ARTICLES
https://doi.org/10.1038/s43588-020-00002-x

nature computational science

Learning properties of ordered and disordered materials from multi-fidelity data

Chi Chen, Yunxing Zuo, Weike Ye, Xiangguo Li and Shyue Ping Ong ✉

Predicting the properties of a material from the arrangement of its atoms is a fundamental goal in materials science. While machine learning has emerged in recent years as a new paradigm to provide rapid predictions of materials properties, their practical utility is limited by the scarcity of high-fidelity data. Here, we develop multi-fidelity graph networks as a universal approach to achieve accurate predictions of materials properties with small data sizes. As a proof of concept, we show that the inclusion of low-fidelity Perdew–Burke–Ernzerhof band gaps greatly enhances the resolution of latent structural features in materials graphs, leading to a 22–45% decrease in the mean absolute errors of experimental band gap predictions. We further demonstrate that learned elemental embeddings in materials graph networks provide a natural approach to model disorder in materials, addressing a fundamental gap in the computational prediction of materials properties.

In silico predictions of the properties of materials can most reliably be carried out using ab initio calculations. However, their high computational expense and poor scalability have mostly limited their application to materials containing fewer than 1,000 atoms without site disorder. A further rule of thumb is that the more accurate the ab initio method, the higher the computational expense and the poorer the scalability[1–3]. It is therefore no surprise that supervised machine learning (ML) of ab initio calculations has garnered substantial interest as a means to develop efficient surrogate models for materials property predictions[4]. State-of-the-art ML models encode structural information (for example, as graphs[5,6] or local environmental features[7–9]) in addition to composition information, allowing them to distinguish between polymorphs that may have vastly different properties.

Frustratingly, although building ML models from high-accuracy calculations or experiments would yield the greatest value, obtaining sufficient data to reliably train such models is extremely challenging. For example, the number of Perdew–Burke–Ernzerhof (PBE)[10] calculations in large, public databases such as the Materials Project[11] and Open Quantum Materials Database[12] is on the order of 10^5–10^6, while the number of more accurate Heyd–Scuseria–Ernzerhof (HSE)[13] calculations is at least two orders of magnitude fewer. Similarly, while Becke, 3-parameter, Lee–Yang–Parr (B3LYP) calculations are available for millions of molecules[14], 'gold standard' coupled-cluster single-, double-, and perturbative triple-excitations (CCSD(T)) calculations are only available for perhaps thousands of small molecules. The number of high-quality experimental data points is even fewer[15].

A potential approach to address this challenge is through multi-fidelity models[16] that combine low-fidelity data with high-fidelity data. From the handful of previous studies utilizing this approach in materials properties ML, most utilize two-fidelity models with a co-kriging[17] approach, which assumes an approximately linear relationship between targets of different fidelity. The training of co-kriging models scales as $O(N^3)$ (where N is the number of data points), which becomes prohibitively expensive when N exceeds 10,000. Further, these efforts have been limited to specific properties of single structure prototypes[18,19]. Similarly, transfer learning and Δ-learning[20] are either two-fidelity approaches or non-trivial[21] to extend to more than two fidelities. Multi-task neural network models[22] can handle multi-fidelity data and scale linearly with the number of data fidelities, but require homogeneous data that have all properties labeled for all the data, which is rarely the case in materials property data sets.

Graph networks constitute a general, composable deep learning framework that supports both relational reasoning and combinatorial generalization[23]. Previously, we have shown that Materials Graph Network (MEGNet) models substantially outperform prior ML models in predicting the properties of both molecules and crystals[5]. However, previous graph network and other state-of-the-art models, such as the crystal graph convolutional neural networks[6] and SchNet[24], are single fidelity, typically trained on a large PBE-computed data set, and have not been extended to multi-fidelity data sets of various sizes. Further, all prior models have been limited to ordered materials only. Here, we develop multi-fidelity graph networks as a generalized framework for materials property prediction across computational methodologies and experiments for both ordered and disordered materials.

Results

Multi-fidelity graph networks. Figure 1 depicts a schematic of the multi-fidelity graph network framework. The starting point is a natural graph representation of a material, where the atoms are the nodes and the bonds between them are the edges. The input atomic attributes are simply the atomic numbers of the elements passed to a trainable elemental embedding matrix to obtain a length-16 elemental embedding vector, and the bond attributes are the Gaussian-expanded distances. The state attribute vector provides a portal for structural-independent features to be incorporated into the model. Here, the data fidelity level (for example, computational method or experiment) is encoded as an integer and passed to a trainable fidelity embedding matrix to get a length-16 fidelity embedding vector, forming the input state attributes. A graph network model is built by applying a series of graph convolutional layers that sequentially exchange information between atoms, bonds and the state vector. In the final step, the latent features in the output graph are read out and passed into a neural network to arrive at a property prediction. Further details are available in the Methods section.

Department of NanoEngineering, University of California, San Diego, CA, USA. ✉e-mail: ongsp@eng.ucsd.edu

