Quantum Deep Field: Data-Driven Wave Function, Electron Density Generation, and **Atomization Energy Prediction and Extrapolation with Machine Learning**

Masashi Tsubaki

National Institute of Advanced Industrial Science and Technology, 2-3-26 Aomi, Koto-ku, Tokyo 135-0064, Japan

Teruyasu Mizoguchi[®]

Institute of Industrial Science, University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan

(Received 16 March 2020; accepted 23 September 2020; published 10 November 2020)

Deep neural networks (DNNs) have been used to successfully predict molecular properties calculated based on the Kohn-Sham density functional theory (KS-DFT). Although this prediction is fast and accurate, we believe that a DNN model for KS-DFT must not only predict the properties but also provide the electron density of a molecule. This Letter presents the quantum deep field (QDF), which provides the electron density with an unsupervised but end-to-end physics-informed modeling by learning the atomization energy on a large-scale dataset. QDF performed well at atomization energy prediction, generated valid electron density, and demonstrated extrapolation.

DOI: 10.1103/PhysRevLett.125.206401

Quantum chemical simulations, such as Kohn-Sham density functional theory (KS-DFT) calculations have been recently approximated by machine learning (ML) techniques such as kernel methods [1-3]. Very recently, deep neural networks (DNNs) [4,5] have been used to successfully predict molecular properties, such as the atomization energy, HOMO, and LUMO, on large-scale datasets. Although this prediction is fast and accurate, there is a problem: a DNN model for KS-DFT must be consistently based on an understanding of the underlying physics and must not only predict the properties but also provide the fundamental quantum characteristics (i.e., the wave function or orbital and electron density) of a molecule. Most existing DNN models, however, cannot provide the electron density because they consider only the atomic coordinates and ignore the molecular field. Furthermore, they are mainly interested in predicting the final output, i.e., the interpolation accuracy of a molecular property within a benchmark dataset, and do not focus on capturing the fundamental characteristics of molecules. This does not lead to learning a physically meaningful model and extrapolation, i.e., prediction for totally unknown molecules in terms of its size and structure that do not appear in the dataset. Extrapolation is important in not only molecular science but also real applications for transferring to other molecules or crystals and predicting their properties [6,7] in materials informatics.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

In this Letter, we present a simple framework called the quantum deep field (QDF) that provides the electron density of molecules by learning their atomization energies on a large-scale dataset. Crucially, our data-driven QDF framework requires only the molecule-energy pairs (e.g., the QM9 dataset [8]) and does not require the molecule-density pairs for training; in other words, QDF generates the electron density indirectly or in an unsupervised fashion with end-toend physics-informed modeling. The QDF model involves three linear and nonlinear components: (i) a linear combination of atomic orbitals (LCAO): $\phi \rightarrow \psi$, where ϕ is the atomic basis function given by the Gaussian-type orbital (GTO) and ψ is the KS molecular orbital, (ii) a nonlinear energy functional: $\psi \to E$, where E is the atomization energy, and (iii) a nonlinear Hohenberg-Kohn (HK) map: $\rho \to V$, where ρ is the electron density and V is the external potential. We optimize all parameters of the LCAO, energy functional, and HK map, in which the latter two are implemented by simple DNNs, simultaneously using the backpropagation and stochastic gradient descent (SGD) [9].

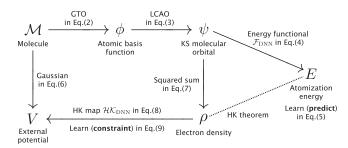


FIG. 1. An overview of the computational flow of our proposed QDF framework. The details of each arrow in this figure are described in the corresponding equation.

