Differentiable Optimization for the Prediction of Ground State Structures (DOGSS)

Junwoong Yoon and Zachary W. Ulissi
Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA

(Received 7 June 2020; accepted 15 September 2020; published 21 October 2020)

Ground state or relaxed inorganic structures are the starting point for most computational materials science or surface science analyses. Many of these structure relaxations represent systematic changes to the structure, but there are currently no general methods to improve the initial structure guess based on past calculations. Here we present a method to directly predict the ground state configuration using differentiable optimization and graph neural networks to learn the properties of a simple harmonic force field that approximates the ground state structure and properties. We demonstrate this flexible open source tool for improving the initial configurations for large datasets of inorganic multicomponent surface relaxations across 32 elements and the relaxation of adsorbates (H and CO) on these surfaces. Using these improved initial configurations reduces the expensive adsorbate-covered surface relaxations by approximately 50% and is complementary to other approaches to accelerate the relaxation process.

DOI: 10.1103/PhysRevLett.125.173001

Fast and accurate optimization methods that provide ground state inorganic structure and properties are crucial for materials science, surface science, and catalysis, among many other fields. Electronic structure methods have made remarkable progress in predicting inorganic structure-property relationships based on relaxed or ground state structures [1–8]. In most cases, these ground state structures are identified by relaxing initial structures guessed by either experts or automated software packages. The structural relaxations of these initial configurations often require high computational cost that has impeded large-scale explorations of new inorganic materials. Unfortunately, there are currently no general methods to predict better initial configurations for the relaxation process, leading to massive redundancy in calculations.

One approach to accelerating the relaxation is the development of machine learning potentials (MLPs) that mimic commonly used electronic structure methods such as density functional theory (DFT) [9–14]. MLPs approximate the true potential energy surface (PES) by fitting numerous DFT single-point energies and forces. A perfect approximation to the ground-truth DFT PES is obviously desirable, but residual force errors in the fitted force field often yield different ground state structures, or worse, lead to unphysical configurations. We note that there has been only very limited success in developing MLPs that work across the elemental composition. Despite the considerable excitement about these methods for practical simulations, no MLP has been demonstrated for the specific challenge of predicting ground state inorganic structures.

In this Letter, we present a deep learning framework, differentiable optimization for the prediction of ground-sate structures (DOGSS) that incorporates differentiable optimization around an end-to-end trainable graph neural network (GNN) to approximate ground state structures of inorganic materials. Over the last few years, there has been a surge of interest in GNN for investigating properties of materials due to their unique ability to model irregular structures [15–22]. We propose that our GNN-based method can be used as a preprocessing tool that provides any electronic structure method with a starting configuration near the ground state and thereby accelerates the relaxation as illustrated in Fig. 1. DOGSS first learns the optimal properties of the harmonic force fields between atoms to predict the PES of the structure. The predicted harmonic PES is then minimized and the relaxed structure is generated. The harmonic force field is not evaluated on how well it fits energies and forces; instead, it is evaluated by how well the structure predicted using the force field matches the ground state structure. Finally, the relaxation of

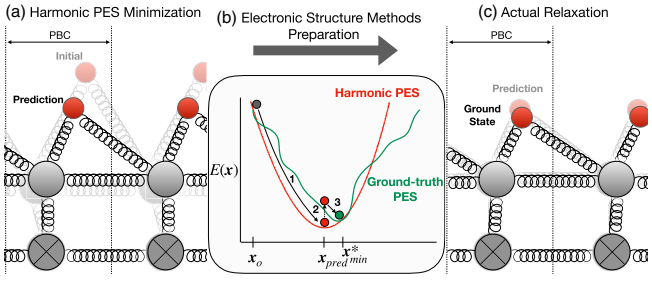

FIG. 1. Workflow to accelerate electronic structure methods using DOGSS as a preprocessing model. (a) DOGSS predicts the ground state structure by relaxing the initial structure using the learned harmonic force field parameters. (b),(c) Electronic structure methods employ the predicted structure as a starting configuration, thereby converge to the ground state in fewer optimization steps.

