To Standardize Charge Density Distributions for Atomic Nuclei through Information Backflow of Neural Networks

Zu-Xing Yang (杨祖星), 1 Xiao-Hua Fan (范小华), 2 Tomoya Naito (内藤智也), 3, 4 Zhong-Ming Niu (牛中明), 5 Zhi-Pan Li (李志攀), 2 and Haozhao Liang (梁豪兆) 4, 3, 5

1 RIKEN Nishina Center, Wako 351-0198, Japan
2 School of Physical Science and Technology, Southwest University, Chongqing 400715, China
3 RIKEN Interdisciplinary Theoretical and Mathematical Sciences Program, Wako 351-0198, Japan
4 Department of Physics, Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan
5 School of Physics and Materials Science, Anhui University, Hefei 230601, China

Based on the back-propagation principle of feedforward neural networks, we design three charge density generators considering physical factors and network structure to make the residual between theory and experiment reverse flow from the root mean square charge radii to the nuclear charge distributions. It is found that about 96% of the nuclei on the validation set falling within 2 standard deviations of the predicted charge radii for CDG-3, which shows the impressive predictive power. In particular, CDG-3 reduces significantly the error interval for radii of \(^{157}\)Dy. For \(^{48}\)Tb, the predictions are larger than the experimental average by almost 0.1 fm and different from DDHF. According to the Hohenberg-Kohn theorem, the charge densities are further mapped to the matter densities and binding energy, which introduces the correlation of the different observables residuals and enhances interpretability to the residual analysis. The corrections basing charge radii residual almost provide a better description for the binding energy of the Ca isotopes excluding \(^{47, 48, 49}\)Ca, which supports an indispensable beyond-mean-field effect near \(^{48}\)Ca.

Introduction—The charge density distribution is essential for extracting shell structure information, including the evolution of shell structure, shape coexistence, shape transition, and neutron-skin thickness.\(^{14}\) Experimentally, the relative radii of neighboring nuclei are determined using muonic-atom spectra\(^5\) and electron scattering experiments\(^6, 7\), as well as isotope shifts. Given the strong connection between charge density and matter density, one generally calculates the charge density distribution by merging the form factor\(^8, 9\) with the matter density determined by proton-nucleon scattering experiments\(^10\), proton-knockout experiments, or even heavy-ion collisions\(^11\).

Early on, the Fermi distributions\(^12, 13\) and Fourier-Bessel expansion\(^13, 14\) were used to describe the matter and charge densities of atomic nuclei roughly. With the development of computational power, various branching models\(^15\)\(^\sim\)24 were derived from density functional theory (DFT) and shell models have become more popular among theorists. However, due to the complexity of nuclear many-body systems, the calculations of the theory are still inadequate with the existence of beyond-mean-field effects and nucleon-nucleon correlations\(^25\)\(^\sim\)28.

For nuclear complex systems, back propagation neural networks have achieved brilliant success in various aspects, such as nuclear masses\(^29\)\(^\sim\)31, nuclear spins and parities\(^32\), charge radii\(^33\)\(^\sim\)34, excited states\(^34\), extrapolation problems in \textit{ab initio} method\(^36\), \(\alpha\)-decay half-lives\(^37\), \(\beta\)-decay half-lives\(^38\), fission yields\(^39\)\(^\sim\)40, and so on. Most present neural networks learn and predict the residuals between theoretical and experimental values, which has the advantage that the corrected predictions are stronger than the existing theoretical model\(^29\)\(^\sim\)38.\(^41\) However, the corrections are difficult to self-consistent for different observables and have weak physical interpretability. Google proposed a hybrid quantum-classical machine learning model for training beyond classical data types, where back-propagation is used to tune the quantum logic gate parameters, allowing for deep integration of physics and neural networks\(^42\). Based on the Hohenberg-Kohn maps\(^43\) of DFT and the proven strong generalization ability of neural networks in describing density distributions\(^44\), we collectively constrain the charge density distributions by back-propagation with experimental charge radii data, which makes the residual information flow back from radii to densities. We construct the charge densities to binding energy map to achieve the further transfer of information from radii to binding energies.

The charge density generators—Previously, a multilayer feed-forward neural network with a back-propagation algorithm of error has been elaborated to perform the maximum likelihood estimation\(^45\) in the process of generating density distributions\(^44\) approximating the theoretical calculation. A network of about 200–300 nuclei trained quickly is sufficient to describe the density distribution of all nuclei on the nuclear chart and has a powerful extrapolation capability.

