Taming Nuclear Complexity with a Committee of Multilayer Neural Networks

Raphaël-David Lasseri[*]

ESNT, CEA, IRFU, Département de Physique Nucléaire, Université Paris-Saclay, F-91191 Gif-sur-Yvette

David Regnier[†]

Centre de mathématiques et de leurs applications, CNRS, ENS Paris-Saclay, Université Paris-Saclay, 94235, Cachan cedex, France and CEA, DAM, DIF, 91297 Arpajon, France

Jean-Paul Ebran[‡]

CEA, DAM, DIF, 91297 Arpajon, France

Antonin Penon[§]

Magic LEMP, 2 Rue Jean Rostand, 91400 Orsay, France

We demonstrate that a committee of deep neural networks is capable of predicting the ground-state and excited energies of more than 1800 atomic nuclei with an accuracy akin to the one achieved by state-of-the-art nuclear energy density functionals (EDFs) and with significantly less computational cost. An active learning strategy is proposed to train this algorithm with a minimal set of 210 nuclei. This approach enables future fast studies of the influence of EDF parametrizations on structure properties over the whole nuclear chart and suggests that for the first time a machine learning framework successfully encoded several correlated aspects of nuclear deformation.

Introduction.—Today, more than 3000 atomic nuclei have been identified, revealing the wide diversity of nuclear phenomenology (deformation, superfluidity, clustering, halo, etc.). Predicting nuclear properties over the whole range of known mass and charge and beyond is therefore a daunting challenge yet essential to unveil new exotic states of matter, foster our quest to the superheavy island of stability, and answer the fundamental questions of nucleosynthesis. Among the different microscopic approaches of nuclear structure, only the framework of energy density functionals (EDFs) [1] is currently capable of providing a complete and accurate description of ground- and excited-state properties [2–4]. Large-scale deployment of nuclear EDFs is however associated with a high computational cost, especially when it is implemented at its multireference level (MR EDFs), also referred to as beyond mean field. Such a cost is prohibitive to our understanding of the variations of global nuclear features with different EDFs. Ultimately, fitting an effective interaction at the beyond-mean-field level is a tremendous task that was only undertaken once [5] and yielded a root mean square error (rms) on the experimentally known masses of 790 keV. As a result, most of our beyond-mean-field calculations are based on EDFs fitted at the mean-field level, which brings a double counting bias in the predictions. Attacking the problem from a different angle, Athanassopoulos *et al.* built a neural network capable of predicting the whole nuclear

table of mass with a rms of 950 keV [6]. This idea has then been further explored by training neural networks, Bayesian networks, or Gaussian processes to predict the residual between an existing theory and experimental data. It was applied on different models and observables (masses, charge radii, and two neutron separation energies) and typically reduces the binding energy rms to a few hundreds of keV [7–13]. In all these studies, the quality of the predictions is obtained (i) by the knowledge of an initial model with good performances (typically 1–2 MeV rms on the ground-state mass) and (ii) by training the artificial intelligence (AI) on a vast amount of experimental data (especially masses and radii), typically 80% of one of the atomic mass evaluations (AME) [14], i.e., more than 1800 nuclei. This large training set, as well as the fact that these algorithms can only predict one observable, severely restricts the predictive capability of such fast approaches as compared to the EDF approach. In this Letter, we propose a new strategy where an algorithm learns not one observable but several intermediate quantities (potential energy surfaces and inertia) involved in a multi-reference EDF approach. The idea is that while speeding up drastically the calculation of these quantities, the AI will encompass the correlations between several aspects of nuclear deformation. After a training step, this approach also enables us to compute from the AI's predictions multiple low-energy observables such as the ground-state and excited energies.