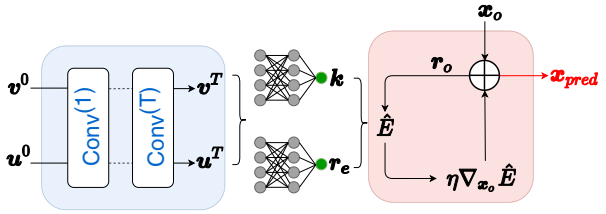

FIG. 2. Structure of the graph neural network with embedded differentiable optimization. Initial node $\boldsymbol{v}^0$ and edge $\boldsymbol{u}^0$ features are fed to T convolutional layers. Two separate hidden networks take the updated features to learn the properties of distinct springs connecting every atom and neighbor atom pairs: spring constants $\boldsymbol{k}$ and equilibrium spring distances $\boldsymbol{r}_e$. With these learned harmonic force field parameters, harmonic PES is computed and minimized using the gradient descent to approximate the ground state structures.

the predicted structure using electronic structure methods can be performed in fewer optimization steps compared to the relaxation of the initial structure.

Similar to the GNN method developed by Xie and Grossman [15], DOGSS represents an atomic structure by an undirected graph consisting of nodes and edges. Figure 2 illustrates how DOGSS maps the graph representation of the input structure to the ground state structure. Each node i is described by a node feature $\boldsymbol{v}_i$, a stack of one-hot encoded atomic and periodic properties of atom i (properties used to define the node features can be found in the Supplemental Material [23]). We define the local environment for each atom i by sampling M nearest neighbor atoms connected by distinct edges. Each edge (i, j) is represented by an edge feature $\boldsymbol{u}_{i,j}$ that encodes the interatomic distance between atom i and atom j and the sum of their covalent radii. Unlike case-by-case human-engineered atomic representations [29–34], the molecular graphs are defined by raw node ($\boldsymbol{v} \in \mathbb{R}^{N\times F_v}$) and edge ($\boldsymbol{u} \in \mathbb{R}^{N\times M\times F_u}$) features that will be further engineered during convolutions. We note that N is the number of nodes, and F_v and F_u are the lengths of the node features and edge features, respectively.

Convolutional layers for the molecular graphs can be expressed as a message-passing scheme [35–37]. For each convolution during T iterations, messages exchanged between node i and node j along the edges (i, j) are generated. These messages capture the pairwise interactions between atoms within the local chemical environment. The node feature $\boldsymbol{v}_i$ and edge feature $\boldsymbol{u}_{i,j}$ are then separately updated using the node messages ϕ_v and edge messages ϕ_u, respectively. We update the node feature $\boldsymbol{v}_i$ according to

$$\phi_v^{t+1} = s(\boldsymbol{z}_{i,j}^t \boldsymbol{W}_{s,v}^t) \odot \sigma(\boldsymbol{z}_{i,j}^t \boldsymbol{W}_{f,v}^t), \tag{1}$$

$$\boldsymbol{v}_i^{t+1} = s\left(\boldsymbol{v}_i^t + \sum_j \phi_v^{t+1}\right), \tag{2}$$

where σ and s denote a sigmoid function and a softplus activation function, respectively, and $\odot$ denotes elementwise multiplication. $\boldsymbol{z}_{i,j}^t \in \mathbb{R}^{N\times M\times F_z}$ represents a local feature ($\boldsymbol{z}_{i,j}^t = \boldsymbol{v}_i^t \oplus \boldsymbol{v}_j^t \oplus \boldsymbol{u}_{i,j}^t$) that is used with weights $\boldsymbol{W}_{s,v}^t$: $\mathbb{R}^{F_z} \to \mathbb{R}^{F_v}$ and $\boldsymbol{W}_{f,v}^t$: $\mathbb{R}^{F_z} \to \mathbb{R}^{F_v}$ to learn a self-feature $s(\cdot)$ and a filter $\sigma(\cdot)$, respectively. Once messages ϕ_v^{t+1} for each pair of node i and node j for the $(t+1)$th convolution are generated, they are aggregated (summed) to compute an overall message for node i from its local environment. The node feature $\boldsymbol{v}_i^t$ is then updated using the aggregated message and a softplus activation. Similarly, the edge feature $\boldsymbol{u}_{i,j}$ can be updated as

