Machine learning density functional theory for the Hubbard model

James Nelson,* Rajarshi Tiwari,† and Stefano Sanvito‡
School of Physics, AMBER and CRANN Institute, Trinity College, Dublin 2, Ireland

(Received 30 October 2018; revised manuscript received 30 January 2019; published 15 February 2019)

The solution of complex many-body lattice models can often be found by defining an energy functional of the relevant density of the problem. For instance, in the case of the Hubbard model the spin-resolved site occupation is enough to describe the system's total energy. Similarly to standard density functional theory, however, the exact functional is unknown, and suitable approximations need to be formulated. By using a deep-learning neural network trained on exact-diagonalization results, we demonstrate that one can construct an exact functional for the Hubbard model. In particular, we show that the neural network returns a ground-state energy numerically indistinguishable from that obtained by exact diagonalization and, most importantly, that the functional satisfies the two Hohenberg-Kohn theorems: for a given ground-state density it yields the external potential, and it is fully variational in the site occupation.

DOI: 10.1103/PhysRevB.99.075132

I. INTRODUCTION

Density functional theory (DFT) [1] is today the most widely used method for computing the electronic structure of solids and molecules, and it finds widespread applications in physics, chemistry, biology, and materials science. The success of DFT has to be attributed to its solid theoretical foundation, contained in the Hohenberg-Kohn theorems [2], in a clear pathway to a practical implementation, formulated in the Kohn-Sham equations [3], and to the possibility of constructing a hierarchical ladder of approximations for the energy functional [4]. Although DFT has been developed to solve the problem of N electrons and M nuclei interacting through the long-range Coulomb potential, one can formulate DFT-type approaches also for many-body lattice models [5]. In this case, the electron density is replaced by an appropriate density defined on the lattice, which becomes the fundamental quantity of the theory. For instance, the local on-site occupation, $\{n_{i\sigma}\}$, is the relevant density of the Hubbard model. This is the expectation value of the number operator, $\hat{n}_{i\sigma} = c^{\dagger}_{i\sigma} c_{i\sigma}$, with $c^{\dagger}_{i\sigma}$ ($c_{i\sigma}$) being the fermionic creation (annihilation) operator at site i for spin $\sigma = \uparrow, \downarrow$. In the lattice DFT framework, the equivalents of the two Hohenberg-Kohn theorems can be demonstrated.

The interest in lattice DFT is twofold. On the one hand, it provides a scalable numerical platform to investigate strongly correlated inhomogeneous systems as they approach the thermodynamic limit [6,7]. This is because the DFT computational effort scales as the local density, i.e., it is linear with the number of sites. On the other hand, it allows one to explore fundamental questions common to any density functional theory, such as the choice of the reference system in the construction of the energy functional [8], the origin of the Mott gap [9], the effects of strong correlation in quantum transport [10–12], and the extension to the time domain [13–15]. For some lattice models, exact solutions exist. These typically concern the homogeneous case [16], but they do not generalize to the inhomogeneous one [17]. Importantly, lattice DFT offers the ideal theoretical framework to treat on the same footing both the homogeneous and the inhomogeneous problem, as we are now going to explain.

Let us consider, as an example, the standard one-dimensional Hubbard model. The Hamiltonian operator reads

$$\hat{H}_U = \hat{T} + \hat{U} + \sum_{i\sigma} \hat{n}_{i\sigma} v_i, \tag{1}$$

where $\hat{T} = -t \sum_{i\sigma} (\hat{c}^{\dagger}_{i+1,\sigma} \hat{c}_{i\sigma} + \hat{c}^{\dagger}_{i\sigma} \hat{c}_{i+1,\sigma})$ is the kinetic energy with hopping parameter t, $\hat{U} = U \sum_{i=1} \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}$ is the Coulomb repulsion of strength $U > 0$, and the last term describes the interaction with an external potential $\{v_i\}$. If $v_i = v_0$ for every site i, one has the homogeneous case. The first Hohenberg-Kohn theorem establishes the existence of an energy functional,

$$E[\{n_{i\sigma}\}] = F[\{n_{i\sigma}\}] + \sum_{i\sigma} n_{i\sigma} v_i, \tag{2}$$

$$F[\{n_{i\sigma}\}] = \langle \Psi | \hat{T} + \hat{U} | \Psi \rangle, \tag{3}$$