Method.—In this work, nuclear structure properties are tackled within the five-dimensional collective Hamiltonian (5DCH) approach [15–18]. First, constrained Hartree-Fock-Bogoliubov (HFB) calculations in the space spanned by both the axial and triaxial quadrupole mass moments capture the static correlations associated to quadrupole deformation and pairing. The generation of this manifold of states labeled by the quadrupole deformation variables (β, γ) is by far the most demanding in terms of numerical resource. Then, the effect of the quantum-mechanical fluctuations of the order parameters β and γ around the minimal energy region is accounted for through the construction and solving of the 5DCH. More precisely, we first compute from the HFB constrained states the corresponding HFB energy surface $E_{\mathrm{HFB}}(\beta, \gamma)$, the 2×2 symmetric matrix $\boldsymbol{B}(\beta, \gamma)$ that stands for the vibrational inertia, the rotational inertia $\mathcal{I}_k(\beta, \gamma)$ associated with the three axes k of the intrinsic frame, and a zero-point energy correction $\Delta V(\beta, \gamma)$. We then build the 5DCH as recalled in Ref. [19] and seek its eigensolutions $g_i(\beta, \gamma, \boldsymbol{\Omega})$:

$$(\hat{\mathcal{H}}_{K,\mathrm{rot}} + \hat{\mathcal{H}}_{K,\mathrm{vib}} + \hat{\mathcal{H}}_V)g_i(\beta, \gamma, \boldsymbol{\Omega}) = E_i g_i(\beta, \gamma, \boldsymbol{\Omega}). \qquad (1)$$

The collective Hamiltonian contains (i) a kinetic term $\hat{\mathcal{H}}_{K,\mathrm{rot}}$ associated with rotation that couples the quadrupolar degrees of freedom to the Euler angles $\boldsymbol{\Omega}$, (ii) a vibrational kinetic term $\hat{\mathcal{H}}_{K,\mathrm{vib}}$, and (iii) a potential term $\hat{\mathcal{H}}_V$ that only depends on the quadrupolar deformations. The eigensolutions of Eq. (1) directly yield the correlated ground-state energy as well as the typical rotational and vibrational bands of the excitation spectrum.

The main idea of this work is to simultaneously teach the eight functions, E_{HFB}, ΔV, $\boldsymbol{B}_{00}$, $\boldsymbol{B}_{01}$, $\boldsymbol{B}_{11}$, and $\mathcal{I}_k$ ($k = 1, 2, 3$), defining the collective Hamiltonian to an AI so that it learns their underlying correlations. Our AI consists of a committee of multilayer neural networks [20] that undertake the regression of these functions. Each neural network (NN) takes as input the number of neutrons N and protons Z and returns the values of the eight functions on a discretized mesh of the deformation space. After a learning stage involving a random initialization of each member, the prediction of the committee is obtained by averaging the outputs of its members. The benefit of using a committee is twofold: (i) reducing the variance of the prediction associated with the random initialization of the members and (ii) providing a simple estimation of this variance which we can leverage in an active learning procedure.

The members of the committee all have the same network architecture. Their input is the number of neutrons and protons encoded in a 600 bit string as detailed in Ref. [19]. Note that contrary to Ref. [11], we chose an encoding of the inputs that is totally agnostic of any *a priori* knowledge of the physics (i.e., shell effects in the vicinity of the valley of stability). This typically avoids imposing a structure of the inputs based on hard coded magic numbers

that may not be relevant in exotic mass and charge regions [21]. Internally, the neural networks contain five hidden layers defined by the sequence $600 - 300 - 150 - 100 - 75$ of their number of neurons. The first part of the network embeds the information of the nucleus into a neck of 75 neurons only while its second part predicts from this embedding the output functions. We attempted to fine-tune some hyperparameters of this architecture such as the number, size, and types (dense, convolutions, etc) of the hidden layers with a grid-search approach. Our results seem quite stable in the neighborhood of the chosen hyperparameters (cf. Refs. [19,22,23]).

With this choice of architecture, we perform a supervised training on a set of nuclei for which we know the targeted functions from previous constrained HFB calculations. To maximize the quality of the committee while minimizing the number of HFB calculations required for its training, we implemented an active learning procedure inspired by Ref. [24]. It consists of an iterative algorithm which can be summarized by these few steps. (1) Sample an initial training set of nuclei and compute their collective functions with constrained HFB. (2) Train each member of the committee on this set. (3) Query from the committee a set of additional nuclei that are likely to improve the predictions of the committee if added in the next training step. (4) Compute the collective functions of these new nuclei with constrained HFB and add them into the training set. (5) Reiterate from step (2) up to some stopping criteria.