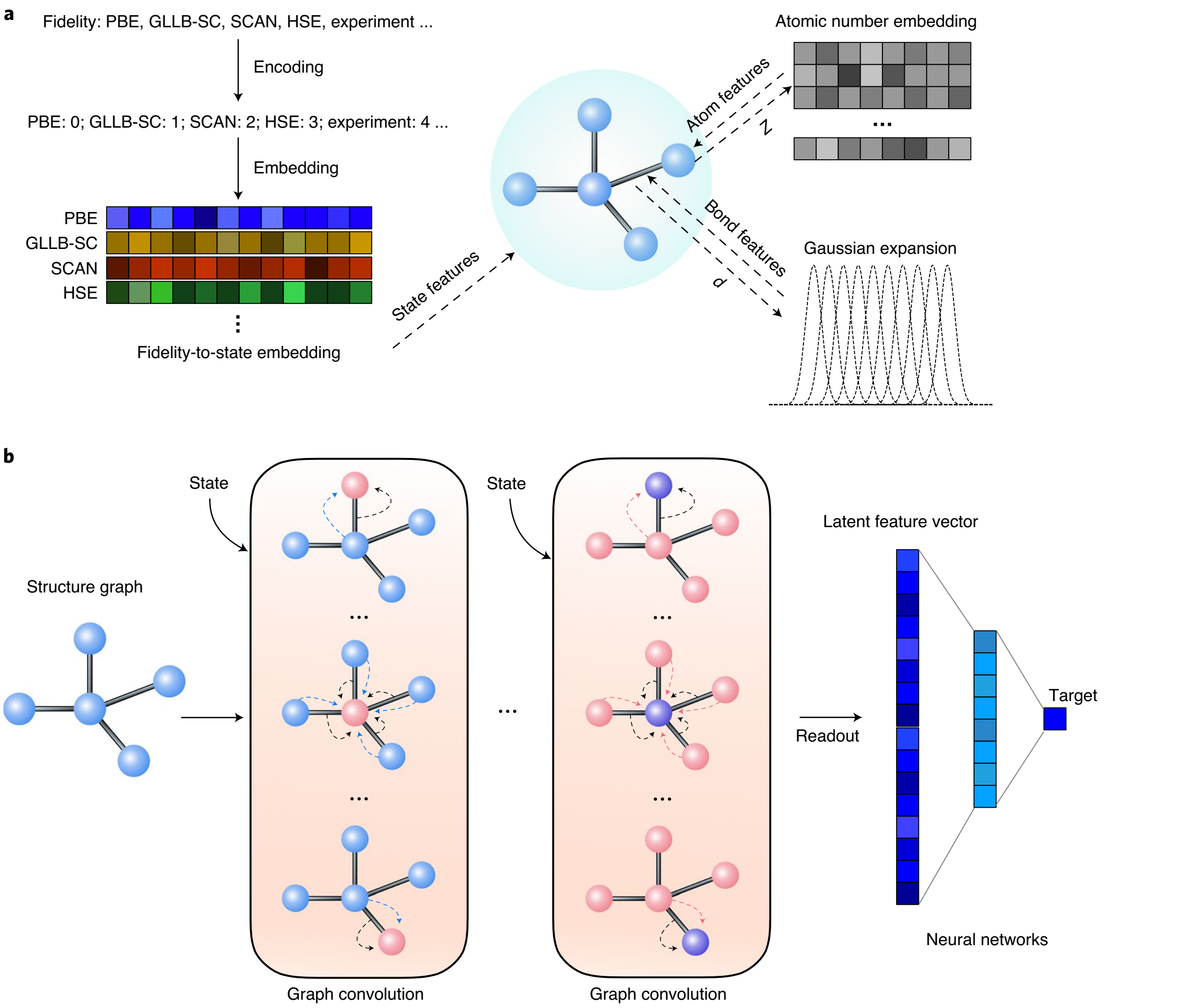


Fig. 1 | Multi-fidelity materials graph networks. a, Representation of a material in a graph network model, with atoms as the nodes, bonds as the edges coupled with a structure-independent global state. The input atomic feature is embedded atomic number of the element. The bond feature vector is the Gaussian-expanded distance. The fidelity of each data is encoded as an integer (for example, 0 for PBE, 1 for GLLB-SC, 2 for SCAN, 3 for HSE and 4 for experiment). **b**, A materials graph network model is constructed by stacking graph convolution layers. In each graph convolution layer, sequential updates of atomic, bond and state features are performed using information from connected neighbors in the graph. The output graph in the last layer is then read out and processed in a neural network to arrive at the final prediction.

We have selected the prediction of the band gap (E_g) of crystals as the target problem because of its great importance in a broad range of technological applications—including photovoltaics and solar water splitting—as well as the availability of multi-fidelity data. To demonstrate the transferability of the approach, another application to the prediction of multi-fidelity molecular energies is demonstrated in Supplementary Section 2. The low-fidelity (low-fi) data set comprises 52,348 PBE band gaps from the Materials Project[11]. The high-fidelity (high-fi) computed data sets comprise 2,290 Gritsenko–Leeuwen–Lenthe–Baerends with solid correction (GLLB-SC)[25–27], 472 strongly constrained and appropriately normed (SCAN)[28,29] and 6,030 HSE[13,30] band gaps. Experimental band gaps for 2,703 ordered crystals and 278 disordered crystals[31] are considered as a separate high-fi data set. The least computationally expensive PBE functional is well known to systematically underestimate the band gap[32], and the high-fi functionals correct this to varying degrees. The data within each fidelity was randomly split into 80% training, 10% validation and 10% test data and repeated six times for all models in this work. The statistics (mean and distribution) of the mean absolute errors (MAEs) on the test data sets are reported, to provide an accurate assessment of model reliability.

Band gaps of ordered structures. Single-fidelity, or 1-fi, graph network models for the band gaps of ordered crystals were first developed for each fidelity in isolation. The MAEs of the 1-fi models (Fig. 2a) are related to the data size as well as the mean absolute deviation (MAD, see Supplementary Data 1) within each data set. The PBE data set is the largest with a small MAD, and the 1-fi PBE models have the smallest average MAE of 0.27 eV. The average MAEs for the computed 1-fi models increase in the order PBE < HSE < GLLB-SC < SCAN, which is inverse to the data-set size order. The lower average MAE of the 1-fi experimental models compared with the 1-fi HSE models, despite having a smaller data set, could be attributed to the relatively large fraction of metals (with zero band gap) in that data set, which leads to a smaller MAD.

Multi-fidelity graph network models utilizing the large PBE data set with data from other fidelities can mitigate this trade-off between data quality/quantity and performance. Much lower average MAEs are achieved across all high-fi computations and

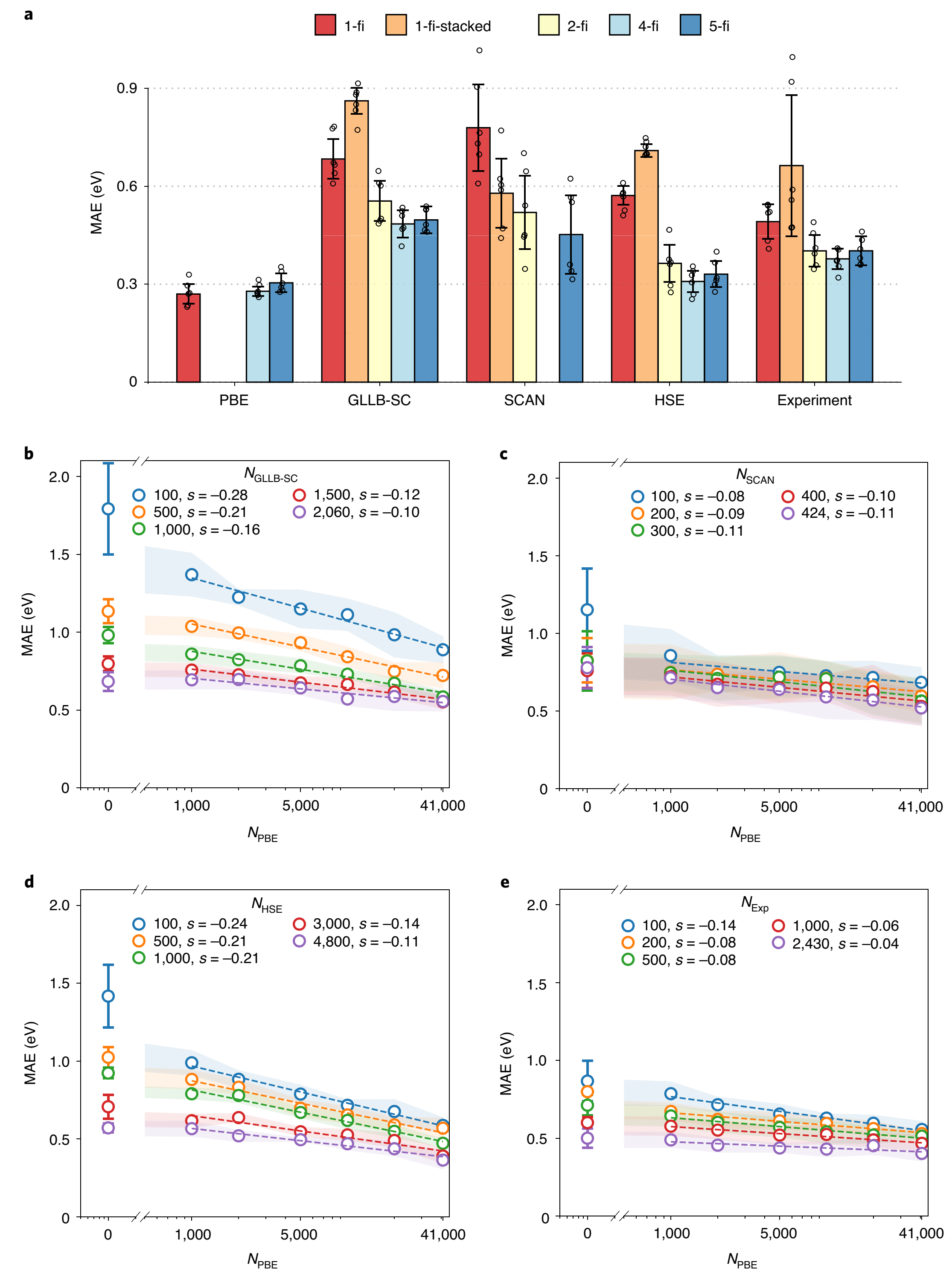


Fig. 2 | Test MAEs of multi-fidelity graph network model predictions on ordered crystal band gaps. a, Performance of graph network models with different fidelity combinations. The 4-fi models used the PBE, GLLB-SC, HSE and experimental data sets, that is, the very small SCAN data set is excluded. All errors were obtained on the corresponding test sets of the fidelity. The error bars show one standard deviation and the dots are the individual model errors. **b**–**e**, Average MAEs of GLLB-SC (**b**), SCAN (**c**), HSE (**d**), and experimental (**e**) band gaps of 2-fi models trained using sampled data sets for high-fidelity data and PBE data. For each sub-plot, the error line is lowered with more high-fidelity data. The x-axis is plotted on a log scale and the shaded areas indicate one standard deviation of the MAE. s indicates the slope for a linear fit of MAE to $\log_{10}N_{PBE}$.

experimental predictions (Fig. 2a). The 5-fi models, that is, the models fitted using all available data, substantially improve on the average MAE on the high-fi predictions over the 2-fi models, at the expense of a small increase in the MAE of the low-fi PBE predictions. The error distributions broken down in metals and non-metals for the 5-fi models are shown in Extended Data Fig. 1.

