我々のデータセットが多様な密度を含んでいることを示唆している。その多様性のあらわれとして, 潜在次元を増やしたとしても密度は急速には減少しない。潜在変数の最大値である $d_z = 512$ は, オリジナル ResNet [68] の出力層のみを変更した場合に実現可能な

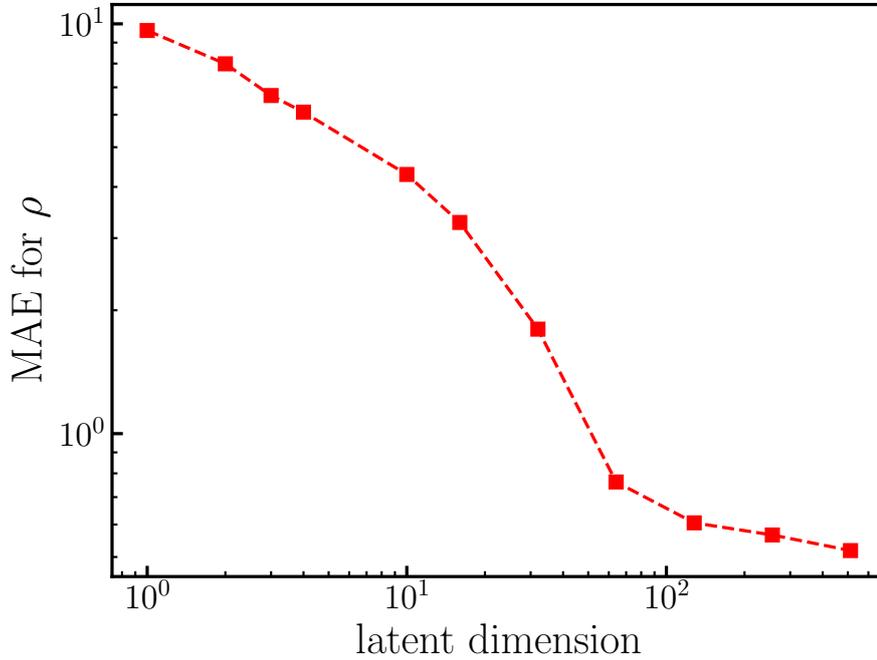


図 4.5 autoencoder を用いた密度復元タスクに対する平均絶対誤差 (MAE). 潜在次元として, $d_z = 1, 2, 3, 4, 10, 2^4, 2^5, 2^6, 2^7, 2^8, 2^9 = 512$ を選び計算を行った.

潜在次元の最大値である. この時 MAE は 0.5 と非常に小さい.

4.4.3 多重極モーメントによる推定

伝統的に, 原子核変形ダイナミクスを記述する上で, しばしば多重極モーメント $\{Q_{\lambda,k}\}_{\lambda,k}$ が用いられてきた. したがって MTL との比較の上で, 我々のデータセットにおいてこれらの量がどの程度の情報を持っているかは興味深い問題である. そこで多重極モーメントの集合を入力変数とし, エネルギーを出力とした $E(Q_{20}, Q_{30}, \dots, Q_{L0})$ の教師あり学習を行い, 多重極モーメントの情報からエネルギーをどの程度推定可能か調べた. ここでは neural network として我々の MTL モデルのデコーダー D_E を用い, $\{Q_{\lambda,k}\}_{\lambda,k}$ を潜在変数とみなす. 図 4.6 は入力変数として $S_L = \{Q_{20}, Q_{30}, \dots, Q_{L0}\}$ で定義されるいくつかの集合を用いた場合, MAE がどのように変化するかを示している. 結果として, 4.4.1 節で得られた 64 keV という基準値に比べ, 著しく大きい MAE が得られた. モデルの表現力が不十分であった可能性を考慮し, より多層のモデルで同様の計算を行ったが, 結果に改善は見られなかった. 比較のため, neural network の出力が完全にランダムで, 我々のデータセットのエネルギーの範囲から一様乱数としてサンプリングされると仮定した場合に MAE を計算した. つまり, エネルギーに関する MAE の定義式 (3.24) の $E_{\text{pred}}^{(i)}$ を乱数で与えるということであり, 結果として 5.0 MeV を得た. このことから, 低い次数の多重極モーメントによるエネルギー推定タスクのエラーは, ランダムな結果の半分程度であることがわかる. この意

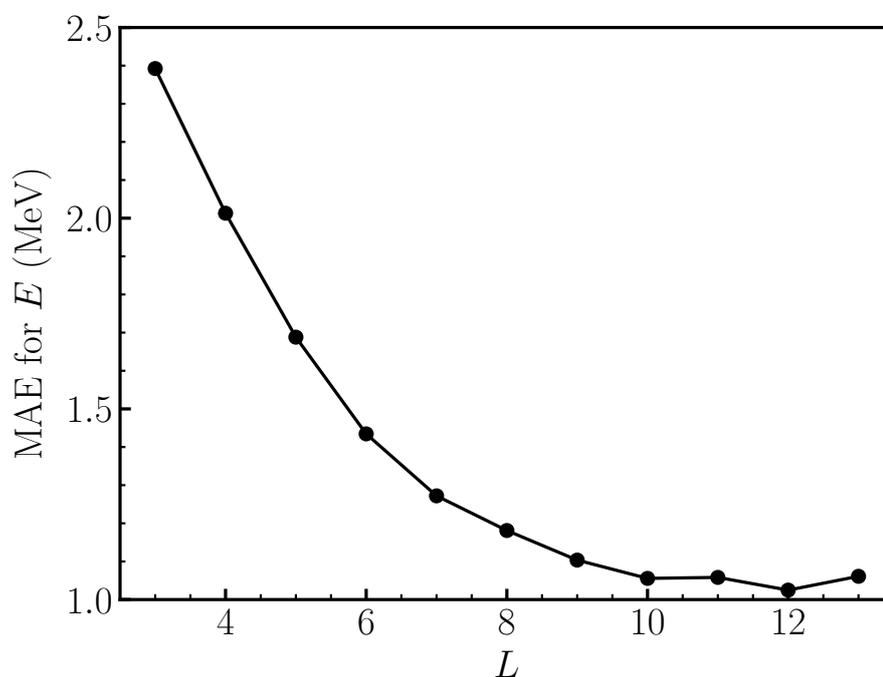


図 4.6 多重極モーメントを入力変数とした場合のエネルギー推定タスクにおける平均絶対誤差 (MAE). 横軸の L は $S_L = \{Q_{20}, Q_{30}, \dots, Q_{L0}\}$ に含まれる最大の多重極モーメントの量子数を意味している. MTL のモデルと比較した場合, 多重極モーメントはある種の潜在変数と見なすことができる.

味で多重極モーメントは, 少なくとも我々のデータセットに対して, ほとんど情報を持っていない.

エネルギー推定タスクと同様に, デコーダー D_ρ を用いた多重極モーメントによる密度の復元タスクも行なった. しかし損失関数の減少が非常に悪く, MAE は質量数 $A = 236$ に対し数倍も大きかった*6. したがって, 少なくとも我々のデータセットに対し, 多重極モーメントは密度を復元する能力を持っていないといえる. 以上の非常に大きい誤差は, 多重極モーメントの密度に対する情報不足に起因すると考えられ, そのために neural network が適切な方向に学習を進めることができなかつたためであると考えられる. そこで, 大規模なデータなどを用いて事前学習したモデルを用いて学習を行えば, 結果は改善する可能性があることに注意が必要である. ネット上に豊富にデータセットが存在する通常の画像認識タスクとは異なり, 本研究のような極めて特殊なドメインに対して使用可能な事前学習用データセットはおそらく存在しないため, 本研究では事前学習を用いた性能の向上には取り組まない.

*6 本研究では ^{24}Mg の場合のような粒子数に関する要請は課していない. したがって, MAE に上限はない.

4.5 結果： ^{236}U の潜在変数

4.5.1 潜在変数と次元

MTL の結果と抽出された潜在変数に関して論じる。MLT の損失関数は1つのハイパーパラメータ α を持っており、これによりエネルギー推定タスクと密度復元タスクのどちらを優先するかをコントロールできる ((4.11) 式を参照)。本研究では、2つの損失関数 \mathcal{L}_E と \mathcal{L}_ρ の大きさが同程度のスケールになるように α を決めた。具体的には $\alpha = 10^{-2}, 10^{-3}, 10^{-4} \text{ fm}^{-3}$ という異なる3つの値を用いて、MLT モデルを最適化した。

図 4.7 は左側にエネルギー推定タスクの MAE, 右側に密度復元タスクの MAE を示している。ただし、全て対数軸であることに注意が必要である。両者共に潜在変数の増加に伴い MAE は減少する傾向にあるが、エネルギー推定タスクでは急速に、密度復元タスクでは比較的なだらかにエラーが減少する。また、2つの異なるタスクは互いに競合するため、 α を増減させた時の MAE の増減において、逆の振る舞いを見せる様子が見て取れる。つまり、 α を大きく取ると密度の MAE は減少し、その一方でエネルギーのエラーは増大する。

ここで、少数の潜在変数によりエネルギーが数百 keV の精度で推定可能であることは、特筆すべきである。特に、潜在次元を1から2に増やした時、 $\alpha = 10^{-2}, 10^{-3} \text{ fm}^{-3}$ においてエネルギーのエラーが急速に減少する。その一方で密度復元タスクに関しては、基準である autoencoder (AE) (ダッシュ線) と比べることにより、我々の MTL モデルは潜在次元が2の場合であっても、十分な精度で密度を復元可能であることがわかる。

結果を比べると、 $\alpha = 10^{-3} \text{ fm}^{-3}$ というハイパーパラメータは良い選択であると考えられる。 α を大きく取るとエネルギーのエラーが大きくなり、小さく取ると密度のエラーが大きくなるという競合の中で、2つの異なるタスク共に最良の性能を出すような学習は実現不可能であると考えられる。一方で、 $\alpha = 10^{-3} \text{ fm}^{-3}$ は2つのタスク共にほどよい性能をもつため、今後はこの α の場合に議論を限定する。ここで、潜在次元を2から3に増やした場合に、エネルギーの MAE はそれほど大きく減少しない。その一方で密度のエラーは、AE の結果に沿うように減少する。このことは、我々のデータセットに含まれるダイナミクスの大部分は、2つのパラメータで特徴付けられ得ることを示唆している。ここで我々の潜在変数は、少なくとも我々のデータセットに対して、伝統的な多重極モーメントよりも多量の情報を持っていることを強調しておく。この事実は、多重極モーメントが ^{236}U といった重い原子核の変形ダイナミクスを記述する上で、適当な集団座標ではないことの根拠を定量的に与える。ただし、我々のデータセットは外場のない場合の基底状態近傍の変形領域であることに注意が必要である。

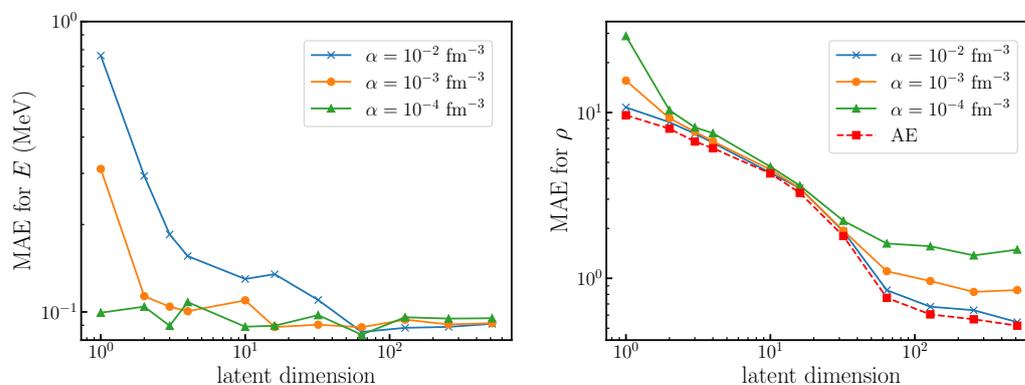


図 4.7 様々な潜在次元とハイパーパラメータ α に対するエネルギー推定タスクの MAE (左図) と密度復元タスクの MAE (右図). 右図のダッシュ線は図 4.5 に示されている autoencoder (AE) の結果と同じである.

