Graph-based machine learning beyond stable materials and relaxed crystal structures

Filip Ekström Kelvinius, Rickard Armiento, and Fredrik Lindsten

1Division of Statistics and Machine Learning, Department of Computer and Information Science (IDA), Linköping University, Linköping 581 83, Sweden
2Theoretical Physics, Department of Physics, Chemistry, and Biology (IFM), Linköping University, Linköping 581 83, Sweden

(Received 30 June 2021; revised 21 September 2021; accepted 18 February 2022; published 11 March 2022)

There has been a recent surge of interest in using machine learning to approximate density functional theory in materials science. However, many of the most performant models are evaluated on large databases of computed properties of, primarily, materials with precise atomic coordinates available, and which have been experimentally synthesized, i.e., which are thermodynamically stable or metastable. These aspects provide challenges when applying such models on theoretical candidate materials, for example for materials discovery, where the coordinates are not known. To extend the scope of this methodology, we investigate the performance of the crystal graph convolutional neural network on a data set of theoretical structures in three related ternary phase diagrams (Ti,Zr,Hf)-Zn-N, which thus include many highly unstable structures. We then investigate the impact on the performance of using atomic positions that are only partially relaxed into local energy minima. We also explore options for improving the performance in these scenarios by transfer learning, either from models trained on a large database of mostly stable systems, or a different but related phase diagram. Models pretrained on stable materials do not significantly improve performance, but models trained on similar data transfer very well. We demonstrate how our findings can be utilized to generate phase diagrams with a major reduction in computational effort.

DOI: 10.1103/PhysRevMaterials.6.033801

I. INTRODUCTION

Discovering new materials is a driving force for new technologies. With the increase in computational resources, there has been a surge in available data from automated high-throughput materials simulations, obtained using supercomputers. The computed properties have been made available in online databases where one can search for materials with specific desired properties, which enables new applications.

A central concern in the design of a new material is its thermodynamical stability. This is determined by the formation energy of a phase in relation to that of other phases at the same composition, as well as the most stable combination of competing phases it can decompose into. Hence, an extensive, in principle exhaustive, search of the corresponding chemical space is needed to properly determine stability. A common computational method for this purpose is density functional theory (DFT) [2,3], which on the semilocal level of theory is usually considered as having a fairly low computational cost compared to other quantum-mechanics-based methods, but nevertheless limits how many materials that can be screened. Using machine learning as an approximation of conventional simulation methods such as DFT has recently gained interest [4–6]. Many methods are reported with impressive results with prediction errors close to or lower than that of DFT [7], and their low computational effort makes them highly interesting for complementing or replacing quantum-mechanical-based calculations for, e.g., computing formation energies.

Much effort has been put into finding suitable representations, so called descriptors, of crystal structures for use in machine learning methods. Early works include, for example, smooth overlap of atomic positions [8], three descriptors inspired from the Coulomb matrix description of molecules [9], descriptors based only on stoichiometry [10], and the many-body tensor representation [11].

An alternative approach to predicting material properties from a descriptor of the crystal structure is to learn an interatomic potential (see, e.g., Refs. [12–15]). The use of interatomic potentials are perhaps more associated with molecular dynamics simulations, but they can also be used to relax the atomic positions and predict the formation energies of competing phases in a phase diagram [16–18].

Recently, much work has been directed toward so called feature learning which aims to find machine learning models that can learn both suitable descriptors and the target value from these descriptors. In particular, the use of neural networks for graph data, so called graph neural networks (GNNs), has gained interest in chemistry and physics. One motivation for this is that molecules can naturally be described as graphs [19,20], but a pioneering work by Xie and Grossman [21] shows that this type of model is also very appealing.
for crystals. Xie and Grossman demonstrate the predictive performance of their crystal graph convolutional neural network (CGCNN) by predicting multiple properties of crystals, including the formation energies of 9350 crystals from the Materials Project database [22]. A number of works have followed (see, e.g., Refs. [23–25]) which all develop GNNs for crystal structures and which are evaluated using data from the Materials Project. However, most of the available databases of computational materials properties, such as the Materials Project, are based on structures found experimentally, e.g., as reported in the Inorganic Crystal Structure Database (ICSD) [26] or Crystallography Open Database [27–29]. This gives a bias toward stable, or close to stable, materials. When searching for new materials by screening a large number of theoretically generated structures, we need a method which can reliably predict the formation energy also for highly unstable materials. Indeed, a concurrent study by Pandey et al. [30] shows that models trained on ICSD data perform poorly when applied to unstable materials, and we aim to investigate further how to cope with this issue.

Candidate materials for a phase diagram are typically created by substituting atoms into the crystal structures of other, already known materials. The atomic positions of these candidate materials can be far away from the local energy minimum. If one seeks the formation energy of the relaxed structure, the model has to, in principle, perform two tasks. Not only does it have to predict the formation energy, but it needs to do so by overcoming the unrelaxed structure. Previous studies indicate that for GNNs, which are relying on information about the geometry of the structure, predicting formation energies from unrelaxed structures indeed affects the performance [25,31]. Zuo et al. [32] try to avoid this by using machine learning to relax the structures before predicting their energies, while Pal et al. [33] have designed a model which is invariant to the scaling of the structures with the purpose of making it more robust to unrelaxed structures.