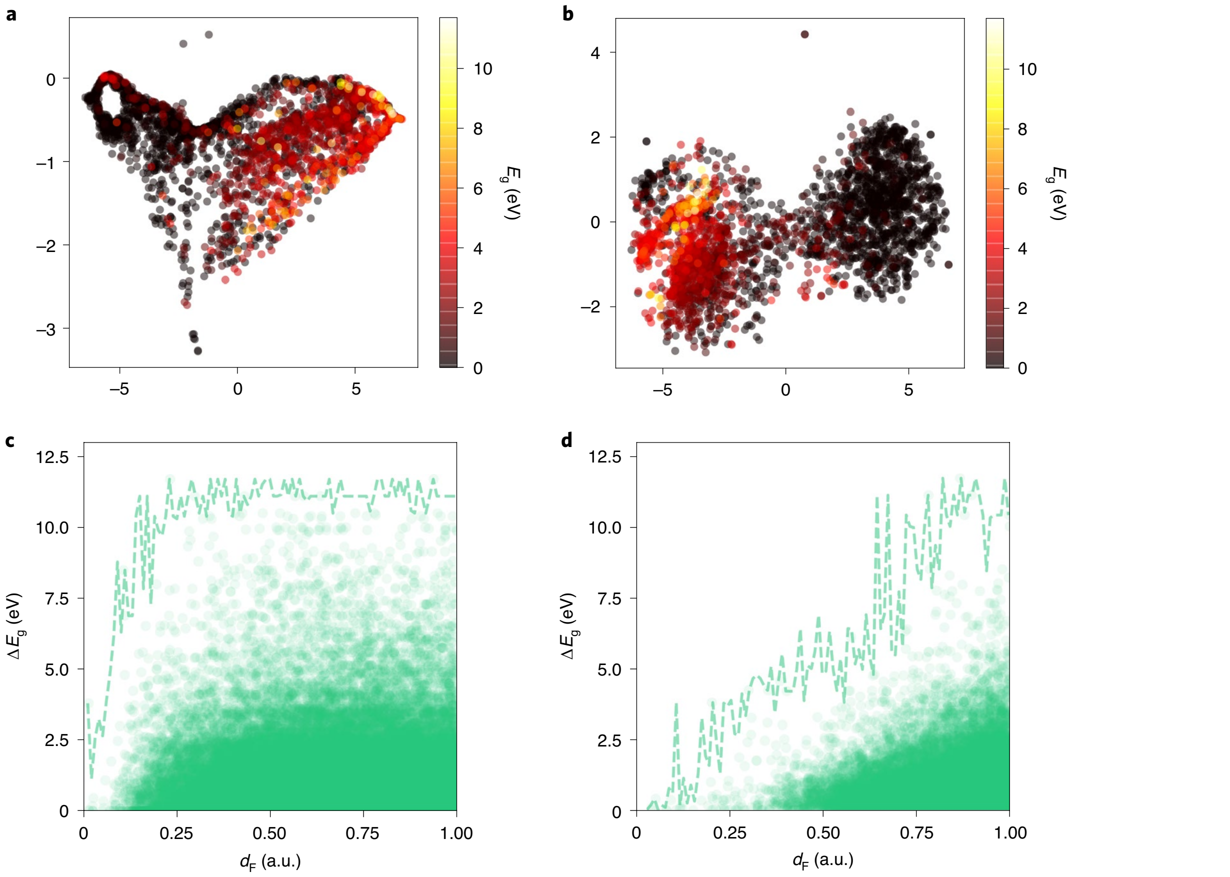


Fig. 3 | Effect of including low-fidelity PBE data on latent structural features. a,b, Two-dimensional *t*-distributed stochastic neighbor embedding (complexity = 1,000) projection of features for 1-fi (**a**) and 2-fi (**b**) models trained using 100 experimental data points and the entire PBE data set. The markers are colored according to the experimental band gap. **c,d**, Plots of the experimental band gap difference (ΔE_g) against normalized latent structural feature distance (d_F) in arbitrary units (a.u.) for the 1-fi (**c**) and 2-fi (**d**) PBE models trained on all available experimental data. The dashed lines indicate the envelope of the maximum ΔE_g at each d_F. The scattering points are sub-sampled by a factor of 15.

With the exception of the extremely small set of SCAN data on metals (17 data points), all model errors have a Gaussian-like distribution centered at zero.

We also explored other 2-fi, 3-fi and 4-fi models, with and without PBE (Supplementary Data 1). The multi-fidelity models without PBE generally have higher MAEs than the multi-fidelity models with PBE, though they typically still outperform the 1-fi models. The 4-fi models that exclude the very small SCAN data set, that is, models trained on PBE, GLLB-SC, HSE and experimental data, outperform the 5-fi models across all non-SCAN fidelities, which indicates that the poor quality of the SCAN data set may have degraded performance. The reduction in average MAE of the 4-fi models over the 1-fi models range from ~22% for the experimental band gap to ~45% for the HSE band gap. Further, an increase in the number of fidelities also tends to improve model consistency, that is, lower the standard deviation in the MAE.

The multi-fidelity graph network models substantially outperform prior ML models in the literature. The best reported GLLB-SC model has a root-mean-squared error (RMSE) of 0.95 eV (ref. [33]), much higher than the average 4-fi RMSE of 0.68 eV on the GLLB-SC predictions. For experimental band gaps, Zhuo et al.[31] reported an MAE of 0.75 eV and an RMSE of 1.46 eV for a test set of ten compounds using a support vector regression model; the average MAE and RMSE for the 4-fi models on the experimental band gap of these compounds are 0.65 eV and 1.39 eV, respectively. Zhuo et al.[31] also reported an RMSE of 0.45 eV on the entire experimental data set, but the data set contains duplicated band gaps for the same composition and thus is an inaccurate benchmark of model performance.

To compare alternative approaches, we have also constructed baseline 1-fi-stacked models, where a linear model is fitted for each high-fi data set to the optimized 1-fi PBE model. This is akin to a model stacking approach and can be justified based on the relatively strong correlation between the high-fi computed and PBE band gaps (Extended Data Fig. 2)[34]. The multi-fidelity models outperform the 1-fi-stacked models, with especially large reductions in average MAEs of up to 38% on arguably the most valuable experimental band gap predictions and 44–56% on the GLLB-SC and HSE predictions. Another alternative approach we explored is transfer learning, whereby the graph convolution layers from the PBE 1-fi models were fixed and the final output layers were retrained with higher fidelity data. Compared with the 2-fi models, the transfer learning models have somewhat lower model errors on the GLLB-SC and SCAN data sets but much higher errors on the most valuable experimental data set, as shown in Supplementary Data 2. It should be noted that transfer learning is fundamentally a two-fidelity approach and requires a two-step training process. The 4-fi models outperform the transfer learning models on the experimental data set even further. These results indicate that multi-fidelity graph network architecture is able to capture complex relationships between data sets of different fidelities. The errors of other combinations of fidelities are listed in Supplementary Data 1.