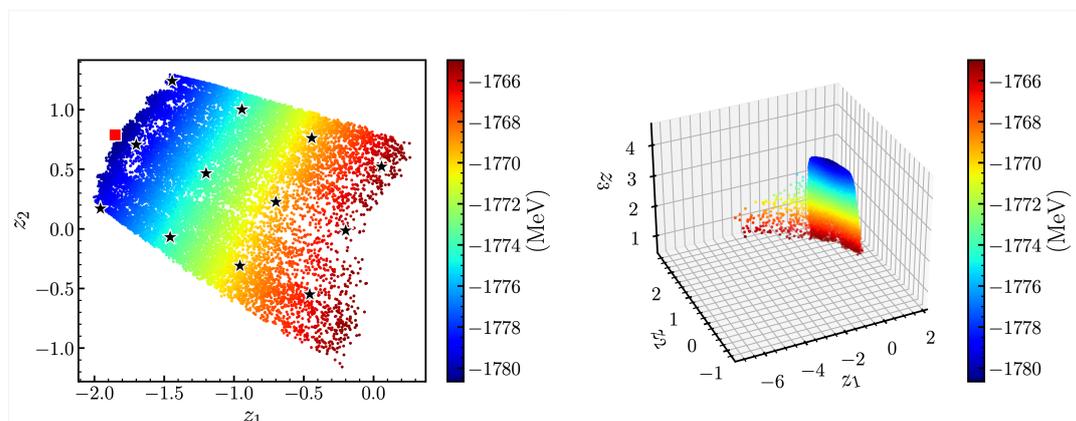


図 4.8 検証データ (5%) とテストデータ (5%) に対する潜在変数の散布図. 異なる 2 つのデータセットに対して性能は同程度であったため, 図に表示する点を増やし, 見やすくするという目的で検証データに関しても同時にプロットしている. MTL 損失関数のハイパーパラメータが $\alpha = 10^{-3} \text{ fm}^{-3}$ の場合に対して, 潜在次元が 2 (左図), 潜在次元が 3 (右図) の場合の潜在変数を表示している. また, 各データの binding energy は点の色で表している. 左図の赤い四角の点は, 外場のない場合の基底状態のデータ点に相当し, 対応する潜在変数は $(z_1, z_2) = (-1.85, 0.79)$ である. また黒い星型の点に関しては, 図 4.9 上に, エンコーダーで復元した密度を表示している.

潜在次元が3以下の場合、潜在変数の分布は容易に可視化することができる。特に、2次元と3次元の場合は4.4節で定義した基準と比べても性能が良いため、潜在変数の振る舞いも興味深い。図4.8は $\alpha = 10^{-3} \text{ fm}^{-3}$ の場合に対する $d_z = 2$ (左図) と $d_z = 3$ (右図) の潜在変数の散布図である。ここで、潜在変数は物理的な次元を持たず、そのスケールと中心位置には意味はないことに注意が必要である。色軸は各点の binding energy を表している。エラーが同程度であったため、各パネルには検証データとテストデータ両方を同時にプロットしている*7。外場がない状況での ^{236}U の基底状態は $(z_1, z_2) = (-1.85, 0.79)$ に対応し、図中では四角い点で示されている。ここで、左図に対称軸が存在する様子が見て取れる。具体的な密度の振る舞いを見るために、対称軸に沿って選ばれた12個の点（左図上の星形の点）に対して、デコーダーで復元した密度を図4.8に示した。復元した密度であるため、一定のエラーを含んでいることに注意が必要であるが、対称軸から見て逆側にある点がおおよそ z 軸に対して反転した密度に相当していることがわかる。この結果は我々のMTLモデルが、ダイナミクスが持つ z 軸方向の parity 対称性を認識するのに成功したということを示唆している。

ここで潜在空間における対称軸の傾きや切片に物理的な意味はなく、neural network のモデルパラメータの初期値で決まることに注意が必要である。実際、全く同じセットアップのもと、別の初期値を用いてMTL計算を行ったところ似たようなプロットが得られたが、 z 軸の原点やスケールが異なっていたため、対称軸も異なっていた。

$d_z = 2$ に焦点を当て、さらに解析を行う。図4.10は各潜在変数の色として、四重極モーメント(左図)と八重極モーメント(右図)を選んだ場合の散布図である。これらに対しても同様に、対称軸の存在が確認できる。また、色のパターンが滑らかな単純な関数になっていないことから、 Q_{20} と Q_{30} が、我々のデータセットに対して特別な役割を果たしていないことが伺える。我々の潜在変数の利点は、多重極モーメントの空間において励起状態に対応するような全ての点が、基底状態として扱えるという点である。データは全てDFTの枠組みで計算されているため、何らかの外場のもとでの基底状態であることは自明である。一方で、多重極モーメントのような限られた変数空間で描かれたエネルギー局面を考えると、一般には多価関数になる。この意味で、DFTの枠組みであっても多重極モーメントに対して励起状態が存在し得る*8。実際、図4.2に示されているように、我々のデータセットは四重極モーメントと八重極モーメントの空間で見た場合励起状態を含む。一方で我々の潜在空間では、少なくとも我々のデータセットに対して、一価のエネルギー関数を与える。

図4.2の多重極モーメント(Q_{20}, Q_{30})が2次元の潜在変数空間に対してどのようにマップされるかを調べるため、拘束条件付きHFB計算を行った*9。図4.11の左のパネルで点同士を結ぶ矢印は、色を含め右のパネルと共通である。矢印により示される多重極モーメントのトラジェクトリーを (z_1, z_2) 平面で見ると、極めて複雑であり、 (Q_{20}, Q_{30}) から (z_1, z_2) への写像が非自明なものであることがわかる。このことから、MTLモデルにより捉えられた我々のデータセットの特徴量は、伝統的な多重極モーメントにより容易に解釈できるようなものではなく、全く新しいものである。

*7 MAEといった定量的な結果には検証データは一切使用していないことを強調しておく。

*8 原子核理論では、このような文脈で励起状態という言葉を使うことがある。

*9 計算にはHFBTHO (v3.00) [65]に付属している機能を用いた。ただし、HFBTHOと本研究の四重極モーメントの定義は、球面関数の規格化定数の分だけ定数倍異なっていることに注意が必要である。

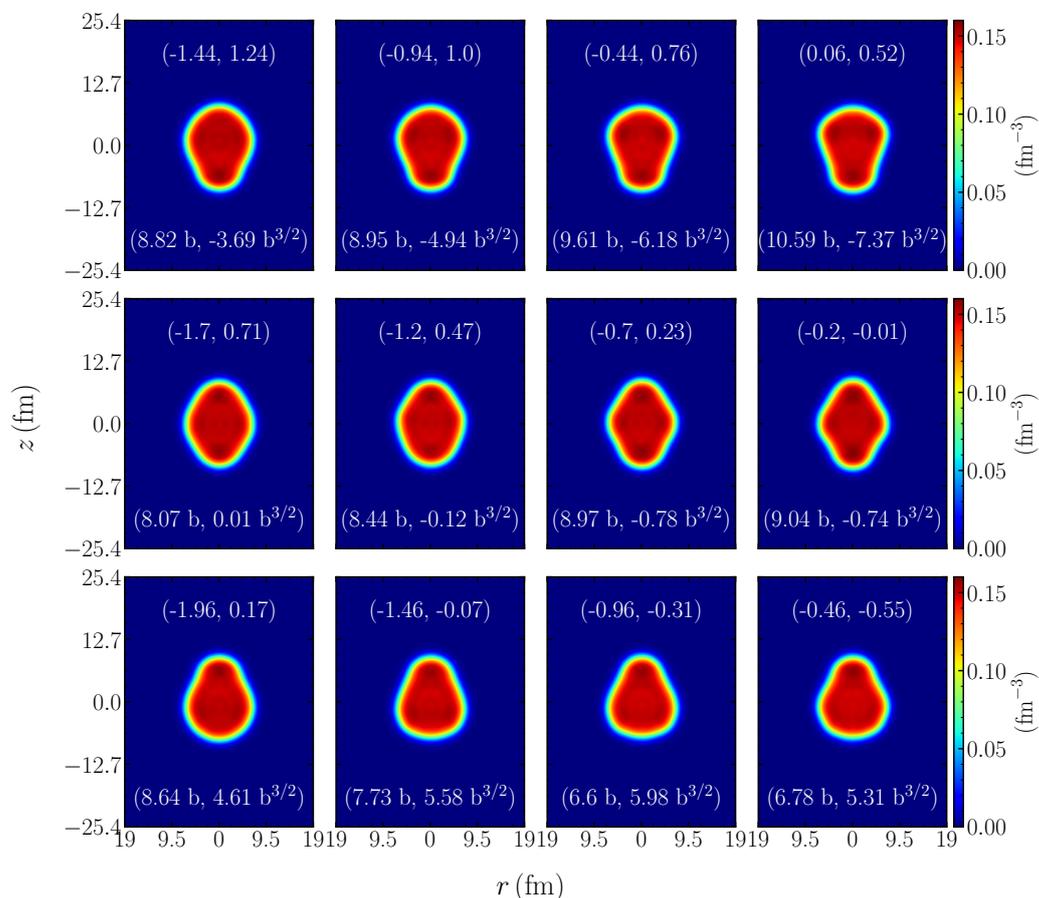


図 4.9 図 4.8 に示された点から復元された粒子数密度のサンプル. 各パネルの上部に書かれた 2 つの数字は, 潜在変数 (z_1, z_2) の値を示している. 一方で, 下部に表示された数字は四重極モーメントと八重極モーメント (Q_{20}, Q_{30}) である.

4.5.2 ドメインシフト

2.3 節で既に述べたように, テストデータに対するモデルの性能は, 未知のデータセットに対する予言能力を保証しない. 前節では, 同一のドメインからサンプリングされたデータをランダムに分割し, hold-out validation 法により学習を行った. このような “traditional machine learning method” ではソースドメインとターゲットドメインが同じであることを暗に仮定しているため, 異なるドメインに対してどの程度モデルが汎化しているかは不明である. このようなドメインに関する詳細な議論は, 物理学において deep learning がまだ黎明期であることから, あまり見かけない. しかし実社会へのアプリケーションにおいてドメインの問題は極めて深刻であり, data driven ア

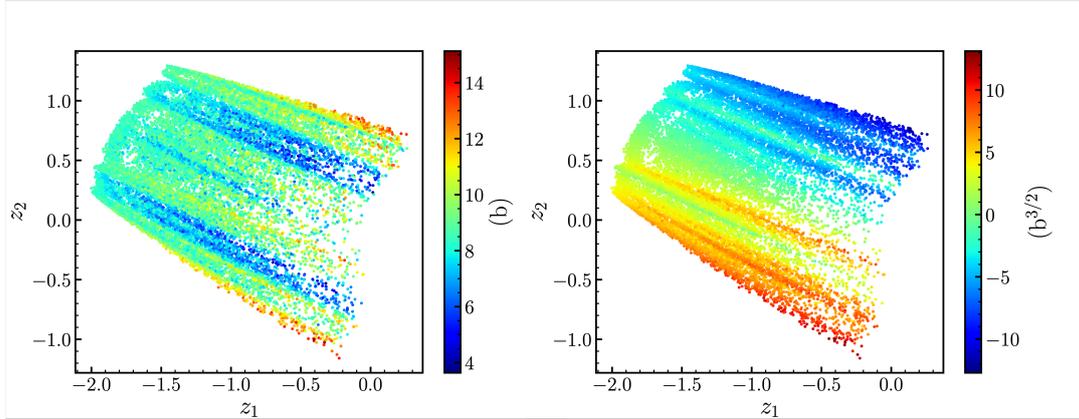


図 4.10 図 4.8 の左側と色を除いて同じ潜在変数の散布図. ただし, 色軸として四重極モーメント (左図) と八重極モーメント (右図) を選んだ.