$$\phi_u^{t+1} = s(\boldsymbol{z}_{i,j}^t \boldsymbol{W}_{s,u}^t) \odot \sigma(\boldsymbol{z}_{i,j}^t \boldsymbol{W}_{f,u}^t), \tag{3}$$

$$\boldsymbol{u}_{i,j}^{t+1} = s(\boldsymbol{u}_{i,j}^t + \phi_u^{t+1}), \tag{4}$$

where $\boldsymbol{W}_{s,u}^t$: $\mathbb{R}^{F_z} \to \mathbb{R}^{F_u}$ and $\boldsymbol{W}_{f,u}^t$: $\mathbb{R}^{F_z} \to \mathbb{R}^{F_u}$ are self-feature generating weights and filter generating weights, respectively, for edge update. Unlike the node update, each edge feature is directly updated based on the preaggregated pairwise messages ϕ_u^{t+1} during each $(t+1)$th convolution.

After T convolutions, both updated node and edge features are fed to two distinct hidden networks to learn the harmonic force field parameters: (i) pairwise equilibrium spring distances for atom-neighbor pairs $\boldsymbol{r}_e \in \mathbb{R}^{N\times M\times 1}$ and (ii) spring constants $\boldsymbol{k} \in \mathbb{R}^{N\times M\times 1}$ for springs connecting each atom-neighbor pair. Once the force field parameters are learned by the hidden networks, the harmonic PES is constructed as

$$\hat{E}(\boldsymbol{r}_o(\boldsymbol{x}_o), \boldsymbol{r}_e, \boldsymbol{k}) = \sum_i \sum_j \boldsymbol{k}_{i,j}(\boldsymbol{r}_{o,(i,j)}(\boldsymbol{x}_o) - \boldsymbol{r}_{e,(i,j)})^2, \tag{5}$$

followed by an energy minimization based on the following optimization problem:

$$\boldsymbol{x}_{\text{pred}} = \underset{\boldsymbol{x}_o \notin \boldsymbol{x}_{\text{con}}}{\arg\min}\, \hat{E}(\boldsymbol{r}_o(\boldsymbol{x}_o), \boldsymbol{r}_e, \boldsymbol{k}), \tag{6}$$

where $\boldsymbol{x}_{\text{pred}}$ is a set of atomic positions at local minima in the harmonic PES, $\boldsymbol{x}_{\text{con}}$ is a set of positions of constrained atoms that are not allowed to move, and $\boldsymbol{r}_o(\boldsymbol{x}_o) \in \mathbb{R}^{N\times M\times 1}$ represents pairwise distances between atoms and their neighbor atoms in the initial structures.

The argmin operation over the unconstrained surface atoms is optimized using the gradient descent algorithm,

$$\boldsymbol{x}_o^{p+1} = \boldsymbol{x}_o^p + \eta \nabla_{\boldsymbol{x}_o^p} \hat{E}(\boldsymbol{r}_{o(\boldsymbol{x}_o^p)}, \boldsymbol{r}_e, \boldsymbol{k}). \tag{7}$$

After the gradient descent has converged to local minima in P steps, the model outputs the updated atomic structures $\boldsymbol{x}_{\text{pred}}$, which we consider as the approximated ground state structures. To measure the closeness between $\boldsymbol{x}_{\text{pred}}$ and the

TABLE I. Summary of the ground state structure prediction performance of DOGSS and SchNet on test sets in the three different datasets. Best prediction losses in bold.