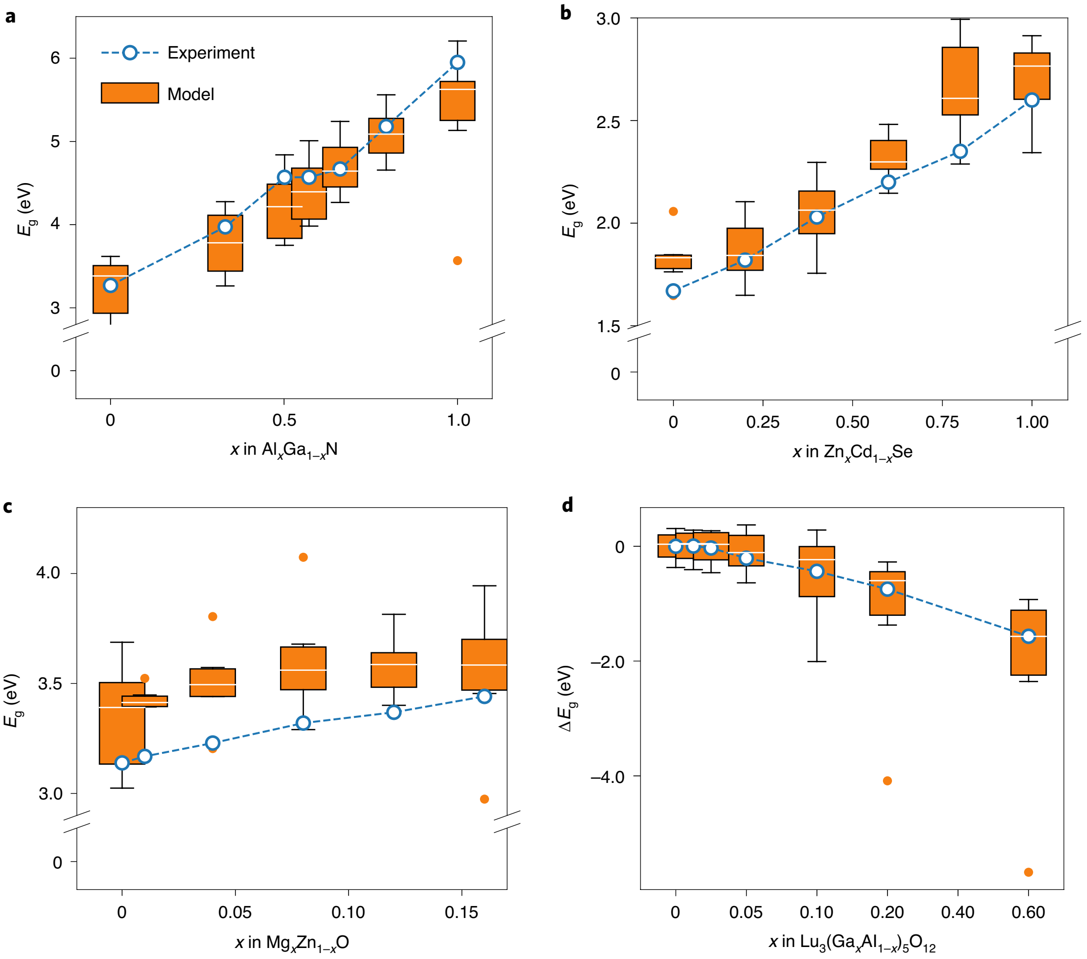


Fig. 4 | Performance of disordered multi-fidelity graph network models. a–c, Predicted and experimental band gaps (E_g) as a function of composition variable x in $Al_xGa_{1-x}N$ (**a**), $Zn_xCd_{1-x}Se$ (**b**) and $Mg_xZn_{1-x}O$ (**c**). **d**, Comparison of the change in band gap with respect to $Lu_3Al_5O_{12}$ (ΔE_g) with x in $Lu_3(Ga_xAl_{1-x})_5O_{12}$. The error bars indicate one standard deviation.

To gain insights into the effect of low-fi and high-fi data size on model accuracy, 2-fi models were developed using sampled subsets of each high-fi computed/experimental data set ($N_{\text{high-fi}}$) together with different quantities of data from the low-fi PBE data set (N_{PBE}). From Fig. 2b–e, it may be observed that adding low-fi PBE data results in a marked decrease in the average MAEs of the high-fi predictions in all cases. The average MAEs of the 2-fi models follow an approximately linear relationship with $\log_{10}N_{\text{PBE}}$. With the exception of the SCAN 2-fi models, the magnitude of the slope decreases monotonically with an increase in $N_{\text{high-fi}}$, that is, the largest improvements are observed in the most data-constrained models. The nearly constant slope for the 2-fi SCAN models may be attributed to the extremely small size of the SCAN data set as well as its strong correlation to the PBE data set (Extended Data Fig. 2).

Latent space visualization. We compared the latent structural features extracted from the 1-fi and the 2-fi models trained using 100 experimental data points without and with PBE data, respectively. The *t*-distributed stochastic neighbor embedding (t-SNE)[35] 2D projections of the latent structure features (Fig. 3a,b) show that the inclusion of the large PBE data set in the 2-fi model results in superior structural representations that clearly separate structures with large band gap differences.

This separation can be further quantified by plotting the band gap difference ΔE_g against the distance in the normalized structural features d_F between all 3,651,753 unique pairs of crystals in the experimental data (as shown in Fig. 3c,d). The 1-fi model for experimental band gaps has poor resolution, especially for large ΔE_g. A wide d_F range from 0.25 to 1 corresponds to a max ΔE_g of ~10 eV, and the correspondence between d_F and max ΔE_g is extremely noisy at low values. By contrast, the 2-fi model exhibits an almost linear correspondence between d_F and max ΔE_g when d_F is less than 1. Our conclusion is therefore that the inclusion of a large quantity of low-fidelity PBE data greatly assists in the learning of better latent structural features, which leads to substantially improved high-fidelity predictions. It should be noted, however, that a prerequisite for such improvements is that the low-fidelity data are of sufficient size and quality. For example, the 2-fi models without PBE perform worse than the 2-fi models with PBE (Supplementary Data 1).

Band gaps of disordered materials. The multi-fidelity graph network architecture also provides a natural framework to address another major gap in the computational materials property predictions—disordered crystals. The majority of known crystals exhibit site disorder. For example, of the ~ 200,000 unique crystals reported in the Inorganic Crystal Structure Database[36], more than 120,000 are disordered crystals. Typically, the properties of disordered crystals are estimated by sampling low-energy structures among a combinatorial enumeration of distinct orderings within a supercell.

In the graph network approach, we can use the learned length-16 elemental embeddings $\mathbf{W}_Z$ directly as the node features. In such a scheme, disordered sites can be represented as a linear combination of the elemental embeddings as $\mathbf{W}_{\text{disordered}} = \sum_{i=1} x_i \mathbf{W}_{Z_i}$, where x_i is the occupancy of species i in the site and $\mathbf{W}_{Z_i}$ is the element embedding for atomic number Z_i. Using the 4-fi models for the ordered crystals without further retraining, the average MAE of the predicted band gaps of the 278 disordered crystals in our experimental data set is a respectable 0.63 ± 0.14 eV, similar to the average MAE of the 1-fi-stacked model on the experimental band gaps of ordered crystals. By retraining with the disordered experimental band gap data set, the average MAE on the experimental band gaps of disordered crystals decreases to 0.51 ± 0.11 eV, while that of the ordered crystals remains approximately the same (0.37 ± 0.02 eV). The average MAEs for a retrained 1-fi model on the experimental band gaps of disordered and ordered crystals are 0.55 ± 0.13 eV and 0.50 ± 0.07 eV, respectively. Clearly, the multi-fidelity approach improves on the performance on disordered crystals as well as enumerated ordered crystals.