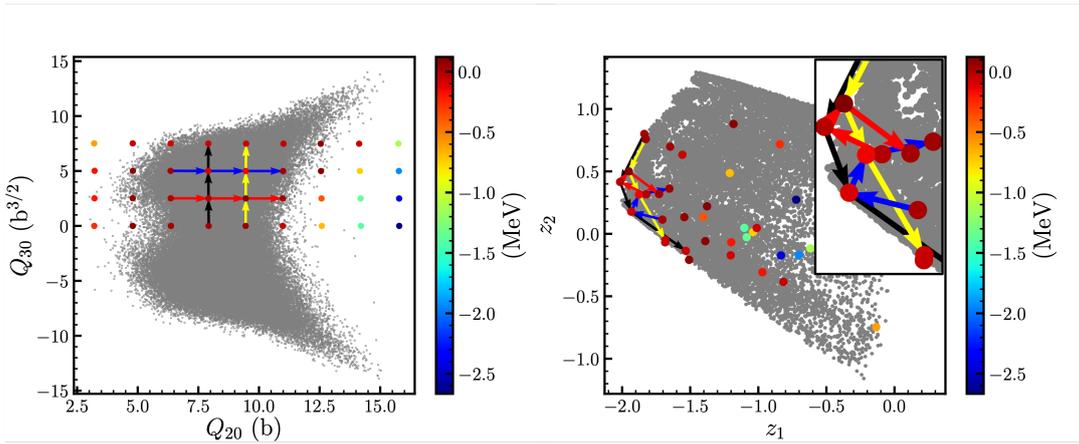


図 4.11 多重極モーメントの空間 (図 4.2) から 潜在空間 (図 4.8 の左側) へのマップ. 各点は色のついた矢印で結ばれ, 左右のパネルで共通している. 各点の色は Kohn-Sham で得られた真のエネルギーと MTL のデコーダーによる推定値の差を表している. ただし, $\alpha = 10^{-3} \text{ fm}^{-3}$, $d_z = 2$ に対する MTL モデルの結果を用いている.

アプローチに基づく物理学の研究においても, 注意深く議論を行わなければならない.

第 2 章で既に見たように, ドメインは入力変数の空間 \mathcal{X} とその上で定義される確率分布 $P(X), X \in \mathcal{X}$ を用いて, $\mathcal{D} = \{\mathcal{X}, P(X)\}$ で定義される [33, 34]. 本研究では \mathcal{X} は核子数密度全体の為す空間であり, $P(X)$ はランダムポテンシャルの生成及び過剰なデータの除去という 2 つの操作によって決まり, その具体形は非自明である. これらによって決まるドメインが本研究におけるソースドメイン \mathcal{D}^s であり, これまではターゲットドメインにも同じドメインを用いて誤差の評価を行った. したがって, MTL モデルの性能をより公平に調べるため, $\mathcal{D}^t \neq \mathcal{D}^s$ となるターゲットドメイン \mathcal{D}^t からサンプリングされたデータが必要とされる.

幸いなことに, 本研究では多重極モーメントの性質を解析するため, 多重極モーメントに対する拘束条件付き HFB 理論に基づくデータの生成を既に行っている. 我々のランダムポテンシャルが基

表 4.2 MTL モデルにおける対称性を調べるための 2 つの評価関数 M_E, M_ρ の値. 両者はそれぞれ (4.15) 式と (4.16) 式で定義されている. これらの値はテストデータ (5%) のみを用いて計算されている. 比較のため, エネルギーと密度の MAE も同時に表示している. 表において対称性が “False” であるとは, 図 4.7 などで計算された対称性を考慮していないモデルの結果を表す. 一方で対称性が “True” とは, data augmentation とエンコーダーの改良を施していることを意味している.

symmetry	d_z	M_E / MeV	M_ρ	MAE (E) / MeV	MAE (ρ)
False	2	0.0842	5.0618	0.1133	9.2352
True	2	0.0798	5.4921	0.1250	9.2756
False	3	0.0675	3.7661	0.1043	7.6681
True	3	0.0716	3.3035	0.1003	7.7055

底展開法に基づいている以上, 原理的には拘束条件付き HFB 計算はソースドメインに含まれる. しかし, 我々の 18 万という数のデータセットにそれらが含まれる確率は極めて小さい. そこで多重極モーメントの拘束条件付き HFB 理論により定義されるドメインはソースドメインと大きく異なる. したがってこのドメインをターゲットドメイン \mathcal{D}^t とし, ドメインシフト $\mathcal{D}^t \neq \mathcal{D}^s$ に対するモデルのロバストネスを評価する^{*10}.

図 4.11 の各点の色は, グレーのものを除き, Kohn-Sham エネルギーと MTL のデコーダーにより推定されたエネルギーの差を表している. 多重極モーメントという観点でデータセットとオーバーラップがある領域とそうでない領域でエラーに大きな違いがある様子が見て取れる. 前者ではエネルギーはよく推定されており, MTL モデルが多重極モーメントでは見えないような予期せぬ特徴量に, 過剰に適合していない一つの根拠となる. 後者では全体的に真のエネルギーを過大に見積もっており, 多重極モーメントを入力としたエネルギー推定タスクの場合と同程度の, 非常に大きいエラーを生じる点もある. したがって, ソースドメインから大きく離れた点に対して我々の MTL モデルは予測性能が悪い. このような点に対してもある程度の予測性能を求める場合は, 単純に学習データに大変形した原子核のデータを多数追加するか, domain adaptation (DA) [33, 34] や domain generalization (DG) [76] といった学習テクニックを用いる必要がある.

4.5.3 対称性と data augmentation

本研究で得られた潜在変数は 2 つのタスクに対して十分良い性能を示したものの, データセットの対称性を考慮することで, より良い潜在変数を得られる可能性がある. これを見るため, この節では系の対称性を考慮したデータセットを構築し, MTL 計算を再び行う.

本研究では local density に対する時間反転対称性と軸対称性を要請した DFT 計算を行った. 原子核のハミルトニアンは以上の陽に要請された対称性の外, z 軸方向の parity 変換に対しても不変

^{*10} 不適切な最適化を行うと, データセットに対して過剰に適合したモデルが得られる. このような場合, データセットに含まれる予期しないバイアスを読み取り, それを推定に利用することで性能を向上させている可能性がある. つまり, 原理的に 2 つのドメインの確率分布にオーバーラップがあるからといって, ターゲットドメインに対して汎化しているかは注意深く見る必要がある.

である。その性質を反映し、各々の密度は一般には parity 変換に対して対称ではないが、原子核の binding energy は密度の parity 変換に対して不変である。この物理的な対称性を学習に反映させるため、データセットに工夫が必要である。ソースドメインは z 軸方向の parity 変換に対して不変であるため、極めて多くのデータをサンプリングした場合、データセットもまた parity 変換に対して不変とみなせることが期待される。現実的にそのような量のデータを生成するのは困難であるため、本研究では画像の反転を利用した data augmentation を用いる。つまり密度を z 方向に反転させたものと、エネルギーはそのままとしたものと考えた場合、対称性によりそれも Skyrme DFT の解となる。実際の計算では、学習の際に z 軸に沿って確率 0.5 でランダムに密度を反転させる。加えて、この研究では MTL モデルのエンコーダー部分に改良を加える。本研究でエンコーダーに利用したオリジナルの ResNet は、stride が 2 の場合に奇数の kernel size を使用しており、これは画像の中心のシフトを引き起こす可能性がある。そこで、エンコーダーの stride が 2 の convolutional layer に関して、kernel size を 1 だけ増やした。

この改良された MTL モデルを、 $d_z = 2, 3$ かつ $\alpha = 10^{-3} \text{ fm}^{-3}$ の場合に対して、data augmentation を施したデータセットで訓練した。ただし、モデルの対称性を評価するため、2つの評価関数を新たに導入する：

$$M_E = \sum_{i=1}^{|D|} \frac{1}{|D|} \left| \mathcal{D}_E(\mathcal{E}[\rho^{(i)}]) - \mathcal{D}_E(\mathcal{E}[\rho_{\mathcal{F}}^{(i)}]) \right|, \quad (4.15)$$

$$M_\rho = \sum_{i=1}^{|D|} \frac{1}{|D|} \int d^3r \left| \mathcal{D}_\rho(\mathcal{E}[\rho^{(i)}]) - \mathcal{F}[\mathcal{D}_\rho(\mathcal{E}[\rho_{\mathcal{F}}^{(i)}])] \right|. \quad (4.16)$$

ただし、 \mathcal{F} は画像の反転操作であり、 $\rho_{\mathcal{F}}$ は $\mathcal{F}[\rho]$ で定義される。モデルが対称性を完全に再現できていれば、これら2つの指標は厳密にゼロになる。

表 4.2 は M_E , M_ρ の結果を示しており、対称性の考慮による MTL モデルの改善は見られないことがわかる。これは ResNet18 が十分な表現能力を持っており、データセットも対称性を学習するのに十分対称であったことを示唆している。実際2つの評価関数を MAE を比較すると、評価関数の方がエラーよりも小さいため、エラーの範囲で対称性が正しく再現できていることがわかる。本研究ではむしろモデルパラメータの初期値の方が結果に影響を与える。一方で、改良された ResNet18 と data augmentation を用いて密度を入力、binding energy を出力としたエネルギー推定タスクを行なったところ、MAE は 52 keV に減少した。したがって、data augmentation 自体は精度の向上に役立つが、我々の MTL モデルのような敵対的な振る舞いをするタスクに対しては、必ずしも寄与する必要はないと結論づけられる。

4.6 まとめ

4.6.1 結果

本研究では multi-task learning (MTL) を ^{236}U の基底状態近傍の変形状態に適用することで、密度とエネルギーの common feature representation の抽出を行なった。学習データの生成のた

め, ランダムポテンシャルのもとでの Skyrme Kohn-Sham Density Functional Theory (Skyrme KS-DFT) の計算を行った. 我々の MTL モデルでは入力された密度は潜在変数に圧縮され, デコーダーにより潜在変数から入力密度の復元とエネルギーの推定が行われる. 多様体仮説と集団座標の仮定の類似性から, 得られた潜在変数は集団座標の一種であると考えられる. 我々の研究はテストデータのエネルギーと密度が, たった 2 次元程度の潜在変数から同時に再現できることを明らかにした. その一方で伝統的な多重極モーメントは, 少なくとも我々のデータセットに対して, ほとんど情報を持っていないことも示された. 以上の結果は, 集団座標をダイナミクスを適切に反映させて選ぶことの重要性を表している.

4.6.2 展望

本研究では, 潜在変数の抽出という目的で autoencoder を用いた. そこで MTL のデコーダーを生成モデルとして利用したい場合は, Variational Autoencoder (VAE) [35, 36] といった, 確率的モデルを採用すべきである. 画像生成において汎化性能を獲得し, 訓練データに含まれない新規な画像を生成することは容易ではない. この分野では, Generative Adversarial Networks (GANs) [77, 78] や, 最近の研究では, Diffusion Models (DMs) [79, 80, 81] が成功を収めてきた. 特に, DM はテキストから画像を生成するタスクにおいて state-of-the-art (SOTA) を達成している. したがって, このような最新の生成モデルを応用することで, 我々の MTL モデルを改良できると考えられる.

核分裂を一貫して記述する潜在変数を抽出するためには, 本研究で用いたものより大きな変形状態のデータを, 数多く収集する必要がある. この時, 潜在次元は 2 から増大する可能性があることに注意が必要である. しかし, 特にバリアトップのような場所で計算に不安定性が生じるため, ランダムポテンシャルに対して大変形した状態を効率よく得るのは難しい. したがって, 人件費をかけることでデータを収集するか, DFT の収束アルゴリズムを工夫する必要がある. また DFT のフレームワークにおいては, 多体の波動関数の情報は利用できないため, MTL から得られた情報を直接 GCM などに結びつけるのは難しい. したがって, 潜在変数に対する慣性質量を与える手法といった, 本研究と集団模型を結びつける新しいアイデアが必要である.

近年の deep learning の著しい発展を支える形で, big data を解析するための優れたツールが数多く登場してきた [22, 23, 24]. その一方で, 原子核理論においてはデータセットは依然として不足している. したがって, ImageNet [70] のような高品質で大規模なデータセットの登場が熱望される. このようなデータに対して, Vision Transformer (ViT) [41] は Conventional Neural Network (CNN) のモデルより優れた性能を示すことが期待される. しかし, big data に見合うだけの十億を超えるようなパラメータを持つ大規模モデルを学習するには, 高価な GPU が複数台必要になる. 原子核物理ではこれまで CPU を用いた計算が主に行われてきたため, 現状でそのような GPU サーバーにアクセスするのは難しい. したがって, 将来的に大規模データを本研究のような手法で解析するためには, GPU の確保は極めて重要な課題であることを強調しておく.