Based on the high computational efficiency and generalization ability of neuron networks, we introduced the correlation between theoretically calculated charge density and experimental root mean square charge radius (RMSCR) to a new hybrid neural network. The structure of the network is shown in Fig\(^1\) where \(\rho_{c,i} = \rho_{c}(r_i)\)
is the density on the lattice with \( r_i = 0.1 \times i \) fm. The input is the number of protons and neutrons of a certain nuclide, \( x = \{P, N\} \). The outputs are the charge density \( \rho_c \) and the RMSCR \( (R_c) \) gained by the customized integration layer \( \hat{l}_{\rho_c \rightarrow R_c} \),

\[
\hat{l}_{\rho_c \rightarrow R_c} : R_{c,\text{pre}} = \sqrt{\frac{\int \rho_{c,\text{pre}} r^4 \, dr}{\int \rho_{c,\text{pre}} r^2 \, dr}},
\]

where

\[
O_p = P \mod 2, \quad O_n = N \mod 2, \quad \delta = \frac{[-1]^P + (-1)^N}{2}, \quad P = v_p v_n / (v_p + v_n).
\]

The charge density distribution calculated by physical models \( \rho_{c, \text{theo}} \) is quite accurate. The RMSCR residuals between theory and experiment can be eliminated by assuming a correction \( \delta \rho_c \) to charge density with \( \int_0^\infty \delta \rho_c(r) r^2 \, dr = 0 \). We record this process as standardization. The standardized charge density distribution \( \rho_{c, \text{std}} \) satisfying \( \hat{l}_{\rho_c \rightarrow R_c} (\rho_{c, \text{std}}) \simeq R_{c, \exp} \) can be obtained by superimposing the correction,

\[
\rho_{c, \text{std}} = \rho_{c, \text{theo}} + \delta \rho_c,
\]

here the experimental RMSR \( R_{c, \exp} \) obtained by various means with a high accuracy are taken from Ref. \[5\]. Additionally, we add an assumption to make the correction \( \delta \rho_c \) reasonable: \textit{make the smallest possible corrections to the theory.}

To this end, we design a composite loss function. The normalized mean square error (NMSE) \[44\] is employed as a valid assessment of the density distribution \( L_d \):

\[
L_d = \frac{1}{N_g} \int \left( \rho_{c, \text{pre}}(r) - \rho_{c, \text{theo}}(r) \right)^2 \, dr
\]

where \( \lambda \) is a normalization factor constrained by nuclear charge number \( Z_e \). \( N_g = 150 \) indicates the number of grid points. \( \rho_{c, \text{theo}}(r) \) denotes charge density distribution calculated by density-dependent Hartree-Fock (DDHF), where the SkM* interaction are adopted and Bardeen-Cooper-Schrieffer (BCS) effect is also taken into account. The charge density is obtained from the charge form factor \( F_C \) by the inverse Fourier-Bessel transform,

\[
\rho_{c, \text{theo}}(r) = \frac{1}{2\pi^2} \int dk k^2 j_0(kr) F_C(k),
\]

where \( j_0 \) is the spherical Bessel function of zeroth order. The contributions of matter density and spin-orbit current are folded in \( F_C \) (see SM for details). RMSCR has a large range of variation (about 1 - 6 fm) and therefore Pearson \( \chi^2 \) divergence is picked as its loss function \( L_r \),

\[
L_r = \frac{(R_{c, \text{pre}} - R_{c, \exp})^2}{R_{c, \text{pre}}}.
\]

Combining the two loss functions ensures that the assumption holds:

\[
\text{Loss}(W, w) := \frac{1}{B_s} \sum_{nu} \left( \sum_{i=1}^{B_s} (1 - W) \times L_d^{nu}(w) + W \times L_r^{nu}(w) \right),
\]

where \( W \) is the weighting factor. \( B_s = 64 \) is batch size, which means that 64 nuclei \( (nu) \) are constrained simultaneously for each training session. Actually, the corrections for different nuclei are derived from the same
parameter updates $\delta w$, i.e., *the correction of each nucleus is uniformly constrained by the other nuclei*. We randomly take 640 nuclei (10 batches) of about 900 nuclei measured to date in laboratories as the train set. The remaining nuclei are recorded as the validation set (see SM for details). It is recorded as an epoch when all nuclei on the train set have been trained once.