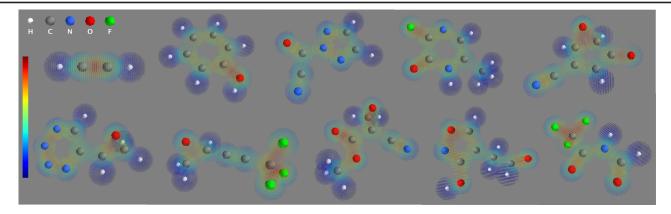


FIG. 2. The electron density maps of some molecules generated by QDF. Each map can be obtained within a second.

We emphasize that learning $\psi \to \rho \to V$ serves as the physical constraint on learning $\psi \to E$ (Fig. 1); this allows QDF to provide valid ψ and ρ (Fig. 2) and leads to high prediction and better extrapolation performance on E.

Initially, a molecule is denoted by $\mathcal{M} = \{(a_1, \mathbf{R}_1), (a_2, \mathbf{R}_2), ..., (a_M, \mathbf{R}_M)\} = \{(a_m, \mathbf{R}_m)\}_{m=1}^M$, where a_m is the mth atom, \mathbf{R}_m is the 3D coordinate of a_m , and M is the number of atoms in \mathcal{M} . Given \mathcal{M} , a set of the molecular orbitals (or wave functions) is denoted by $\{\psi_1(\mathbf{r}), \psi_2(\mathbf{r}), ..., \psi_N(\mathbf{r})\} = \{\psi_n(\mathbf{r})\}_{n=1}^N$, where \mathbf{r} is a position in the field and N is the number of orbitals. The LCAO (or the superposition of wave functions) provides the nth molecular orbital:

$$\psi_n(\mathbf{r}) = \sum_{i=1}^{N} c_{ni} \phi_i(\mathbf{r} - \mathbf{R}_i) \quad \text{s.t.} \sum_{i=1}^{N} c_{ni}^2 = 1, \quad (1)$$

where c_{ni} is the *i*th coefficient, $\phi_i(\mathbf{r} - \mathbf{R}_i)$ is the *i*th atomic basis function whose origin is \mathbf{R}_i , and N is the number of basis functions. As the basis function, we use the GTO:

$$\phi_i(\mathbf{r} - \mathbf{R}_i) = \frac{1}{Z(q_i, \zeta_i)} D_i^{(q_i - 1)} e^{-\zeta_i D_i^2}, \tag{2}$$

where $D_i = ||\mathbf{r} - \mathbf{R}_i||$, q_i is the principle quantum number, ζ_i is the orbital exponent, and $Z(q_i, \zeta_i)$ is the normalization term. Additionally, we use the 6-31G basis set and represent $\{\psi_n(\mathbf{r})\}_{n=1}^N = \psi(\mathbf{r})$ with the *N*-dimensional vector

$$\psi(\mathbf{r}) = \sum_{i=1}^{N} \mathbf{c}_{i} \phi_{i}(\mathbf{r} - \mathbf{R}_{i}), \tag{3}$$

where $\psi(\mathbf{r}) \in \mathbb{R}^N$ has $\psi_n(\mathbf{r})$ as its nth element and $\mathbf{c}_i \in \mathbb{R}^N$ has c_{ni} as its nth element. Note that the orbital exponents $\{\zeta_1, \zeta_2, ..., \zeta_N\} = \{\zeta_i\}_{i=1}^N$ and the coefficient vectors $\{\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_N\} = \{\mathbf{c}_i\}_{i=1}^N$ are randomly initialized and then learned or optimized for predicting the atomization energy using the backpropagation and SGD.

For ψ , we express a DNN-based energy functional \mathcal{F}_{DNN} as follows:

$$E_{\mathcal{M}}' = \mathcal{F}_{\text{DNN}}[\boldsymbol{\psi}],\tag{4}$$

where $E'_{\mathcal{M}}$ is the predicted atomization energy of \mathcal{M} . Herein we use a simple feed-forward architecture for the implementation of \mathcal{F}_{DNN} , in which the DNN models the interaction between ψ_n and ψ_m . Finally, we minimize the loss function

$$\mathcal{L}_E = ||E_{\mathcal{M}} - E_{\mathcal{M}}'||^2, \tag{5}$$

where $E_{\mathcal{M}}$ is the atomization energy of \mathcal{M} in the QM9 dataset. Details about \mathcal{F}_{DNN} and its optimization are described in the Supplemental Material [10].