	Sample size		L_{surf} (Å)		L_{ads} (Å)		L_{total} (Å)	
Dataset	Training	Test	Initial	Prediction	Initial	Prediction	Initial	Prediction
Bare surface								
DOGSS	5524	683	0.1337	**0.0635**	N/A	N/A	0.1337	**0.0635**
SchNet	21 034	683	0.1337	1.1449	N/A	N/A	0.1337	1.1449
H adsorption								
DOGSS	9242	1,142	0.0855	**0.0508**	0.7841	**0.3301**	0.1358	**0.0709**
SchNet	37 346	1,142	0.0855	0.6104	0.7841	1.0594	0.1358	0.6428
CO adsorption								
DOGSS	6682	825	0.0844	**0.0548**	0.7286	**0.3237**	0.1706	**0.0908**
SchNet	43 787	825	0.0844	0.4239	0.7286	0.9116	0.1706	0.4891

ground state structures $\boldsymbol{x}^*_{\text{min}}$, we define a loss function for the GNN model as follows:

$$L = \frac{1}{N}\sum_i \|\boldsymbol{x}_{i,\text{pred}} - \boldsymbol{x}^*_{i,\text{min}}\|_2. \tag{8}$$

This loss represents the average per-atom distance that the unconstrained atoms in the predicted structures need to move to arrive at the ground states. Initial loss between $\boldsymbol{x}_o$ and $\boldsymbol{x}^*_{\text{min}}$ was also computed to be compared with the prediction loss. In case an atom crosses the boundary of its unit cell to reach the ground state, the loss invokes the minimum-image distance convention that considers the nearest periodic image of the ground state atom. The entire network was trained using end-to-end backpropagation with the AdamW optimizer [24] to minimize the prediction loss. Hyperparameters were optimized using SigOpt [25], a model tuning software package, and the optimized hyperparameters are summarized in Table S3 in the Supplemental Material [23].

To demonstrate the effectiveness of our proposed framework, we predict the ground state structures of diverse surfaces of pure metals, metal alloys, and intermetallic alloys covering 32 different elements, with and without the presence of adsorbates (H and CO) collected from our open-source database, the generalized adsorption simulation for PYTHON (GASpy) [26]. We collected 6821 DFT-relaxed diverse bare metal surfaces, 11 411 surfaces with H, and 8250 surfaces with CO. Details about the data preparation are available in the Supplemental Material [23]. Each of the three datasets was randomly split into train, validation, and test (80, ten, and ten) sets for training and testing DOGSS on structures covering 32 surface elements. We also sampled about four to six intermediate structures from the DFT relaxation of each of the training structures to train force fields using SchNet [14]. More details about training SchNet on our datasets can be found in Fig. S2 in the Supplemental Material [23].

After training DOGSS, we computed the prediction loss on the test sets to evaluate how close the predictions were to the ground state structures. The same test structures were also relaxed using the trained SchNet force fields to generate ground state structures, and the same loss was used to compute the prediction loss on those relaxed structures for comparison. We summarize the performance of DOGSS and SchNet for approximating the ground state structures for the three different datasets in Table I. We include initial losses to quantitatively compare the relative closeness between the predicted structures and the ground state structures. We further separate the total loss (L_{total}) into the loss for the surface atoms (L_{surf}) and the loss for the adsorbate atoms (L_{ads}) if applicable. For all three datasets, the force field method using SchNet could not achieve enough prediction accuracy and resulted in different ground state structures. We note that we provided only limited examples (four to six examples of DFT relaxation of each structure) for training SchNet on our datasets. SchNet previously demonstrated accurate force field predictions for small organic molecules using 1000 to 50 000 training examples and a separate model for each molecule. However, training a separate force field model using the same amount of training examples for each of the unique inorganic multicomponent structures in our datasets is computationally infeasible. By comparison, DOGSS only requires the initial and the final relaxed structures for directly predicting ground state structures instead of using thousands of example structures to fit every part in the PES. In addition, one DOGSS model can be used to approximate ground state structures of diverse materials instead of training a separate force field model for each material.