The performance of graph neural networks on highly unstable material phases, as needed to predict phase diagrams, has so far not been considered in great detail. Neither has it been thoroughly studied how the predictive performance of a model is affected when predicting formation energies of structures which are not completely relaxed, and how it is improved when the structure moves closer to the relaxed geometry. To investigate this, we perform an empirical evaluation, using the CGCNN model and a data set with theoretical structures from three related ternary phase diagrams, Hf-Zn-N, Ti-Zn-N, and Zr-Zn-N generated by Tholander et al. [34]. This data set was generated by substituting chemical elements into known crystal structures, and these materials therefore span a wide range of formation energies and are for the most part highly unstable. We first investigate the performance when CGCNN is trained with this data set of mostly unstable materials. Then, we investigate how the performance changes when the atomic positions have only been partially relaxed into local minima.

To cope with the computational cost related to obtaining training data, we explore transfer learning. Previous applications of transfer learning in materials science include using multifidelity data to improve predictive performance on high-fidelity properties, like predicting experimental band gaps by using information about DFT band gaps [35] and combining calculations from different exchange-correlation functionals [36]. Graph neural networks have been trained to predict properties where the availability of data is high, and have then been transferred to predict properties with lower data availability [24,37]. In our work, we instead envision the use of transfer learning when one would like to explore compounds from a limited chemical space by incorporating data from outside this space. For this setting, we train a model using data from one phase diagram, and then fine-tune this with data from the phase diagram of interest. We also compare this with pretraining a model on a large database with a variety of mostly stable materials.

We conclude our findings by using CGCNN to speed up the generation of phase diagrams. CGCNN predicts the formation energies of partially relaxed structures between steps in the relaxation process, and structures far from the convex hull are discarded and excluded from further relaxation. Thus, the total computational cost is reduced.

II. BACKGROUND

A. Composition phase diagrams and thermodynamical stability

A composition phase diagram shows the stable phase of a material as a function of its composition. A phase is stable if its formation energy is lower than any linear combination of competing phases. This means that the stability of each phase can be determined by constructing a convex hull of the phases in the phase diagram [38,39]. If a phase belongs to the convex hull it is stable; otherwise, it will decompose into the linear combination of phases on the hull that decreases the total energy. To generate a phase diagram without any prior knowledge of which phases are stable, a large number of potential materials needs to be considered since one in principle needs to determine the formation energy of every possible phase at every composition.

These types of phase diagrams are highly relevant in materials design since finding materials that are stable is a central aspect. Additionally, they solve the crystal structure prediction problem; i.e., for a given composition, the corresponding crystal structure can be determined.

Computational methods, in particular DFT, are often used to predict formation energies of unknown phases. To generate the potential materials, other elements are substituted into already known materials [40,41]. This procedure generates hypothetical materials for which the exact geometry is unknown. Hence, before their formation energy can be calculated, they have to go through a relaxation procedure which typically consists of a number of steps where the cell volume, cell shape, and atom positions are changed into the local minimum of the formation energy.

B. Graph neural networks

Graph neural networks (GNNs) are a special kind of artificial neural network designed to handle graph structured data. Let G be a graph defined by \((V, E)\) where \(V\) is the set of nodes and \(E\) is the set of edges. A node \(v \in V\) is associated with a feature vector \(x_v\), and an edge \((v, w) \in E\) with a feature vector \(e_{vw}\). A GNN is then a parametrized function \(f_\theta(\cdot)\) that
to lack of knowledge. This arises in machine learning since a model is trained from a finite training data set, resulting in (epistemic) uncertainty about the parameter values themselves. For example, an ensemble could consist of $N$ different neural networks, trained on the same data but with random initializations and stochastic optimization. Thus, $N$ plausible parameter configurations $\{\theta_i^{\text{N}}\}$, are obtained, which gives $N$ different predictions. Using the spread of the predictions of the ensemble members has proven to be an efficient way of quantifying epistemic uncertainty for neural networks [44,45].