Unfortunately, only minimizing \mathcal{L}_E does not lead to learning a physically meaningful model; because the DNN has a strong nonlinearity, such a powerful $\mathcal{F}_{\text{DNN}}[\psi]$ will output the correct atomization energy $E_{\mathcal{M}}$ even if ψ is not valid. In other words, the model does not guarantee the KS orbitals $\psi(\mathbf{r})$, which means that $\psi(\mathbf{r})$ cannot provide the correct electron density by $\rho(\mathbf{r}) = \sum_{n=1}^{N} |\psi_n(\mathbf{r})|^2$. To address this problem, we impose a constraint on $\psi(\mathbf{r})$ based on the HK theorem, which ensures that the external potential $V(\mathbf{r})$ is a unique (i.e., a nonlinear but one-to-one correspondence) function of the electron density $\rho(\mathbf{r})$, i.e., $V(\mathbf{r}) \leftrightarrow \rho(\mathbf{r})$. Specifically, we implement the constraint by a nonlinear map $\rho(\mathbf{r}) \rightarrow V(\mathbf{r})$, which we refer to as the HK map [2,11] and learn the nonlinearity using a simple DNN.

Formally, we consider a Gaussian external potential [2,12]

$$V_{\mathcal{M}}(\mathbf{r}) = -\sum_{i=m}^{M} Z_m e^{-||\mathbf{r} - \mathbf{R}_m||^2}, \tag{6}$$

where Z_m is the nuclear charge of a_m . We assume $V_{\mathcal{M}}(\mathbf{r})$ to be the correct external potential of \mathcal{M} ; that is, $V_{\mathcal{M}}(\mathbf{r})$ is used as a target for minimizing loss in the model. Additionally, the electron density is given by

$$\rho(\mathbf{r}) = \sum_{n=1}^{N} |\psi_n(\mathbf{r})|^2. \tag{7}$$

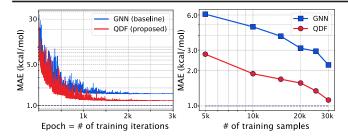


FIG. 3. (a) The learning curves of GNN and QDF for atomization energy prediction, which were performed on the QM9under14atoms dataset. (b) The number of training molecules versus MAE. The number of test samples was fixed at 10 000 molecules and we varied the number of training molecules, which were extracted from the QM9 dataset.

For $\rho(\mathbf{r})$, we express a DNN-based HK map \mathcal{HK}_{DNN} as follows:

$$V'_{\mathcal{M}}(\mathbf{r}) = \mathcal{H}\mathcal{K}_{\text{DNN}}(\rho(\mathbf{r})),$$
 (8)

where $V_{\mathcal{M}}'(\mathbf{r})$ is the predicted external potential of \mathcal{M} . This Letter uses a simple feed-forward architecture for the implementation of \mathcal{HK}_{DNN} . Finally, we minimize the loss function:

$$\mathcal{L}_V = ||V_{\mathcal{M}}(\mathbf{r}) - V_{\mathcal{M}}'(\mathbf{r})||^2. \tag{9}$$

Details about \mathcal{HK}_{DNN} and its optimization are described in the Supplemental Material [10].

Note that the HK map used in Ref. [2] learns $V(\mathbf{r}) \to \rho(\mathbf{r})$ by a supervised kernel method, where $V(\mathbf{r})$ is the input potential and $\rho(\mathbf{r})$ is the target density to be learned. In contrast, our HK map is different; that is, the direction is opposite, i.e., $\rho(\mathbf{r}) \to V(\mathbf{r})$, where the input density is $\rho(\mathbf{r}) = \sum_{n=1}^{N} |\psi_n(\mathbf{r})|^2$, $\psi_n(\mathbf{r})$ is obtained by LCAO, and $V(\mathbf{r})$ is the target potential to be learned.