To demonstrate the power of the disordered multi-fidelity graph network models, band gap engineering data were extracted from the literature for $Al_xGa_{1-x}N$ (ref. [37]), $Zn_xCd_{1-x}Se$ (ref. [38]), $Mg_xZn_{1-x}O$ (ref. [39]) and $Lu_3(Ga_xAl_{1-x})_5O_{12}$ (ref. [40]). The band gaps of $Lu_3(Ga_xAl_{1-x})_5O_{12}$ were not present and only the band gaps of the stoichiometric endpoints for the other systems were present in our experimental data set. The 4-fi model performs remarkably well, reproducing qualitative trends in all instances and achieving near quantitative accuracy for most systems, as shown in Fig. 4. The 4-fi model reproduces the concave relationship between x and the change in band gap ΔE_g for $Lu_3(Ga_xAl_{1-x})_5O_{12}$ (Fig. 4d) reported previously[40]. For $Zn_{1-x}Mg_xO$, a more pronounced concave relationship is predicted by the 4-fi model compared with the experimental measurements[39]. The band gap of ZnO is notoriously poorly estimated by DFT techniques[41], and even experimental measurements range from 3.1 to 3.4 eV across publications[42]. An additional proof of concept for $Ba_ySr_{1-y}Co_xFe_{1-x}O_{3-\delta}$ perovskite[43], a highly promising catalyst for the oxygen reduction reaction that exhibits disorder on multiple sites, is given in Extended Data Fig. 3.

Discussion

Data quality and quantity constraints are major bottlenecks to materials design. Multi-fidelity graph networks enable the efficient learning of latent structural features using large quantities of cheap low-fidelity computed data to achieve vastly improved predictions for more costly computational methods and experiments. While crystal band gaps have been selected as the model problem in this work, the multi-fidelity graph network framework is universal and readily applicable to other properties and to molecules. Two examples are provided in Extended Data Fig. 4, where a large number of low-fidelity molecular energies are shown to lead to vast improvements in the high-fidelity energy predictions.

The ability to predict band gaps of disordered materials suggests that the learned elemental embeddings in graph network models encode chemistry in a way that naturally interpolates between elements. In a traditional ML model development, each element is represented by a vector of atomic features—for example atomic number, electronegativity and atomic radius. A disordered site, for example Al_xGa_{1-x} in $Al_xGa_{1-x}N$, cannot be naturally represented as a linear combination of such feature vectors. Therefore, this is a unique attribute of our graph network model that is not present in feature-engineered ML models. The interpolation approach bears some similarity to that used in the virtual crystal approximation[44]. A limitation of the virtual crystal approximation is that it may fail in systems that do not exhibit full disorder. In the case of multi-fidelity graph networks, this limitation can be mitigated to a certain extent by applying the highly efficient models to the interpolated disordered crystal as well as ordered crystals at the same composition to arrive at a range of property predictions. Thus, multi-fidelity graph network models provide an alternative approach to in silico materials design for the large class of disordered materials that is extremely difficult to treat with existing ab initio computations or ML techniques.

One potential limitation of the proposed approach is its reliance on large, low-fidelity data sets for learning effective latent structural representations. The low-fi data set needs to reproduce at least general qualitative trends in the target property between different materials for such learning to be effective. For some properties, even low-fi data sets may not be available in sufficiently large quantities for effective learning. Under such instances, a transfer learning approach, that is, where portions of a model trained on a different property are retrained on the target property, may be more appropriate.

Methods

Data collection and processing. The PBE[10] data set comprising 52,348 crystal structures with band structure calculations was obtained from the Materials Project[11] on 1 June 2019 using the Materials Application Programming Interface in the Python Materials Genomics (pymatgen) library[45,46]. The GLLB-SC band gaps from Castelli et al.[27] were obtained via MPContribs[47]. The total number of GLLB-SC band gaps is 2,290 after filtering out materials that do not have structures in the current Materials Project database and those that failed the graph computations due to abnormally long bonds (>5 Å). The GLLB-SC data all have positive band gaps due to the constraints applied in the previous structure selection[27]. The SCAN[28] band gaps for 472 nonmagnetic materials were obtained from a previous study by Borlido et al.[29]. The HSE[13] band gaps with corresponding Materials Project structures were downloaded from the MaterialGo website[30]. After filtering out ill-converged calculations and those with a much smaller HSE band gap compared with the PBE band gaps, 6,030 data points remain, of which 2,775 are metallic. Finally, the experimental band gaps were obtained from work by Zhuo and colleagues[31]. As this data set only contains compositions, the experimental crystal structure for each composition was obtained by looking up the lowest energy polymorph for a given formula in the Materials Project, followed by cross-referencing with the corresponding Inorganic Crystal Structure Database entry[36]. Further, as multiple band gaps can be reported for the same composition in this data set, the band gaps for the duplicated entries were averaged. In total, 2,703 ordered (938 binary, 1,306 ternary and 459 quaternary) and 278 disordered (41 binary, 132 ternary and 105 quaternary) structure–band gap pairs were obtained. All data sets are publicly available[48].

Materials graph networks construction. In materials graph networks, atoms and bonds are represented as nodes and edges in an undirected graph as (V, E, $\boldsymbol{u}$). The atom attributes V are the atomic numbers $Z \in \mathbb{N}$. Each atom attribute is associated with a row vector $\mathbf{W}_{Z_i} \in \mathbb{R}^{16}$ in the element embedding matrix $\mathbf{W}_Z = [\mathbf{W}_0; \mathbf{W}_1; \mathbf{W}_2; \ldots \mathbf{W}_{94}]$ where $\mathbf{W}_0$ is a dummy vector. The bond attribute is the set of Gaussian-expanded distances. For the k-th bond in the structure, the attributes are

$$e_{k,m} = \exp - \frac{(d_k - \mu_m)^2}{\sigma^2}, \forall d_k \leq R_c$$

where d_k is the length of the bond k formed by atom indices r_k and s_k; R_c is the cutoff radius and $\mu_m = \frac{m}{n_{bf}-1}\mu_{max}$ for $m = \{0, 1, 2, \ldots n_{bf} - 1\}$ and n_{bf} is the number of bond features. In this work, $R_c = 5$ Å, $\mu_{max} = 6$ Å, and $n_{bf} = 100$. The graphs are constructed using an edge list representation, and the edge set of the graph is represented as $E = \{(\mathbf{e}_k, r_k, s_k)\}$. The state attributes $\mathbf{u}$ are fidelity levels $F \in \mathbb{N}$. Similar to atom attributes, fidelity F_i is associated with a row vector $\mathbf{W}^f_{F_i}$ in the fidelity embedding matrix $\mathbf{W}_F = [\mathbf{W}^f_0; \mathbf{W}^f_1; \mathbf{W}^f_2; \mathbf{W}^f_3, \mathbf{W}^f_4]$. Both embedding matrices $\mathbf{W}_Z$ and $\mathbf{W}_F$ are trainable in the models, except in disordered models where the elemental embedding matrix $\mathbf{W}_Z$ is fixed to previously obtained values.

