Materials discovery for high-temperature, clean-energy applications using graph neural network models of vacancy defects and free-energy calculations†

Matthew Witman,[a‡*] Anuj Goyal,[b‡] Tadashi Ogitsu,[c] Anthony McDaniel,[a*] Stephan Lany,[b*]

We present a graph neural network modeling approach that fully automates the prediction of the DFT-relaxed vacancy formation enthalpy of any crystallographic site from its DFT-relaxed host structure. Applicable to arbitrary structures with an accuracy limited principally by the amount/diversity of the data on which it is trained, this model accelerates the screening of vacancy defects by many orders of magnitude by replacing the DFT supercell relaxations required for each symmetrically unique crystal site. It can thus be used off-the-shelf to rapidly screen 10,000s of crystal structures (which can contain millions of unique defects) from existing databases of DFT-relaxed crystal structures. This modeling approach therefore provides a significant screening and discovery capability for a plethora of applications in which vacancy defects are the primary driver of a material's utility. For example, by high-throughput screening the Materials Project's metal oxides, we rapidly "re-discover" and identify new high potential candidate materials for hydrogen generation via solar thermochemical water splitting and energy storage, for CO_2 conversion via reverse water gas shift chemical looping, and for cathodes in solid oxide fuel cells. Thermodynamic modeling on the basis of the high-throughput screening results allows us to connect the predicted defect energies to high temperature process conditions relevant to the different application areas, and we extract the reduction entropies as an additional selection criterion for high-performance materials. Further model development and accumulation of additional training data will only serve to expand the significant utility of this generalizable defect model to solving materials discovery problems in clean energy applications and beyond.

1 Introduction

High-accuracy calculations of vacancy defect formation enthalpies elucidate the primary and critical figure of merit needed to assess a material's utility across a large variety of applications. These can range anywhere from catalysis (e.g., oxides for water splitting[1–3]), to degradation resistance in extreme environments (e.g., radiation hardness of transition metal dichalcogenides[4,5]), to neuromorphic computing (e.g., tuning metal-to-insulator transition with oxygen vacancy formation[6,7]), to multiferroics (e.g., oxygen vacancy induced magnetic phase transitions[8]).

Density functional theory (DFT) is the method of choice to compute these vacancy formation enthalpies in a high throughput fashion. However, given the need for supercell construction, atomic force relaxation, and the general presence of multiple non-equivalent atomic sites, the computational effort of defect calculations far exceeds that of the computation of the ideal defect-free material in the primitive cell. Thus, explicit DFT defect calculations exist so far only for a small fraction of the O(100) compounds contained in existing computational databases like the Materials Project (MP),[9,10] Open Quantum Materials Database (OQMD),[11] NREL Materials Database (NRELMatDB),[12–14] Joint Automated Repository for Various Integrated Simulations (JARVIS),[15] and the Quantum Point Defect database (QPOD).[16] Additional, potentially combinatorial, complexity arises from the desire to predict defect behavior in non-ideal materials, e.g., in the presence of atomic site disorder. A successful surrogate modeling approach that avoids the cost requirements of DFT is therefore critical in order to efficiently facilitate materials discovery efforts in these applications.

Previous efforts to predict vacancy formation enthalpies span various methods and material classes within which the models are applicable. Notable examples include modeling vacancy formation enthalpy using a simple hand-derived or machine learning (ML) model based on hypothesized important features[17–21] and descriptor derived properties to train ML regression models for defect property prediction in semiconductors.[22,23] For 2D materials consisting of TMDs, hexagonal boron nitride, and other selected wide band gap 2D materials, similar utilization of hand-engineered features and random forests predicted vacancy defect formation enthalpies.[24] These examples provide highly useful models in specific situations by capitalizing on physical intuition regarding important material descriptors. But they do not provide a generalized solution that (1) automatically predicts the vacancy formation enthalpy of any crystal site in any material class and (2) negates the need for manual feature-engineering of material descriptors.

Various deep learning techniques, such as graph or convolutional neural networks, can circumvent such limitations. Rather than requiring the hand-crafting of an input feature vector to describe a training example, the model directly learns a feature representation from iterative convolutions and non-linear transformations of the input data, i.e. a graph representation of the crystal structure. The general success of graph neural networks at performing property prediction on crystal structures is evident from their recent explosion in popularity across a host of materials science applications.[25–36] In this work, we extend the graph neural network concept for directly predicting vacancy defect formation enthalpies (abbreviated dGNN for short). The only re-

[a]Sandia National Laboratories, Livermore, California 94551, United States; E-mail: mwitman@sandia.gov, amcdani@sandia.gov
[b]National Renewable Energy Laboratory, Golden, CO 80401, USA; E-mail: stephan.lany@nrel.gov
[c]Lawrence Livermore National Laboratory, Livermore, California 94550, USA
‡These authors contributed equally.

quired input is the non-defected, DFT-relaxed host structure, and the model output is the predicted relaxed vacancy formation enthalpy of any site in the structure. The surrogate model therefore replaces the need to do an expensive DFT supercell relaxation for each symmetrically nonequivalent atomic site in the host, of which there can be a sizable number (up to about 100 in the present work) in compositionally and structurally complex materials. Other quantities typically derived from a DFT-relaxed structure like band gap (E_g), effective electron mass (m^*), compound formation enthalpy (ΔH_f), and oxidation state (s) can readily be encoded into the material's graph if available to further improve prediction accuracy. Existing computational databases containing 10,000s of compounds with potentially up to millions of non-equivalent defect sites can be screened in an automated fashion using desktop computational resources with a prediction accuracy limited mainly by the amount and diversity of the vacancy defect training data.

As a specific use case, we first focus on the discovery of new solar thermochemical (STCH) water splitting oxides, widely considered a promising route for renewable hydrogen production.[37–39] DFT investigations play a critical role in understanding these materials and helping to guide experiments,[40–45] but can only be accomplished on a limited scale. Due to the complexity of the associated vacancy defect calculations using DFT, approximately one year's work was required to build our training database ($\sim$ 200 host oxide compounds consisting of $\sim$1500 unique defects). Our highly generalizable dGNN model then extends upon capabilities of previous work that requires carefully hand-engineered features in specific crystal systems (e.g., perovskites)[21,46] and obtains a similar expected mean absolute error in oxygen vacancy formation enthalpy (MAE $<$ 450 meV) in unseen compounds (assuming its cations are represented in the training data). Depending on the model's defect predictions, oxide stability, and the stringency of these down-selection criteria, we then narrow down a small number of priority candidates for experimental efforts from 10,000s of possible MP oxides (among which are known STCH oxides we "re-discover" through our screening procedure).

Finally, optimal candidates for other clean energy applications are highlighted where target ranges for oxygen vacancy formation enthalpies have been proposed as a primary metric for material down-selection, such as catalysts for CO_2 conversion and cathodes for solid oxide fuel cells.[47,48] We even extend our analysis to predict defect density for all materials at finite temperatures, a capability that can only be achieved due to our method's ability to rapidly predict vacancy formation enthalpies of all sites in the crystal structure (i.e., account for configurational entropy). Thus materials can be assessed beyond the static picture of single vacancy defect calculations at 0K for a limited set of materials. With continued accumulation of training data, this success paves the way for automated materials discovery in other vacancy defect dependent applications and lays the groundwork for more complicated machine learning tasks such as correlated vacancy formation enthalpy or vacancy mobility predictions. All data, code, and scripts needed to reproduce this study are provided open source.[49]

2 Results and Discussion

2.1 Developing a diverse database of vacancy defects

We developed an automated workflow for DFT vacancy defect calculations (Figure 1) as follows: (1) import the stoichiometric host oxide crystal structures from the ICSD[50] that are available in NRELMatDB; (2) relax the host supercell in ferromagnetic and different possible anti-ferromagnetic spin configurations; (3) choose the minimum energy host atomic and magnetic structure; and (4) perform point defect calculations on the supercell using the automated defect framework.[51] In the data acquisition step, we extract host properties (from step 3) and calculate vacancy formation energies (after step 4), which then supplies input and target properties, respectively, for training the machine learning model. Further details on the DFT settings are provided in Section S1.[14,44,52–58] DFT is currently the preferred method for high-throughput supercell defect calculations. True benchmark calculations require total-energy methods beyond DFT, such as quantum Monte Carlo or the random phase approximation, which are currently available only for few defect systems and with restrictions in cell size and atomic relaxations.[44,59,60]. Given the DFT-relaxed crystal structure, $\mathscr{C}_h$, of the host oxide with total energy E_h and the DFT-relaxed defected structure, $\mathscr{C}_d$, with total energy E_d (calculated using the standard supercell approach of Ref. 61), we compute the enthalpy of defect formation via

$$\Delta H_d = E_d - E_h + \sum_i n_i \mu_i^{\text{ref}}. \tag{1}$$

Here the reference chemical potential of added or removed atoms ($n_i = -1$ and $+1$, respectively) are taken as the fitted elemental reference energies (FERE)[14,57], $\mu_i^{\text{ref}} = \mu_i^{\text{FERE}}$, which improve the description of thermochemical properties in DFT calculations[62]. For an oxygen vacancy, Equation (1), simplifies to $\Delta H_{V_O} = E_{V_O} - E_h + \mu_O^{\text{ref}}$.