第5章

結論

密度汎関数理論 (Density Functional Theory, DFT) は外場の下での基底状態ダイナミクスが密度のみによって記述可能であるということを主張した理論で, 量子多体系の極めて困難な数値計算を, ダイナミクスの自由度を大幅に減らすことにより実現可能にする. 実際の計算では粒子数密度以外の局所密度量に依存した Kohn-Sham 型のエネルギー密度汎関数 (KS-EDF) が用いられ, 数値解法には平均場理論と同じアルゴリズムが用いられる. 原子核理論における DFT では KS-EDF に分類される Skyrme EDF が用いられており, 原子核構造研究において多大な成功をもたらした. その一方で Skyrme EDF が本来の密度汎関数理論が主張するような密度のみに依存した汎関数 (Orbital-free EDF) と等価であるかは, 自明ではない. これは原子核理論における DFT が, 簡易的に定義はされているものの, 数学的に厳密に定義されているわけではないからである. また, KS-EDF に基づく計算には粒子数の冪乗で増大するような計算コストがかかるため, 重い核の解析には不利である.

KS-EDF では外場を入力とし, 最終的な出力としてその外場の元での基底状態の密度とエネルギーを得る. しかし電子系とは異なり, 原子核物理において外場は多くの場合単なる数学的な道具であり, 外場の選び方に人間のバイアスが生じる. 伝統的に原子核の変形状態の解析に用いられてきた多重極モーメントはその際たる例であり, 特に重い核において, それらが適切である保証はない. 原子核の重要な変形モードが何かという問いは, 核分裂の記述において重要となるため, ダイナミクスに根差した議論が行われるべきである. しかし, KS-EDF において我々ができることは外場を入力することであり, バイアスから逃れることは難しい. この“外場の呪い”とも言うべきバイアスにより, 原子核理論では長らくの間, 多重極モーメントのような仮定された変形モードが標準的に用いられてきた.

以上で述べた OF-EDF の実在性と KS-EDF の計算コストの問題, 及び仮定された変形モードの問題は未解決問題である. 本研究では, 深層学習を用いた新しいアプローチにより, これらの問題解決に向けて取り組んだ.

まず, 第3章で Skyrme EDF を再現するような OF-EDF を得るために, 電子系で開発された手法 [39] を原子核系に応用した. つまりランダムに生成されたポテンシャルのもとで Skyrme Kohn-Sham DFT を解き, 対応する密度と binding energy からなるデータセットを生成した. ただし, surface type の DDDI 汎関数を用いることで, pairing の効果も考慮している. Coulomb term

を無視した上で、系として軸対称性と時間反転対称性を要請された ^{24}Mg で解くことにより、得られた密度はグレースケールの画像と等価な形のデータとして与えられる。これらを用いて、10 層の隠れ層を持った Multi-layer perceptron (MLP) により密度を入力、binding energy を出力とする教師あり学習を行ったところ、テストデータに対する推定誤差は 43 keV であった。これらは核分裂への適用を考える上で無視できるほど小さい。このことは、リーズナブルなエラーの範囲内で KS-EDF を再現するような OF-EDF が、少なくとも我々のデータセットのドメインに対して存在することを意味している。同時に外場から binding energy や密度を推定するタスクに関して教師あり学習を行った結果、それぞれ現象、粒子数に比べて十分小さなエラーで推定することに成功した。また電子系の先行研究 [39] では調べられていない、原子核をはじめとした引力系で重要とされる pairing energy に関しても、密度の汎関数として調べた。テストデータに対する推定誤差は 157 keV であり、pairing のみに関しても密度から予言できることが明らかになった。計算時間に関して、与えられた密度や外場からエネルギーを推定する時間は従来の方法の 10^5 倍から 10^6 倍高速であった。この超高速なエネルギー推定により、核分裂の最適なポテンシャルエネルギー平面の探索といった未解決問題への応用が期待され、この意味で原子核構造研究を加速させ得る。また、ランダムな調和振動子型の外場から生成された、異なるデータセットに関して学習済みのモデルを適用したところ、推定誤差は 306 keV であった。これは、我々の学習済みモデルが異なるドメインのデータに対して一定の予言能力を持つことを意味している。

Skyrme EDF を再現するような OF-EDF の存在を保証する計算結果に動機付けられ、第 4 章では仮定された変形パラメータの正当性を、Skyrme DFT の枠組みで解析した。軸対称性と時間反転対称性のみを仮定した ^{236}U に対して、基底状態の変形ダイナミクスを調べるため、ランダムポテンシャルを用いて密度と binding energy からなるデータセットを生成し、multi-task learning (MTL) により潜在変数抽出を行った。我々の MTL モデルでは、エンコーダーにより密度は潜在的な特徴量に圧縮され、デコーダーは入力された密度自身の復元とエネルギーの推定タスクを行う。この時、得られた潜在変数は ^{236}U のダイナミクスの情報を豊富に含んだ量であり、多様体仮説と集団座標の仮定の類似性から、一種の集団座標であることが期待される。我々の MTL モデルは、潜在変数の次元が 2 の場合であっても、112 keV の誤差でエネルギーを推定し、密度も autoencoder の結果と遜色ない程正確に復元することに成功した。一方で、伝統的な変形パラメータである多重極モーメントからエネルギーを推定する教師あり学習を行ったところ、2 MeV 程度の非常に大きいエラーが得られた。つまり、少なくとも我々のデータセットに対して、多重極モーメントはダイナミクスの情報をほとんど持っていない。以上の結果により、ダイナミクスに根差した集団座標導出の重要性は、ますます強調されたといえる。

本論文では data-driven アプローチに基づく解析により、Skyrme EDF を再現する OF-EDF の存在や計算速度の大幅な加速を示し、そして仮定された変形モードの問題点の定量的な評価とダイナミクスに根差した集団座標を与える新しいアイデアを与えた。本研究の意義は上記の結果のみならず、ランダムポテンシャルという人間のバイアスの影響を受けにくい手法で生成されたデータに基づく deep learning の解析の有用性を示した点にも存在する。deep learning においてはデータが主役であり、その質と量が結果の良し悪しの大半を決めるといっても過言ではない。しかし原子核物理において deep learning はまだ黎明期であり、データは依然として不足している。本論文で示

した結果は、原子核コミュニティにおいてデータ生成を促すものになっている。したがって今後本研究に動機付けされ、様々な原子核に対して人間のバイアスを極力排除したデータセットが生成されることが期待される。これらのデータを統合することで、十分な量のデータが集まれば、様々な原子核を同時に記述する OF-EDF の開発なども可能になる。さらに学習済みの OF-EDF と実験データを組み合わせ、転移学習のテクニックを用いることで、新しい OF-EDF を与えることもできる。また、集まったデータを deep learning の多様な解析手法により調べることで、多重極モーメントといった人間のバイアスが入ったパラメータでは捉えることのできない、原子核の新しい姿が見えてくると考えられる。本論文で示した研究は、このような人間のバイアスを排除した data-driven アプローチに基づく原子核物理の新展開を誘発し、そして先駆ける内容である。

Appendix A

Artificial Neural Network

A.1 Convolutional neural network

第2章で解説した Multi-layer perceptron (MLP) は、典型的には入力ベクトルのサイズの自乗に比例してパラメータ数が増大するため、Full HD のような大きい画像を取り扱うには不適切である。一方で、画像のような空間的な構造があるデータを扱う場合、局所的な特徴量に着目するというアイデアが考えられる。また、画像のある場所で局所的な特徴を捉えるように学習されたフィルターを考えると、別の場所でも役に立つことが期待される。以上のようなアイデアを反映させ、大きい画像に対してもパラメータ数がそれほど多くなならない ANN が畳み込みニューラルネットワーク (Convolutional Neural Network, CNN) である。

CNN では画像を主に扱うため、入力ベクトルを $\mathbf{x} \in \mathbb{R}^{C_{\text{in}} \times H_{\text{in}} \times W_{\text{in}}}$ というテンソル形式で与えるとする。ただし、 $H_{\text{in}}, W_{\text{in}}$ はそれぞれ入力画像の縦、横のピクセルの数であり、 C_{in} は channel 数である*1 この時、2次元の convolutional layer の基本形は以下で与えられる：

$$y_{c', h', w'} := b_{c'} + \sum_{c=1}^{C_{\text{in}}} \sum_{k_h=1}^{K_h} \sum_{k_w=1}^{K_w} W_{c', c; k_h, k_w} x_{c, S_h \times (h' - 1) + k_h, S_w \times (w' - 1) + k_w}. \quad (\text{A.1})$$

ただし、 $W \in \mathbb{R}^{C_{\text{out}} \times C_{\text{in}} \times K_h \times K_w}$, $\mathbf{b} \in \mathbb{R}^{C_{\text{out}}}$ は CNN のモデルパラメータであり、それぞれ重み行列とバイアス項である。一方で、出力ベクトルは $\mathbf{y} \in \mathbb{R}^{C_{\text{out}} \times H_{\text{out}} \times W_{\text{out}}}$ で与えられるテンソルである。ここで、整数 $1 \leq K_h \leq H_{\text{in}}, 1 \leq K_w \leq W_{\text{in}}$ は kernel size, 整数 $S_h, S_w \geq 1$ は stride と呼ばれるハイパーパラメータある。この時、和が \mathbf{x} の添え字の範囲を越えないという条件

$$S_h(h' - 1) + K_h \leq H_{\text{in}}, \quad (\text{A.2})$$

$$S_w(w' - 1) + K_w \leq W_{\text{in}}, \quad (\text{A.3})$$

*1 フルカラー画像の場合は $C_{\text{in}} = 3$ であるが、CNN ではこの数を隠れ層を順伝播する過程で変化させるため、一般には正の整数値を考える必要がある。

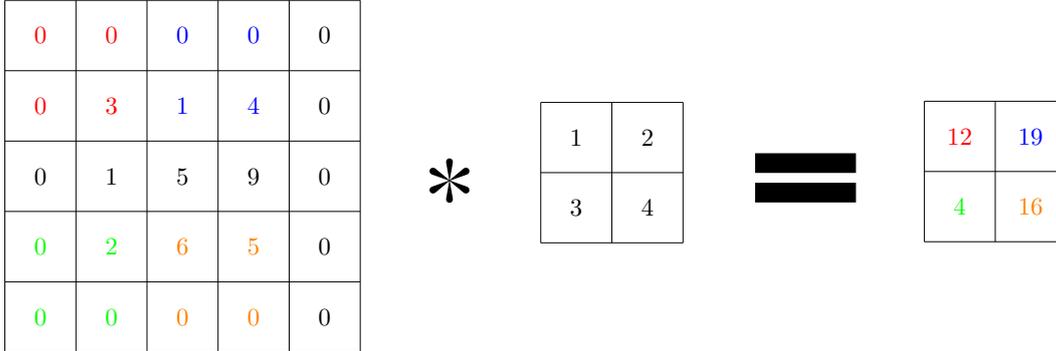


図 A.1 CNN の具体例. 入力は $C = 1, H = W = 5$ の画像であり, カーネルサイズは $C' = 1, K_h = H_w = 1$. stride に関しては $S_h = 3, S_w = 2$ と縦横で異なるものを使い, バイアス項は 0 とした. 各色は図の左右で対応している.

から出力と入力の次元の間関係が与えられる. したがって, Gauss 記号 $[\cdot]$ を用いることにより,

$$H_{\text{out}} = \left[\frac{H_{\text{in}} - K_h}{S_h} + 1 \right], \quad (\text{A.4})$$

$$W_{\text{out}} = \left[\frac{W_{\text{in}} - K_w}{S_w} + 1 \right], \quad (\text{A.5})$$

が得られる. 計算の具体例を図 A.1 に示した.

定義から明らかなように, convolutional layer は fully-connected layer の特別な場合である. 具体的には重み関数の作用する範囲の制限と, 異なるニューロン間におけるバイアス項と重み関数の共有を行なった fully-connected layer である. したがって, MLP の場合と同様に backpropagation を用いることができる.