In each graph convolution layer, the graph networks are propagated sequentially as follows:

- The attributes of each bond k in the graph are updated as

  $$\mathbf{e}'_k = \phi_e(\mathbf{v}_{s_k} \oplus \mathbf{v}_{r_k} \oplus \mathbf{e}_k \oplus \mathbf{u})$$

  where ϕ_e is the bond update function, $\mathbf{v}_{s_k}$ and $\mathbf{v}_{r_k}$ are the atomic attributes of the two atoms forming the bond k, and ⊕ is the concatenation function.
- Each atom i is then updated as

  $$\mathbf{v}'_i = \phi_v(\bar{\mathbf{v}}^e_i \oplus \mathbf{v}_i \oplus \mathbf{u})$$

  where ϕ_v is the atomic update function, and $\bar{\mathbf{v}}^e_i = \text{average}_k(\mathbf{e}'_k), \forall r_k = i$ is the averaged bond attributes from all bonds connected to atom i.

- Finally, the state attributes are updated as

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

where ϕ_u is the state update function, and $\bar{\mathbf{u}}^e = \mathrm{average}_k(\mathbf{e}'_k)$ and $\bar{\mathbf{u}}^v = \mathrm{average}_i(\mathbf{v}'_i)$ are the averaged attributes from all atoms and bonds, respectively.

Experimental set-up. The models were constructed using TensorFlow[49]. Three graph convolution layers and the Set2Set readout function with two steps are used in the model training[5]. The update functions in each graph convolutional layer were chosen to be multi-layer perceptron models with [64, 64, 32] hidden neurons and shifted softplus function $\ln(e^x + 1) - \ln(2)$ as the non-linear activation function.

We split the data into 80:10:10 train:validation:test ratios randomly for each fidelity independently and repeated the splitting six times. It should be noted that each structure–band gap data point is considered as unique, and there are instances where the same structure with different fidelity band gaps is present in the training and validation/test data. An alternative data-splitting strategy wherein structural overlaps in the training/validation/test sets are disallowed is presented in Supplementary Data 3. It was found that such a data-splitting strategy results in significantly higher model errors, which indicates that information from multi-fidelities is necessary for the models to learn the relationships between different fidelities.

A learning rate of 10^{-3} was used with the Adam optimizer with mean squared error as the loss function. The MAE was used as the error metric on the validation and test data sets, and the batch size for the training was set to 128. All models were trained on the corresponding training data for a maximum of 1,500 epochs. During the training process, the MAE metrics were calculated on the validation data set after each epoch. The model weights were saved after each epoch if the validation MAE reduced. Fitting was deemed to have converged if the validation MAE did not reduce for a consecutive 500 steps. An automatic training recovering mechanism was also implemented by reloading the saved model weights and reducing the learning rate by half when gradient explosion happens. For multi-fidelity model fitting, only the high-fidelity data sets without the PBE data were used in the validation set. The final model performances were evaluated on the test data sets and reported in this work.

Data availability

Multi-fidelity band gap data and molecular data are available at https://doi.org/10.6084/m9.figshare.13040330[48]. The data for all figures and extended data figures are available in Source Data.

Code availability

Model fitting and results plotting codes are available at https://github.com/materialsvirtuallab/megnet/tree/master/multi-fidelity. MEGNet is available at https://github.com/materialsvirtuallab/megnet. The specific version of the package can be found at https://doi.org/10.5281/zenodo.4072029[50].

Received: 20 July 2020; Accepted: 2 November 2020;
Published online: 14 January 2021

References

1. Chevrier, V. L., Ong, S. P., Armiento, R., Chan, M. K. Y. & Ceder, G. Hybrid density functional calculations of redox potentials and formation energies of transition metal compounds. *Phys. Rev. B* **82**, 075122 (2010).
2. Heyd, J. & Scuseria, G. E. Efficient hybrid density functional calculations in solids: assessment of the heyd-scuseria-ernzerhof screened coulomb hybrid functional. *J. Chem. Phys.* **121**, 1187–1192 (2004).
3. Zhang, Y. et al. Efficient first-principles prediction of solid stability: towards chemical accuracy. *npj Comput. Mat.* **4**, 9 (2018).
4. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
5. Chen, C., Ye, W., Zuo, Y., Zheng, C. & Ong, S. P. Graph networks as a universal machine learning framework for molecules and crystals. *Chem. Mat.* **31**, 3564–3572 (2019).
6. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).
7. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
8. Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).
9. Zuo, Y. et al. Performance and cost assessment of machine learning interatomic potentials. *J. Phys. Chem. A* **124**, 731–745 (2020).
10. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
11. Jain, A. et al. Commentary: the materials project: a materials genome approach to accelerating materials innovation. *APL Mat.* **1**, 011002 (2013).
12. Kirklin, S. et al. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Comput. Mat.* **1**, 15010 (2015).
13. Heyd, J., Scuseria, G. E. & Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **118**, 8207–8215 (2003).
14. Hachmann, J. et al. The Harvard Clean Energy Project: large-scale computational screening and design of organic photovoltaics on the World Community Grid. *J. Phys. Chem. Lett.* **2**, 2241–2251 (2011).
15. Hellwege, K. H. & Green, L. C. Landolt-Börnstein, numerical data and functional relationships in science and technology. *Am. J. Phys.* **35**, 291–292 (1967).
16. Meng, X. & Karniadakis, G. E. A composite neural network that learns from multi-fidelity data: application to function approximation and inverse PDE problems. *J. Comput. Phys.* **401**, 109020 (2020).
17. Kennedy, M. C. & O'Hagan, A. Predicting the output from a complex computer code when fast approximations are available. *Biometrika* **87**, 1–13 (2000).
18. Pilania, G., Gubernatis, J. E. & Lookman, T. Multi-fidelity machine learning models for accurate bandgap predictions of solids. *Comput. Mat. Sci.* **129**, 156–163 (2017).
19. Batra, R., Pilania, G., Uberuaga, B. P. & Ramprasad, R. Multifidelity information fusion with machine learning: a case study of dopant formation energies in hafnia. *ACS Appl. Mat. Interfaces* **11**, 24906–24918 (2019).
20. Ramakrishnan, R., Dral, P. O., Rupp, M. & vonLilienfeld, O. A. Big data meets quantum chemistry approximations: The Δ-machine learning approach. *J. Chem. Theory Comput.* **11**, 2087–2096 (2015).
21. Zaspel, P., Huang, B., Harbrecht, H. & von Lilienfeld, O. A. Boosting quantum machine learning models with a multilevel combination technique: Pople diagrams revisited. *J. Chem. Theory Comput.* **15**, 1546–1559 (2019).
22. Dahl, G. E., Jaitly, N. & Salakhutdinov, R. Multi-task neural networks for QSAR predictions. Preprint at https://arxiv.org/abs/1406.1231 (2014).
23. Battaglia, P. W. et al. Relational inductive biases, deep learning, and graph networks. Preprint at https://arxiv.org/abs/1806.01261 (2018).
24. 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.* **148**, 241722 (2018).
25. Gritsenko, O., van Leeuwen, R., van Lenthe, E. & Baerends, E. J. Self-consistent approximation to the Kohn-Sham exchange potential. *Phys. Rev. A* **51**, 1944–1954 (1995).
26. Kuisma, M., Ojanen, J., Enkovaara, J. & Rantala, T. T. Kohn-Sham potential with discontinuity for band gap materials. *Phys. Rev. B* **82**, 115106 (2010).
27. Castelli, I. E. et al. New light-harvesting materials using accurate and efficient bandgap calculations. *Adv. Energy Mat.* **5**, 1400915 (2015).
28. Sun, J., Ruzsinszky, A. & Perdew, J. P. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.* **115**, 036402 (2015).
29. Borlido, P. et al. Large-scale benchmark of exchange-correlation functionals for the determination of electronic band gaps of solids. *J. Chem. Theory Comput.* **15**, 5069–5079 (2019).
30. Jie, J. et al. A new MaterialGo database and its comparison with other high-throughput electronic structure databases for their predicted energy band gaps. *Sci. China Technol. Sci.* **62**, 1423–1430 (2019).
31. Zhuo, Y., Mansouri Tehrani, A. & Brgoch, J. Predicting the band gaps of inorganic solids by machine learning. *J. Phys. Chem. Lett.* **9**, 1668–1673 (2018).
32. Perdew, J. P. & Levy, M. Physical content of the exact Kohn-Sham orbital energies: band gaps and derivative discontinuities. *Phys. Rev. Lett.* **51**, 1884–1887 (1983).
33. Davies, D. W., Butler, K. T. & Walsh, A. Data-driven discovery of photoactive quaternary oxides using first-principles machine learning. *Chem. Mat.* **31**, 7221–7230 (2019).
34. Morales-García, Á., Valero, R. & Illas, F. An empirical, yet practical way to predict the band gap in solids by using density functional band structure calculations. *J. Phys. Chem. C* **121**, 18862–18866 (2017).
35. van der Maaten, L. & Hinton, G. Visualizing data using T-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
36. Hellenbrandt, M. The Inorganic Crystal Structure Database (ICSD)–present and future. *Crystallogr. Rev.* **10**, 17–22 (2004).
37. Chen, H., Chen, K., Drabold, D. A. & Kordesch, M. E. Band gap engineering in amorphous $Al_xGa_{1-x}N$: experiment and *ab initio* calculations. *Appl. Phys. Lett.* **77**, 1117–1119 (2000).
38. Santhosh, T. C. M., Bangera, K. V. & Shivakumar, G. K. Band gap engineering of mixed $Cd_{(1-x)}Zn_{(x)}$ Se thin films. *J. Alloys Compd.* **703**, 40–44 (2017).
39. Rana, N., Chand, S. & Gathania, A. K. Band gap engineering of ZnO by doping with Mg. *Phys. Scr.* **90**, 085502 (2015).
40. Fasoli, M. et al. Band-gap engineering for removing shallow traps in rare-earth $Lu_3Al_5O_{12}$ garnet scintillators using Ga^{3+} doping. *Phys. Rev. B* **84**, 081102 (2011).
41. Harun, K., Salleh, N. A., Deghfel, B., Yaakob, M. K. & Mohamad, A. A. DFT+U calculations for electronic, structural, and optical properties of ZnO wurtzite structure: a review. *Results Phys.* **16**, 102829 (2020).