D. Data sets

In this work we use a data set of Tholander et al. [34] (TAATA, after the name of the authors). This data set contains DFT calculations, including structure relaxation, for a large number of materials phases in the phase diagrams Hf-Zn-N, Ti-Zn-N, and Zr-Zn-N. These three phase diagrams make up the three subsets TAATA-Hf, TAATA-Ti, and TAATA-Zr. The three subsets consist of 4839 (TAATA-Hf), 4328 (TAATA-Ti), and 3650 (TAATA-Zr) materials, each with three different relaxation steps: preprerelax, prerelax, and final relax, with final relax being the most accurate level. These relaxations were performed using the projector augmented wave method (PAW) [46] and Perdew-Burke-Ernzerhof generalized gradient approximation (PBE-GGA) [47] as exchange-correlation functional, implemented in the Vienna Ab-initio Simulation Package [48,49] (VASP) (v5.4.1). All calculations used a 600 eV plane wave energy cutoff. In the first relaxation step, preprerelax, only the unit cell volume is relaxed, which requires on average approximately 5 core minutes/phase. In the prerelax and final relaxation steps, volume, cell shape, and the internal degrees of freedom describing the atomic coordinates are relaxed, with the final step done with higher precision (the VASP-setting PREC set to “Accurate” instead of “Normal,” which controls the number of grid points in the Fourier transform grids and the accuracy of the PAW projector representation in real space). These two steps require approximately 60 core minutes/phase (prerelax) and 540 core minutes/phase (final).

We have also used a model trained on a data set extracted from the Materials Project by Xie and Grossman [34] (the model can be found together with the CGCNN code [50]). This training data set consists of roughly 37 000 materials with completely relaxed structures, covering 87 elements, and each material consisting of up to seven different elements. Apart from covering a different chemical space, this data set has not been produced with the same computational parameters as TAATA.

E. Transfer learning

A machine learning model requires data from DFT calculations to learn to approximate DFT, and typically improves with more data. This means that there is a trade-off between the computational expense of producing DFT simulations (training data) and model accuracy. Relying on training data that, in some sense, is expensive to obtain is common in machine learning (e.g., images annotated by humans according to what they depict), and much effort has therefore been
put into investigating methods to reduce the need for data. One such method is transfer learning [51]. The idea is simple: a model is pretrained on one problem, after which it is fine-tuned by training on data from a different, but similar, problem. This can be seen from the perspective of feature learning as assuming that the two problems could be represented well with more or less the same features. If the problem of interest is suffering from low availability of training data, these features can be learned better from a different problem where more data are available, and then transferred to the new problem domain.

III. METHOD

We use an ensemble of five \((N = 5)\) CGCNNs. Each network is trained on the different subsets of the TAATA data set using an L1 loss function and the Adam optimizer [52]. Similarly to the original work, the optimized hyperparameters are the learning rate, number of layers, and weight decay, using a grid search with completely relaxed structures from all three TAATA subsets. The combination of hyperparameters with the best mean absolute error (MAE) on a validation set was then chosen for all subsequent investigations. For the other hyperparameters, the default values provided in the CGCNN code [50] were used.

The individual ensemble members are trained with access to the same training data. However, some of the training data are reserved to be used only as validation data to avoid overfitting by early stopping. To capture the uncertainty of model parameters (i.e., the epistemic uncertainty), the split of training/validation data is different for the different members, in addition to initializing the model parameters randomly and using stochastic-gradient-based training. When using CGCNN to aid in the generation of a phase diagram (described later in this section), we estimate the total uncertainty as \(\sigma = \sqrt{\sigma_a^2 + \sigma_e^2}\), where \(\sigma_a^2\) and \(\sigma_e^2\) are the aleatoric and epistemic uncertainty respectively. We estimate \(\sigma_a^2\) as the average of the mean squared errors of the ensemble members when predicting formation energies of the validation set during training, and \(\sigma_e^2\) as the variance of the predictions made by the individual ensemble members.

To investigate transfer learning, we have compared two different scenarios. In the first scenario, data are assumed to be available for a phase diagram that is related to an unknown phase diagram of interest. We want to use these data to improve the predictive performance for the unknown phase diagram. This situation could appear, for example, when a study is expanded into a larger chemical space. As a second scenario, related phase diagrams are not available, but we instead try to utilize data from a large database of materials involving all elements like the Materials Project. We explore and compare both of these scenarios by (i) transferring a model pretrained on one TAATA phase diagram by fine-tuning with data of another; (ii) transferring a model trained by Xie and Grossman [21] on Materials Project data (available online [50]) by fine-tuning with data from one of the TAATA phase diagrams.

Finally, we illustrate how our findings can be used to accelerate the generation of a phase diagram with CGCNN using the following scheme: Start by creating a set of candidate structures, \(S\), e.g., by substituting desired chemical elements in structure prototypes available in present materials databases. This set is divided into a training set \(S_{\text{train}}\) and an evaluation set \(S_{\text{eval}}\). The structures in \(S_{\text{train}}\) are fully relaxed, through the prepre-, pre-, and final relaxation steps, and their formation energies are obtained. Using the partially relaxed training structures and their final formation energies, two (ensembles of) CGCNN models are trained for the prepre- and prerelease steps, respectively. Then, all the structures in \(S_{\text{eval}}\) are relaxed to just the prerelease step. The prerelease-level CGCNN model predicts the final formation energies of the partially relaxed evaluation structures and a phase diagram is constructed based on these predictions together with all the structures in the training set. Then, all structures in \(S_{\text{eval}}\) which, in the predicted phase diagram, has an energy above the resulting convex hull of stability \(E_{\text{hull}} > 2\sigma\) are removed (with \(\sigma\) being the total uncertainty of the prediction described previously). The remaining structures go through the prerelease step and the procedure of predicting a phase diagram and removing structures far from the convex hull is repeated. The remaining structures are then relaxed through the final step and the phase diagram is constructed using the energies computed by DFT for the structures in \(S_{\text{train}}\) and the nondiscarded structures in \(S_{\text{eval}}\).