Distributions of individual per-atom distances in the predicted structures (blue) by DOGSS and the initial structures (orange) for the three different datasets are presented in Fig. 3. We could clearly observe that the surface atoms in the predicted structures for the bare surface dataset are more closely located to the ground state positions compared to the atoms in the initial

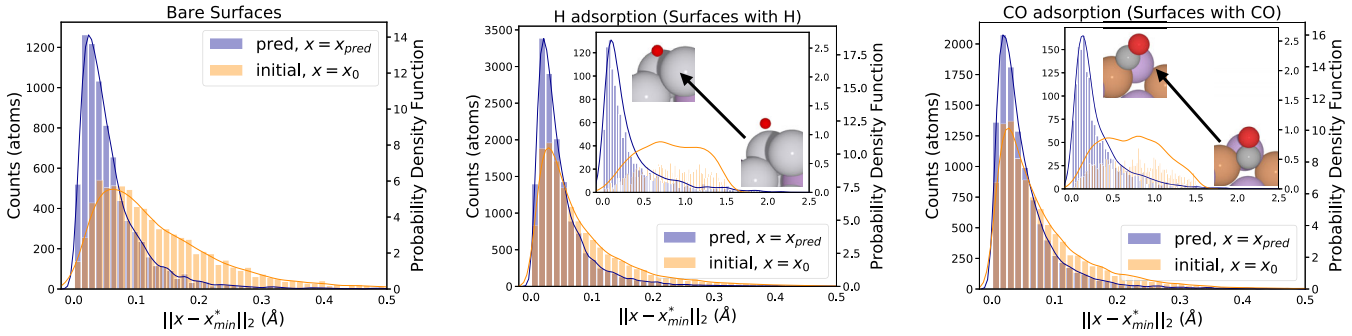

FIG. 3. Histograms showing distributions of the individual per-atom distances $||x - x^*_{\text{min}}||_2$ for surface atoms in each of the three datasets: (left) bare surfaces, (center) H adsorption, and (right) CO adsorption. For all three datasets, the distributions for atoms in the initial structures are shown in orange, and the distributions for atoms in the predicted ground state structures by DOGSS are shown in blue. The insets of the center and the right figures show the per-atom distance distributions for adsorbate atoms in H and CO adsorption datasets, respectively, with visualized example configurations.

structures. Quantitatively, DOGSS could reduce the initial loss for the surface atoms in this dataset by 54%. We acknowledge that the amount of reduction in the loss does not guarantee the same amount of reduction in the DFT relaxation steps. However, we will shortly demonstrate that the ability of our method to reduce the loss can effectively attain a reduction in the computational cost required for the actual DFT relaxations.

For H and CO adsorption datasets, DOGSS could reduce the total losses for both datasets by approximately 47%. We note that selecting initial positions of adsorbates currently relies on experts' guesses, and the insets in the center and the right figures in Fig. 3 show this randomness in the adsorbates' positions in the initial structures (orange). Also, most of the metal surfaces in both datasets were initially relaxed before the adsorbates were added, but the presence of adsorbates near the surfaces changes the local environment of the surface atoms and their positions. About 35%–40% reduction in the losses for surface atoms in both datasets shows that DOGSS was able to capture the surface rearrangements induced by the presence of adsorbates. More importantly, DOGSS could approximate ground state positions of adsorbates for given initial structures and reduce the initial losses for H and CO adsorbate atoms by approximately 58% and 55%, respectively.

We studied the impact of different potentials such as the Morse potential and the Lennard-Jones (LJ) potential on the performance of DOGSS, and Table S4 in the Supplemental Material [23] summarizes the results. In general, the harmonic potential with the gradient descent we proposed provides a better approximation of the ground state structures. The Morse potential and the LJ potential encounter the vanishing gradient problem that causes a bottleneck in the energy minimization using the simple gradient descent algorithm when atoms are too far apart from their neighbor atoms.