where $F[\{n_{i\sigma}\}]$ is universal and independent from the external potential, v_i. This means that there is a one-to-one correspondence between the density and the external potential, namely the knowledge of the former is enough to uniquely determine the latter. Since the functional is universal, lattice DFT can be applied identically to both the homogeneous and the inhomogeneous problem. Note that for this lattice model $F[\{n_{i\sigma}\}]$ is universal only for a given $\hat{T}$ and $\hat{U}$ operator. This has some important consequences. For instance, even in one dimension, arranging the sites in a ring or in a chain geometry yields two different $\hat{T}$'s. This means that $F[\{n_{i\sigma}\}]$ for a ring is different from $F[\{n_{i\sigma}\}]$ for a chain, even if both t and U

*janelson@tcd.ie
†tiwarir@tcd.ie
‡stefano.sanvito@tcd.ie

remain the same. Finally, the second Hohenberg-Kohn theorem guarantees that the energy functional is minimized at the ground-state density, $\{n_{i\sigma}^{\mathrm{GS}}\}$, to yield the ground-state energy, E^{GS}, namely $E[\{n_{i\sigma}\}] \geqslant E[\{n_{i\sigma}^{\mathrm{GS}}\}] = E^{\mathrm{GS}}$. The advantage of a DFT approach to the Hubbard model is that the ground-state energy is determined only by the spin-resolved site occupation, namely by a $2L$-dimensional vector, where L is the total number of sites. In contrast, the many-body wave function has dimension $d = (\frac{L}{n^{\uparrow}})(\frac{L}{n^{\downarrow}})$ with $(\frac{L}{n}) = \frac{L!}{n!(L-n)!}$, where n^{σ} is the number of electrons with spin σ.

As for standard DFT, even in the lattice case the exact functional is not known and approximations need to be formulated. These vary depending on the specific problem and usually proceed by interpolation from known exact solutions. Here we take a completely different approach, and we construct an *exact* functional by using a machine learning model trained on exact-diagonalization results. Machine learning (ML) models form a large class of algorithms, which have been traditionally used for data processing and analysis. Recently, these numerical techniques have found a common place in many-body problems [18], phase transitions [19–21], Green's functions calculation [22], and in the solution of single-particle problems in an arbitrary potential [23].

The main motivation behind our work is thus to show that such an ML functional can be constructed. This means that we will establish an exact one-to-one relation between the site occupation and the total energy, and we will numerically demonstrate that this satisfies the two Hohenberg-Kohn theorems. Notably, here we will not proceed by mapping the exact system onto an effective noninteracting one, as is done in the conventional Kohn-Sham approach [3], but rather we will construct a so-called orbital-free DFT [24]. The paper is organized as follows. In the next section, we will introduce our procedure for constructing the ML functional, including the definition of the training and test sets and a brief description of the ML algorithms used. Then we will discuss our results, demonstrating the validity of the two Hohenberg-Kohn theorems. Finally, we will conclude and discuss the limitations and outlook of our approach.

II. METHODS

"Supervised" ML algorithms construct a one-to-one correspondence between vectors or between vectors and scalars. We can then formulate the search for the energy functional as a ML problem, i.e., we can search for a function, f_{ML}, which associates to a given density (the vector of site occupation $\{n_{i\sigma}\}$) the corresponding energy. In particular, our strategy consists in solving exactly the many-body problem for a set of different randomly chosen external potentials, $\{v_i\}$, and then to use the calculated ground-state density and ground-state energy to define the ML model. In practice, we compute

$$f_{\mathrm{ML}} : \left\{n_{i\sigma}^{\mathrm{GS}}\right\} \rightarrow E^{\mathrm{GS}} - \sum_{i\sigma} n_{i\sigma}^{\mathrm{GS}} v_i = \langle \Psi^{\mathrm{GS}} | \hat{T} + \hat{U} | \Psi^{\mathrm{GS}} \rangle, \tag{4}$$

where $|\Psi^{\mathrm{GS}}\rangle$ is the many-body ground-state wave function of $\hat{H}_U$. The fact that we use a set of random external potentials effectively allows us to explore a broad range of densities, and hence to map accurately the functional.