42. Kamarulzaman, N., Kasim, M. F. & Chayed, N. F. Elucidation of the highest valence band and lowest conduction band shifts using XPS for ZnO and $Zn_{0.99}Cu_{0.01}O$ band gap changes. *Results Phys.* **6**, 217–230 (2016).
43. Shao, Z. & Haile, S. M. A high-performance cathode for the next generation of solid-oxide fuel cells. *Nature* **431**, 170–173 (2004).
44. Nordheim, L. The electron theory of metals. *Ann. Phys* **9**, 607 (1931).
45. Ong, S. P. et al. Python Materials Genomics (pymatgen): a robust, open-source python library for materials analysis. *Comput. Mat. Sci.* **68**, 314–319 (2013).
46. Ong, S. P. et al. The Materials Application Programming Interface (API): a simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles. *Comput. Mat. Sci.* **97**, 209–215 (2015).
47. Huck, P., Jain, A., Gunter, D., Winston, D. & Persson, K. A community contribution framework for sharing materials data with materials project. In *2015 IEEE 11th International Conference on E-Science* 535–541 (2015).
48. Chen, C., Zuo, Y., Ye, W., Li, X. & Ong, S. P. Learning properties of ordered and disordered materials from multi-fidelity data. *Figshare* https://doi.org/10.6084/m9.figshare.13040330 (2020).
49. Abadi, M. et al. TensorFlow: a system for large-scale machine learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* 265–283 (2016).
50. Chen, C., Ong, S. P., Ward, L. & Himanen, L. materialsvirtuallab/megnet v.1.2.3 https://doi.org/10.5281/zenodo.4072029 (2020).

Acknowledgements

This work was primarily supported by the Materials Project, funded by the US 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. The authors also acknowledge support from the National Science Foundation SI2-SSI Program under award no. 1550423 for the software development portions of the work. C.C. thanks M. Horton for his assistance with the GLLB-SC data set.

Author contributions

C.C. and S.P.O. conceived the idea and designed the work. C.C. implemented the models and performed the analysis. S.P.O. supervised the project. Y.Z., W.Y. and X.L. helped with the data collection and analysis. C.C. and S.P.O. wrote the manuscript. All authors contributed to discussions and revisions.

Competing interests

The authors declare no competing interests.

Additional information

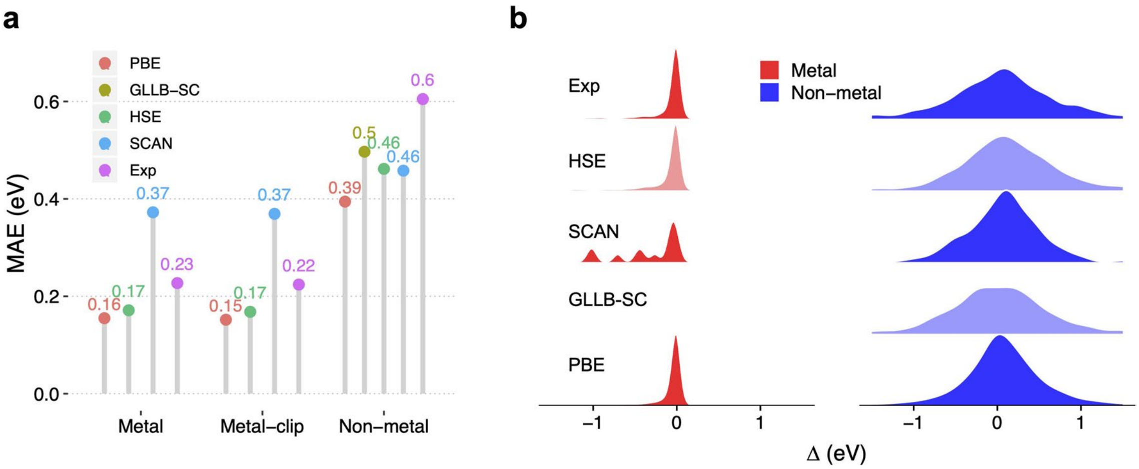
Extended data is available for this paper at https://doi.org/10.1038/s43588-020-00002-x.

Supplementary information is available for this paper at https://doi.org/10.1038/s43588-020-00002-x.