Our final DFT training database consists of $\sim$1500 unique defect sites from $\sim$200 parent oxides, which span 15 cations (Mg, Al, Ca, Ti, Mn, Fe, Co, Ni, Sr, Y, Nb, In, Ba, La, Ce), 63 space groups, and 51 unique stoichiometries. The choice of chemical space (Figure 2) is motivated by previous literature,[58,63–65] which provides guidance on cations that form stable oxides and play an active role in tuning oxygen vacancy formation energy via their redox activity. Previous STCH material searches have targeted perovskite (ABO_3) stoichiometry and related structures systems (cubic, orthorhombic, tetragonal). Our training data is more diverse and spreads across all seven crystal structure systems and includes a wide range of stoichiometries to cover different coordination environments and metal ion oxidation states varying from 2+ to 5+.

2.2 A graph neural network model for vacancy defects

We create a vacancy defect graph neural network (dGNN) surrogate model f_{dGNN}, parameterized by weights θ, of the general form

$$\Delta\hat{H}_d = f_{\text{dGNN}}\left(\mathscr{C}_h, i', \boldsymbol{v}_g, \boldsymbol{s}; \theta\right), \tag{2}$$

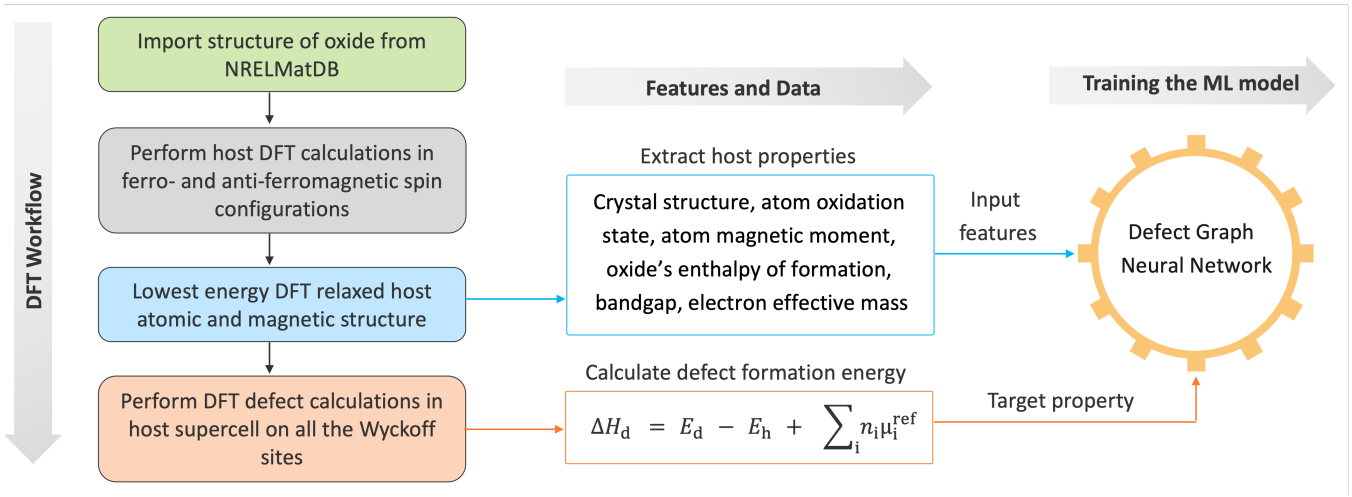

Fig. 1 The automated DFT vacancy defect calculation workflow combined with the data acquisition and curation necessary to develop the training dataset for a machine learning model.

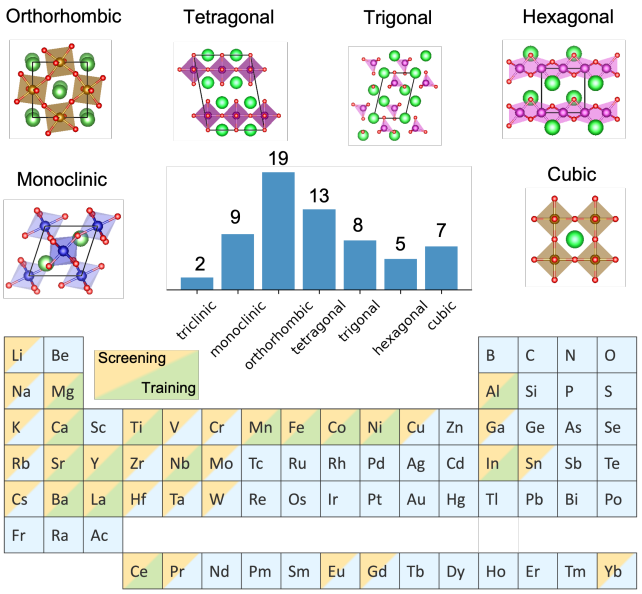

Fig. 2 The chemical and structural search space (cations in binary and ternary oxides and the number of unique space groups in each of the seven crystal systems) explored via DFT defect calculations. Cations considered in training are shown in green, along with ones that should eventually be considered for screening in yellow.

that will drastically reduce the computational cost associated with Equation (1) when performing high-throughput materials screening. Here, $\boldsymbol{v}_g = \{E_g, m^*, \Delta H_f\}$ refers to a set of global input features as derived from the host compound (e.g., the band gap, effective electron mass, and compound formation enthalpy, respectively) while $\boldsymbol{s} = \{s_1, s_2, ...\}$ refers to site specific input features for each atom, (e.g., s_1 is the oxidation state of atom 1 in the host structure). Intuitively, for example, we expect oxides with cations in high oxidation states to form O vacancies (i.e., to reduce) more easily than when cations are in lower oxidation states, hence motivating the inclusion of these properties derived from the relaxed host structure.

Requiring only $\mathscr{C}_h$ and the index of the atom to be defected, i', as an input, the model negates the cost of a DFT supercell relaxation for each unique symmetry site when predicting $\Delta\hat{H}_d$, and executing the ML screening is of negligible computational cost when querying $\mathscr{C}_h$ from existing repositories like MP or NRELMatDB. The inputs for model training (the host's relaxed POSCAR file, formation enthalpy, bandgap, effective electron mass, and oxidation states) and neutral vacancy formation energies (for both oxygen and cations) are provided in our open-source data repository.[49] Our dGNN closely follows the original Crystal Graph Convolutional Neural Network (CGCNN) method of Ref. 31, for which we highlight the necessary modifications to predict defect formation enthalpies. The deep learning framework is composed of three major steps.