The above description gives all steps in the full workflow. For testing the workflow, we used the data from the work by Tholander et al. [34] (the TAATA data set). This means that the first step, substituting chemical elements into structure prototypes, was not done by us. Since the data set also includes structures from the prepre- and prerelease level, we did not need to perform any DFT calculations as part of the present work.

IV. RESULTS

Test sets for evaluating the performance were created by setting aside 20% of the materials for each TAATA subset. For the partially relaxed structures, structures used for training and testing have been relaxed equally far. However, the target formation energy is always the final formation energy, obtained at the last relaxation step.

A. Training on different relaxation levels

Test MAEs for models trained from scratch are shown in Fig. 2. For the largest training set size, i.e., all of the data points not used for testing (2920, 3460, and 3870 materials for TAATA-Zr, TAATA-Ti, and TAATA-Hf, respectively) the test MAE in eV/atom is 0.038 for TAATA-Ti, 0.043 for TAATA-Hf, and 0.046 for TAATA-Zr when using fully relaxed structures. The corresponding test errors for prerelease is 0.11 eV/atom (TAATA-Hf and TAATA-Ti) and 0.12 eV/atom (TAATA-Zr), and for preprerelease 0.20 eV/atom (TAATA-Hf and TAATA-Ti) and 0.21 eV/atom (TAATA-Zr). It is clear that the performance improves with the amount of training data, and in general worsens with more inaccurate descriptions of the atomic positions, as expected.

Comparing these results with using the sine-based Coulomb matrix together with kernel ridge regression (KRR) [9], the Coulomb matrix seems to be very robust with respect...
to the relaxation levels. For example, using the largest training size of 3870 training materials from the TAATA-Hf data set, the test MAE is 0.33 eV/atom for preprerelax, 0.32 eV/atom for prerelax, and 0.31 eV/atom for the final relaxation level. However, these errors are still much larger than those of the CGCNN model. It also seems like the Coulomb matrix with KRR suffers from overfitting for small training sizes with a test MAE of more than 0.9 eV/atom for TAATA-Hf when using 60 training data points (for full details on the results with KRR, see the Supplemental Material [53]). To put the results in some perspective, a model always predicting a formation energy of 0 eV/atom would yield a test MAE of roughly 0.5 eV/atom (for all three subsets, improves that MAE with less than 0.05 eV/atom.)

We further investigate how the predicted formation energies compare to the actual formation energies in the predicted-versus-target scatter plot in Fig. 3 (since the performance is similar between the different phase diagrams, we have only separated the predictions according to relaxation level and not according to phase diagram). Apart from again demonstrating that more accurate structures give better predictions (as can be seen also in Fig. 2), Fig. 3 shows a small tendency of the model to underestimate high formation energies and overestimate lower formation energies for the less accurate relaxation levels.

All predictions using CGCNN have, so far, been the average of the five ensemble members. To demonstrate the improved performance enabled by using an ensemble, we compare the test MAE of the ensemble with those of the individual ensemble members, using the TAATA-Hf data set. The results can be seen in Fig. 4, and demonstrate the power of using ensembles; the MAE of the ensemble is consistently lower than that of the individual members, except for the smaller training sets where individual members occasionally outperform the ensemble. On the other hand, in the case of small training sets, some of the members are also far worse. We obtain the same qualitative results for the other data sets; see the Supplemental Material [53].

B. Transfer learning

Test errors when transferring models to the TAATA-Hf data set are shown in Fig. 5. Apart from pretraining on a different TAATA subset, we have also used a pretrained CGCNN model provided by Xie and Grossman [21], trained on data from the Materials Project (MP) (available online [50]). The results when training a model from scratch are provided as a comparison. For the pretrained models, test MAEs before being fine-tuned are presented in Table I. To put the poor
performance of the pretrained models without fine-tuning into perspective, we reiterate that simply guessing the formation energy as 0 (alternatively, as the mean of the training data) gives a test MAE of roughly 0.5 eV/atom. However, as shown in Fig. 5, as soon as the models are fine-tuned on a small training data set from the actual phase diagram of interest, the errors drop significantly. This is particularly apparent for the models pretrained on the related TAATA data set.

The model pretrained on Materials Project data has been trained on completely relaxed structures but still, before being fine-tuned, performs worse than or similar to guessing 0 eV/atom. This may seem surprising since the model has been trained on a large data set of relaxed structures. It should, however, be kept in mind that the computations for the materials in the Materials Project have been performed with different computational parameters than those used for TAATA. To investigate the performance more closely, the predicted formation energies compared to the actual formation energies for the TAATA-Hf data set are shown in Fig. 6(a). It is clear that without any fine-tuning the model severely underestimates the formation energy for many structures, leading to the high MAE. Fine-tuning with 60 data points [Fig. 6(b)] appears to compensate for this underestimation. We also compare this fine-tuning with training the model from scratch with 60 data points [Fig. 6(c)].