At step (2), we train independently the neural networks corresponding to each member of the committee following a standard technique in machine learning. This procedure, detailed in Refs [19,25–28], minimizes a training loss while avoiding overfitting the network. The training loss consists of a weighted sum of the partial losses $\mathcal{L}_t(N, Z)$ per nucleus (N, Z) and per output function t. The partial losses are themselves defined as the squared error between the AI's prediction t_{AI} and the HFB calculation t_{HFB} averaged on the deformation space $(\beta, \gamma) \in [0, \beta_+ = 0.9] \times [0, (\pi/3)]$ for one nucleus:

$$\mathcal{L}_t(N, Z) = \frac{6}{\pi \beta_+^2} \int_{\beta, \gamma} |t_{\mathrm{AI}}(\beta, \gamma) - t_{\mathrm{HFB}}(\beta, \gamma)|^2 d\beta \beta d\gamma. \qquad (2)$$

After each training stage, five new nuclei are added to the training set. To select them, we improved the method proposed in Ref. [24] in the following way. Each member of the committee makes a prediction for more than 2000 nuclei and we first isolate the 10% for which the standard deviation between members is the highest. Then we use a k-means algorithm to detect five clusters among these nuclei and take in each cluster the nucleus for which members' predictions differ the most. To accelerate the training process, we normalized each output function (cf. Ref. [19]). The HFB energy is, for instance, transformed by first removing a deformed liquid drop formula inspired by Ref. [29] and then
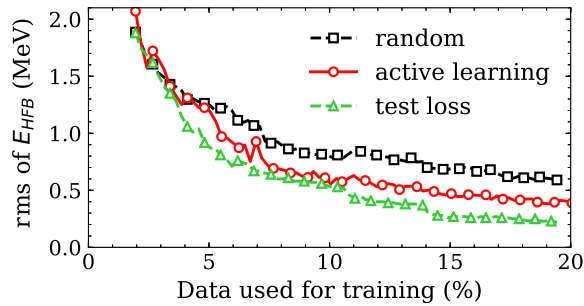
FIG. 1. rms of the HFB energy on the test nuclei as a function of the size of the training set (in percent of the AMEDEE database). We compare results obtained with a training set determined by the active learning (red), by random sampling (black), and by an incremental choice based on the test loss (green).

TABLE I. The rms obtained on the test set at different stages of the active learning. The first column contains the size of the training set in percent of the AMEDEE database while the others highlight the rms of the outputs of the committee of NN. The last column contains the rms associated with the correlated ground-state energy E_{GS} solution of Eq. (1). We emphasize in bold font the training set retained for the rest of this Letter.

Training %	E_{HFB} (keV)	ΔV	$\mathcal{I}_1$	$\mathcal{I}_2$ ($\hbar^2 \times \mathrm{MeV}^{-1}$)	$\mathcal{I}_3$	$\boldsymbol{B}_{00}$	$\boldsymbol{B}_{01}$ (MeV^{-1})	$\boldsymbol{B}_{11}$	E_{GS} (keV)
5	1190	417	1.84	2.80	0.97	13.8	12.0	28.2	1325
10	**557**	**312**	**1.40**	**2.25**	**0.76**	**11.7**	**10.2**	**23.9**	**716**
15	471	247	1.25	2.02	0.69	10.6	9.4	21.9	655
20	388	202	1.22	1.96	0.68	10.2	9.1	21.2	518

performing a linear scaling to obtain a zero mean and unity standard deviation on the training set of nuclei.