We apply our strategy to the one-dimensional Hubbard model for a system of $L = 8$ sites arranged in a ring geometry (periodic boundary conditions). Furthermore, we restrict ourselves to the paramagnetic quarter-filling case, where the total number of electrons is $N = 4$ and we have $N_{\uparrow} = N_{\downarrow} = 2$. To fully determine the model, we set [25] $U = 4$ and $t = 1$, and, as mentioned before, we construct the dataset for the ML model by exact diagonalization [26]. The random external potential is taken according to a uniform distribution with $v_i \in [-W, +W]$. In particular, we have constructed several external potential distributions with W varying between $0.005t$ and $2.5t$. Furthermore, in order to prevent the dataset from having large fluctuations in the total energy, we neglect potentials yielding to total energies $0.15t$ larger than that of the homogeneous case. Note that the SU(2) symmetry of the problem and the condition $N_{\uparrow} = N_{\downarrow}$ also guarantee that the local site occupation remains spin-unpolarized, i.e., we have $n_{i\uparrow} = n_{i\downarrow}$. This means that in the special case investigated here, the functional depends only on one of the two spin densities, for instance on $\{n_{i\uparrow}\}$, which is solely used to construct the ML model.

To increase the size of the dataset without performing further exact-diagonalization steps, we include configurations obtained from the calculated ones by applying the allowed symmetry operations. In particular, for any potential v_i the mirror-symmetric potential $v_i \rightarrow v_{L+1-i}$ yields a mirror-symmetric charge density with identical total energy. A similar situation applies to potentials obtained by translation, namely $v_i \rightarrow v_{i+1}$. Examples of the charge-density profiles obtained with such symmetry operations are presented in the inset of Fig. 1. The addition of such configurations drastically improves the ML model. The improvements have two main origins. On the one hand, the dataset is larger. On the other hand, the inclusion of the degenerate configurations allows the model to learn about the symmetries of the system.

In the construction of the ML model, the dataset is split into four mutually exclusive subsets. The training set, containing 52 500 samples, not including samples generated by applying symmetries, is used to train the model. The validation set, containing 26 250 samples, is employed to select the best ML model. The test set, also containing 26 250 samples, serves to estimate the generalization error of the model, namely how well it performs on new data never seen before. Finally, we set aside 100 configurations to test the gradient descent scheme for the demonstration of the second Hohenberg-Kohn theorem. These latter configurations are chosen uniformly over energy such that the entire range is explored.

The choice of ML model proceeds in two steps. First, we test several different neural network architectures and choose the best performing architecture on the validation set. Second, the validation set is employed again to determine how many epochs the model must be trained for (in each epoch the entire training set is used to update the model). For an overestimated number of epochs, the neural network becomes accustomed to the training set and will not be able to generalize well. In contrast, if the number of epochs is underestimated, the model will fail to learn any complex relationships between the feature vectors and the targets.

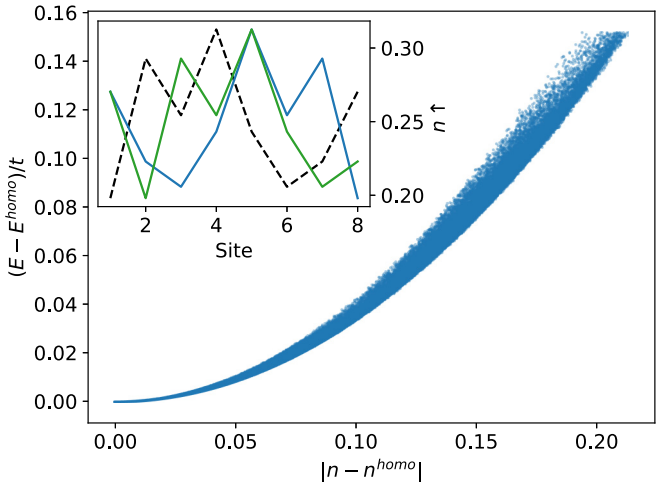

FIG. 1. Structure of the data used to construct the ML model. We plot the ground-state total energy as a function of the Euclidean distance of the corresponding charge density with respect to that of the homogeneous case, $n_{i\sigma} = 1/4$, for a large selection of external potentials, $\{v_i\}$. The total energy is measured with respect of that of the $v_i = 0$ case, which corresponds to $E^{\text{homo}} = F[\{1/4\}]$. Note that the data points in the main plot are those generated numerically, and they do not take into account the symmetries of the system. In the inset, we show examples of charge densities constructed from a given one (dashed line) by applying the symmetry operations of the system: mirror symmetry (blue line) and translational symmetry (green line).