Furthermore, please note that as a total learning algorithm of QDF, we minimize Eq. (5) and Eq. (9)

TABLE I. The model sizes and final prediction errors on the QM9under14atoms dataset. The number of parameters and MAE for SchNet were obtained from the SchNetPack of original paper [5]. We note that these results can vary and SchNet may outperform QDF with careful tuning of its hyperparameters; however, our main aim herein is not to build a competitive model with regard to the interpolation performance within a single benchmark dataset.

Model	No. of parameters	MAE (kcal/mol)
GNN (baseline)	483 631	1.58
DTNN [4]		1.51
SchNet [5]	1 676 133	1.23
QDF (proposed)	495 262	1.21
Chemical accuracy		1.00

alternately in an end-to-end fashion and optimize all parameters of the LCAO, \mathcal{F}_{DNN} , and \mathcal{HK}_{DNN} using the backpropagation and SGD. We believe that this algorithm involving the HK map constraint allows QDF to function as a self-consistent learning machine for KS-DFT.

To evaluate the QDF performance, we first describe the learning and prediction results of atomization energy E. For training and testing, we used the QM9under14atoms dataset, which is a subset (15 000 samples of relatively smallsized molecules) of the full set (130 000 samples) of the QM9 dataset (see Supplemental Material [10]). Figure 3(a) displays the prediction errors, mean absolute error (MAE, where lower is better) in units of kcal/mol, of our QDF and a baseline model in the form of a learning curve. As the baseline, we implemented a variant of graph neural networks (GNNs) [13], which uses the molecular structure (i.e., the types and 3D coordinates of constituent atoms) alone, does not consider the molecular field, and has hierarchical structure and strong nonlinearity in the modeling phase of the molecular structure. Compared to the GNN, our QDF was able to predict E and its error was much closer to chemical accuracy (1.0 kcal/mol). Additionally, Fig. 3(b) describes how the accuracy improves as the number of training samples increases. The resulting curve was almost linear, and this result is useful for estimating how many training samples are required to achieve a desired accuracy.

Table I shows the final prediction errors and model sizes of the GNN, QDF, and others [4,5] as references. SchNet, which is a variant of the earlier proposed deep tensor neural network (DTNN), is now a standard state-of-the-art deep learning model. We argue that our implemented GNN is not a weak baseline because it achieved a reasonable performance that was competitive with that of DTNN. Additionally, we believe that QDF outperformed (or was competitive with) SchNet in terms of the prediction error; both MAEs were close to 1.20 kcal/mol. In terms of the model size, however, SchNet has more than 1.5×10^6 learning parameters; in contrast, QDF, which has less than half a million parameters, is much more compact.

From the test dataset, we extracted 10 molecules and visualized their electron density maps in Fig. 2 using Mayavi [14]. We first observed that the electron density ρ is higher at points around C, N, O, and F, which have high electronegativity (red to green) than that at points around H, which has low electronegativity (blue), indicating that ρ transfers from H to C, N, O, and F were qualitatively reproduced by learning the atomization energy E. Additionally, ρ increases between two atoms; for example, ρ between two carbon atoms in C_2H_2 , C_6H_5OH , and other molecules are high, indicating that the formation of double and triple bonding was also reproduced. Furthermore, ρ is much higher on the bonding of F atoms. We believe that QDF has the potential to generate valid ρ even if the model is trained only with respect to E.

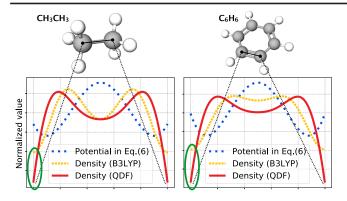


FIG. 4. The electron densities, generated by our QDF and calculated by the B3LYP simulation, on the chemical bonds. We extracted 1000 points between the two atoms on the x axis and displayed the intensity (i.e., the normalized density or potential value) on the y axis.