Results.—We consider a dataset of 2100 even-even nuclei taken from the AMEDEE database [30] and with the charge and neutron ranges $Z \in [10, 120]$ and $N \in [10, 260]$. For these nuclei, we dispose of the eight functions, E_{HFB}, ΔV, $\boldsymbol{B}_{00}$, $\boldsymbol{B}_{01}$, $\boldsymbol{B}_{11}$, and $\mathcal{I}_k$ ($k = 1, 2, 3$), calculated at 94 deformation points with the Gogny D1S effective interaction [31]. For each nucleus, we first interpolated these raw data on a 64×64 regular grid with splines of degree 2. We start the active learning of the committee of NN with 2% of the nuclei sampled randomly from an adapted jittered sampling [32], that ensures a certain uniformity in the N-Z plane. We then run the active learning up to a point where the training set contains 20% of the 2100 nuclei. At each learning step, we evaluate the quality of the committee's predictions on the M_{test} test nuclei not present in the training set ($M_{test} > 80\%$ of the database). To do so we determine for each output function t its associated rms defined as

$$\mathrm{rms}(t) = \left(\frac{1}{M_{test}} \sum_{i}^{M_{test}} \mathcal{L}_t(N_i, Z_i) \right)^{1/2}. \qquad (3)$$

We show in Fig. 1 the evolution of the rms associated with the HFB energy as a function of the size of the training set. Starting above 2 MeV, it follows an exponential-like decrease to reach less than 400 keV at 20%. We compare these results with the ones obtained if we train the same committee of NN on a set of nuclei that is (i) randomly chosen at each step (ii) incremented at each step with the 5 nuclei that maximize the global test rms (computed on the test set). Note that this last procedure requires the *a priori* knowledge of the HFB results for all nuclei. The active learning approach outperforms the naive random selection of the training set by roughly 200 keV as soon as more than 5% of the dataset is used for training. In addition, the training based on the test loss gives even better results in this region. This shows the possibility that a more

sophisticated algorithm of active learning could still improve our current results. Choosing the size of the training set is a trade-off between the accuracy of the resulting AI and the numerical cost associated with the HFB calculations of the training nuclei. In Table I, we report the rms of the eight output functions obtained at four different steps of the active learning. A striking result is the quality of the committee's prediction already achieved with only 10% of the total dataset. The HFB energy, which is a key feature in the determination of the correlated energies, is reproduced within 557 keV over the 1890 nuclei of the test set. In the following we therefore show the results obtained with this 10% training set.