When transferring to the other two data sets, we observe the same qualitative behavior; see the Supplemental Material [53].

C. Using CGCNN for generating a phase diagram

We now demonstrate the scheme described in Sec. III for predicting the full phase diagram of Hf-Zn-N, using a model transferred from Zr-Zn-N by fine-tuning with 60 data points. The development of the phase diagram during structure relaxation is shown in Fig. 7. The final convex hulls of stability are the same as one would arrive at by relaxing and computing formation energies for all candidate structures with DFT. However, by following the suggested scheme, there is a significant reduction in the number of such calculations that are needed. There are 4839 structures in TAATA-Hf, and 60 of them are used for training the model, meaning there are 4779 structures in the initial evaluation set, which are relaxed at the prerelax step. After predicting the initial phase diagram [Fig. 7(a)], 2227 structures are kept and go through the prerelax step. From the predicted phase diagram at this level [Fig. 7(b)], 1210 structures are kept and are completely relaxed. Including the training data, 1270 structures are completely relaxed, instead of all of the 4839 candidate structures. Using the average computation times for the different steps, an equivalent of approximately 1417 complete structure relaxations have been performed, which is less than 30% of the initial number of candidate structures.

We also tried the same scheme for predicting the Hf-Zn-N phase diagram, using a model trained from scratch with 440 training structures. This time, there were 4399 structures in the evaluation set which are relaxed at the prerelax step. The number of materials that continued to the prerelax step was 3408, and 1157 were completely relaxed (meaning 1597 materials were completely relaxed when including the training structures). In total, an equivalent of approximately 1860 complete structure relaxations would have been performed.
FIG. 6. Predicted vs actual formation energies when predicting formation energies of structures from Hf-Zn-N, using a model (a) pretrained on Materials Project, before fine-tuning, (b) pretrained on Materials Project, fine-tuned with 60 data points, and (c) trained from scratch using 60 data points.

which is less than 40% of the total number of candidate structures.

V. DISCUSSION

A. Prediction performance on the TAATA data set

The results in Fig. 2 show that CGCNN performs well when predicting formation energies of the materials in the TAATA data set. The models trained on the largest training sets of relaxed structures show accuracies close to that of the original CGCNN work [21], where a test MAE of 0.039 eV/atom was reported for formation energies of structures in the Materials Project. However, the size of the training set for that model was much larger, roughly 37,000 materials. This indicates that these types of data, where materials are confined to only a few elements and for the most part are highly unstable, are indeed manageable for this model. CGCNN also outperforms KRR with the sine-based Coulomb matrix descriptor, further emphasizing that it is a powerful method, and that GNNs in general are an interesting way to further improve the application of machine learning to materials science.

The worsened performance of the model when the materials are not fully relaxed is not surprising, considering that the model relies on information about the geometry. Nevertheless, the model still seems to be able to extract useful information and learn also from these inaccurate descriptions, since the performance is improving as the training set is growing. Additionally, it seems like the error has not saturated at our largest training sets, indicating that the performance could further improve with more data. This can be seen as an indication that the model is able to overcome the fact that the materials are not relaxed and, in some sense, jointly learn the combination of structure relaxation and prediction of the formation energy.

B. Transfer learning

A model pretrained on one phase diagram performs very poorly when applied to a different phase diagram without fine-tuning, which is perhaps surprising considering the similarity of the phase diagrams in the TAATA data set. We speculate that the reason for this is that the pretrained model has only seen three different atom types, and in particular has not seen data with the new atom type introduced in the data set.

FIG. 7. Development of the predicted phase diagram during structure relaxation after the (a) prere lax, (b) pre relax, and (c) final relaxation level. Blue circles are structures on the convex hull. For the preprocess and pre relax steps, formation energies are predicted using CGCNN for structures not included in the training set. Dark green squares are structures above the convex hull, but with energy above hull within the threshold $2 \sqrt{\sigma^2_e + \sigma^2_p}$ and have therefore been further relaxed. Light red squares are structures above hull with energy above hull outside the threshold, and have thus not been further relaxed. For the final level, all formation energies have been computed with DFT, and red squares indicate structures above the convex hull. The phase diagrams have been plotted with Pymatgen [54].
is transferred to, meaning that it is unable to create a good representation of this element. This suspicion is strengthened by the fact that only a small amount of data for fine-tuning improves the performance considerably.