Additional experiments were also conducted to investigate DOGSS's ability to generalize to new, unseen, inorganic structures that contain new elements using leave-one-out and leave-two-out data splitting methods. Table S5 in the Supplemental Material [23] summarizes the performance of DOGSS on new structures. The prediction loss slightly increases as DOGSS gets tested on structures containing more new elements, but in general, DOGSS could generate reasonably approximated ground state structures of those unseen structures. DOGSS could reduce the initial loss of H and CO adsorbate atoms by approximately 40%–45% and the initial loss of surface atoms in the bare surface dataset by approximately 30%–40% depending on the number of new elements in the test structures.

Further, in Fig. 4 we compare the relative properties of the initial structures in the test sets and the corresponding predicted structures by DOGSS to the ground state

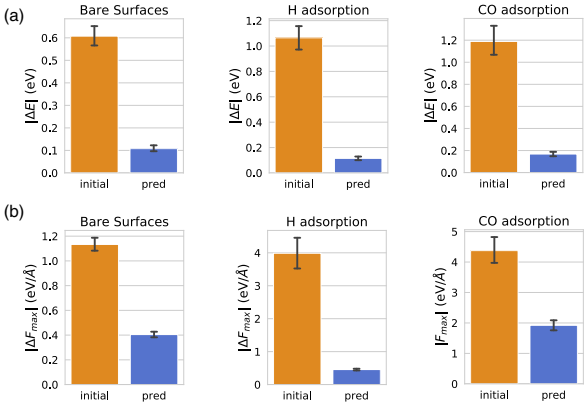

FIG. 4. Comparison of properties of the initial structures (orange) and the predicted ground state structures (blue) by DOGSS for the three datasets: (left) bare surfaces, (center) H adsorption, and (right) CO adsorption. (a) Average absolute differences between total energies of the initial structures or the predicted structures and the ground state energies. (b) Average maximum interatomic forces in each of the initial structures and the predicted ground state structures. Error bars represent the 95% confidence interval.

TABLE II. The average number of steps in DFT relaxations of randomly sampled initial and DOGSS-predicted structures from test sets in the three different datasets.

		Average DFT steps	
Dataset	Sample size	Initial	Prediction
Bare surfaces	601	14.3	10.1
H adsorption	480	27.3	15.0
CO adsorption	540	104.8	50.8

structures. Figure 4(a) shows that DOGSS reduces the difference in the total energies between the initial structures and the ground state structures $|\Delta E_{\text{initial}}|$ for the three datasets by a factor of 5.63 (bare surfaces), 9.41 (H adsorption), and 7.07 (CO adsorption). We find that the total energy differences between the predicted structures and the ground state structures $|\Delta E_{\text{pred}}|$ in all three datasets are close to 0.1 eV. Similarly, Fig. 4(b) shows that the maximum interatomic forces $|F_{\text{max}}|$ in the initial structures are reduced by a factor of 2.80 (bare surfaces), 8.72 (H adsorption), and 2.28 (CO adsorption) after preprocessing the initial structures using DOGSS. Given these comparisons with the initial structures, DOGSS's predicted structures and their properties are fairly close to the ground states.

To show the feasibility of our proposed workflow as illustrated in Fig. 1, we performed DFT relaxations using randomly sampled predicted structures from test sets of the three different datasets. We used the same DFT settings as the initial structures were relaxed for a direct comparison of the number of steps in DFT relaxations until the maximum interatomic force reached below 0.05 eV/Å. Table II summarizes the results from the test DFT simulations, and example structures are visualized in Fig. S3 in the Supplemental Material [23]. For bare surfaces, it takes about 10.1 ionic relaxation steps to relax the predicted structures, which is about 30% fewer steps compared to the relaxation of the initial surfaces. For relaxations of H and CO adsorbates on the metal surfaces, DOGSS could reduce about 45% and 52% of the total DFT relaxation steps, respectively. We observe that bringing randomly placed adsorbate in the initial structure to the adsorption site during DFT relaxations is a time-consuming process as the relaxations of adsorbate-covered surfaces are significantly more expensive than the relaxations of bare surfaces. Therefore, approximating ground state positions of adsorbates prior to the relaxations could effectively reduce the expensive computational cost of DFT for investigating surface science that involves adsorptions.