Crystal embedding. First, $\mathscr{C}_h$ is embedded as a graph with nodes, $V = \{v_0, v_1, ...\}$, corresponding to each atom and edges $B = \{b_{ij}\}$ corresponding to bonds that are defined between atoms i and j below a cutoff radius and up to a maximum number of nearest neighbors. CGCNN One-Hot encodes[66] and concatenates a node's elemental solid properties, $\boldsymbol{v}_e$ = One-Hot(Mendeleev number, atomic weight, melting temperature, covalent radius, electronegativity, ground state volume per atom, ground state band gap, ground state magnetic moment, and space group number) as the initial feature vector, $\boldsymbol{v}_i^{(0)} = \boldsymbol{v}_e$. We additionally concatenate a One-Hot encoding of the site's oxidation state, as calculated in our DFT defect database, to increase model accuracy

$$\boldsymbol{v}_i^{(0)} = \boldsymbol{v}_e \oplus \text{OneHot}(s_i) \,. \tag{3}$$

Two atoms of the same element type are no longer guaranteed to share an identical encoding due to their dependence on s_i since various elements can assume different oxidation states depending on their local environment. Since One-Hot encoding increases sparsity and dimensionality of the initial node feature vectors and removes any quantitative ordering of a physical property, we also investigate whether a purely continuous encoding strategy can improve performance. This alter-

native approach involves the scaling of each elemental property between [0,1], which can be combined with the element's ground state electron configuration and valence electrons (e.g., $\left\{\frac{1s}{2},\frac{2s}{2},\frac{2p}{6},\ldots,\frac{5d}{10},\frac{6p}{6},\frac{v_s}{2},\frac{v_p}{6},\frac{v_d}{10},\frac{v_f}{14}\right\}$),[35] to yield a continuously-scaled elemental representation, $\boldsymbol{v}'_e$. This vector can further be concatenated with s'_i, scaled between [0,1], to produce a $\boldsymbol{v}_i^{(0)}$ that is lower dimensionality, less sparse, and preserves ordering of physical quantities.

CGCNN[31] also utilized a discretized Gaussian filter applied to the distance between two atoms, r_{ij}, to generate the initial edge features, $\boldsymbol{b}_{ij}$, but this representation can be sparse and sensitive to an arbitrary choice of the filter's standard deviation. A more systematic approach is to use a radial basis set expansion (e.g., similar Behler and Parinello's G^{II} functions[67]) to generate the initial feature vector according to a set of gaussian widths (η), centers (R_s), and a cutoff radius (r_c):

$$\boldsymbol{b}_{ij} = \{\exp\left[-\eta(r_{ij}-R_s)^2/r_c^2\right]\}. \quad (4)$$

This simultaneously reduces the sparsity and dimensionality of $\boldsymbol{b}_{ij}$ for a moderate basis set size ($\eta = \{0.5, 1.0, 1.5\}$, $R_c = \{1.0, 2.0, 3.0, 4.0, 5.0\}$). Such modifications facilitate distinguishing between very similar crystal structures.

Convolutions. Automated feature extraction is then performed, whereby the feature vector for each node is iteratively updated via convolutions with its neighboring nodes and edges. Mathematically, the output of the t^{th} convolutional layer can be written as[31]

$$\boldsymbol{v}_i^{(t+1)} = g\left(\boldsymbol{v}_i^{(t)} + \sum_j \sigma\left(\boldsymbol{z}_{ij}^{(t)}\boldsymbol{W}_1^{(t)} + \boldsymbol{b}_1^{(t)}\right) \odot g\left(\boldsymbol{z}_{ij}^{(t)}\boldsymbol{W}_2^{(t)} + \boldsymbol{b}_2^{(t)}\right)\right). \quad (5)$$

Here $\boldsymbol{z}_{ij} = \boldsymbol{v}_i \oplus \boldsymbol{v}_j \oplus \boldsymbol{b}_{ij}$ is the concatenation of information from connected nodes in the graph, $\boldsymbol{W}_1, \boldsymbol{b}_1$ and $\boldsymbol{W}_2, \boldsymbol{b}_2$ represent weights and biases of different learnable weight matrices (i.e., fully connected neural network layers), σ denotes a sigmoid activation function, g denotes a softplus activation, and $\odot$ denotes element-wise multiplication. This step remains unaffected for the dGNN.

Property prediction. Following T total convolutions, the original CGCNN method acquires an overall feature vector representing the crystal by pooling all nodes in the structure

$$\boldsymbol{v}_c = \text{Pool}\left(\boldsymbol{v}_0^{(T)}, \boldsymbol{v}_1^{(T)}, ..., \boldsymbol{v}_N^{(T)}\right), \quad (6)$$

which, for example, consists of a summation operation. Thus crystals of arbitrary size are described by a vector of the same dimensionality. The model predicts some final *global* property by applying one (or more) fully connected layers to $\boldsymbol{v}_c$.

However, we are interested in the defect formation enthalpy and therefore isolate the information contained only on the host node/atom to be defected (specified at index i') following the T total convolutions. We therefore replace eq. (6) with

$$\boldsymbol{v}_d = g\left((\boldsymbol{v}_{i'}^{(T)} \oplus \boldsymbol{v}_g)\cdot\boldsymbol{W} + \boldsymbol{b}\right). \quad (7)$$

We incorporate the global compound features at this step, $\boldsymbol{v}_g = \{E_g, m^*, \Delta H_f\}$, before applying the subsequent fully connected layer. Additional feed-forward layers may be applied before the final property prediction of $\Delta\hat{H}_d$ is then computed through one final output layer. Equation (7) is specifically designed for predicting defect formation enthalpies for the limit of infinite dilution in this work, but could be changed to, for example, expand its applicability to correlated vacancy defects.

All training data, model structures, and hyperparameters used in this study can be found in our Zenodo repository, from which all results can be reproduced when used in conjunction with the dGNN implementation built in Pytorch[68] provided at `https://github.com/mwitman1/cgcnndefect/tree/Paper1` (modified from Ref. 31). In brief, due to the small size of our training data set, a dGNN model of minimal complexity is required to facilitate training. In practice the number of trainable parameters often exceeds the number of training examples in deep learning applications and explicit regularization isn't even always needed to achieve low generalization error.[69] Nonetheless, our minimal complexity architecture consists of $T = 2$ convolution steps, $\boldsymbol{v}_i \in \mathbb{R}^8$, and $\boldsymbol{v}_d \in \mathbb{R}^{16}$, leading to a GNN with only $\sim 2{,}000$ learnable parameters. After fixing the architecture, the learning rate was adjusted to minimize K-fold test set performance (next section), and over-fitting was minimized via early-stopping using the mean absolute error on a 10% validation set within each train fold.

2.3 Defect GNN validation and performance

We executed three different cross-validation (CV) strategies, demonstrated by the toy examples in Figure 3a, to gauge model performance. For defect-wise and compound-wise CV, we utilize $K = 10$-fold cross validation. For each k-fold, 10% of the training data is held as a validation set for early stopping, and the mean absolute error over all n defects in the test set is computed, $\text{MAE}_k^Y = (1/n)\sum_n |\Delta\hat{H}_{d,n}^Y - \Delta H_{d,n}^Y|$. The model's expected prediction error is then estimated across all folds, $\langle \text{MAE}^Y \rangle_K = (1/K)\sum_{k=1}^K \text{MAE}_k^Y$. Here $Y = \text{O}$ or $Y = \text{Other}$ filters evaluation of the MAE by a defect's specific element type to delineate model performance between oxygen and non-oxygen vacancy predictions. The standard deviation in MAE across all K models, $\sigma_K(\text{MAE}^Y)$, should be small once sufficient data has been collected, i.e., the sampled distribution of training data no longer changes significantly between each fold.