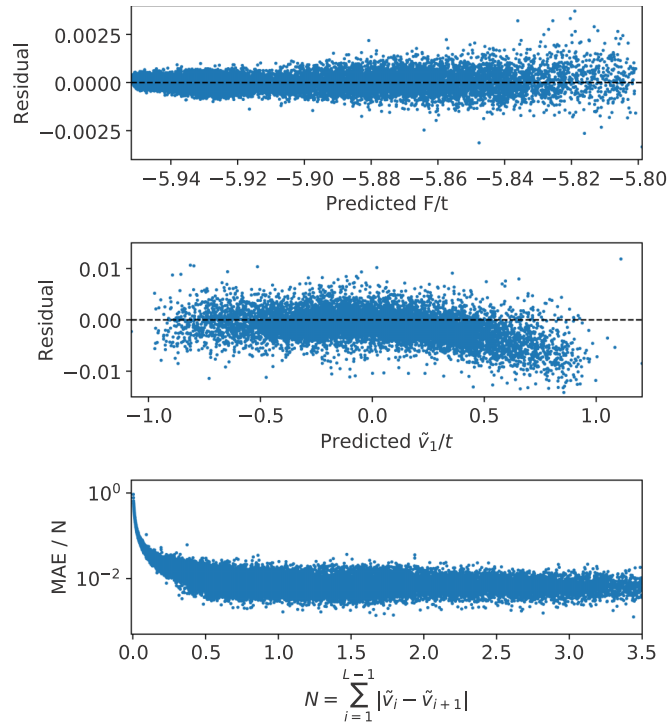

FIG. 2. Numerical accuracy of the ML model measured by the residuals, namely the difference between a predicted quantity and its exact value. In the upper panel, we show the residual of f_{ML} as a function of the corresponding exact-diagonalization result for the entire test set. The model is almost indistinguishable from the exact results, demonstrating that f_{ML} is the exact functional. The mean absolute error is $0.0002t$, namely it is about 0.02%. In the middle panel, we present a numerical demonstration of the first Hohenberg-Kohn theorem. The residual of the external potential at site $i = 2$, measured with respect to that at site $i = 1$, $\tilde{v}_1 = v_2 - v_1$, is plotted against the exact value for the entire test set. The mean absolute error is $0.001t$. Similar curves can be obtained for the other on-site energies. The lower plot shows the relative deviation of the ML potential from the exact one, $\sum_i^{L-1} |\tilde{v}_i^{\text{exact}} - \tilde{v}_i^{\text{ML}}|/N$, as a function of the potential corrugation, $N = ||\tilde{v}^{\text{exact}}||$.

The structure of our dataset is presented in Fig. 1. In the figure, we plot the ground-state total energy as a function of the Euclidean distance of the corresponding charge density with respect to that of the homogeneous case, $n_{i\sigma} = 1/4$. Larger Euclidean distances are associated with the choices of external potential with larger fluctuations. As expected, the distribution of total energies gets broader as the deviation from the homogeneous case gets larger. Note that the homogeneous case, $v_i = 0$, is associated with the minimum of the distribution, where we have $E^{\text{GS}} = F[\{1/4\}]$.

Having tested a few ML models, we have opted for a convolutional neural network, which is found to perform better than standard neural networks [27]. All the machine learning algorithms have been implemented by using the KERAS PYTHON package [28]. In the convolutional neural network, in order to fully capture all the information, we extend each of the occupation vectors by adding their first $k-1$ components to the end of the vector, thus creating an $(L+k-1)$-component-long vector. Here k is the size of the one-dimensional convolution window, in our particular case $k = 3$. Since the convolutional neural network slides the kernel window over the feature vector by choosing k elements at the time, such a prescription guarantees that there are kernel windows, which contain both the first $(k-1)$ elements and the last one of the on-site occupation. The convolutional neural network used has eight convolutional filters, followed by two fully connected layers each with 128 units, and finally an output layer. The loss function used to construct the convolutional neural network is the mean-squared error, and the optimizer is the Adam algorithm.