上記の convolutional layer は, 画像の端の情報に関して参照される回数が少ない. そこで, 入力画像の周囲を何らかの方法で埋め, 画像を大きくすることで, 端の情報を多く参照可能にするという操作がしばしば行われる. これは padding と呼ばれ, 通常は画像の四方をゼロで埋める zero padding が用いられる. 上下方向, 左右方向はそれぞれ同じだけ画像サイズを増やすため, 上下方向の padding size を P_h , 左右方向の padding size を P_w とすると, 有効的に $W_{\text{in}} \rightarrow W_{\text{in}} + 2P_w$, $H_{\text{in}} \rightarrow H_{\text{in}} + 2P_h$ というサイズの画像を padding なしで入力した場合とみなすことができる. したがって, 入力次元と出力次元の間関係は

$$H_{\text{out}} = \left[\frac{H_{\text{in}} + 2P_h - K_h}{S_h} + 1 \right], \quad (\text{A.6})$$

$$W_{\text{out}} = \left[\frac{W_{\text{in}} + 2P_w - K_w}{S_w} + 1 \right], \quad (\text{A.7})$$

となる. padding の具体例は図 A.2 に示した.

最後に dilation という, 広範囲のセルの影響を取り入れるための操作について述べる. dilation を扱った convolution は dilated convolution と呼ばれ, kernel を単純に作用させるのではなく, いくつかのセルをスキップしながら作用させる. つまり, 図 A.3 のようにカーネルに 0 を入れて拡大

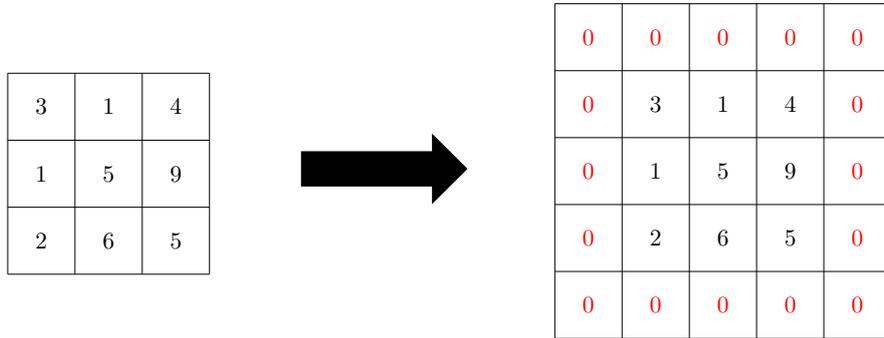


図 A.2 padding の具体例. 入力は $C = 1, H = W = 3$ の画像に対して, $P_w = P_h = 1$ の zero padding を行なっている. 各色は図の左右で対応している.

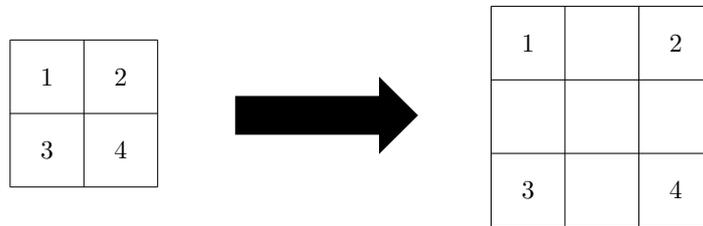


図 A.3 dilation の具体例. サイズが $C' = 1, K_h = K_w = 2$ のカーネルに対して, $D_h = D_w = 2$ の dilation を行なうと, 図の右側のカーネルを使う操作と等価である. ただし, 何も入力していないセルに関しては, 画像に作用させる際に計算を行わないものとする.

し, その新しいカーネルで convolution を行う. そこで上下方向, 左右方向の dilation size をそれぞれ D_h, D_w とすると, 有効的にカーネルサイズを $K_w \rightarrow K_w D_w - D_w + 1, K_h \rightarrow K_h D_h - D_h + 1$ と変更した場合とみなせるため,

$$H_{\text{out}} = \left\lceil \frac{H_{\text{in}} + 2P_h - D_h(K_h - 1) - 1}{S_h} + 1 \right\rceil, \quad (\text{A.8})$$

$$W_{\text{out}} = \left\lceil \frac{W_{\text{in}} + 2P_w - D_w(K_w - 1) - 1}{S_w} + 1 \right\rceil, \quad (\text{A.9})$$

という出力の次元を得る*2

*2 この他にも grouped convolution layer [15] など存在するが, ここでは省略する.

A.2 Vision transformer

A.2.1 Transformer

2017年, Google Reserch と Google Brain により “Attention Is All You Need” という論文が発表され, 機械学習業界にブレイクスルーが起きた [16]. これはタイトル通り, Attention 機構のみを用いて自然言語処理を実行するというアイデアを提示した論文である. Attention 機構とは, 自然言語処理分野でこの論文以前に存在した, 単語間の重要度を測るような操作である. この Attention 機構から構成されたネットワークは原論文 [16] では Transformer と呼ばれ, 現在の大規模言語モデル (Large Language Model, LLM) で標準的に用いられている*3.

Transformer のアーキテクチャーは図 A.4 に示されている. これらのうち, mutli-head attention を説明するため, まず scaled dot-product attention について説明する. これは, 次のような線形変換と softmax 関数を組み合わせた操作である:

$$\text{Attention}(Q, K, V) := \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \quad (\text{A.10})$$

ただし, Q, K, V はそれぞれ query, key, value と呼ばれる量で, $Q, K \in \mathbb{R}^{n \times d_k}$, $V \in \mathbb{R}^{n \times d_v}$ という行列で与えられるものとする. つまり, $QK^T \in \mathbb{R}^{n \times d_n}$ を $\sqrt{d_k}$ でスケールさせ, QK^T の右側の足に関して softmax を作用させ正規化を行なった後, V と行列積を取る. したがって, query と key で重みを作ることで, value のどの値に注意を向けるかを定めるモジュールと解釈される. 入力文章であれば, 文脈を捉えるような操作に相当する.

ここで原論文に従い $Q, K, V \in \mathbb{R}^{n \times d_{\text{model}}}$ と変数を定義し直す. この時, scaled dot-product attention を用いることで multi-head attention は次のように定義される:

$$\text{MultiHead}(Q, K, V) := \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O. \quad (\text{A.11})$$

ただし,

$$\text{head}_i := \text{Attention}(QW_i^Q, KW_i^K, VW_i^V), \quad (\text{A.12})$$

である. 各変数は, $W_i^Q, W_i^K \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{\text{model}} \times d_v}$, $W^O \in \mathbb{R}^{hd_v \times d_{\text{model}}}$ で定義されるパラメータ行列である. また, Concat は行列を列の方向に積み重ねる操作を表す. 通常 $d_k = d_v = d_{\text{model}}/h$ となるように取られる. ただし d_k, d_v が整数になるように h が選ばれることに注意が必要である. このように, 複数の head を用いて異なる attention を同時に行うことで, 多様な注意の向け方を考慮できる. 入力文章であれば, 多様な文脈を同時に捉えるような操作を可能にする.

*3 ChatGPT [17] の GPT は Generative Pre-trained Transformer の略である.

図 A.4 の Transformer ブロックは以下の一連の流れとして定義される。

1. multi-head attention
2. residual connection [68]
3. layer normalization [82]
4. Feed-Forward (原論文では fully-connected layer を使用)
5. residual connection
6. layer normalization

ここで、現代的には layer normalization の位置が異なっていたり、Feed-Forward として Gated Linear Unit (GLU) [83] といった別のモジュールが用いられるなどの違いがある。

A.2.2 ViT

Transformer を画像認識タスクに応用し、CNN 系の SOTA モデルを超えたのが Vision Transformer (ViT) [41] である。画像をパッチに分割し、そのパッチをトークンとみなしてベクトル空間に埋め込むことで、自然言語処理の場合と同様に Transformer ブロックを用いることができる。ViT ではこのアイデアに基づき、モデルが設計されている (図 A.5 を参照)。以下では ViT の計算を入力から順番に解説してゆく。

入力として画像 $\mathbf{x} \in \mathbb{R}^{C \times H \times W}$ を考える。この画像を $P_h \times P_w$ のパッチに分割し、それぞれのパッチを

$$\mathbf{x}_p^1, \mathbf{x}_p^2, \dots, \mathbf{x}_p^N \in \mathbb{R}^{C P_h P_w}, \quad (\text{A.13})$$

というベクトル量で表す。ただし、 $N := HW/(P_h P_w)$ が整数になるような画像サイズやパッチサイズを選ぶものとする。これらの量を単一のパラメータ行列 $E \in \mathbb{R}^{C P_h P_w \times D}$ を用いることで、 D 次元のベクトルに埋め込む。ここで、パッチ間の位置関係の情報を保持するため、position embedding により $E_{\text{pos}} \in \mathbb{R}^{(N+1) \times D}$ を導入する。ただし、これは学習可能なパラメータとして用意される。また、class token と呼ばれるベクトルを $\mathbf{x}_{\text{class}} \in \mathbb{R}^D$ により用意し、学習可能なパラメータとする。これらを用いて

$$\mathbf{z}^{\text{in}} := (\mathbf{x}_{\text{class}}, \mathbf{x}_p^1 E, \mathbf{x}_p^2 E, \dots, \mathbf{x}_p^N E)^T + E \in \mathbb{R}^{(N+1) \times D}, \quad (\text{A.14})$$

とまとめることで、Transformer ブロックの入力とする。

ViT では、オリジナルの Transformer ブロック (図 A.4 を参照) とは異なり、先に layer normalization を行ってから multi-head attention を作用させる。また Feed-Forward 部分に関しても、順番が変わっていることに注意が必要である。この Transformer ブロックを L 回繰り返す。その出力として得られる $\mathbf{z}^{\text{out}} \in \mathbb{R}^{(N+1) \times D}$ の最初の行のベクトルを $\mathbf{z}_0^{\text{out}} \in \mathbb{R}^D$ とし、このベクトルのみを 1 層の隠れ層を持つ MLP に受け渡すことで最終的な出力を得る。

以上のアイデアにより、ViT は Noisy Student [84] という当時 ImageNet [70] に対して state-of-the-art (SOTA) を達成していたモデルを打ち破った。ただ、ViT で良い性能を出すには大規模なデータによる事前学習が必要であることに注意が必要である。実際に 2.2 節で行ったような

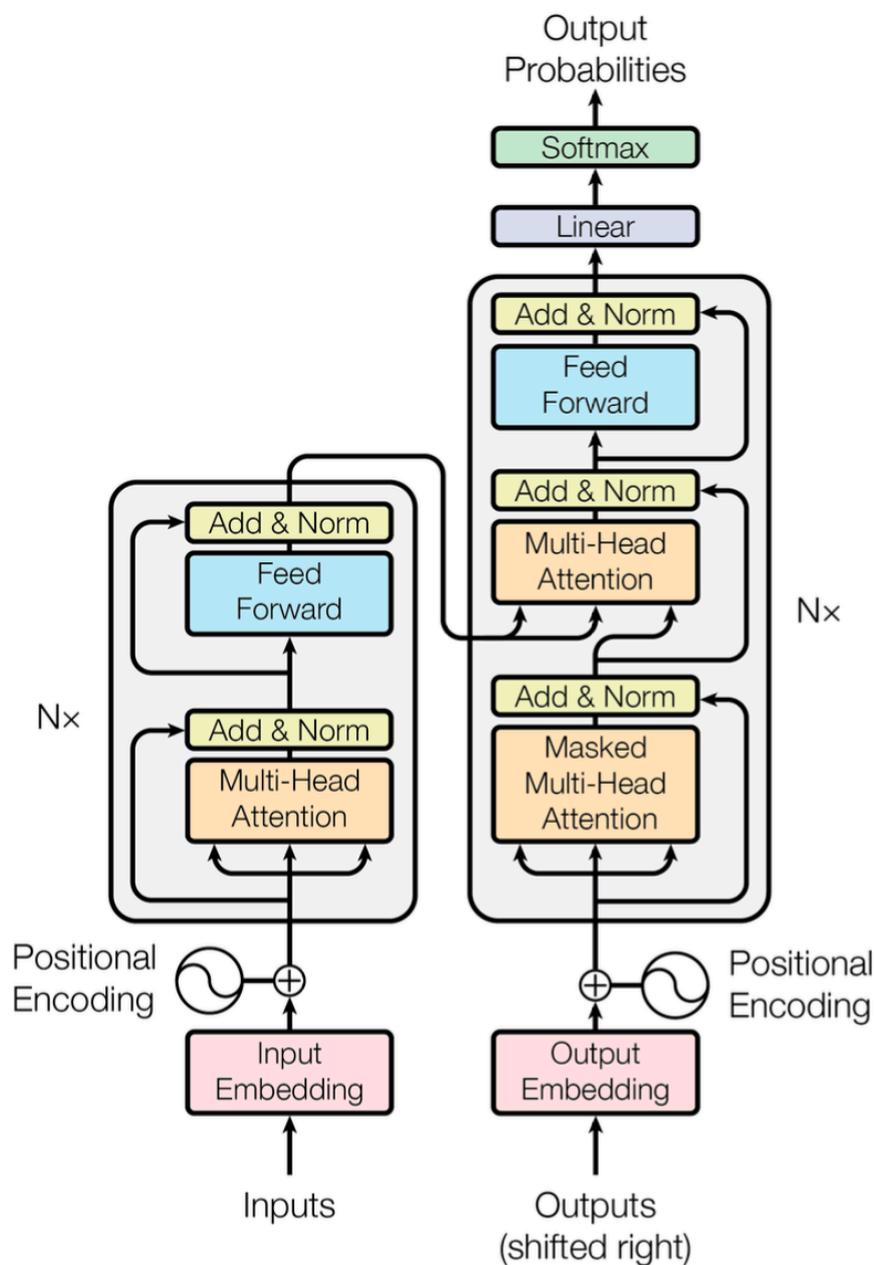


図 A.4 Transformer のモデルアーキテクチャー. 入力左側のグレーの領域で表される Transformer ブロックの繰り返しによりエンコードされ、右側のデコーダーに用いられることで、目的のタスクに必要なとされる確率分布の出力を得る。例えば翻訳タスクでは、日本語の文章を学習済みの Transformer に入力すると、対応する英文の一つ目の単語として最も確率の高いものがデコーダーの出力である確率分布から分かり、その結果を逐次的にデコーダーに入力することで、英語訳した文章全体を得ることができる。Transformer の原論文 [16] から抜粋。