The quality of the committee's prediction varies with N and Z. To assess what parts of the nuclear chart are correctly grasped by the committee of neural networks, we emphasize in Fig. 2 the individual rms per nucleus $\mathcal{L}_t^{1/2}(N, Z)$ for three different kinds of outputs. For the HFB energy, the AI captures very well the vast majority of heavy nuclei but struggles in the medium and light sectors ($N < 50$). An especially high rms is found close to the $N = Z$ line where HFB calculations are known to predict a strong energy cusp. The difficulty to reproduce the HFB energies in the light sector with neural networks was already encountered in Ref. [11] and is related to the sharp variations present in this region. Figure 2(a) shows that the active learning procedure automatically densified the training set in this region to mitigate this difficulty. Concerning the vibrational and rotational inertia, the error of the AI globally increases with the mass and some of the error peaks can be identified close to shell closures, e.g., for the vibrational inertia close to the neutron number $N = 80$. We compare in Fig. 3 the AI and reference HFB predictions for three targeted functions for ^{178}Os. We choose this nucleus because (i) its excitation spectrum is known experimentally and (ii) its partial root mean square rms(E_{HFB}) is 409 keV, which lies just above the median of this quantity over the test set. It is therefore representative of how the committee of NN performs for most of the test nuclei. Once again, the
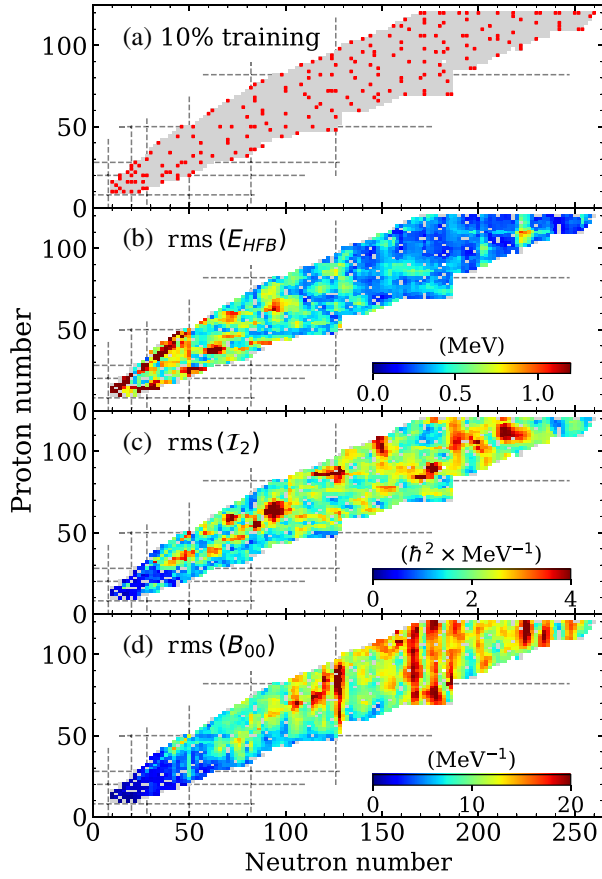
FIG. 2. (a) The AMEDEE database nuclei are plotted in gray as a function of N and Z. The red squares stand for nuclei included in the 10% training set obtained by the active learning. Panels (b)–(d) display the resulting AI versus HFB root mean square per nucleus $[\mathcal{L}_t^{1/2}(N,Z)]$ for the three outputs E_{HFB}, $\mathcal{I}_2$, and $\boldsymbol{B}_{00}$, respectively.
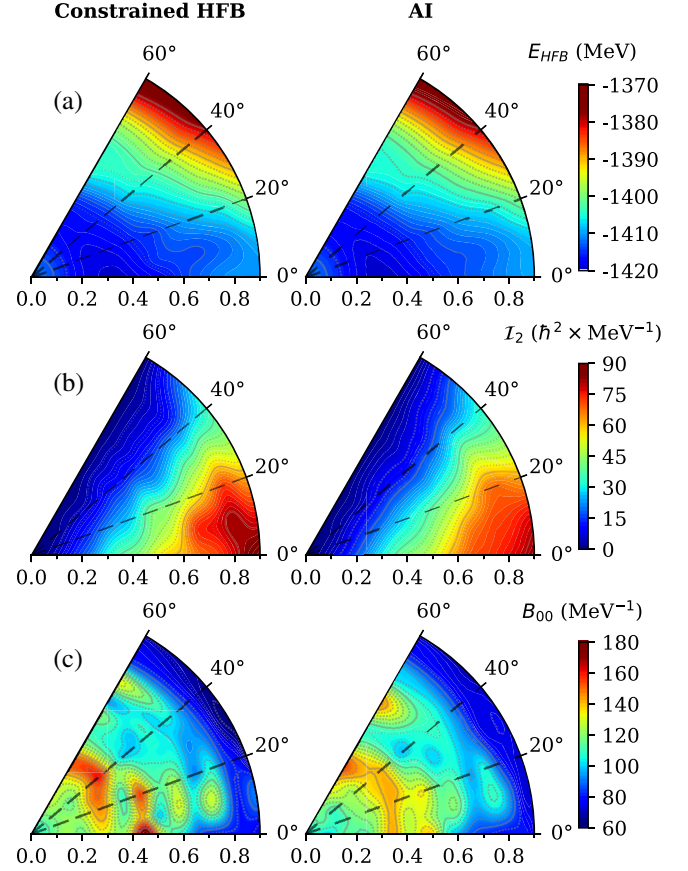

FIG. 3. HFB energy (a), rotational inertia along the principal axis (b), and vibrational inertia related to elongation (c) resulting from both constrained HFB calculations and the AI. These functions are plotted for ^{178}Os in the standard polar representation where β is the radial coordinate and γ the polar angle.

overall topology of E_{HFB}, $\mathcal{I}_2$, and $\mathcal{B}_{00}$ are very well grasped by the committee despite the fact that the closest nucleus in the training set is the ^{180}W, which has 4 additional neutrons and 2 protons less.