![Graph](image)

**FIG. 2.** (Color online) Upper panel: The evaluation functions NMSE($W = 0$) for density distributions as a function of epochs on the train set and validation set. Lower panel: The evaluation functions Pearson $\chi^2$ divergence ($W = 1$) for charge radii as a function of epochs on the train set and validation set. The inner panels: The errors of the densities (upper) and radii (lower) of CDG-1, CDG-2 and CDG-3 on validation set labeled as G1, G2 and G3.

*Machine learning processes.* The process of machine learning is divided into two stages, simulating the DDHF theory and correcting with experimental data. The evaluation of these processes is shown in Fig. 2. The first 3000 epochs are the stage of simulating the DDHF theory, during which the maximum likelihood estimation of the objective function $Loss(0, w) = L_d$ is performed. One can see that the loss functions of the train set and validation set almost completely overlap. It means neither for charge density (upper panel) nor RMSCR (lower panel) is overfitting, which shows the generalization ability of the network. The short magenta line is the Pearson $\chi^2$ divergence $L_{r,\text{theo}}$ on validation set between charge radius $R_{c,\text{theo}}$ of the DDHF and the experimental value $R_{c,\exp}$, defined as

$$L_{r,\text{theo}} = \frac{1}{V} \sum_{\text{nuc\in val.}} (\frac{R_{c,\text{theo}}}{R_{c,\exp}})^2,$$

where $V$ is the nuclide number on validation set. Obviously, after a short training, the loss values $L_r$ overlap with $L_{r,\text{theo}}$. This indicates that the network naturally captures the RMSCR information well in the process of learning the density distribution.

After 3000 epochs, the pre-trained model is further tuned with objective function $Loss(W = 0.7)$, which allows the importance of experimental data to slightly exceed that of theoretical calculations. Thus, the charge density distributions under collective constraints are spontaneously corrected by the network. Since the correction makes the predicted charge radius close to the experimental value and the distribution naturally deviates from the physical model, the loss value $L_r$ falls and the loss value $L_d$ jumps as shown in Fig. 2.

We define the degree of optimization for residuals between theoretical and experimental values $\Delta$:

$$\Delta = L_{r,\text{theo}} - L_{r,\text{s}}$$

where $L_{r,\text{s}} = Loss(1, w_{\text{theo}}^{\text{MLE}} + \delta w)$ indicates the loss value $L_r$ of the trained network. $w_{\text{theo}}^{\text{MLE}}$ achieves the maximum likelihood estimate to the theory. For any other theoretical models, the correction is still valid if $\Delta > 0$. Noteworthy is that the training costs only 10 GPU minutes.

![Graph](image)

**FIG. 3.** (Color online) The trained nuclides’ position on nuclear chart (blue square box) and the charge radii predicted by CDG-3(a) and the prediction accuracy of different generators(b). The accuracy indicates the share of experimental data on validation set falling within the different predicted standard deviations (red for 1 std., blue for 1.5 std. and pink for 2 std.) .

*Standardized charge density and charge radius.* To explore the performance of the network, the errors of the densities (upper) and radii (lower) of the retained CDG-1, CDG-2 and CDG-3 on validation set are plotted in the inner panels of Fig. 3, labeled as G1, G2 and G3. It
can be observed that the validation set errors of CDG-3 is minimal for both density and radius. The predictions of CDG-3 are closest to the experimental radii, and the corrections to the theory of CDG-3 are smallest, which agrees with our assumption. It is due to a strong connection between odd-even staggering and pair, shell effects taken account in CDG-3, which is consistent with the mass research [29]. Therefore, we conclude that the network structures without considering the physical factors is insufficient to improve the prediction accuracy, compared with CDG-1 and CDG-2.

Accuracy is a more intuitive representation of CDGs’ performance than error. Fig. 4(b) shows the prediction accuracy of the three networks, where the accuracy indicates the share of experimental data on validation set falling within the different predicted standard deviations (red for 1 std., blue for 1.5 std. and pink for 2 std.). The accuracy are consistent with errors on validation set shown in Fig. 2 for different networks, i.e., the smaller the error, the higher the accuracy. In particular, the share of experimental data falling within 2 standard deviations did not exceed 80% for CDG-1 and CDG-2. Meanwhile, the accuracy of CDG-3 (1 std.) is already close to 80%, which even comes to 96% when expanded to 2 std. It can be speculated that CDG-3 can predict the radii of the remaining about 2000 unmeasured nuclides with high precision. We present the predictions for nuclear charge radii in Fig. 4(a) with colorful squares.