In this Letter, we present a deep learning framework that incorporates differentiable optimization within graph neural networks to approximate ground state structures. We demonstrate that the framework provides reliable ground state structure predictions of diverse inorganic multicomponent surfaces with and without the presence of adsorbates (H and CO) and reduces the computational cost of DFT by allowing relaxations of the predicted structures that require fewer optimization steps. Further, the framework is flexible so that one can easily modify the molecular graph representations, neural network architecture, embedded optimization method, and type of potential energy to investigate different systems and achieve different prediction performance. This method is complementary to other published methods to accelerate DFT relaxations using on-the-fly neural network potentials as it learns a better initial guess for the process.

The code for this work is available from Ref. [38]. This research used resources of the National Energy Research Scientific Computing Center, a Department of Energy Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research. In addition, we thank Chrysanthos Gounaris, Isaac Tamblyn, and Seoin Back for helpful comments and discussions.

[1] J. K. Nørskov, T. Bligaard, J. Rossmeisl, and C. H. Christensen, Towards the computational design of solid catalysts, Nat. Chem. **1**, 37 (2009).
[2] S. M. Woodley and R. Catlow, Crystal structure prediction from first principles, Nat. Mater. **7**, 937 (2008).
[3] A. Jain, Y. Shin, and K. A. Persson, Computational predictions of energy materials using density functional theory, Nat. Rev. Mater. **1**, 15004 (2016).
[4] B. Hammer, L. B. Hansen, and J. K. Nørskov, Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals, Phys. Rev. B **59**, 7413 (1999).
[5] A. J. R. Hensley, K. Ghale, C. Rieg, T. Dang, E. Anderst, F. Studt, C. T. Campbell, J.-S. McEwen, and Y. Xu, DFT-based method for more accurate adsorption energies: An adaptive sum of energies from RPBE and vdW density functionals, J. Phys. Chem. C **121**, 4937 (2017).
[6] Z. W. Seh, J. Kibsgaard, C. F. Dickens, I. Chorkendorff, J. K. Nørskov, and T. F. Jaramillo, Combining theory and experiment in electrocatalysis: Insights into materials design, Science **355**, eaad4998 (2017).
[7] A. R. Oganov, C. J. Pickard, Q. Zhu, and R. J. Needs, Structure prediction drives materials discovery, Nat. Rev. Mater. **4**, 331 (2019).
[8] V. Milman, B. Winkler, J. A. White, C. J. Pickard, M. C. Payne, E. V. Akhmatskaya, and R. H. Nobes, Electronic structure, properties, and phase stability of inorganic crystals: A pseudopotential plane-wave study, Int. J. Quantum Chem. **77**, 895 (2000).
[9] V. L. Deringer, M. A. Caro, and G. Csányi, Machine learning interatomic potentials as emerging tools for materials science, Adv. Mater. **31**, 1902765 (2019).