Finally, we focus on the correlated ground-state energy and excitation spectra obtained from Eq. (1). To solve the eigenproblem, we discretized the Euler angle space on the basis of Wigner rotational wave functions as in Ref. [15], whereas the deformation space is discretized on a finite element basis implemented with the FELIX-2.0 library [33]. We perform the comparison between the AI and the HFB reference for the 1666 nuclei of the test whose minimum of the potential energy lies below $\beta = 0.8$. This simple filter removes the superheavy nuclei with an open fission channel in our deformation space, for which the ground state would be spuriously predicted at too high deformations. As reported in Table I, we obtain a rms of 716 keV for the ground-state energy. Although a direct comparison is not sound, note that this number has the same order of magnitude than the rms difference between experiments and predictions

from state-of-the-art nuclear EDFs (500–800 keV for the Skyrme HFB mass models [34,35]). Figure 4 displays the excitation spectra of ^{178}Os obtained from both the HFB and committee predictions and gives for the sake of completeness
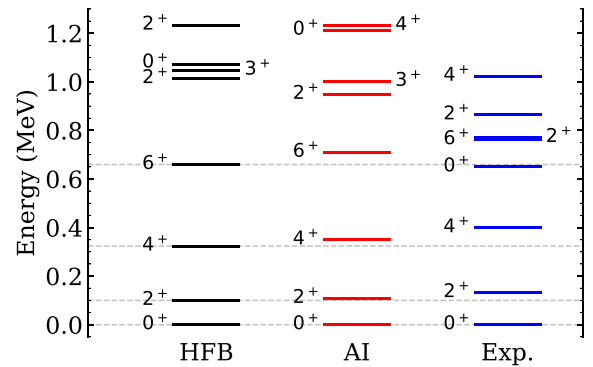

FIG. 4. Excitation spectrum of ^{178}Os obtained from both the AI (AI) and the constrained HFB calculations (HFB). The experimental spectrum (Exp.), taken from the ENSDF database [36], is also displayed.

the experimental values taken from the ENSDF database [36]. The rotational band predicted by the committee of NN impressively matches the HFB data with only a 8% deviation for the first 2^+, 4^+, and 6^+ states. Finally, the first excited 0^+ level is reproduced within 13% despite the complexity of the vibrational inertia topology.

Conclusions.—In this work we built, for the first time, a machine learning framework capable of estimating the low-energy structure of all nuclei from a given EDF. Stunning performances are achieved, viz., a rms of 716 keV on the correlated ground-state energy with respect to the MR EDF calculation for a training on only ∼200 nuclei. Further improvements seem in reach, e.g., by (i) refining the selection of the training set of nuclei (cf. Fig. 1) and (ii) exploring more involved active learning techniques such as negative correlation learning [37] or more sophisticated kinds of neural networks. This fast framework opens the opportunity to quickly test the impact of new parametrizations of EDFs in the context of astrophysics and superheavy production. On top of this, it paves the way toward fitting new EDFs at the multireference level and with multiple observables (ground-state masses, radii, and spectroscopic features). Finally, the success of this approach is a first proof of principle that a committee of NN is able to encode several correlated aspects of nuclear deformation. The neural networks involved likely possess a satisfying nontrivial internal representation of the physics of the system. Studying this representation may unveil new physical concepts grasped during the active learning.

[*] raphael.lasseri@cea.fr
[†] david.regnier@cea.fr
[‡] jean-paul.ebran@cea.fr
[§] penon@magic-lemp.com

[1] N. Schunck, M. Bender, A. Bulgac, T. Duguet, J.-P. Ebran, J. Engel, M. F. Michael, M. Kortelainen, and T. Nakatsukasa, *Energy Density Functional Methods for Atomic Nuclei* (Institute of Physics Publishing, Bristol, 2019), https://iopscience.iop.org/book/978-0-7503-1422-0.pdf.

[2] M. Bender, Eur. Phys. J. Special Topics **156**, 217 (2008).

[3] T. Nikšić, D. Vretenar, and P. Ring, Prog. Part. Nucl. Phys. **66**, 519 (2011).

[4] J. L. Egido, Phys. Scr. **91**, 073003 (2016).

[5] S. Goriely, S. Hilaire, M. Girod, and S. Péru, Phys. Rev. Lett. **102**, 242501 (2009).