We compared the electron densities obtained by our QDF and a hybrid-functional (B3LYP) simulation, in which we focused on the density profiles along the chemical bonds of ethane (CH₃CH₃) and benzene (C₆H₆) shown in Fig. 4. We first found that the QDF density (red line) could capture the characteristic double peak features of the B3LYP density (yellow dotted line) on the C–C bond in CH₃CH₃ and the C = C bond in C₆H₆. Additionally, comparing CH₃CH₃ and C₆H₆, the difference with respect to the sharpness of these double peaks could also be reproduced.

Furthermore, we compared the net charge of an atom in a molecule obtained by our QDF with that obtained by the B3LYP simulation and estimated by Bader analysis [15]. Indeed, Bader analysis cannot be directly applied to the current QDF because we generated the grid points around the atoms assuming a spherical distribution (see Fig. 2 and the Supplemental Material [10]). Therefore, we estimated the net charge of the C atom by summing the electron densities inside a sphere with a specific radius, which is half the length of the C-C bond in CH₃CH₃ (0.76 Å) and the C = C bond in C_6H_6 (0.70 Å). Each estimated net charge is as follows: C in CH₃CH₃: 4.05 (B3LYP by Bader), 3.87 (QDF by sphere), and its error is -4.44%; C in C_6H_6 : 4.11 (B3LYP by Bader), 3.89 (QDF by sphere), and its error is -5.35%. These indicate that our ODF could quantitatively reproduce the electron density distribution when compared with that calculated by the B3LYP.

Here, note that we simply followed the work of Brockherde *et al.* [2] and used the Gaussian external potential as a target for learning the HK map. Although the external potential V is not limited to a Gaussian, the current V between two atoms (blue dotted line in Fig. 4) seems to be reasonable for reproducing the double peak of the B3LYP density. We also believe that this Gaussian potential should be improved because our results could not reproduce the density close to the nucleus (green circle in Fig. 4). Hence, another V and atomic orbital, e.g., a Slater-

type orbital (STO), would be more suitable. Overall, Fig. 4 demonstrates the viability of QDF for generating valid ρ in an unsupervised fashion.

Lastly, using a more practical evaluation setting, we present evidence that QDF can capture the molecular orbital or wave function and electron density, i.e., the fundamental characteristics of molecules, from a large-scale dataset. We believe that the following result is an interesting finding of this study that has the potential to facilitate extrapolation, which is difficult to solve by general ML approaches in principle.

We assume that if an ML model could capture the fundamental characteristics of data, the model can be used to conduct a prediction for totally unknown data, i.e., perform an extrapolation. This study evaluated an extrapolation as follows: we trained a model with small molecules and then tested it with large molecules. Specifically, we trained the QDF model with small molecules consisting of fewer than 14 atoms [15 000 samples; the prediction performance was already shown in Fig. 3(a) and Table I] and then tested it with large molecules consisting of more than 15 atoms (115 000 samples), which is the remainder of the QM9 dataset [see Fig. 5(a)].

Figure 5(b) shows the results of this large-scale extrapolation evaluation. The accuracy achieved by GNN is the same as (or superior to) that of the QDF in interpolation; however, the QDF could maintain this accuracy even when the molecular size increases in extrapolation, whereas the GNN could not. Actually, the GNN and its variants have strong nonlinearity in the modeling phase of the molecular structure. By eliminating such nonlinearity using LCAO and imposing the HK map constraint, QDF does not suffer from overfitting. We believe that this is evidence that QDF can capture the fundamental quantum characteristics of molecules that are independent of the system.