III. RESULTS AND DISCUSSION

The accuracy of our ML model can be appreciated by looking at the upper panel of Fig. 2, where we present the residual of the predicted $F[\{n_{i\sigma}^{\text{GS}}\}]$, namely $F_{\text{exact}}^{\text{GS}} - f_{\text{ML}}$, calculated over the test set against the predicted results. From the figure it can be clearly appreciated that the ML model is almost indistinguishable from the exact functional. Its mean error over the entire test set is in fact $0.0002t$, i.e., it is of the order of 0.02%. This remarkable accuracy confirms that the functional has been fully learned by our ML model.

We now proceed to demonstrate that the ML functional, f_{ML}, satisfies both of the Hohenberg-Kohn theorems. The first Hohenberg-Kohn theorem states that the ground-state density uniquely determines the external potential up to a constant [2]. To numerically demonstrate the theorem, we take the convolutional neural network constructed before and we generate as an output a seven-component vector. This contains the external potential, namely the on-site energies $\{v_i\}$, which are measured by imposing that the potential of the first site

is $v_1 = 0$. Such a constraint ensures that all the potentials are determined with the same constant. In other words, our constant-rescaled external potential is the vector

$$\tilde{\mathbf{v}} = (v_2 - v_1, v_3 - v_1, \ldots, v_L - v_1). \tag{5}$$

The numerical proof of the first Hohenberg-Kohn theorem is presented in the middle panel of Fig. 2, where we compare the residual of the first component of $\tilde{\mathbf{v}}$, namely $\tilde{v}_1 = v_2 - v_1$, against the exact results. As before, the agreement over the entire test set is nearly perfect with a mean absolute error (MAE) of $0.001t$. Similar agreements are found for the other on-site energies. Furthermore, in the lower panel of Fig. 2 we show the relative deviation of the entire external potential as a function of the potential corrugation. This is defined as

$$\mathrm{MAE}/N = \frac{\sum_i^{L-1} \left|\tilde{v}_i^{\mathrm{exact}} - \tilde{v}_i^{\mathrm{ML}}\right|}{||\tilde{v}^{\mathrm{exact}}||}, \tag{6}$$

where the normalization constant is the potential corrugation $N = ||\tilde{v}^{\mathrm{exact}}|| = \sum_i^{L-1} |\tilde{v}_i^{\mathrm{exact}} - \tilde{v}_{i+1}^{\mathrm{exact}}|$. Clearly our ML functional delivers a potential that is within 1% of the exact one. The only deviations are found near $||\tilde{v}^{\mathrm{exact}}|| = 0$, namely around the homogeneous case, where the error arises from the vanishingly small normalization factor. Overall one has to conclude that $0.001t$ is the absolute precision of our ML model to determine the external potential and 1% is the relative one.

Finally, we proceed to demonstrate the second Hohenberg-Kohn theorem, i.e., we show that the ground-state energy can be found as the minimum of the universal functional at the ground-state density. Strictly speaking, this information was not explicitly used in the construction of our ML functional, since f_{ML} interpolates only at the ground state and not around it. However, since our dataset includes a vast number of external potentials, it includes a vast number of different densities. As such, f_{ML} has been *de facto* constructed by extensively exploring the entire density landscape. The minimization of the energy functional (2) is carried out by gradient descent [27]. The generic derivative of $E[\{n_{i\sigma}\}]$ reads

$$\frac{\partial E}{\partial n_{j\sigma'}} = v_j + \frac{\partial}{\partial n_{j\sigma'}} F[\{n_{i\sigma}\}], \tag{7}$$

where the gradient of the ML model, $F[\{n_{i\sigma}\}] = f_{\mathrm{ML}}$, is estimated by using second-order finite differences. The search for the minimum must also satisfy two conditions, namely $0 \leqslant n_{i\sigma} \leqslant 1$ and particle conservation, $\sum_{i=1}^{L} n_{i\sigma} = N_\sigma$. The first condition is imposed by simply halting the gradient descent algorithm whenever it is violated. In contrast, the second one is enforced by normalizing the occupations after each update.