However, we see the same behavior with poor performance before fine-tuning of the model pretrained on the Materials Project data set, which has seen a large variety of atoms. In Fig. 6(a) we see that in this case, most of the prediction error stems from the model underestimating the formation energy. We believe that one reason for this is the fact that the model is pretrained with mostly stable materials, which in general will have lower formation energies. This would then explain the underestimation in the predicted formation energies for TAATA, which mostly contain unstable materials. However, another reason could be the difference in computational parameters used to produce the data, which could lead to systematic errors. Either way, using only a small amount of training data for fine-tuning improves the performance by compensating for the underestimation of formation energies. We therefore conclude that while CGCNN is capable of handling also unstable structures, it is important that it is exposed to this during training.

Another observation that can be made in Fig. 6 is that after fine-tuning, the errors seem higher for materials with high formation energies, and that this difference is more prominent for the partially relaxed structures [Fig. 6(b)]. This does not seem to be the case when training from scratch [Fig. 6(c)], and a reason for this could be that the pretrained on relaxed structures with lower formation energies is lingering even after fine-tuning. One could think of combining a model trained from scratch with a model pretrained on the Materials Project to obtain the best performance for all formation energies.

The fact that a small amount of extra data gives these significant improvements suggests that, if possible, it can be worthwhile to generate a small annotated training data set for the specific task at hand, instead of relying solely on models trained on some generic database. It seems that when the target data set is a certain phase diagram, pretraining on a data set like the Materials Project improves over training a model from scratch, at least when the data availability is small. However, if possible, pretraining on a similar data set to the one under study improves the performance substantially. For example, when predicting formation energies of relaxed structures from TAATA-Hf the test MAE is 0.39 eV/atom when the model was trained from scratch, using 60 data points. When using a model pretrained on the Materials Project and fine-tuning it with 60 data points, the error lowers to 0.23 eV/atom. If instead using a model pretrained on the TAATA-Zr data set and fine-tuned with 60 data points, the test MAE is as low as 0.049 eV/atom. Screening a large chemical space of similar compounds is not an uncommon application, and our results show that there is a possibility of very high gains for such applications by using transfer learning.

C. Using CGCNN for generating phase diagrams

Our method for generating a phase diagram with the help of carefully constructed CGCNN models is an example how our findings can be utilized. We acknowledge that the methodology in this example can be further improved; we have, for example, just chosen a training set at random, and our threshold for when a structure is removed from the relaxation process is chosen more or less arbitrarily.

Additionally, the workflow does not depend on CGCNN; any predictive model can be used, and using a model such as the one by Pal et al. [33] which is designed to be invariant to scaling could improve the performance in the early stages of the relaxation process.

Either way, we think that the results are promising and indeed demonstrate the potential in using machine learning in combination with DFT in high-throughput screening.

VI. CONCLUSIONS

We have demonstrated that CGCNN can be trained to predict formation energies of materials that belong to a single ternary phase diagram and which for the most part are far from the convex hull. In this scenario, we obtain test MAEs in the range 0.038 to 0.046 eV/atom for three different ternary phase diagrams when using less than 4000 training data points from the respective phase diagram. We have also shown that CGCNN can make informative predictions also when the structures are not fully relaxed, although with loss of performance compared to fully relaxed structures. Using transfer learning, we are able to transfer a model from one phase diagram to another, using very little additional data; in our case 60 materials are sufficient to get highly accurate predictions. We also compare this with transferring a model pretrained on a much larger Materials Project data set. Interestingly, a model pretrained on a different data set performs very poorly, often worse than predicting a constant formation energy for all materials. However, fine-tuning the pretrained model with 60 training data points can improve the performance drastically, especially when the data set used for fine-tuning is similar to the target data. From this we also conclude that even though CGCNN can handle unstable materials from a single ternary phase diagram, it needs to be exposed to such materials during training to perform well. Finally, we have illustrated how these insights could be used to aid in the generation of a phase diagram by, during structure relaxation, using carefully constructed CGCNN models to predict the final formation energies and only continue relaxation of the structures that are predicted to lie close to the convex hull of stability. Even though it is a simple approach, it emphasizes that machine learning has a great potential for speeding up high-throughput screening.

ACKNOWLEDGMENTS

R.A. acknowledges support from the Swedish Research Council (VR) Grant No. 2020-05402 and the Swedish e-Science Centre (SeRC). The computations for the TAATA data set were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at NSC partially funded by the Swedish Research Council through Grant Agreement No. 2018-05973. F.L. acknowledges support from the Swedish Research Council (VR) Grant No. 2020-04122.
the Swedish Foundation for Strategic Research (SSF) Grant No. ICA16-0015, the Wallenberg AI, Autonomous Systems, and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation, and Excellence Center at Linköping Lund in Information Technology (ELLIIT), a strategic research environment funded by the Swedish government.

[1] R. Armiento, Database-driven high-throughput calculations and machine learning models for materials design, in *Machine Learning Meets Quantum Physics*, Lecture Notes in Physics, edited by K. T. Schütt, S. Chmiela, O. A. von Lilienfeld, A. Tkatchenko, K. Tsuda, and K.-R. Müller (Springer International Publishing, Cham, 2020), pp. 377–395.