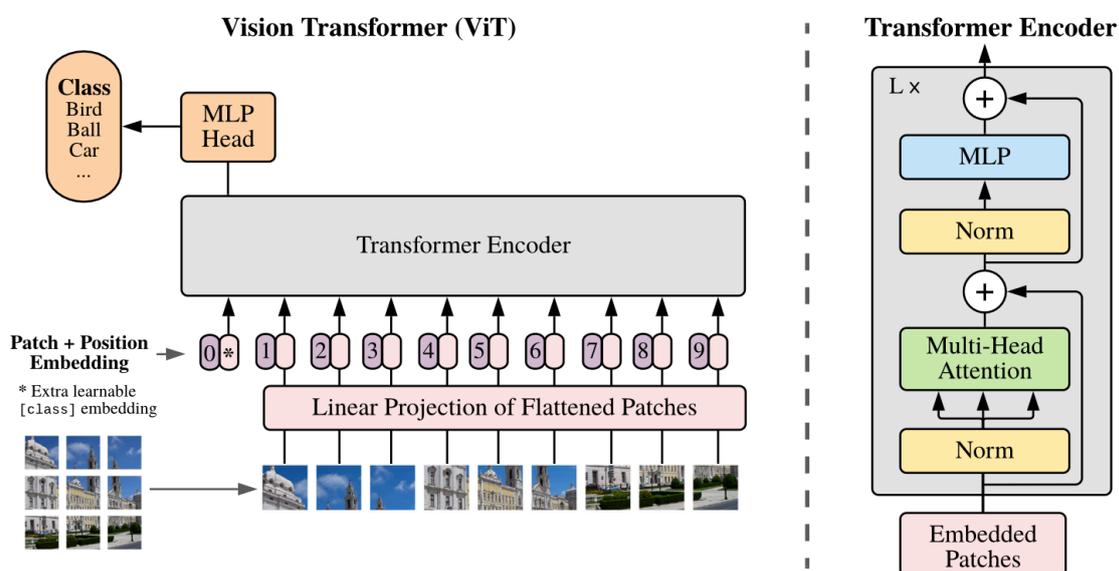


図 A.5 Vision Transformer のアーキテクチャ。入力画像はパッチに分割され、それぞれベクトル空間に埋め込まれることで、Transformer ブロックが使用可能になる。ただし、パッチ間の配置の情報を入れ込むため position embedding を行う。Transformer ブロックを L 回繰り返す、最終的な出力の一部に MLP を作用させ、出力層に渡すことで多クラス分類タスクなどを行う。

MNIST に対して単純に ViT を用いると、CNN 系のモデルよりも悪い結果が得られる。

近年 ViT のアイデアは一般化され、MetaFormer という概念が導入された [85]。MetaFormer の主張では、入力ベクトルを layer norm などで正規化した後、multi-head attention のようなトークンを混ぜる操作を行ってから、入力ベクトルに対する skip connection を行うこと自体が重要であり、attention そのものがモデル性能向上に寄与したわけではないという。彼らはその考えに基づき、PoolFormer という attention の代わりに pooling layer を利用したモデルで計算を行い、ViT と遜色のない性能を得ている。

Appendix B

軸対称系の Skyrme-Hartree-Fock-BCS

B.1 Kohn-Sham 方程式

B.1.1 時間反転対称性

軸対称性と時間反転対称性を要請した場合の Kohn-Sham 方程式に関して、具体的な数値計算を再現する上で必要な方程式について述べる。以下では、BCS ansatz が正当化されるために必要な時間反転対称性のみを要請する。最小化問題 (1.25) は変分により解くことができる。しかし、BCS ansatz のもとの Kohn-Sham 軌道やその占有率には

$$\int d^3r \varphi_{q,k}^*(\mathbf{r}) \varphi_{q,k'}(\mathbf{r}) = \delta_{k,k'}, \quad (\text{B.1})$$

$$\int d^3r, \rho_q(\mathbf{r}) = \sum_{k>0} v_{q,k}^2 = N_q, \quad (\text{B.2})$$

という条件が必要である。ただし、陽子数を N_p 、中性子数を N_n とした。したがって KS-DFT を解くには、拘束条件付き変分問題を扱う必要がある：

$$\delta_{v_{q,k}^2} \left\{ E_{\text{Skyrme}}[\rho_q, \tau_q, \mathbf{J}_q] + E_{\text{DDDI}}[\rho, \tilde{\rho}_q] + \sum_{q=p,n} \int d^3r v_q(\mathbf{r}) \rho_q(\mathbf{r}) - F \right\} = 0, \quad (\text{B.3})$$

$$\delta_{\varphi_{q,k}^*} \left\{ E_{\text{Skyrme}}[\rho_q, \tau_q, \mathbf{J}_q] + E_{\text{DDDI}}[\rho, \tilde{\rho}_q] + \sum_{q=p,n} \int d^3r v_q(\mathbf{r}) \rho_q(\mathbf{r}) - F \right\} = 0. \quad (\text{B.4})$$

ただし、

$$F = \sum_k \tilde{\epsilon}_{q,k} \left(\int d^3r \varphi_{q,k}^*(\mathbf{r}) \varphi_{q,k}(\mathbf{r}) - 1 \right) - \sum_{q=n,p} \lambda_q \left(\sum_{k>0} v_{q,k}^2 - N_q \right), \quad (\text{B.5})$$

である。ここで、後ほど分かるように、Kohn-Sham 軌道の直交性は自動的に課すことができるため、議論を簡単化するため、この段階で規格化のみ拘束条件として考慮する。(B.3) 式の変分問題は、

$$v_{q,k}^2 h_q \varphi_{q,k} - v_{q,k}^2 \frac{V_q \gamma}{4\rho_0} \left(\frac{\rho(\mathbf{r})}{\rho_0} \right)^{\gamma-1} \tilde{\rho}_q(\mathbf{r})^2 \varphi_{q,k} + v_{q,k}^2 v_q(\mathbf{r}) \varphi_{q,k} = \tilde{\epsilon}_{q,k} \varphi_{q,k}, \quad (\text{B.6})$$

で与えられる。ただし,

$$h_q := -\nabla \cdot \frac{\hbar^2}{2m^*(\mathbf{r})} \nabla + U_q(\mathbf{r}) - i\mathbf{B}_q(\mathbf{r}) \cdot (\nabla \times \boldsymbol{\sigma}) \quad (\text{B.7})$$

$$\frac{\hbar^2}{2m^*(\mathbf{r})} = \frac{\hbar^2}{2m} \left(1 - \frac{1}{A}\right) + b_1\rho - b'_1\rho_q, \quad (\text{B.8})$$

$$U_q(\mathbf{r}) := b_0\rho - b'_0\rho_q + b_1\tau - b'_1\tau_q \frac{b_3}{3}(\alpha+2)\rho^{\alpha+1} - \frac{b'_3}{3} \left\{ \alpha\rho^{\alpha-1} \sum_q \rho_q^2 + 2\rho^\alpha \rho_q \right\} \\ - b_4\nabla \cdot \mathbf{J} - b'_4\nabla \cdot \mathbf{J}_q + b'_2\nabla^2\rho_q - b_2\nabla^2\rho, \quad (\text{B.9})$$

$$\mathbf{B}_q := b'_1\mathbf{J}_q + b_4\nabla\rho + b'_4\rho_q, \quad (\text{B.10})$$

である。ここで, $\tilde{\epsilon}_{q,k} := v_{q,k}^2\epsilon_{q,k}$ と Lagrange の未定乗数をスケール変換することで, (B.7) 式は pairing のない場合の Kohn-Sham 方程式と等価になる。また, 固有値問題の形式になっているため, 各 Kohn-Sham 軌道の直交性は保証される。実際の計算では, しばしば (B.7) 式左辺第 2 項と, (B.10) 式右辺第 1 項は無視される [66, 86]。本論文でもこれらの項は無視している。

一方, $v_{q,k}^2$ に関する変分 (B.4) 式は, (B.6) 式を用いることで,

$$\epsilon_{q,k} - \frac{1 - 2v_{q,k}^2}{2\sqrt{v_{q,k}^2(1 - v_{q,k}^2)}} \Delta_{q,k} = \lambda_q \quad (\text{B.11})$$

と計算される。ただし,

$$\Delta_{q,k} := \frac{V_q}{2} \int d^3r \left\{ 1 - \left(\frac{\rho}{\rho_0} \right)^\gamma \right\} \tilde{\rho} |\varphi_{q,k}|^2, \quad (\text{B.12})$$

である。以上の式を $v_{q,k}^2$ について解くことで

$$v_{q,k}^2 = \frac{1}{2} \left(1 - \frac{\epsilon_{q,k} - \lambda_q}{\sqrt{(\epsilon_{q,k} - \lambda_q)^2 + \Delta_{q,k}^2}} \right), \quad (\text{B.13})$$

を得る。ただし拘束条件から,

$$\frac{1}{2} \sum_{k>0} \left(1 - \frac{\epsilon_{q,k} - \lambda_q}{\sqrt{(\epsilon_{q,k} - \lambda_q)^2 + \Delta_{q,k}^2}} \right) = N_q, \quad (\text{B.14})$$

となる。

B.1.2 軸対称性

軸対称性がある場合, つまり 1 粒子 Hamiltonian が z 方向の全角運動量演算子 \hat{J}_z と可換になる場合を考える。すると Kohn-Sham 軌道は \hat{J}_z の固有状態となるので,

$$\varphi_{q,k\Omega,\Omega}(r, z, \phi) = \frac{1}{\sqrt{2\pi}} \begin{pmatrix} e^{i(\Omega-1/2)\phi} \varphi_{q,k\Omega,\Omega}^-(r, z) \\ e^{i(\Omega+1/2)\phi} \varphi_{q,k\Omega,\Omega}^+(r, z) \end{pmatrix} \quad (\text{B.15})$$