Validating the dGNN with defect-wise CV is a less challenging task, since train and test sets may contain defects from the same material in similar chemical environments (i.e., just above the symmetry tolerance). Validating with compound-wise CV is more challenging, since the test set contains all the defects from a given material (none of which may appear in the training set). This better reflects the practical performance for materials discovery, because one is usually interested predicting for materials for which no DFT vacancy calculations have been performed and thus not a single defect site could exist in training set. Finally, element-wise, leave-one-out CV proves the hardest challenge, where all defects for a compound containing the held-out element are placed in the

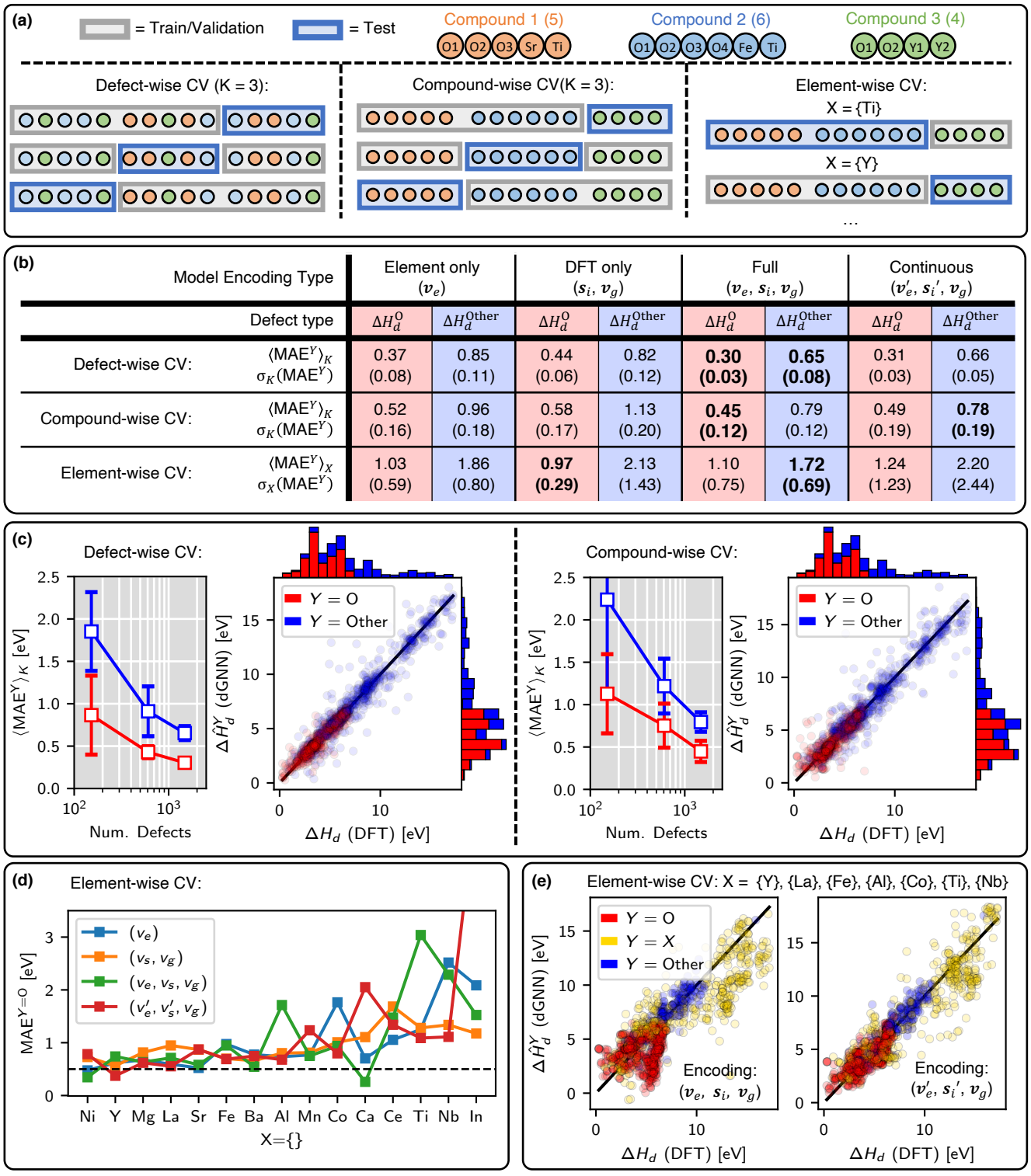

Fig. 3 (a) Demonstration of different CV strategies in a toy dataset. In defect-wise CV, different defects from the same compound can appear in both train/validation and test splits, whereas in compound-wise CV, the train/validation and test stratification is performed such that all of compound's defects can only appear in either train/validation or test splits. The number of compounds is kept constant between folds, leading to small variations in the number of unique defects contained per K-fold test set. In element-wise, leave-one-out CV, all defects in any compound containing the test element, X, are placed in the test set. (b) Summary of expected prediction errors and a heuristic uncertainty metric (mean MAE and standard deviation of MAE across K-fold test sets) for different cross-validation (rows) and graph encoding strategies (columns). (c) MAE averaged across the CV test sets ($K = 10$) models as a function of training data size and the parity plots of the test set predictions with DFT for the "Full" encoding strategy. (d) $\mathrm{MAE}^{Y=O}$ for the element-wise, leave-one-out CV, separated by test element and encoding strategy. (e) Test set predictions for the X ={Y},{La},{Fe},{Al},{Co},{Ti},{Nb} models.

test set (thus there are $X = 15$ models and test sets, corresponding to the number of cation types in the training data). At inference time, the ML model is supplied with a graph encoding containing node features that may not be represented by the training data.

Figure 3b shows the evaluation metrics for each CV strategy, with the MAE separated by predictions on oxygen and all other vacancy types. Cation vacancies tend to have higher enthalpies, but the mean relative errors between O and non-O defects are quite similar. This CV analysis was repeated for four different encoding strategies. "Element-only" encoding uses $\boldsymbol{v}_e$, while $\boldsymbol{s}_i$ and $\boldsymbol{v}_g$ are empty. Conversely, "DFT-only" encoding keeps $\boldsymbol{s}_i$ and $\boldsymbol{v}_g$ but $\boldsymbol{v}_e$ is empty. The "Full" encoding strategy incorporates all information, while the "Continuous" encoding replaces the one-hot encoding of elemental properties and oxidation state with $\boldsymbol{v}'_e$ and s'_i. Three key observations arise from the defect-wise and compound-wise CV. First, our strategy to encode both elemental and DFT data via the "Full" and "Continuous" models provides improved accuracy; nonetheless, "Element-only" and "DFT-only" encoding strategies both lead to reasonable accuracy on their own. Second, the best accuracy for compound-wise CV is obtained with the "Full" model with a $\langle \text{MAE} \rangle_K^{\text{O}} < 0.45$ eV. Third, the ability of the model to predict $\Delta H_d^{\text{Other}}$ validates the generality of this approach and its usefulness in other applications requiring predictive modeling beyond oxygen vacancies.

Figure 3c demonstrates the continued decrease in the CV MAE as more defects are added to the training data. While the MAE decrease with defect-wise CV starts to plateau, the compound-wise CV still benefits from a log-linear decrease in MAE with increasing data, highlighting the model can still be significantly improved as more training data is acquired. This highlights the need for continued, high-quality DFT defect calculations with automated workflows[51] (see Section S1 for recommended settings to build a larger training dataset consistent with this work). The parity plots correspond to the test set results concatenated across all CV models, with good performance between both oxygen and non-oxygen vacancy defect predictions.

Figure 3d shows the results of the element-wise, leave-one-out CV for each of the four different encoding strategies. While this is not a task that one might expect a deep learning approach to succeed at, several elements are well predicted and close to the target error of 0.5 eV, regardless of encoding type. Yet other elements are poorly predicted for all encoding types. Without any *a priori* knowledge of which held-out elements are well-predicted, it is not possible to assume that predictions on materials with unseen cation types can achieve $\langle \text{MAE} \rangle_K^{\text{O}} < 0.97$ eV (Figure 3). We therefore recommend this current generation of models only be used on compounds containing the fifteen cations spanned by the training database. Nonetheless, Figure 3e demonstrates how the continuous encoding strategy significantly improves property prediction across selected element types. This could be due to the continuous encoding's preservation of quantitative periodic trends (e.g., electronegativity, mean volume per atom, etc.). In order to expand the quantitative applicability of the model for materials discovery outside of the current cation set, we plan to include a larger chemistry space in the future.