[2] P. Hohenberg and W. Kohn, Inhomogeneous electron gas, *Phys. Rev.* **136**, B864 (1964).

[3] W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects, *Phys. Rev.* **140**, A1133 (1965).

[4] 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).

[5] G. R. Schleder, A. C. M. Padilha, C. M. Acosta, M. Costa, and A. Fazio, From DFT to machine learning: Recent approaches to materials science—a review, *J. Phys.: Mater.*, **2**, 032001 (2019).

[6] C. Chen, Y. Zuo, W. Ye, X. Li, Z. Deng, and S. P. Ong, A critical review of machine learning of energy materials, *Adv. Energy Mater.* **10**, 1903242 (2020).

[7] 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, Prediction errors of molecular machine learning models lower than hybrid DFT error, *J. Chem. Theory Comput.* **13**, 5255 (2017).

[8] A. P. Bartók, R. Kondor, and G. Csányi, On representing chemical environments, *Phys. Rev. B* **87**, 184115 (2013).

[9] F. Faber, A. Lindmaa, O. A. von Lilienfeld, and R. Armiento, Crystal structure representations for machine learning models of formation energies, *Int. J. Quantum Chem.* **115**, 1094 (2015).

[10] F. A. Faber, A. Lindmaa, O. A. von Lilienfeld, and R. Armiento, Machine Learning Energies of 2 Million Elpasolite (ABC2D6) Crystals, *Phys. Rev. Lett.* **117**, 135502 (2016).

[11] H. Huo and M. Rupp, Unified representation of molecules and crystals for machine learning, arXiv:1704.06439.

[12] W. J. Szlachta, A. P. Bartók, and G. Csányi, Accuracy and transferability of Gaussian approximation potential models for tungsten, *Phys. Rev. B* **90**, 104108 (2014).

[13] D. Dragoni, T. D. Daff, G. Csányi, and N. Marzari, Achieving DFT accuracy with a machine-learning interatomic potential: Thermomechanics and defects in bcc ferromagnetic iron, *Phys. Rev. Mater.*, **2**, 013808 (2018).

[14] A. V. Shapeev, Moment tensor potentials: A class of systematically improvable interatomic potentials, *Multiscale Model. Simul.* **14**, 1153 (2016).

[15] E. V. Podryabinkin and A. V. Shapeev, Active learning of linearly parametrized interatomic potentials, *Comput. Mater. Sci.* **140**, 171 (2017).

[16] 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).

[17] N. Bernstein, G. Csányi, and V. L. Deringer, De novo exploration and self-guided learning of potential-energy surfaces, *npj Comput. Mater.* **5**, 99 (2019).

[18] K. Gubaev, E. V. Podryabinkin, G. L. W. Hart, and A. V. Shapeev, Accelerating high-throughput searches for new alloys with active learning of interatomic potentials, *Comput. Mater. Sci.* **156**, 148 (2019).

[19] 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 28*, edited by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett (Curran Associates, Inc., Red Hook, NY, 2015), pp. 2224–2232.

[20] S. Kearnes, K. McCloskey, M. Berndl, V. Pande, and P. Riley, Molecular graph convolutions: Moving beyond fingerprints, *J. Comput.-Aided Mol. Des.* **30**, 595 (2016).

[21] 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).

[22] 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, The Materials Project: A materials genome approach to accelerating materials innovation, *APL Mater.* **1**, 011002 (2013).

[23] S. Sanyal, J. Balachandran, N. Yadati, A. Kumar, P. Rajagopalan, S. Sanyal, and P. Talukdar, MT-CCGNN: Integrating graph crystal convolutional neural network with multitask learning for material property prediction, arXiv:1811.05660.

[24] C. Chen, W. Ye, Y. Zuo, C. Zheng, and S. P. Ong, Graph networks as a universal machine learning framework for molecules and crystals, *Chem. Mater.* **31**, 3564 (2019).

[25] C. W. Park and C. Wolverton, Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery, *Phys. Rev. Mater.* **4**, 063801 (2020).

[26] G. Bergerhoff, R. Hundt, R. Sievers, and I. D. Brown, The inorganic crystal structure data base, *J. Chem. Inf. Comput. Sci.* **23**, 66 (1983).

[27] R. T. Downs and M. Hall-Wallace, The American Mineralogist crystal structure database, *Am. Mineral.* **88**, 556 (2003).

[28] S. Gražulis, D. Chatteigner, R. T. Downs, A. F. T. Yokochi, M. Quirós, L. Lutterotti, E. Manakova, J. Butkus, P. Moeck, and A. Le Bail, Crystallography Open Database: An open-access collection of crystal structures, *J. Appl. Crystallogr.* **42**, 726 (2009).

[29] S. Gražulis, A. Daškevič, A. Merkys, D. Chatteigner, L. Lutterotti, M. Quirós, N. R. Serebryanaya, P. Moeck, R. T. Downs, and A. Le Bail, Crystallography Open Database
(COD): An open-access collection of crystal structures and platform for world-wide collaboration, Nucleic Acids Res. 40, D420 (2012).