[10] V. L. Deringer, C. J. Pickard, and G. Csányi, Data-Driven Learning of Total and Local Energies in Elemental Boron, Phys. Rev. Lett. **120**, 156001 (2018).
[11] J. Behler, Constructing high-dimensional neural network potentials: A tutorial review, Int. J. Quantum Chem. **115**, 1032 (2015).
[12] E. V. Podryabinkin, E. V. Tikhonov, A. V. Shapeev, and A. R. Oganov, Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning, Phys. Rev. B **99**, 064114 (2019).
[13] K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko, and K.-R. Müller, Schnetpack: A deep learning toolbox for atomistic systems, J. Chem. Theory Comput. **15**, 448 (2019).
[14] K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, and K.-R. Müller, SchNet: A continuous-filter convolutional neural network for modeling quantum interactions, arXiv:1706.08566.
[15] T. Xie and J. C. Grossman, Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties, Phys. Rev. Lett. **120**, 145301 (2018).
[16] S. Back, J. Yoon, N. Tian, W. Zhong, K. Tran, and Z. W. Ulissi, Convolutional neural network of atomic surface structures to predict binding energies for high-throughput screening of catalysts, J. Phys. Chem. Lett. **10**, 4401 (2019).
[17] G. H. Gu, J. Noh, S. Kim, S. Back, Z. Ulissi, and Y. Jung, Practical deep-learning representation for fast heterogeneous catalyst screening, J. Phys. Chem. Lett. **11**, 3185 (2020).
[18] J. Klicpera, J. Groß, and S. Günnemann, Directional message passing for molecular graphs, arXiv:2003.03123.
[19] B. Chen, R. Barzilay, and T. Jaakkola, Path-augmented graph transformer network, arXiv:1905.12712.
[20] T. Xie and J. C. Grossman, Hierarchical visualization of materials space with graph convolutional neural networks, J. Chem. Phys. **149**, 174111 (2018).
[21] T. Xie, A. France-Lanord, Y. Wang, Y. Shao-Horn, and J. C. Grossman, Graph dynamical networks for unsupervised learning of atomic scale dynamics in materials, Nat. Commun. **10**, 2667 (2019).
[22] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams, Convolutional networks on graphs for learning molecular fingerprints, in Advances in Neural Information Processing Systems Vol. 28, edited by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett (Curran Associates, Inc., 2015), pp. 2224–2232, http://papers.nips.cc/paper/5954-convolutional-networks-on-graphs-for-learning-molecular-fingerprints.pdf.
[23] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.125.173001 for detailed description of the molecular graph representation, the hyperparameters of the graph neural networks, and datasets preparation. Additional experiments with different types of potential and different data splitting methods are also included in the Supplemental Material. In addition, a summary of training a force field model (SchNet) and example initial, prediction, and ground state structures are visualized, which includes Refs. [14,15,24–28].
[24] I. Loshchilov and F. Hutter, Decoupled weight decay regularization, arXiv:1711.05101.
[25] S. Clark and P. Hayes, https://sigopt.com (2019).
[26] K. Tran and Z. W. Ulissi, Active learning across intermetallics to guide discovery of electrocatalysts for CO_{2_i} reduction and H_2 evolution, Nat. Catal. **1**, 696 (2018).
[27] Z. Cao, Y. Dan, Z. Xiong, C. Niu, X. Li, S. Qian, and J. Hu, Convolutional neural networks for crystal material property prediction using hybrid orbital-field matrix and magpie descriptors, Crystals **9**, 191 (2019).
[28] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, Commentary: The materials project: A materials genome approach to accelerating materials innovation, APL Mater. **1**, 011002 (2013).
[29] J. Behler and M. Parrinello, Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces, Phys. Rev. Lett. **98**, 146401 (2007).
[30] A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials, J. Comput. Phys. **285**, 316 (2015).
[31] M. Gastegger, L. Schwiedrzik, M. Bittermann, F. Berzsenyi, and P. Marquetand, wACSF–weighted atom-centered symmetry functions as descriptors in machine learning potentials, J. Chem. Phys. **148**, 241709 (2018).
[32] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons, Phys. Rev. Lett. **104**, 136403 (2010).
[33] A. P. Bartók, R. Kondor, and G. Csányi, On representing chemical environments, Phys. Rev. B **87**, 184115 (2013).
[34] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, Machine learning of accurate energy-conserving molecular force fields, Sci. Adv. **3**, e1603015 (2017).
[35] J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques, Recent advances and applications of machine learning in solid-state materials science, npj Comput. Mater. **5**, 83 (2019).
[36] M. Fey and J. E. Lenssen, Fast graph representation learning with PyTorch Geometric, in *Proceedings of the ICLR Workshop on Representation Learning on Graphs and Manifolds* (2019) [arXiv:1903.02428].
[37] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, Neural message passing for quantum chemistry, in *Proceedings of the 34th International Conference on Machine Learning, ICML'17* (PMLR, 2017), Vol. 70, pp. 1263–1272, http://proceedings.mlr.press/v70/gilmer17a/gilmer17a.pdf.
[38] https://github.com/Open-Catalyst-Project/baselines.