2.4 Comparison with previous modeling efforts

Our cross-validation performance with the current training dataset is comparable to the performance of linear models derived via careful feature engineering to predict neutral oxygen vacancy formation enthalpies. Deml et al. trained a model on 45 binary and ternary oxides and achieved around 0.4 eV MAE on a small test set of 18 oxides that had not been included in model development.[17] Wexler et al. computed SCAN+U vacancy formation enthalpies of 341 ABO_3 perovskites and derived a linear model that globally achieved an MAE of 0.7 eV and an MAE of 0.45 eV for the subset (142 materials) with hull energies less than 0.025 eV/atom. Another in depth validation of these studies with ours is presented in Section S3. These linear models are more interpretable than our approach due to the small number of manually derived features, but are less generalizable. For example, they are not capable of predicting neutral cation vacancies, whereas our trained model predicts either. Our model architecture was not designed specific to oxides and could be applied to any material class since it operates generally on any crystal structure, whereas the linear models contain features that can only be calculated if the structure contains O. Figure 3c also demonstrates strong evidence for continued significant model improvement with more training data, which is unclear for the linear models. This pattern of comparable accuracy but increased generalizability also holds true for our method relative to other machine learning efforts for defect predictions, e.g., Frey et al.'s model for transition metal dichalcogenides[24] with MAE=0.67 eV and Cheng et al.'s model for amorphous GeTe.[70] Finally, in a concurrent preprint with ours,[71,72] Choudary et al. used graph neural networks models to predict total energy of a host structure and with an atom removed to estimate vacancy formation enthalpies, but this neglects the relaxation of the host upon vacancy formation and yields an MAE prediction of 1.5 eV for a single test set, including 2.3 eV for oxides.

2.5 Predictions on known STCH materials.

Before utilizing the dGNN model to screen potential candidate materials for STCH water splitting, it is instructive to first validate against additional DFT calculations for experimentally known STCH oxides. The materials behavior for the STCH redox processes can be expressed in terms of reduction enthalpies and entropies.[40,73] For workable and economic thermodynamic conditions, these considerations lead to a desirable value for the oxygen vacancy defect formation energy in an interval of about [2.3, 4.0] eV.[21,41,64] Lower formation energies impair the ability to produce hydrogen in the oxidation step, while higher energies prevent significant changes of the O stoichiometry in the reduction step. To predict the defect formation energies for these "unseen" materials, we utilize the expectation across K-fold models, $\langle \Delta \hat{H}_d^Y \rangle_K = K^{-1} \sum_{k=1}^{K} \hat{H}_{d,k}^Y$, and the standard deviation, $\sigma_K(\Delta \hat{H}_d^Y) = \sqrt{K^{-1} \sum_{k=1}^{K} (\hat{H}_{d,k}^Y - \langle \hat{H}_{d,k}^Y \rangle)^2}$, as a heuristic estimate of the uncertainty in the property prediction.[74,75]

As a first test case, we consider the family of BXM oxides (B=Ba; X=Ce,Nb,Pr; M=Mn)[76] which have a higher degree of compositional complexity than the materials in the training

dataset. Figure 4a compares the $\langle \Delta \hat{H}_d^O \rangle_K$ with DFT for BXM oxides for both "Full" and "DFT-only" encoded models. On average, "DFT-only" predictions are only more accurate when structures contain unseen cation types missing from the training set (e.g., Pr), but are generally less accurate if all of a structure's cation types are in the train set (Figure 3b,d). Nonetheless, both models are quite similar in qualitatively ranking and quantitatively capturing ΔH_d^O. We additionally tested our model on newly discovered disordered perovskite oxides, $Sr_{1-x}Ce_xMnO_3$ (SCM), which demonstrate suitable STCH performance.[77] In contrast to the BXM family which are line compounds, the SCM family can accommodate a wide range of Ce doping, thereby providing better control over water splitting capabilities by varying Ce concentration. To model SCM alloy structure we employ the special quasirandom structure (SQS) approach[78] and generate two 80-atom supercell structures representing random alloying with differing Ce content. Figure 4b shows the "Full" encoding model predictions on both SQS's. While the absolute value of the predictions are slightly below the generally accepted optimal range of [2.3, 4.0] eV, the ML model predicts the SQS with higher Ce content to have oxygen vacancy enthalpies closer to the target range, which is experimentally consistent with its improved water splitting capabilities from the increased Ce content.[77]

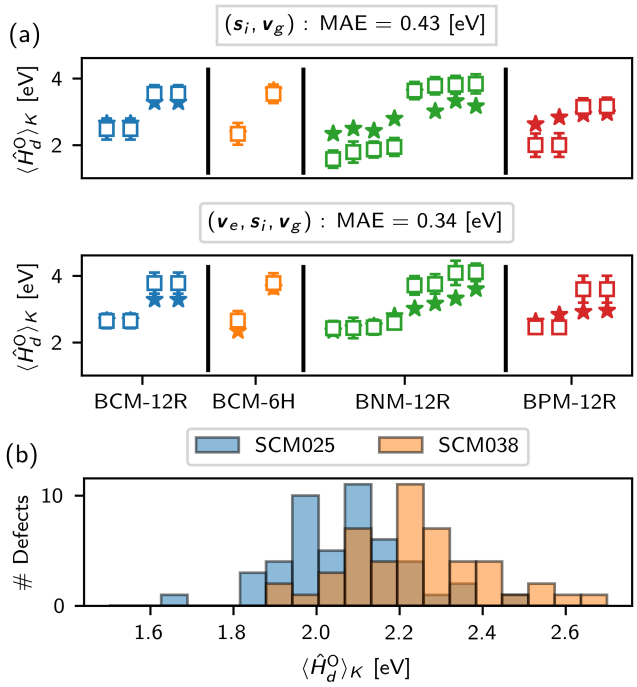

Fig. 4 (a) Defect-wise CV model predictions (open squares) and DFT values (stars) for each O vacancy in NRELMatDB structures of known 4 known STCH materials including BCM-12R = $Ba_4CeMn_3O_{12}$ (blue), BCM-6H = $Ba_3CeMn_2O_9$ (orange), BNM-12R = $Ba_4NbMn_3O_{12}$ (green), and BPM-12R = $Ba_4PrMn_3O_{12}$ (red). Error bars correspond to $\sigma_K(\Delta \hat{H}_d^O)$. (b) Distribution of $\langle \Delta \hat{H}_d^O \rangle_K$ for the ($Sr_{1-x}Ce_xMnO_3$) SCM family of alloys with differing Ce concentrations with x = 0.25 and 0.38 in SCM025 and SCM038, respectively.

2.6 High-throughput screening for new STCH materials.

Our "element-only" encoding models can perform consistent predictions on input host crystal structures in which the DFT settings for host relaxations are not identical to those used to create the training set (see Section S2 and Figure S2 for details). Thus we can now efficiently predict ΔH_d across 10,000s of DFT-relaxed crystal structures included in open source repositories beyond NRELMatDb, such as Materials Project (v2021.03.22). We employ the "Element-only" encoding model at the expense of slightly lower accuracy since Materials Project (MP) data doesn't necessarily contain all the features needed for the "Full" encoding models. Figure 5a shows that the space of ~35,000 oxides (excluding non-metals) is reduced to about ~2,200 structures by setting a maximum energy above the hull requirement, $E_H < 0.1$ eV/atom, and discarding any materials with cations not present in the training set. From these remaining ~2,200 host oxides, we predict $\langle \Delta \hat{H}_d \rangle_K$ for the ~48,000 symmetrically unique defect sites. Before proceeding with candidate down-selection, we perform yet another hold-out validation by comparing our MP screening predictions with existing first principles calculations and model predictions that could be easily mined from the literature (see Section S3).

Although we omit host structures with cations outside the training set in this study, adding a small number of training structures in the future (the full search space from Figure 2) could further expand the model's applicability. Figure 5b shows the predicted $\langle \Delta \hat{H}_d^O \rangle_K$ vs. ΔH_f, and, although correlation is evident across the entire enthalpy range, there is little correlation within [2.3, 4.0] eV. Furthermore, a simple model using features derived only from the host composition would clearly be insufficient due to an inability to distinguish individual oxygen vacancies. Within a single structure, these can span a very large range as shown for the min, median, and max predictions for MP structure mp-1247717 ($Ca_4Mn_3AlO_{11}$).