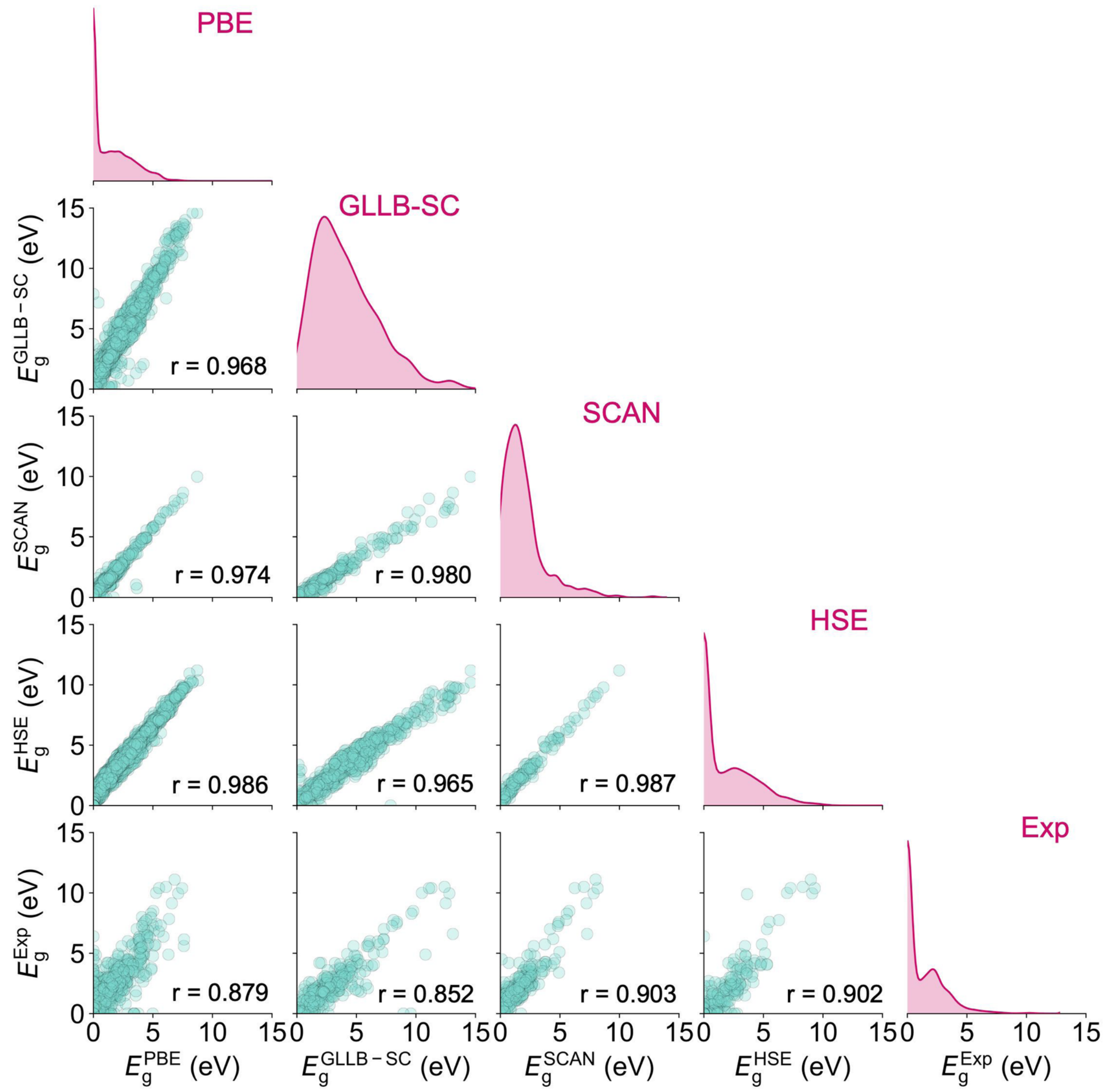
Correspondence and requests for materials should be addressed to S.P.O.

Reprints and permissions information is available at www.nature.com/reprints.

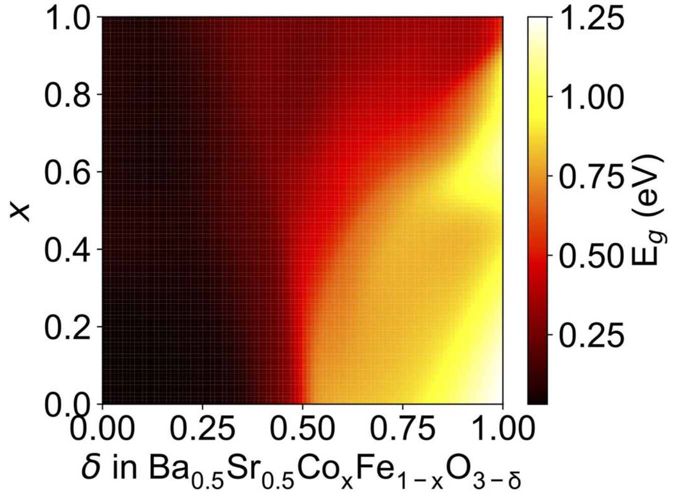
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



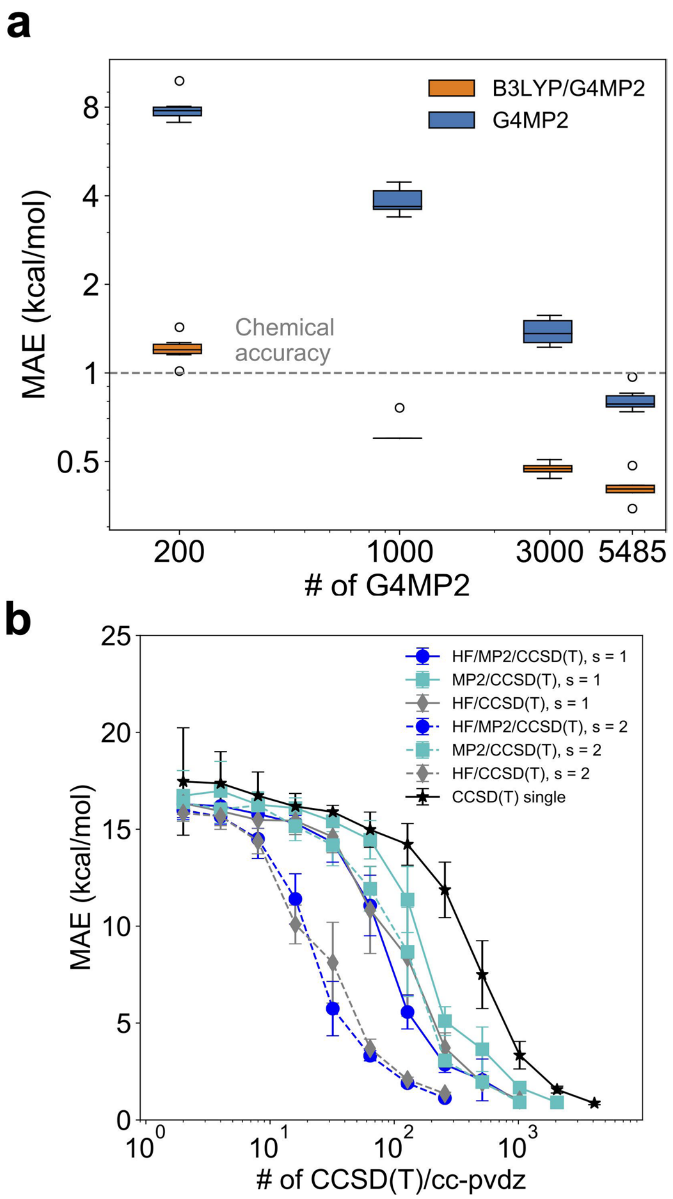
Extended Data Fig. 1 | Five-fidelity model test error distributions. a, The model errors decomposed into metals vs non-metals and (**b**) the test error distributions. The 'metal-clip' category means that the predicted negative band gaps are clipped at zero.



Extended Data Fig. 2 | Band gap data distribution and correlation. Plots of the pairwise relationship between band gaps from different fidelity sources. The band gap distribution in each data set is presented along the top diagonal, and the Pearson correlation coefficient r between each pair of data are annotated in each plot.



Extended Data Fig. 3 | Predicted experimental band gaps of $Ba_{0.5}Sr_{0.5}Co_xFe_{1-x}O_{3-\delta}$ using 4-fi models. Both the Co ratio x and oxygen non-stoichiometry δ are changed to chart the two dimension band gap space.



Extended Data Fig. 4 | Multi-fidelity modeling of energies of molecules. **a**, Average MAE in G4MP2 energy predictions for the QM9 data set using 1-fi G4MP2 models and 2-fi B3LYP/G4MP2 models trained with different G4MP2 data sizes. **b**, Average MAE in CCSD(T) energy predictions for the QM7b data set using 1-fi CCSD(T) models, 2-fi HF/CCSD(T) and MP2/CCSD(T) models, and 3-fi HF/MP2/CCSD(T) models. *s* is the ratio of data sizes. *s* = 1 and 2 correspond to CCSD(T):MP2:HF ratios of 1:2:4 and 1:4:16, respectively. The error bars indicate one standard deviation.