Figure 5(c) shows the error analysis on QDF. We found that the molecules with large errors (over 10 kcal/mol) are often sterically strained or have a cagelike structure and are polarized. However, the number of such molecules in the QM9 dataset is very small, e.g., the ratio of polarized molecules is less than 0.3%. In contrast, although the ratio of molecules including F atom(s) is less than 1.5%, their errors were not large, approximately 0.1 kcal/mol, because these molecules have a stable ring structure and are not polarized; this shows the robustness of QDF.

This study proposed the QDF framework, which is different from other deep learning approaches; in other words, QDF is not an extension of existing DNN models. Our aim was to design a simple DNN model without incorporating complicated techniques and architectures. To model the electron density, some ML approaches have been proposed [16–20]; in particular, we were inspired by Grisafi *et al.* [16]. They (i) created an electron density dataset including 1000 configurations of a few kinds of small

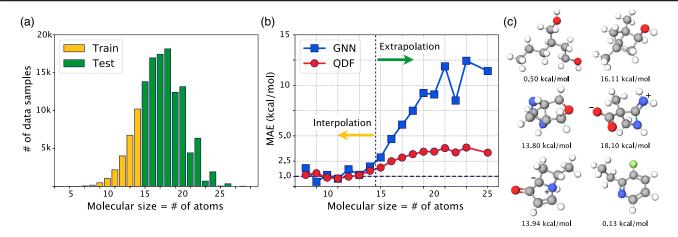


FIG. 5. (a) The data distribution of interpolation (train) and extrapolation (test) samples. The number of test molecules of the QM9over15atoms dataset is 115 000, which is 10 times more than that of training molecules of the QM9under14atoms dataset. (b) The MAEs on this large-scale extrapolation evaluation. The extrapolation error (the right side of this figure) of our QDF is approximately 3.0 kcal/mol. (c) Examples of predicted molecules by QDF and their errors.

molecules (C_4H_6 and C_4H_{10}), (ii) learned a Gaussian process (GP) model for ρ (i.e., supervised learning) of these molecules, and (iii) evaluated the transferability of the learned model by predicting the ρ of large molecules (C_8H_{10} and C_8H_{18}). In contrast, our QDF can be viewed as an unsupervised model to reproduce ρ using only a large-scale dataset of atomization energy E (not ρ). Additionally, we demonstrated the large-scale extrapolation in predicting E. A supervised model as GP would be relatively easy to train and superior to QDF in predicting ρ because the unsupervised QDF is a model closer to generative adversarial networks [21] and would be unstable when generating ρ .

Furthermore, QDF can also be viewed as one of the approaches such as the physics informed, Hamiltonian, Fermionic neural networks [22–25]; these solve the physical problems and equations using physically meaningful modeling. QDF is designed as a self-consistent machine to solve the KS equation with minimal (three) learning and physical constraint components: LCAO, $\mathcal{F}_{DNN}[\psi]$, and $\mathcal{HK}_{DNN}(\rho)$. We believe that integrating a supervised model with a dataset of the electron density [26] (i.e., ρ in Fig. 1 is given as a target) and an unsupervised but physically informed and meaningful model with a dataset of the atomization energy, HOMO-LUMO gap, and other properties [27] will yield an interesting hybrid ML model. QDF will admit many extensions (e.g., for crystals [28]) and applications (e.g., for transfer learning to solve more practical problems in materials informatics [6,29,30]), and our research position and future directions could prove useful (see Supplemental Material [10]).

This study was supported by the Grant-in-Aid for Early-Career Scientists (Grant No. 20K19876) from the JSPS and the Grant-in-Aid for Scientific Research (Grants No. 19H05787 and No. 19H00818) from the MEXT.