Assessing an oxide's STCH potential first requires determining ΔH_d^O of all sites and computing the fraction above the minimum threshold of 2.3 eV, denoted x_{min}, and the fraction of defects in the optimal range of [2.3, 4.0] eV, denoted x_{rng}. If the material contains any defects below the target range, the reduced metal oxide cannot readily be regenerated at the oxidation conditions relevant for STCH, and thus we require $x_{\text{min}} = 1$. It may be ideal for all defects to fall within in the target range ($x_{\text{rng}} = 1$) to increase capacity and defect mobility, but these considerations are beyond the scope of this study. In practice, only one defect needs to fall within the targeted range to be considered a promising STCH material ($x_{\text{rng}} > 0$).

Recent total energy calculations for defects in hercynite $FeAl_2O_4$ have found good agreement between DFT+U, hybrid functional, and the random phase approximation, but uncertainties on the order of a few tenths of an eV should be expected for DFT calculations in transition metal oxides.[44] We therefore extend the defect screening metrics to be uncertainty inclusive ($x_{\text{min},1}$, $x_{\text{rng},1}$), agnostic ($x_{\text{min},2}$, $x_{\text{rng},2}$), or exclusive ($x_{\text{min},3}$, $x_{\text{rng},3}$) for increasingly strict down-selection. Given a host's N_s symmetry sites and the set of all predictions $\{\langle \Delta \hat{H}_d^O \rangle_K\} \equiv H = \{H_1 \dots H_{N_s}\}$

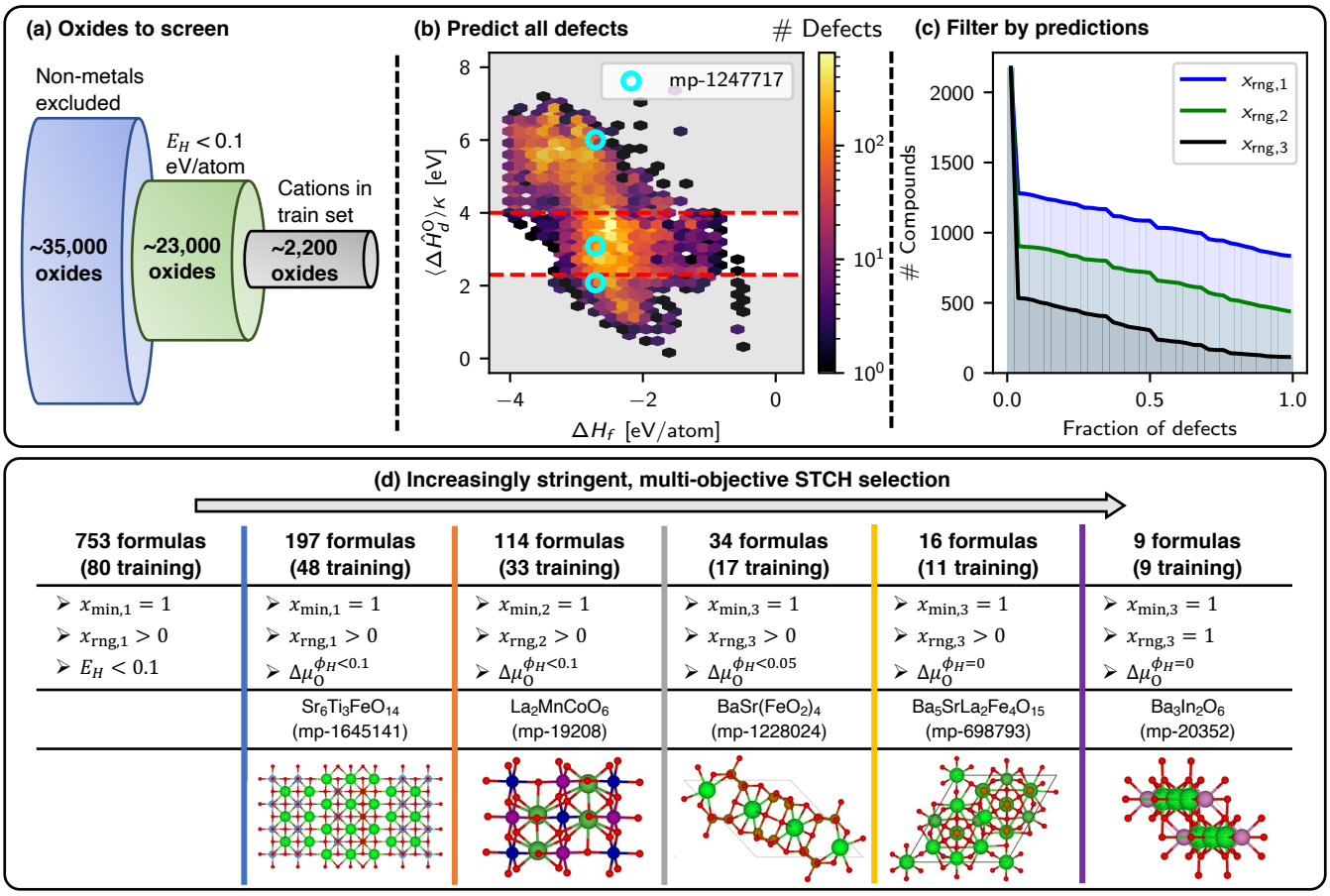

Fig. 5 (a) Removing MP oxides containing non-metals, those with $E_H > 0.1$ eV/atom, and those with elements outside the training set yields ~2,200 structures for screening. (b) Screening predictions using the "Element-only" encoding, defect-wise CV models vs. ΔH_f. (c) The cumulative histogram of compounds that contain a specific fraction of defects predicted to satisfy the range criteria, x_{rng}, as defined in Table 1. (d) Utilization of increasingly stringent defect and host compound criteria (x_{rng}, x_{min}, E_H, and $\Delta\mu_{O_2}^{\phi_H<X}$) to narrow the list of potential STCH candidates. At each down-selection stage, we show the number of unique remaining formulas, the number of which overlap with the training data, and an exemplar structure.

Uncertainty	C for x_{min}	C for x_{rng}
Inclusive	$H_i + U_i > 2.3$	$[H_i - U_i, H_i + U_i] \cap [2.3, 4.0] \neq \emptyset$
Agnostic	$H_i > 2.3$	$H_i \in [2.3, 4.0]$
Exclusive	$H_i - U_i > 2.3$	$[H_i - U_i, H_i + U_i] \in [2.3, 4.0]$

Table 1 Criteria for Equation (8) to determine the defect fractions, x_{min} and x_{rng}, for increasingly strict uncertainty inclusive, agnostic, or exclusive down-selection.

and uncertainties $\{\sigma_K(\Delta\hat{H}_d^O)\} \equiv U = \{U_1 \ldots U_{N_s}\}$, these defect fractions can be calculated subject to a criteria C,

$$x = \frac{1}{N_s} \sum_{i=1}^{N_s} \begin{cases} 1 & if\ C \\ 0 & otherwise \end{cases} \tag{8}$$

which is summarized in Table 1. Figure 5 shows the impact of applying these increasingly strict defect criteria and that a significant number of candidates remain even when using an uncertainty exclusive filter and requiring $x_{\text{rng}} \to 1$. See Section S4 for a more detailed discussion on uncertainty.