Neural networks give different nuclear radii from DDHF. We display the radii obtained by CDG-3 prediction with statistical error (1 std.), experiment and DDHF on different isotope chains in Fig. 4. It is found that the trained network with parity, paring and shell effects can describe well the odd-even staggering of Ca isotopes from Ca(a), Pb(b), Dy(c), and Tb(d) isotope chains. For Pb isotopes (b), the performance of CDG-3 is comparable to the theoretical model, because the nature of Pb is a benchmark for SkM* interaction. The present predictions almost match the experimental data in the case of training several Pb isotopes. The predictive power is still significant, even if Ca and Pb isotopes are removed from the train set (see SM). Given the excellent performance of CDG-3, we make predictions for untrained \( ^{A}\text{Dy}(c) \) and \( ^{A}\text{Tb}(d) \). For \( ^{A}\text{Dy} \), our predictions are in agreement with the experiment, but with much narrower error intervals than the experiment. And for \( ^{A}\text{Tb} \), our predictions are larger than the experimental average by almost 0.1 fm. The findings from CDG-3 can clearly be tested in laboratories worldwide.

FIG. 4. (Color online) The radii obtained by CDG-3 prediction with statistical error (1 std.), experiment and DDHF on Ca(a), Pb(b), Dy(c), and Tb(d) isotope chains.

FIG. 5. (Color online) Schematic diagram of the structure of feedforward neural network for the CDTBE map.
equation. Details of the parameters for M2 and M3 can be found in the SM. As the charge density updates in CDG-3, the residual information flows to other observables, $\delta R_c \rightarrow \delta \rho_c \rightarrow \{\delta \rho_n, \delta \rho_p\} \rightarrow \delta (E/A)$.

The binding energy is one of the most intriguing observables, and the effect from radii residuals for Ca isotopes is visualized in Fig. 6. We note that the propagated corrections provide a better description for the binding energy of the Ca isotopes excluding $^{47,48,49}$Ca. Especially for $^{40,42,44}$Ca, both radii (see Fig. 4(a)) and binding energies coincide well with the experimental values. This illustrates that the physical mechanisms contained in CDTBE map based on density functional theory are adequate for these nuclei, i.e., other effects such as relativistic effects, deformatons, and mixed configurations do not matter to them. For $^{48}$Ca, the radius is corrected to smaller than DDHF, leading to an increase of nuclear densities and a further decrease of binding energy, which is consistent with equation of state. However, it results in further deviations from the experiment, which supports an indispensable beyond-mean-field effect near $^{48}$Ca. Employing more advanced physical models to train neural networks with more physical mechanisms, such as relativistic effects, deformations, and mixed configurations, is conducive to connect more observables and enhance the predictions.

**SUMMARY**

In this study, supervised learning of DDHF model data is performed firstly to generate the nuclear charge density distributions through deep neural networks, CDG-1, CDG-2 and CDG-3. The pre-trained networks are further tuned by adding the experimental root mean square radii at the outputs, and the collective constraint make the neuron weights be corrected. CDG-3 has a higher accuracy compared to CDG-1 and CDG-2, that 80% of the untrained nuclei fall within one standard deviation of prediction, while 96% of the nuclei fall within two standard deviations, since a feature capture layer with parity, pairing and shell effects is added in CDG-3 to tame odd-even staggering. CDG-3 reduces significantly the error interval for radii of $^{\Lambda}$Dy. For $^{\Lambda}$Tb, the predictions are larger than the experimental average by almost 0.1 fm and different from DDHF. These findings can be proven by worldwide experiments.

Correlation of physical quantities is essential for the development of nuclear many-body models and the measurement of experimental data. The residuals in the charge radii are further propagated to charge density, matter density and binding energy. The propagated corrections provide a better description for the binding energy of the Ca isotopes excluding $^{47,48,49}$Ca, which supports an indispensable beyond-mean-field effect near $^{48}$Ca.