と書くことができる。ただし、 \hat{J}_z の固有値を $\hbar\Omega$, $\Omega \in \mathbb{Z} + \frac{1}{2}$ とし、Kohn-Sham 軌道の量子数を $k = (k_\Omega, \Omega)$ と書いた。この時、 $\rho_q, \tau_q, \nabla \cdot \mathbf{J}_q, \tilde{\rho}_q$ が r, z のみに依存していることがわかる^{*1}:

$$\rho_q(r, z) = \frac{1}{\pi} \sum_{\Omega > 0} \sum_{k_\Omega} v_K^2 \sum_{s=\pm} |\varphi_K^s|^2, \quad (\text{B.16})$$

$$\tau_q(r, z) = \frac{1}{\pi} \sum_{\Omega > 0} \sum_{k_\Omega} v_K^2 \sum_{s=\pm} (\Omega_s^2 |\varphi_K^s|^2 + |\partial_r \varphi_K^s|^2 + |\partial_z \varphi_K^s|^2), \quad (\text{B.17})$$

$$\nabla \cdot \mathbf{J}_q = -\frac{2}{\pi} \sum_{\Omega > 0} \sum_{k_\Omega} v_K^2 \sum_{s=\pm} s \{ \partial_r \varphi_K^s \partial_z \varphi_K^{-s} + \Omega_s \varphi_K^s (\partial_r \varphi_K^s + s \partial_z \varphi_K^{-s}) \}, \quad (\text{B.18})$$

$$\tilde{\rho}_q(r, z) = -\frac{1}{\pi} \sum_{\Omega > 0} \sum_{k_\Omega} u_K v_K \sum_{s=\pm} |\varphi_K^s|^2. \quad (\text{B.19})$$

ただし、量子数を $K = (q, k_\Omega, \Omega)$ と略記した。また、

$$\Omega_\pm := \frac{\Omega \pm 1/2}{r}, \quad (\text{B.20})$$

である。軸対称な場合の Kohn-Sham 軌道と local density の表式を (B.6) 式に代入することで、

$$\begin{pmatrix} t_- + U_q + B_q^r \Omega_- + v_q & -B_q^z \Omega_+ - B_q^z \partial_r + B_q^r \partial_z \\ -B_q^z \Omega_- + B_q^z \partial_r - B_q^r \partial_z & t_+ + U_q - B_q^r \Omega_+ + v_q \end{pmatrix} \begin{pmatrix} \varphi_K^- \\ \varphi_K^+ \end{pmatrix} = \epsilon_K \begin{pmatrix} \varphi_K^- \\ \varphi_K^+ \end{pmatrix}, \quad (\text{B.21})$$

を得る。ただし、

$$t_\pm := \frac{\hbar^2}{2m^*} \left\{ \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} - \Omega_\pm^2 + \frac{\partial^2}{\partial z^2} \right\} + \left(\frac{\partial f}{\partial r} \frac{\hbar^2}{2m^*} \right) \frac{\partial}{\partial r} + \left(\frac{\partial f}{\partial z} \frac{\hbar^2}{2m^*} \right) \frac{\partial}{\partial z}, \quad (\text{B.22})$$

$$B_q^r := \frac{\partial}{\partial r} (b_4 \rho + b'_4 \rho_q), \quad (\text{B.23})$$

$$B_q^z := \frac{\partial}{\partial z} (b_4 \rho + b'_4 \rho_q), \quad (\text{B.24})$$

である。

B.2 数値計算アルゴリズム

実際の計算では、量子数の和は有限で打ち切られる:

$$\sum_{\Omega > 0} \sum_{k_\Omega} \rightarrow \sum_{\Omega > 0}^{\Omega_{\max}} \sum_{k_\Omega, \epsilon_K < \epsilon_{\text{cut}}} . \quad (\text{B.25})$$

^{*1} 軸対称性を課するという条件に関して、1 粒子 Hamiltonian から出発するのではなく、この local density に対する条件から始めても同じ結果を得ることができる。

ただし, Ω_{\max} は角運動量量子数の最大値であり, ϵ_{cut} は Kohn-Sham の軌道エネルギーに対するカットオフである. また, 解くべき方程式を以下のように表示する:

$$h_{q,\Omega}\varphi_{q,\Omega,k\Omega}(r,z) = \epsilon_{q,\Omega,k\Omega}\varphi_{q,\Omega,k\Omega}(r,z), \quad (\text{B.26})$$

$$\Delta_{q,\Omega,k\Omega} = \int d^3r \Delta_q(\mathbf{r})|\varphi_{q,\Omega,k\Omega}|^2 \quad (\text{B.27})$$

$$v_{q,\Omega,k\Omega}^2 = \frac{1}{2} \left(1 - \frac{\epsilon_{q,\Omega,k\Omega} - \lambda_q}{\sqrt{(\epsilon_{q,\Omega,k\Omega} - \lambda_q)^2 + \Delta_{q,\Omega,k\Omega}^2}} \right). \quad (\text{B.28})$$

ただし,

$$h_{q,\Omega} := \begin{pmatrix} t_- + U_q + B_q^r \Omega_- + v_q & -B_q^z \Omega_+ - B_q^z \partial_r + B_q^r \partial_z \\ -B_q^z \Omega_+ + B_q^z \partial_r - B_q^r \partial_z & t_+ + U_q - B_q^r \Omega_+ + v_q \end{pmatrix}, \quad (\text{B.29})$$

$$\Delta_q(\mathbf{r}) := \frac{V_q}{2} \int d^3r \left\{ 1 - \left(\frac{\rho}{\rho_0} \right)^\gamma \right\} \tilde{\rho}, \quad (\text{B.30})$$

$$\varphi_{q,\Omega,k\Omega}(r,z) := \begin{pmatrix} \varphi_{q,\Omega,k\Omega}^-(r,z) \\ \varphi_{q,\Omega,k\Omega}^+(r,z) \end{pmatrix} \quad (\text{B.31})$$

とする. ここで, 方程式が Kohn-Sham 軌道と占有率に対して非線形であるため, iterative に解かなければならない. この時, $h_{q,\Omega}$ や Δ_q を iteration の過程で適切に更新する必要がある. 最も単純な方法は simple mixing で, これは一つ前の iteration で計算した $h_{q,\Omega}^{\text{old}}, \Delta_q^{\text{old}}$ と現在の iteration で求めた $h_{q,\Omega}, \Delta_q$ を一定の割合で混ぜ合わせる手法である. ただし, $h_{q,\Omega}$ の初期条件として, 典型的には Woods – Saxon ポテンシャルや変形調和振動子ポテンシャルが用いられる. 一方で Δ_q の初期値には $\Delta_q = 12/\sqrt{A}$ が用いられる. この更新則に従い, binding energy や平均場ポテンシャルが収束するまで計算を続ける. 以上の軸対称系における Skyrme Kohn-Sham DFT のアルゴリズムは Algorithm 1 にまとめている. ただし, 本研究における計算では iteration の収束がより速い modified Broyden algorithm [87] を用いている. しかし, 研究の再現という観点では本質的ではないため, 省略する.

Algorithm 1 軸対称系における Kohn-Sham BCS スキーム**Require:** $b_{0-4}, b'_{0-4}, \alpha, \gamma, \rho_0, V_q$: Model parameters**Require:** $\Omega_{\max}, c_{\text{cut}}$: cut-off parameters**Require:** $\beta_{\text{mix}} \in (0, 1)$: Mixing parameter

```

1:  $h_{q,\Omega}, \Delta_q, \lambda_q \leftarrow$  Initial values
2: repeat
3:   for  $q = n, p$  do
4:     for  $\Omega = 1/2, 3/2, \dots, \Omega_{\max}$  do
5:       Calculate  $h_{q,\Omega} \varphi_{q,\Omega,k_\Omega}(r, z) = \epsilon_{q,\Omega,k_\Omega} \varphi_{q,\Omega,k_\Omega}(r, z)$  ▷ Eq. (B.26)
6:       for all  $k_\Omega$  do
7:         Calculate  $\Delta_{q,\Omega,k_\Omega}$  ▷ Eq. (B.27)
8:       end for
9:     end for
10:    repeat
11:      for all  $\Omega, k_\Omega$  do
12:        if  $\epsilon_{q,\Omega,k_\Omega} < \epsilon_{\text{cut}}$  then
13:          Calculate  $v_{q,\Omega,k_\Omega}^2$  ▷ Eq. (B.28)
14:        else
15:           $v_{q,\Omega,k_\Omega}^2 \leftarrow 0$ 
16:        end if
17:      end for
18:       $\tilde{N}_q \leftarrow 2 \sum_{\Omega>0}^{\Omega_{\max}} \sum_{k_\Omega} v_{q,\Omega,k_\Omega}^2$ 
19:       $\lambda_q \leftarrow \lambda_q + (N_q - \tilde{N}_q) / (dN_q/d\lambda_q)$ 
20:    until  $\tilde{N}_q$  is converged
21:    Calculate  $\rho_q, \tau_q, \nabla \cdot \mathbf{J}_q, \tilde{\rho}_q$  ▷ Eqs. (B.16, B.17, B.18, B.19)
22:  end for
23:  for  $q = n, p$  do
24:    for  $\Omega = 1/2, 3/2, \dots, \Omega_{\max}$  do
25:       $h_{q,\Omega}^{\text{old}} \leftarrow h_{q,\Omega}$ 
26:      Calculate  $h_{q,\Omega}$  ▷ Eq. (B.29)
27:       $h_{q,\Omega} \leftarrow \beta_{\text{mix}} h_{q,\Omega} + (1 - \beta_{\text{mix}}) h_{q,\Omega}^{\text{old}}$  ▷ simple mixing
28:    end for
29:     $\Delta_q^{\text{old}} \leftarrow \Delta_q$ 
30:    Calculate  $\Delta_q$  ▷ Eq. (B.30)
31:     $\Delta_q^{\text{old}} \leftarrow \beta_{\text{mix}} \Delta_q + (1 - \beta_{\text{mix}}) \Delta_q^{\text{old}}$  ▷ simple mixing
32:  end for
33:   $E_{\text{bin}} \leftarrow E_{\text{Skyrme}}[\rho_q, \tau_q, \mathbf{J}_q] + E_{\text{DDDI}}[\rho_q, \tilde{\rho}_q]$ 
34: until  $E_{\text{bin}}$  is converged

```

謝辞

本論文は、著者が京都大大学院理学研究科物理学・宇宙物理学専攻博士後期課程在学中に、原子核理論研究室において行った2編の研究をまとめたものである。これらの研究を遂行するにあたり、指導教員の萩野浩一教授と共同研究者であり RCNP 所属の吉田賢市准教授には多大な支援及びご指導を賜った。特に萩野氏には修士課程と博士課程の計5年間、原子核物理だけでなく、文章の構成や発表の技能も含め、様々なことをご教授いただいた。この場を借りて、深く感謝の意を表す。加えて本論文の執筆にあたり、有用なコメントを頂いた土居孝寛助教授にも深く感謝する。また、学びの機会を与えてくださった原子核理論研究室や核多体ゼミのメンバー、研究活動を支えてくださった秘書室の方々、計算資源を供与してくださった基礎物理学研究所にも感謝の意を伝えたい。最後に、課題番号 22KJ1697 を通じて、博士課程の3年間生活及び研究活動を支援していただいた日本学術振興会にも深く感謝の意を表す。

参考文献

- [1] M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.*, Vol. 378, pp. 686–707, 2019.
- [2] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, Vol. 136, pp. B864–B871, Nov 1964.
- [3] Mel Levy. Electron densities in search of hamiltonians. *Phys. Rev. A*, Vol. 26, pp. 1200–1208, Sep 1982.
- [4] Werner Kutzelnigg. Density functional theory in terms of a legendre transformation for beginners. *J. Mol. Struct.*, Vol. 768, No. 1, pp. 163–173, 2006. Coupled-cluster Methods: Theory and Applications. A Collection of Invited Papers in Honor of Debashis Mukherjee on the Occasion of his 60th Birthday.
- [5] Mel Levy, John P. Perdew, and Virahit Sahni. Exact differential equation for the density and ionization energy of a many-particle system. *Phys. Rev. A*, Vol. 30, pp. 2745–2748, Nov 1984.
- [6] Elliott H. Lieb. Density functionals for coulomb systems. *Int. J. Quantum Chem.*, Vol. 24, No. 3, pp. 243–277, 1983.
- [7] L. H. Thomas. The calculation of atomic fields. *Math. Proc. Camb. Philos. Soc.*, Vol. 23, No. 5, p. 542–548, 1927.
- [8] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, Vol. 140, pp. A1133–A1138, Nov 1965.
- [9] John E. Harriman. Orthonormal orbitals for the representation of an arbitrary density. *Phys. Rev. A*, Vol. 24, pp. 680–682, Aug 1981.
- [10] P. Ring and P. Schuck. *The nuclear many-body problem*. Springer-Verlag, New York, 1980.
- [11] Michael Bender, Paul-Henri Heenen, and Paul-Gerhard Reinhard. Self-consistent mean-field models for nuclear structure. *Rev. Mod. Phys.*, Vol. 75, pp. 121–180, Jan 2003.
- [12] P.-G. Reinhard, D. J. Dean, W. Nazarewicz, J. Dobaczewski, J. A. Maruhn, and M. R. Strayer. Shape coexistence and the effective nucleon-nucleon interaction. *Phys. Rev. C*, Vol. 60, p. 014316, Jun 1999.
- [13] A. Bohr and B.R. Mottelson. *Nuclear Structure*. No. 1 in Nuclear Structure. World