Down-selection must also consider host oxide stability under STCH relevant conditions. For oxygen chemical potential $\mu_O = \mu_O^{ref} + \Delta\mu_O$, typical STCH operating conditions necessitate that the host is stable in a target range $\Delta\mu_O^{\text{target}} = [-3.0, -2.5]$ eV.[40,44] Given the compound's energy above the hull in the grand ensemble, $\phi_H(\Delta\mu_O)$, we define the chemical potential range over which the host stability is below some threshold X,

$$\Delta\mu_O^{\phi_H \leq X} = [\Delta\mu_O | \phi_H(\Delta\mu_O) \leq X]. \tag{9}$$

Setting $X \leq 0.1$ eV/atom, for example, helps avoid false negatives during materials' selection due to the synthesizability of metastable structures[79] or due to uncertainties originating from the specific DFT approach and the convex hull analysis. Our final down-selection criteria requires that the target and stability chemical potential ranges intersect,

$$\Delta\mu_O^{\phi_H \leq X} \cap \Delta\mu_O^{\text{target}} \neq \varnothing. \tag{10}$$

The vacancy defect fractions and host stability criteria can be tuned for custom down-selection using our open access data and post-processing scripts.[49]

Figure 5d shows how increasingly stringent x_{min}, x_{rng}, and $\Delta\mu_O^{\phi_H \leq X}$ criteria can narrow the candidate space from thousands of oxides to just a handful. At each down-selection criteria

we have highlighted one material among many that appear interesting due to relatively low prediction uncertainty, relatively wide stability range, high cation/structure complexity, verified experimental synthesis,[80–84] and a lack of any STCH-specific experimental investigations that we are aware of. Importantly, this screening "rediscovers" the known STCH material BCM-12R. However, it is eliminated by stricter down-selection criteria since the "Element-only" encoding model under-predicts the oxygen vacancy enthalpies with $\{2.0 \pm 0.3, 2.9 \pm 0.4\}$ eV for the two O sites in BCM, compared to the "Full" encoding predictions, $\{2.6 \pm 0.3, 3.8 \pm 0.5\}$ eV, and DFT predictions, $\{2.66, 3.29\}$ eV[85] (Figure 4). This emphasizes the prudence of considering materials inclusive of their uncertainty, especially since the "Element-only" model necessitated by MP screening has slightly higher CV error. Some oxides with a relatively simple composition satisfying strict down-selection criteria include: Mn_3O_4,[86] Fe_3O_4,[87] Fe_2O_3,[87] $Ba_2Fe_2O_5$,[88] Mn_2CoO_4,[89,90] $Mn(FeO_2)_2$,[91] $Sr_2Mn_2O_5$,[84] $Sr_3(FeO_3)_2$,[92] $Ba(FeO_2)_2$,[93] $Ba_3In_2O_6$,[94] Fe_2NiO_4,[87] and $Sr_5Mn_5O_{13}$,[84] All have been experimentally synthesized in the literature. Notably, several have already been investigated in the context STCH performance[86,87,89] or other water splitting approaches.[90] Rediscovering these known STCH materials further validates our approach, and we have now identified many new candidates from which promising STCH materials can be experimentally targeted (see Table S6 for details and the comprehensive list). Raw data for all predicted defect properties and customizeable open source scripts for reproducing or modifying the down-selection criteria can be found in the project's Zenodo repository.[49]

2.7 Beyond STCH: materials discovery across diverse energy applications

Our dGNN methodology and high-throughput database of vacancy predictions can be used to rapidly screen candidate materials in other important clean energy applications. The reverse water gas shift-chemical looping (RWGS-CL) approach operates in a concept very similar to STCH. H_2 gas is used to reduce a metal oxide, after which oxidation with CO_2 produces CO for downstream hydrogenation to carbon-based fuels. Like STCH, the process is driven by thermodynamics of oxygen vacancy formation. Ref. 47 strongly correlated the energetics of vacancy formation to other computed properties like O_2 surface-adsorbate binding energy to conclude, "It is thus best said that E_{vac} can solely describe the RWGS-CL process and is capable of predicting the CO_2 conversion ability of perovskite oxides." Based on a previously known CO_2-splitting perovskite ($La_{0.75}Sr_{0.25}FeO_3$) for which they computed an average oxygen vacancy formation enthalpy of 3.4 eV, they concluded that candidate materials with similar thermodynamics (e.g., $\Delta\hat{H}_d^O \in [3.0, 4.0]$ eV) would be highly active. This was confirmed by synthesizing new perovskites with the desired vacancy thermodynamics and measuring their outstanding activity. By this metric, one can readily use our model to identify optimal candidates for RWGS-CL. Similar to our STCH screening, we "re-discover" experimentally known CO_2-splitting oxides, e.g., the La_2MnCoO_6 system.[95], while discovering new ones (Figure 6).

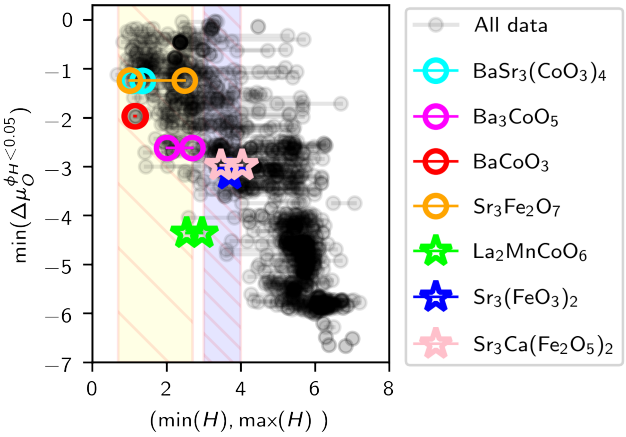

Fig. 6 Identification of candidate materials for CO_2 conversion via RWGS-CL (stars) and SOFC cathodes (circles). For each host oxide we plot the lower bound on stability vs the min and max of all vacancy enthalpies. Some top candidates are noted whose vacancy enthalpy ranges fall within or close to the RWGS-CL and SOFC target ranges (blue and yellow shading, respectively) while simultaneously displaying stability to the most reducing conditions.

Oxygen vacancy formation enthalpy has also been correlated with the critical performance metrics for perovskite oxide cathodes in solid oxide fuel cells. Ref. 48 discovered a simple linear relationship for SOFC perovskite cathodes between $\Delta\hat{H}_d^O$ and a metric for the macroscale oxygen-transfer performance, the area-specific resistance (ASR). The authors noted that successful materials should approximately be targeted with ASR $\in [0.02\ \Omega\ \text{cm}^2, 0.24\ \Omega\ \text{cm}^2]$, or between the ASR values in the optimized $Ba_{0.5}Sr_{0.5}Co_{0.75}Fe_{0.25}O_{3-\delta}$ (BSCF) and $La_{0.625}Sr_{0.375}Co_{0.25}Fe_{0.75}O_{3-\delta}$ (LSCF) systems, respectively. This essentially represents the trade-off between the correlated stability (high defect formation enthalpy) and low operating temperature (low defect formation enthalpy). Using DFT to compute the average $\Delta\hat{H}_d^O$ in model BSCF and LSCF crystal structures, the authors established guidelines that $\Delta\hat{H}_d^O \in [0.7, 2.7]$ eV should be targeted.

Once again, we can rapidly target such materials with our approach. BCSF-like $BaSr_7Fe_6(CoO_{12})_2$ (mp-1099936) with $\Delta\hat{H}_d^O \in [0.7, 1.5]$ and LCSF-like are $Sr_4LaFe_2(CoO_5)_3$ (mp-1218676) with $\Delta\hat{H}_d^O \in [0.7, 1.5]$ are "re-discovered" in the screening to further validate our approach. Interestingly, the BCSF-like structure is only metastable with $\min\left(\Delta\mu_O^{\phi_H<0.1}\right) = -0.88$ eV, while the LCSF-like is stable with $\min\left(\Delta\mu_O^{\phi_H=0}\right) = -0.35$ eV, an observation consistent with the reduced stability of BCSF. For example, some Fe-lacking analogs not discussed in Ref. 48 but computationally investigated elsewhere[96] maintain low predicted vacancy formation enthalpy but improve upon phase stability according to MP phase diagrams. Furthermore, non-simple perovskite compounds can be identified that display similarly desirable vacancy properties and stability (Figure 6).

2.8 High Temperature Defect Thermodynamics

Thus far we have discussed materials selection purely in terms of zero kelvin predictions of the defect formation enthalpies. However, our approach additionally permits the rapid assessment of defect densities at high temperatures, which is critical in predicting their behavior under realistic process conditions. The defect formation energies ΔH_{V_O} were defined above for the reference energy μ_O^{ref}, i.e., the zero temperature limit. At finite temperatures T and partial pressures pO_2, a corresponding chemical potential $\Delta\mu_O(T, pO_2)$ is added to obtain the formation energy under the respective thermodynamic condition (*cf.* eq. 1). In a given oxide with multiple O sites i and respective reference formation energies $\Delta H_{V_O,i}$, minimization of the free energy of defect formation [40] yields the dimensionless fractional concentration of O vacancies relative to the nominal O stoichiometry of the respective oxide,

$$[V_O] = \sum_i g_i \frac{\exp[-(\Delta H_{V_O,i} + \Delta\mu_O)/k_B T]}{1 + \exp[-(\Delta H_{V_O,i} + \Delta\mu_O)/k_B T]}. \quad (11)$$

Here, k_B is the Boltzmann constant, and $g_i = m_i/\sum_i m_i$ are the normalized degeneracies of the different oxygen sites with their respective multiplicities m_i.