[6] S. Athanassopoulos, E. Mavrommatis, K. A. Gernoth, and J. W. Clark, AIP Conf. Proc. **831**, 397 (2006).

[7] R. Utama and J. Piekarewicz, Phys. Rev. C **96**, 044308 (2017).

[8] H. F. Zhang, L. H. Wang, J. P. Yin, P. H. Chen, and H. F. Zhang, J. Phys. G **44**, 045110 (2017).

[9] R. Utama and J. Piekarewicz, Phys. Rev. C **97**, 014306 (2018).

[10] L. Neufcourt, Y. Cao, W. Nazarewicz, and F. Viens, Phys. Rev. C **98**, 034318 (2018).

[11] Z. M. Niu and H. Z. Liang, Phys. Lett. B **778**, 48 (2018).

[12] S. Akkoyun, T. Bayram, S. O. Kara, and A. Sinan, J. Phys. G **40**, 055106 (2013).

[13] R. Utama, W.-C. Chen, and J. Piekarewicz, J. Phys. G **43**, 114002 (2016).

[14] M. Wang, G. Audi, F. G. Kondev, W. J. Huang, S. Naimi, and X. Xu, Chin. Phys. C **41**, 030003 (2017).

[15] K. Kumar and M. Baranger, Nucl. Phys. **A92**, 608 (1967).

[16] J. Libert, M. Girod, and J.-P. Delaroche, Phys. Rev. C **60**, 054301 (1999).

[17] J. P. Delaroche, M. Girod, J. Libert, H. Goutte, S. Hilaire, S. Péru, N. Pillet, and G. F. Bertsch, Phys. Rev. C **81**, 014303 (2010).

[18] Y. Fu, H. Mei, J. Xiang, Z. P. Li, J. M. Yao, and J. Meng, Phys. Rev. C **87**, 054305 (2013).

[19] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.124.162502 for technical details on conditioning the dataset and training the neural networks.

[20] C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed. (Springer-Verlag, New York, 2006).

[21] O. Sorlin and M. G. Porquet, Prog. Part. Nucl. Phys. **61**, 602 (2008).

[22] Autonomio Talos (computer software), http://github.com/autonomio/talos.

[23] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and J. Han, arXiv:1908.03265.

[24] J. Yao, Y. Wu, and H. Zhai, Phys. Rev. Research **2**, 013287 (2020).

[25] F. Chollet, *Deep Learning with Python* (Manning Publications, Shelter Island, NY, 2017).

[26] M. Abadi *et al.*, in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation* (USENIX Association, Berkeley, 2015), https://dl.acm.org/doi/proceedings/10.5555/3026877.

[27] S. J. Reddi, S. Kale, and S. Kumar, in *6th International Conference on Learning Representations (ICLR) Vancouver, BC, Canada, 2018* (OpenReview.net, 2018), https://openreview.net/forum?id=ryQu7f-RZ.

[28] D. P. Kingma and J. Ba, arXiv:1412.6980.

[29] N. Wang, M. Liu, and X. Wu, Phys. Rev. C **81**, 044322 (2010).

[30] S. Hilaire and M. Girod, Eur. Phys. J. A **33**, 237 (2007).

[31] J. F. Berger, M. Girod, and D. Gogny, Comput. Phys. Commun. **63**, 365 (1991).

[32] F. Pausinger, M. Rachh, and S. Steinerberger, Stat. Probab. Lett. **132**, 55 (2018).

[33] D. Regnier, N. Dubray, M. Verrière, and N. Schunck, Comput. Phys. Commun. **225**, 180 (2018).

[34] N. Chamel, S. Goriely, and J. M. Pearson, Nucl. Phys. **A812**, 72 (2008).

[35] S. Goriely, N. Chamel, and J. M. Pearson, Phys. Rev. Lett. **102**, 152503 (2009).

[36] E. Achterberg, O. A. Capurro, and G. V. Marti, Nucl. Data Sheets **110**, 1473 (2009).

[37] Y. Liu and X. Yao, Neural Netw. **12**, 1399 (1999).