- *tsubaki.masashi@aist.go.jp †teru@iis.u-tokyo.ac.jp
- M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. Von Lilienfeld, Phys. Rev. Lett. 108, 058301 (2012).
- [2] F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, Nat. Commun. 8, 872 (2017).
- [3] F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley, and O. A. Von Lilienfeld, J. Chem. Theory Comput. 13, 5255 (2017).
- [4] K. T. Schütt, F. Arbabzadah, S. Chmiela, K.-R. Müller, and A. Tkatchenko, Nat. Commun. 8, 1 (2017).
- [5] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, J. Chem. Phys. 148, 241722 (2018).
- [6] T. D. Huan, A. Mannodi-Kanakkithodi, C. Kim, V. Sharma, G. Pilania, and R. Ramprasad, Sci. Data 3, 160012 (2016).
- [7] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder *et al.*, APL Mater. **1**, 011002 (2013).
- [8] R. Ramakrishnan, P.O. Dral, M. Rupp, and O. A. Von Lilienfeld, Sci. Data 1, 140022 (2014).
- [9] D. P. Kingma and J. Ba, arXiv:1412.6980.
- [10] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.125.206401 for details of the dataset, method, limitations, and future directions.
- [11] J. R. Moreno, G. Carleo, and A. Georges, Phys. Rev. Lett. 125, 076402 (2020).
- [12] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. 104, 136403 (2010).
- [13] S. Kearnes, K. McCloskey, M. Berndl, V. Pande, and P. Riley, J. Comput.-Aided Mol. Des. **30**, 595 (2016).
- [14] P. Ramachandran and G. Varoquaux, Comput. Sci. Eng. 13, 40 (2011).
- [15] W. Tang, E. Sanville, and G. Henkelman, J. Phys. Condens. Matter **21**, 084204 (2009).
- [16] A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf, and M. Ceriotti, ACS Cent. Sci. 5, 57 (2018).
- [17] Y. Zhang, C. Hu, and B. Jiang, J. Phys. Chem. Lett. **10**, 4962 (2019).

- [18] S. Gong, T. Xie, T. Zhu, S. Wang, E. R. Fadel, Y. Li, and J. C. Grossman, Phys. Rev. B 100, 184103 (2019).
- [19] A. Chandrasekaran, D. Kamal, R. Batra, C. Kim, L. Chen, and R. Ramprasad, npj Comput. Mater. 5, 22 (2019).
- [20] K. Ryczko, D. A. Strubbe, and I. Tamblyn, Phys. Rev. A 100, 022512 (2019).
- [21] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, in *Advances in Neural Information Processing Systems* 27 (NIPS 2014) (Curran Associates, Inc., Vancouver, Canada, 2014), pp. 2672–2680, http://papers.nips.cc/paper/5423-generative-adversarial-nets.
- [22] M. Raissi, P. Perdikaris, and G. E. Karniadakis, J. Comput. Phys. 378, 686 (2019).
- [23] G. P. Pun, R. Batra, R. Ramprasad, and Y. Mishin, Nat. Commun. 10, 2339 (2019).

- [24] S. Greydanus, M. Dzamba, and J. Yosinski, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., Montréal, Canada, 2019), pp. 15379–15389.
- [25] D. Pfau, J. S. Spencer, A. G. d. G. Matthews, and W. M. C. Foulkes, Phys. Rev. Research 2, 033429 (2020).
- [26] A. V. Sinitskiy and V. S. Pande, arXiv:1809.02723.
- [27] G. Chen, P. Chen, C.-Y. Hsieh, C.-K. Lee, B. Liao, R. Liao, W. Liu, J. Qiu, Q. Sun, J. Tang *et al.*, arXiv:1906.09427.
- [28] T. Xie and J. C. Grossman, Phys. Rev. Lett. 120, 145301 (2018).
- [29] S. A. Lopez, E. O. Pyzer-Knapp, G. N. Simm, T. Lutzow, K. Li, L. R. Seress, J. Hachmann, and A. Aspuru-Guzik, Sci. Data 3, 160086 (2016).
- [30] C. Kim, T.D. Huan, S. Krishnan, and R. Ramprasad, Sci. Data 4, 170057 (2017).