[30] S. Pandey, J. Qu, V. Stevanović, P. S. John, and P. Gorai, Predicting energy and stability of known and hypothetical crystals using graph neural network, Patterns 2, 100361 (2021).

[31] J. Noh, G. H. Gu, S. Kim, and Y. Jung, Uncertainty-quantified hybrid machine learning/density functional theory high throughput screening method for crystals, J. Chem. Inf. Model. 60, 1996 (2020).

[32] Y. Zuo, M. Qin, C. Chen, W. Ye, X. Li, J. Luo, and S. P. Ong, Accelerating materials discovery with Bayesian optimization and graph deep learning, Mater. Today 51, 126 (2021).

[33] K. Pal, C. W. Park, Y. Xia, J. Shen, and C. Wolverton, Scale-invariant machine-learning model accelerates the discovery of quaternary chalcogenides with ultralow lattice thermal conductivity, arXiv:2109.03751.

[34] C. Tholander, C. B. A. Andersson, R. Armiento, F. Tasnádi, and B. Alling, Strong piezoelectric response in stable TiZnN2, ZrZnN2, and HfZnN2 found by \textit{ab initio} high-throughput approach, J. Appl. Phys. 120, 225102 (2016).

[35] M. L. Hutchinson, E. Antono, B. M. Gibbons, S. Paradiso, J. Ling, and B. Meredig, Overcoming data scarcity with transfer learning, arXiv:1711.05099.

[36] C. Chen, Y. Zuo, W. Ye, X. Li, and S. P. Ong, Learning properties of ordered and disordered materials from multi-fidelity data, Nat. Comput. Sci. 1, 46 (2021).

[37] J. Lee and R. Asahi, Transfer learning for materials informatics using crystal graph convolutional neural network, Comput. Mater. Sci. 190, 110314 (2021).

[38] S. P. Ong, L. Wang, B. Kang, and G. Ceder, Li-Fe-P-O2 phase diagram from first principles calculations, Chem. Mater. 20, 1798 (2008).

[39] S. P. Ong, A. Jain, G. Hautier, B. Kang, and G. Ceder, Thermal stabilities of delithiated olivine MPO4 (\(M = \text{Fe, Mn}\)) cathodes investigated using first principles calculations, Electrochem. Commun. 12, 427 (2010).

[40] S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. RühI, and C. Wolverton, The Open Quantum Materials Database (OQMD): Assessing the accuracy of DFT formation energies, npj Comput. Mater. 1, 15010 (2015).

[41] G. Hautier, C. Fischer, V. Ehrlicher, A. Jain, and G. Ceder, Data mined ionic substitutions for the discovery of new compounds, Inorg. Chem. 50, 656 (2011).

[42] 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, edited by D. Precup and Y. W. Teh (PMLR, Sydney, NSW, Australia, 2017), Vol. 70, pp. 1263–1272.

[43] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, SchNet: A deep learning architecture for molecules and materials, J. Chem. Phys. 148, 241722 (2018).

[44] B. Lakshminarayanan, A. Pritzel, and C. Blundell, Simple and scalable predictive uncertainty estimation using deep ensembles, in Advances in Neural Information Processing Systems, Vol. 30, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., Red Hook, NY, USA, 2017), pp. 6402–6413.

[45] Y. Ovadia, E. Fertig, J. Ren, Z. Nado, D. Sculley, S. Nowozin, J. Dillon, B. Lakshminarayanan, and J. Snoek, Can you trust your models uncertainty? Evaluating predictive uncertainty under dataset shift, in Advances in Neural Information Processing Systems, Vol. 32, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., Red Hook, NY, USA, 2019).

[46] P. E. Blöchl, Projector augmented-wave method, Phys. Rev. B 50, 17953 (1994).

[47] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple, Phys. Rev. Lett. 77, 3865 (1996).

[48] G. Kresse and J. Furthmüller, Efficiency of \textit{ab initio} total energy calculations for metals and semiconductors using a plane-wave basis set, Comput. Mater. Sci. 6, 15 (1996).

[49] G. Kresse and J. Furthmüller, Efficient iterative schemes for \textit{ab initio} total-energy calculations using a plane-wave basis set, Phys. Rev. B 54, 11169 (1996).

[50] The code and the pretrained model are from https://github.com/txie-93/cgcnn

[51] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, How transferable are features in deep neural networks? in Advances in Neural Information Processing Systems, Vol. 27, edited by Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger (Curran Associates, Inc., Red Hook, NY, USA, 2014), pp. 3320–3328.

[52] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, in International Conference on Learning Representations (ICLR), San Diego, CA, USA, 2015.

[53] See Supplemental Material at http://link.aps.org supplemental/10.1103/PhysRevMaterials.6.033801 for full details on the results.

[54] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, Python Materials Genomics (pymatgen): A robust, open-source Python library for materials analysis, Comput. Mater. Sci. 68, 314 (2013).