- Scientific, 1998.
- [14] J. Bardeen, L. N. Cooper, and J. R. Schrieffer. Theory of superconductivity. *Phys. Rev.*, Vol. 108, pp. 1175–1204, Dec 1957.
- [15] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, Vol. 25. Curran Associates, Inc., 2012.
- [16] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc., 2017.
- [17] OpenAI. Gpt-4 technical report, 2023.
- [18] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)*, Vol. 2, No. 4, pp. 303–314, December 1989.
- [19] Moshe Leshno, Vladimir Ya. Lin, Allan Pinkus, and Shimon Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, Vol. 6, No. 6, pp. 861–867, 1993.
- [20] Sejun Park, Chulhee Yun, Jaeho Lee, and Jinwoo Shin. Minimum width for universal approximation. In *International Conference on Learning Representations*, 2021.
- [21] Laith Alzubaidi, Jinglan Zhang, Amjad J. Humaidi, Ayad Q. Al-Dujaili, Ye Duan, Omran Al-Shamma, Jesus Santamaría, Mohammed Abdulraheem Fadhel, Muthana Al-Amidie, and Laith Farhan. Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. *Journal of Big Data*, Vol. 8, , 2021.
- [22] PyTorch. <https://pytorch.org>.
- [23] JAX. <https://jax.readthedocs.io/en/latest/index.html#>.
- [24] TensorFlow. <https://www.tensorflow.org/>.
- [25] Sebastian Ruder. An overview of gradient descent optimization algorithms, 2017.
- [26] Samuel L. Smith and Quoc V. Le. A bayesian perspective on generalization and stochastic gradient descent. In *International Conference on Learning Representations*, 2018.
- [27] Ning Qian. On the momentum term in gradient descent learning algorithms. *Neural Networks*, Vol. 12, No. 1, pp. 145–151, 1999.
- [28] Nitish Srivastava Geoffrey Hinton and Kevin Swersky. Neural networks for machine learning lecture 6a overview of mini-batch gradient descent, 2016.
- [29] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.
- [30] Madhu S. Advani, Andrew M. Saxe, and Haim Sompolinsky. High-dimensional dynamics of generalization error in neural networks. *Neural Networks*, Vol. 132, pp. 428–446, 2020.
- [31] Yann LeCun and Corinna Cortes. MNIST handwritten digit database. 2010.

- [32] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 22, No. 10, pp. 1345–1359, 2010.
- [33] Mei Wang and Weihong Deng. Deep visual domain adaptation: A survey. *Neurocomputing*, Vol. 312, pp. 135–153, 2018.
- [34] Gabriela Csurka. Domain adaptation for visual applications: A comprehensive survey, 2017.
- [35] Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2022.
- [36] Michael Tschannen, Olivier Bachem, and Mario Lucic. Recent advances in autoencoder-based representation learning, 2018.
- [37] J. C. Slater. A simplification of the hartree-fock method. *Phys. Rev.*, Vol. 81, pp. 385–390, Feb 1951.
- [38] N. Hizawa, K. Hagino, and K. Yoshida. Applications of the dynamical generator coordinate method to quadrupole excitations. *Phys. Rev. C*, Vol. 105, p. 064302, Jun 2022.
- [39] Kevin Ryczko, David A. Strubbe, and Isaac Tamblyn. Deep learning and density-functional theory. *Phys. Rev. A*, Vol. 100, p. 022512, Aug 2019.
- [40] D. Vautherin. Hartree-Fock Calculations with Skyrme’s Interaction. 2. Axially Deformed Nuclei. *Phys. Rev. C*, Vol. 7, pp. 296–316, 1973.
- [41] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2021.
- [42] Keras official document. <https://keras.io>.
- [43] Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. On the variance of the adaptive learning rate and beyond. In *International Conference on Learning Representations*, 2020.
- [44] Kyle Mills, Michael Spanner, and Isaac Tamblyn. Deep learning and the schrödinger equation. *Phys. Rev. A*, Vol. 96, p. 042113, Oct 2017.
- [45] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.
- [46] R.B. Firestone. Nuclear data sheets for $a = 24$. *Nuclear Data Sheets*, Vol. 108, No. 11, pp. 2319–2392, 2007.
- [47] M. R. Mumpower, R. Surman, G. C. McLaughlin, and A. Aprahamian. The impact of individual nuclear properties on r -process nucleosynthesis. *Prog. Part. Nucl. Phys.*, Vol. 86, pp. 86–126, 2016.
- [48] N. Dubray and D. Regnier. Numerical search of discontinuities in self-consistent potential energy surfaces. *Comput. Phys. Commun.*, Vol. 183, No. 10, pp. 2035–2041, 2012.
- [49] Felix Brockherde, Leslie Vogt, Li Li, Mark E. Tuckerman, Kieron Burke, and Klaus-Robert Müller. Bypassing the kohn-sham equations with machine learning. *Nat. Commun.*, Vol. 8,

-
- No. 1, p. 872, Oct 2017.
- [50] Kaggle. <https://www.kaggle.com>.
- [51] Winners of IceCube machine learning competition announced. <https://icecube.wisc.edu/news/outreach/2023/06/winners-of-icecube-machine-learning-competition-announced/>.
- [52] Rampi Ramprasad, Rohit Batra, Ghanshyam Pilania, Arun Mannodi-Kanakkithodi, and Chiho Kim. Machine learning in materials informatics: recent applications and prospects. *npj Quantum Mater.*, Vol. 3, No. 1, p. 54, 2017.
- [53] Aurel Bulgac, Piotr Magierski, Kenneth J. Roche, and Ionel Stetcu. Induced fission of ^{240}Pu within a real-time microscopic framework. *Phys. Rev. Lett.*, Vol. 116, p. 122504, Mar 2016.
- [54] Toshio Marumori, Toshihide Maskawa, Fumihiko Sakata, and Atsushi Kuriyama. Self-Consistent Collective-Coordinate Method for the Large-Amplitude Nuclear Collective Motion. *Prog. Theor. Phys.*, Vol. 64, No. 4, pp. 1294–1314, Oct 1980.
- [55] Masayuki Matsuo. Treatment of Nucleon-Number Conservation in the Selfconsistent Collective-Coordinate Method: —Coupling between Large-Amplitude Collective Motion and Pairing Rotation—. *Prog. Theor. Phys.*, Vol. 76, No. 2, pp. 372–386, Aug 1986.
- [56] Masayuki Matsuo, Takashi Nakatsukasa, and Kenichi Matsuyanagi. Adiabatic Selfconsistent Collective Coordinate Method for Large Amplitude Collective Motion in Nuclei with Pairing Correlations. *Prog. Theor. Phys.*, Vol. 103, No. 5, pp. 959–979, May 2000.
- [57] Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, Vol. 29, No. 4, pp. 983–1049, October 2016. Funding Information: The first author was supported by NSF grant DMS 1265524, AFOSR grant FA9550-12-1-0425 and U.S.-Israel Binational Science Foundation grant 2014055. The second author was supported by NSF grant EECS-1135843. Publisher Copyright: © 2016 American Mathematical Society.
- [58] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, 2022.
- [59] Steven Weinberg. Phenomenological lagrangians. *Physica A*, Vol. 96, No. 1, pp. 327–340, 1979.
- [60] Marc Verriere, Nicolas Schunck, Irene Kim, Petar Marević, Kevin Quinlan, Michelle N. Ngo, David Regnier, and Raphael David Lasserri. Building surrogate models of nuclear density functional theory with gaussian processes and autoencoders. *Front. Phys.*, Vol. 10, , 2022.
- [61] Michael Crawshaw. Multi-task learning with deep neural networks: A survey. *ArXiv*, Vol. abs/2009.09796, , 2020.

- [62] X.H. Wu, Y.Y. Lu, and P.W. Zhao. Multi-task learning on nuclear masses and separation energies with the kernel ridge regression. *Phys. Lett. B*, Vol. 834, p. 137394, 2022.
- [63] Zu-Xing Yang, Xiao-Hua Fan, Zhi-Pan Li, and Haozhao Liang. A kohn-sham scheme based neural network for nuclear systems. *Phys. Lett. B*, Vol. 840, p. 137870, 2023.
- [64] D. Vautherin and D. M. Brink. Hartree-fock calculations with skyrme’s interaction. i. spherical nuclei. *Phys. Rev. C*, Vol. 5, pp. 626–647, Mar 1972.
- [65] R. Navarro Perez, N. Schunck, R.-D. Lasserri, C. Zhang, and J. Sarich. Axially deformed solution of the skyrme–hartree–fock–bogolyubov equations using the transformed harmonic oscillator basis (iii) hfbtho (v3.00): A new version of the program. *Comput. Phys. Commun.*, Vol. 220, pp. 363–375, 2017.
- [66] E. Chabanat, P. Bonche, P. Haensel, J. Meyer, and R. Schaeffer. A skyrme parametrization from subnuclear to neutron star densities part ii. nuclei far from stabilities. *Nucl. Phys. A*, Vol. 635, No. 1, pp. 231–256, 1998.
- [67] G. F. Bertsch, W. Younes, and L. M. Robledo. Scission dynamics with k partitions. *Phys. Rev. C*, Vol. 97, p. 064619, Jun 2018.
- [68] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR ’16, pp. 770–778. IEEE, June 2016.
- [69] LSVRC2015. <https://www.image-net.org/challenges/LSVRC/2015/index.php>.
- [70] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.
- [71] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1026–1034, 2015.
- [72] Augustus Odena, Vincent Dumoulin, and Chris Olah. Deconvolution and checkerboard artifacts. *Distill*, 2016.
- [73] R. Cipolla, Y. Gal, and A. Kendall. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7482–7491, Los Alamitos, CA, USA, jun 2018. IEEE Computer Society.
- [74] S. Liu, E. Johns, and A. J. Davison. End-to-end multi-task learning with attention. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1871–1880, Los Alamitos, CA, USA, jun 2019. IEEE Computer Society.
- [75] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade W Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, Patrick Schramowski, Srivatsa R Kundurthy, Katherine Crowson, Ludwig Schmidt, Robert Kaczmarczyk, and Jenia Jitsev. LAION-5b: An open large-scale dataset for train-

-
- ing next generation image-text models. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2022.
- [76] Kaiyang Zhou, Ziwei Liu, Yu Qiao, Tao Xiang, and Chen Change Loy. Domain generalization: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 45, No. 4, pp. 4396–4415, 2023.
- [77] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, Vol. 27. Curran Associates, Inc., 2014.
- [78] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, and Jieping Ye. A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 35, No. 4, pp. 3313–3332, 2023.
- [79] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, Vol. 37 of *Proceedings of Machine Learning Research*, pp. 2256–2265, Lille, France, 07–09 Jul 2015. PMLR.
- [80] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, Vol. 33, pp. 6840–6851. Curran Associates, Inc., 2020.
- [81] Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. Diffusion models: A comprehensive survey of methods and applications, 2023.
- [82] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. Layer normalization, 2016.
- [83] Yann N. Dauphin, Angela Fan, Michael Auli, and David Grangier. Language modeling with gated convolutional networks, 2017.
- [84] Qizhe Xie, Minh-Thang Luong, Eduard Hovy, and Quoc V. Le. Self-training with noisy student improves imagenet classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [85] Weihao Yu, Mi Luo, Pan Zhou, Chenyang Si, Yichen Zhou, Xinchao Wang, Jiashi Feng, and Shuicheng Yan. Metaformer is actually what you need for vision. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10809–10819, 2022.
- [86] E. Terán, V. E. Oberacker, and A. S. Umar. Axially symmetric hartree-fock-bogoliubov calculations for nuclei near the drip lines. *Phys. Rev. C*, Vol. 67, p. 064314, Jun 2003.
- [87] D. D. Johnson. Modified broyden’s method for accelerating convergence in self-consistent calculations. *Phys. Rev. B*, Vol. 38, pp. 12807–12813, Dec 1988.