The accuracy of the computed site occupations is measured as

$$\delta n = \sqrt{\sum_{i\sigma} \left|n_{i\sigma} - n_{i\sigma}^{\mathrm{exact}}\right|^2} = \sqrt{2}\sqrt{\sum_{i} \left|n_{i\uparrow} - n_{i\uparrow}^{\mathrm{exact}}\right|^2}, \tag{8}$$

where $n_{i\sigma}^{\mathrm{exact}}$ is the exact site occupation, and the second equality follows from $n_{i\uparrow} = n_{i\downarrow}$. Similarly, we compute the difference in energy from the reference system. In particular, we perform two tests. In the first one, we set the external potential to $v_i = 0$ and search for $\{n_{i\sigma}\}$ and $F[\{n_{i\sigma}\}]$ corresponding to the homogeneous case starting from a random nonhomogeneous occupation. In the second one, we start from the homogeneous occupation and test the convergence to $\{n_{i\sigma}^{\mathrm{GS}}\}$ corresponding to the inhomogeneous potential $v_i \neq 0$. Such an exercise is performed over the gradient descent set (100 samples), and the results are presented in Fig. 3.

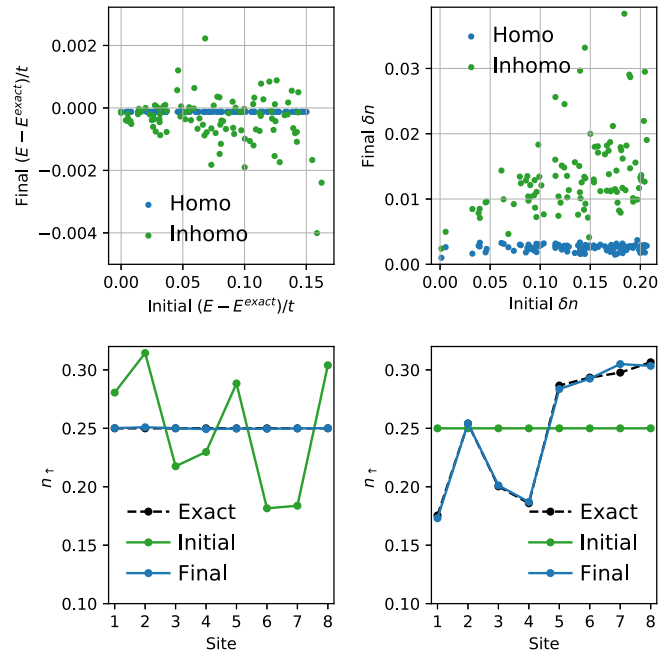

FIG. 3. Numerical demonstration of the second Hohenberg-Kohn theorem. The top two panels show the error in the final converged energy (left) and density (right) [see Eq. (8)] as a function of the distance of the initial quantity from the exact one. Note that when searching for the homogeneous density/energy (blue dots), the gradient descent algorithm appears more accurate than when looking for the inhomogeneous one (green dots). The two lower panels show examples of the converged occupations when searching for the ground-state density of either the homogeneous (left) or the inhomogeneous (right) system.

The two top panels demonstrate that f_{ML} is variational. In these we show the difference in the energy (left-hand side panel) and site occupation (right-hand side panel) between the expected exact values and that found by gradient descent with respect to their initial value. This effectively explores how well f_{ML} has learned about the energy functional landscape, and it looks at how close it converges to the ground-state energy and occupation as a function of how far the initial occupation/energy was. We use blue dots when searching for the ground state of the homogeneous case, and green ones for that of the inhomogeneous case. If f_{ML} is exact, all dots will be on a horizontal straight line at zero. In general, we find that the ground-state quantities (energy and density) are reached within a few percent regardless of their initial value. When searching for the ground state of the homogeneous system, the error is minimal and almost independent from the initial condition. In contrast, the search for $\{n_{i\sigma}^{\mathrm{GS}}\}$ and E^{GS} corresponding to an inhomogeneous potential is less accurate,

and the average error grows linearly with the distance of the initial occupation to the exact one.

The two lower panels instead display two examples of converged site occupation. The left-hand side panel shows that one can recover the homogeneous occupation when the gradient descent starts from an inhomogeneous one, while that on the right-hand side depicts the opposite, namely that an almost exact inhomogeneous occupation can be found by starting from the homogeneous one.

IV. CONCLUSION AND OUTLOOK

We have used here machine learning to construct the exact energy functional for the inhomogeneous Hubbard model. This is provided by a convolutional neural network constructed over exact results obtained by exact diagonalization. The functional appears numerically indistinguishable from the exact solutions and satisfies both of the Hohenberg-Kohn theorems, namely it establishes a one-to-one correspondence between the site occupation and the external potential, and it is variational.