We propose a new phenomenological approach combining theory and experiment, which is available to find the realistic Hohenberg-Kohn map for the complex system. A more realistic map requires connecting more physical quantities by the neural networks. However, as the residuals propagate deeply, the model dependence of the observables grows, which can be eliminate by employing Bayesian model average (BMA) \cite{49,50}. In the current framework, We can not only improve the accuracy of theoretical predictions, but also maintain self-consistency and interpretability.

**ACKNOWLEDGEMENTS**

This work is supported by the National Natural Science Foundation of China under Grants No. 12005175, the Fundamental Research Funds for the Central Universities under Grant No. SWU119076, the JSPS Grant-in-Aid for Early-Career Scientists under Grant No. 18K13549, the JSPS Grant-in-Aid for Scientific Research (S) under Grant No. 20H05648. This work was also partially supported by the RIKEN Pioneering Project: Evolution of Matter in the Universe and the RIKEN Special Post-doctoral Researcher Program.

\[ fanxiaohua@swu.edu.cn \]
\[ haozhao.liang@phys.s.u-tokyo.ac.jp \]

[1] P. Campbell, I. Moore, and M. Pearson, Prog. Part. Nucl. Phys. **86**, 127 (2016).
[2] J. Wood, K. Heyde, W. Nazarewicz, M. Huyse, and P. van Duppen, Phys. Rep. **215**, 101 (1992).
[3] M. Bender, P.-H. Heenen, and P.-G. Reinhard, Rev. Mod. Phys. **75**, 121 (2003).
[4] X.-F. Li, D.-Q. Fang, and Y.-G. Ma, Nucl. Sci. Tech. 27, 1 (2016).
[5] I. Angeli and K. P. Marinova, At. Data Nucl. Data Tables 99, 69 (2013).
[6] R. H. Helm, Phys. Rev. 104, 1466 (1956).
[7] R. Hofstadter, Rev. Mod. Phys. 28, 214 (1956).
[8] B. Dreher, J. Friedrich, K. Merle, H. Rothhaas, and G. Lihrs, Nucl. Phys. A 235, 219 (1974).
[9] J. Friar and J. Negele, Adv. Nucl. Phys., 219 (1975).
[10] R. Hofstadter, Rev. Mod. Phys. 28, 214 (1956).
[11] B. Dreher, J. Friedrich, K. Merle, H. Rothhaas, and G. Lihrs, Nucl. Phys. A 235, 219 (1974).
[12] J. Friar and J. Negele, Adv. Nucl. Phys., 219 (1975).
[13] R. Hofstadter, Rev. Mod. Phys. 28, 214 (1956).
[14] R. Hofstadter, Rev. Mod. Phys. 28, 214 (1956).
[15] T. Naito, X. Roca-Maza, G. Colò, and H. Liang, Phys. Rev. C 99, 024309 (2019).
[16] Y. Zhang, Y. Chen, J. Meng, and P. Ring, Phys. Rev. C 95, 044316 (2017).
[17] C. A. Bertulani and J. Valencia, Phys. Rev. C 100, 015802 (2019).
[18] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. 106, 162 (1957).
[19] N. N. Bogolyubov, Sov. Phys. JETP 7, 41 (1958).
[20] T. Neff, H. Feldmeier, and W. Horiuchi, Phys. Rev. C 92, 024003 (2015).
[21] J. D. Dean, G. Hagen, M. Hjorth-Jensen, T. Papenbrock, and A. Schwenk, Phys. Rev. Lett. 101, 119201 (2008).
[22] R. Roth, Phys. Rev. C 79, 064324 (2009).
[23] T. Dytrych, K. D. Launey, J. P. Draayer, P. Maris, J. P. Vary, E. Saule, U. Catalyurek, M. Sosonkina, D. Langr, and M. A. Caprio, Phys. Rev. Lett. 113, 252501 (2013).
[24] Z. X. Yang, X. L. Shang, G. C. Yong, W. Zuo, and Y. Gao, Phys. Rev. C 100, 054325 (2019).
[25] J. Tanaka, Z. Yang, S. Typel, S. Adachi, S. Bai, et al., Science 371, 260 (2021).
[26] M. Duer et al. (CLAS), Nature 560, 617 (2018).
[27] G. A. Miller, Phys. Rev. Lett. 123, 232003 (2019).
[28] Z. Niu and H. Liang, Phys. Lett. B 778, 48 (2018).