The present functional is constructed for the quarter-filling diamagnetic case, but the same procedure can be applied to other filling factors, including the possibility to describe magnetic ground states. These ML models, when combined, effectively define an exact universal functional, which extends to any filling factor. More difficult is the extension to an arbitrary number of sites. As mentioned in the Introduction, the kinetic energy and Coulomb repulsion operators change as the number of sites L varies. This means that, strictly speaking, one has to construct a new functional for any L. Clearly this is a limitation of our approach. A possible solution to such a problem consists in introducing semilocal functionals, where the energy at a given site depends only on the site occupation at that site and its neighborhood. In this case, one can construct the functional from the dataset of an L-site ring and use it for larger rings.

In conclusion, our results demonstrate that ML can be used to define exact density functional theories, and it may have the potential to be expanded to other lattice and continuous models.

ACKNOWLEDGMENTS

Funding is provided by the Irish Research Council (J.N.) (Grant No. GOIPG/2016/1056) and by the European Research Council project quest (R.T.) (Project No. 307891). We acknowledge the DJEI/DES/SFI/HEA Irish Centre for High-End Computing (ICHEC) and Trinity Centre for High Performance Computing (TCHPC) for the provision of computational resources.

[1] W. Kohn, Rev. Mod. Phys. **71**, 1253 (1999).
[2] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[3] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[4] J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria, and G. I. Csonka, J. Chem. Phys. **123**, 062201 (2005).
[5] K. Schönhammer, O. Gunnarsson, and R. M. Noack, Phys. Rev. B **52**, 2504 (1995).
[6] F. C. Alcaraz and K. Capelle, Phys. Rev. B **76**, 035109 (2007).
[7] M. Saubanére and G. M. Pastor, Phys. Rev. B **84**, 035111 (2011).
[8] N. A. Lima, M. F. Silva, L. N. Oliveira, and K. Capelle, Phys. Rev. Lett. **90**, 146402 (2003).
[9] N. A. Lima, L. N. Oliveira, and K. Capelle, Europhys. Lett. **60**, 601 (2002).
[10] G. Stefanucci and S. Kurth, Phys. Rev. Lett. **107**, 216401 (2011).
[11] V. Vettchinkina, A. Kartsev, D. Karlsson, and C. Verdozzi, Phys. Rev. B **87**, 115117 (2013).
[12] A. Pertsova, M. Stamenova, and S. Sanvito, J. Phys.: Condens. Matter **25**, 105501 (2013).
[13] C. Verdozzi, Phys. Rev. Lett. **101**, 166401 (2008).
[14] D. Karlsson, A. Privitera, and C. Verdozzi, Phys. Rev. Lett. **106**, 116401 (2011).
[15] A. Akande and S. Sanvito, J. Phys.: Condens. Matter **24**, 055602 (2012).
[16] E. H. Lieb and F. Y. Wu, Phys. Rev. Lett. **20**, 1445 (1968).
[17] F. H. L. Essler, H. Frahm, F. Göhmann, A. Klümper, and V. E. Korepin, *The One-Dimensional Hubbard Model* (Cambridge University Press, Cambridge, 2009).
[18] G. Carleo and M. Troyer, Science **355**, 602 (2017).
[19] J. Carrasquilla and R. G. Melko, Nat. Phys. **13**, 431 (2017).
[20] K. Ch'ng, J. Carrasquilla, R. G. Melko, and E. Khatami, Phys. Rev. X **7**, 031038 (2017).
[21] P. Broecker, J. Carrasquilla, R. G. Melko, and S. Trebst, Sci. Rep. **7**, 8823 (2017).
[22] L.-F. Arsenault, A. Lopez-Bezanilla, O. A. von Lilienfeld, and A. J. Millis, Phys. Rev. B **90**, 155136 (2014).
[23] K. Mills, M. Spanner, and I. Tamblyn, Phys. Rev. A **96**, 042113 (2017).
[24] V. L. Lignéres and E. A. Carter, in *An Introduction to Orbital-Free Density Functional Theory*, edited by S. Yip, Handbook of Materials Modeling (Springer, Dordrecht, 2005).
[25] Throughput the paper, we measure all energies in units of t.
[26] M. Sharma and M. A. H. Ahsan, Comput. Phys. Commun. **193**, 19 (2015).
[27] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016), http://www.deeplearningbook.org.
[28] F. Chollet *et al.*, Keras: The Python Deep Learning Library, https://keras.io (2015).