To identify oxides which develop a desired degree of O-deficient off-stoichiometry under high-temperature thermodynamic conditions (T, pO_2), suitable for different application areas, we numerically invert Equation (11) to solve for the chemical potential $\Delta\mu_O$ at a given target concentration $[V_O]$. At the same time, the respective oxide must be stable under this condition and not decompose into other phases. Here, we include the consideration of a stability threshold X as defined above. For any given temperature, the chemical potential $\Delta\mu_O$ can be translated into the corresponding partial pressure pO_2 (or vice versa) using the ideal gas law. Note that many "stoichiometric oxides" do not accommodate high levels of defect concentrations, but instead prefer to form a more reduced, ordered phase with lower O content. In this case, there may not be any (pO_2, T) conditions for the target $[V_O]$. On the other hand, oxides that are able to develop a desired level of off-stoichiometry under suitable conditions are considered as potential candidates for functional O-deficient materials in the different application areas.

For a target vacancy concentration of $[V_O] = 1\%$, Figure 7a shows the pO_2 vs temperature diagram for oxides within $X < 0.05$ eV/at from the convex hull, using the ML screening of the MP data (*cf.* Section 2.6). This concentration is generally considered as a demarcation between the dilute, defect-like, and concentrated, alloy like, limits of non-stoichiometric materials, but a similar analysis can be made for any value of $[V_O]$. Desirable process conditions are indicated in Figure 7a for Solid Oxide Fuel Cells (SOFC) [97,98], Thermochemical Energy Storage (TCES) [99,100], and STCH [38,76], using the reduction step for the latter two. A spreadsheet with the oxides falling into the respective regions of interests is included in the SI. All three applications depend crucially on the formation of O vacancies, [101] although there are of course other materials considerations that we do not address here. Therefore, the present screening should give valuable insights about potential candidate materials at least for the aspect of O deficient off-stoichiometry.

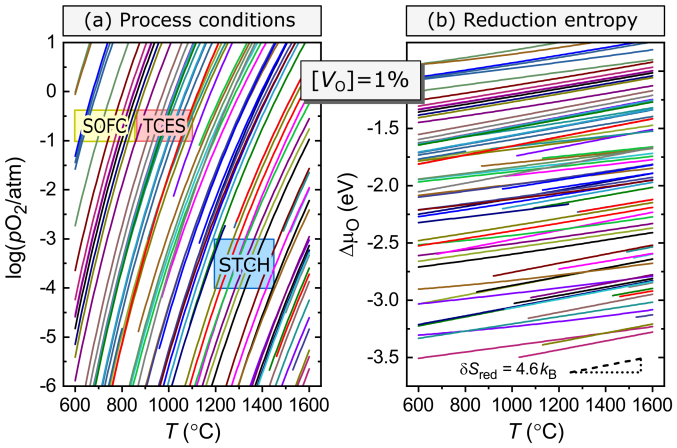

Fig. 7 (a) pO_2 vs temperature diagram for oxides with constant O-deficiency of $[V_O] = 1\%$ between 600-1600 °C. Each line represents one out of the 82 materials out of the high-throughput screening dataset that attain this defect concentration while simultaneously fulfilling a stability criterion of $X = 0.05$ eV/atom. Regions of interest for solid oxide fuell cells (SOFC), thermochemical energy storage (TCES) and solar thermochemical hydrogen (STCH) are highlighted. (b) Same data, except presented as $\Delta\mu_O(T)$. The graphs are approximately linear with a slope corresponding to the reduction entropy δS_{red}. The ideal configurational entropy of mixing (4.6 k_B at $[V_O] = 1\%$) is indicated at the bottom.

The list of STCH oxides contains barium, strontium, and lanthanum manganates, which are previously identified classes of oxides for this application [37,76,77], but also new suggestions like $Ba_2Fe_2O_5$. On the other hand, it also contains $BaMnO_3$, which at first sight appears to be a false-positive, because this oxide is known to reduce too easily and therefore be unable to split water [76]. However, it is just one (mp-19267) out of 9 different $BaMnO_3$ structures in the MP database, which it is not the ground state. With a ML predicted minimum V_O formation energy of $\Delta H_d = 3.0$ eV it would be a useful water splitter, but the corresponding energy is only 2.2 eV for the $BaMnO_3$ ground state (mp-1205336) in the MP database, which is too low. Thus, the ML model is consistent with experimental observations, and this example illustrates the tradeoff in choosing the tolerance for the stability criterion. Finally, we note that the list does not contain BCM, an apparent case of a false-negative, resulting from underestimation of the defect energy in the "element-only" encoding used for the MP screening (see Section 2.6). Using the energy from the direct DFT calculation or the "full" encoding ML model, BCM would indeed fall into the STCH process window indicated in Figure 7a.

Our thermodynamic modeling affords direct access to the reduction entropy, [40,73] which is of great benefit to applications that utilize a temperature swing, like STCH and TCES. For example, a large entropy facilitates high H_2/H_2O ratios in the STCH oxidation step [40]. Figure 7b represents the same data as Figure 7a, but showing the O chemical potential $\Delta\mu_O$ as ordinate instead of the O_2 partial pressure. The relevant quantity is the differential reduction entropy with change in defect concentration (shorthand δS_{red}), which equals the slope of the chemical potential,

i.e., $\delta S_{red} = \partial / \partial T[\Delta\mu_O(T)]$, as described in detail in Ref.[40]. We observe in Figure 7b significant variations in δS_{red} between the different oxides, which originates from the distribution of defect energies over different O sites. The ideal configurational entropy of 4.6 k_B for $[V_O] = 1\%$ is indicated in Figure 7b, and the material-specific, numerically determined values are included in the spreadsheet (SI) for the three process windows, giving additional guidance on materials selection over the enthalpy criterion via ΔH_d alone. We further note that additional electronic entropy effects[102] can arise in certain materials, in particular when the O vacancies assume a charged defect state,[40] where the excess electrons either form small polarons or occupy itinerant conduction band states. Such effects are relatively rare in transition metal oxides, where the redox activity is typically dominated by the transition metal ions close to the O defect. They could, however, play a role in the extraordinary behavior of CeO_2[76,103,104]. Our high-throughput screening and thermodynamic analysis is a valuable starting point for identifying new potential high-entropy materials.

3 Conclusions

We have developed a powerful, generalized GNN approach for predicting vacancy formation enthalpies of relaxed, defected structures using the relaxed host geometry as input. Therefore, only one DFT relaxation of the host is needed to derive the model inputs, and the model efficiently replaces the computationally intensive supercell calculations with numerous defect relaxations (one per symmetry site) needed to obtain the vacancy formation enthalpies. The model's applicability is not limited to structures in specific crystal/symmetry classes or elemental compositions, and it's accuracy is primarily limited by their representation in the training data. Through careful cross validation, we have thoroughly highlighted the advantages and limitations of the model. The best model performance was achieved by integrating DFT-computed host compound properties beyond just the relaxed crystal structure into the featurization process (i.e., oxidation states, compound formation enthalpy, band gap, and effective electron mass) to achieve a expected prediction error below 450 meV for relaxed oxygen vacancy defect formation enthalpy. Nonetheless, models trained only on the crystal structure exhibited just ∼15% higher MAE, since properties like oxidation state and compound formation enthalpy are already indirectly encoded in the crystal structure. ΔH_d of any element/crystal site can be predicted using the same model architecture and learned parameters. Furthermore, so long as two nominally identical materials (relaxed under different DFT settings) have very similar structures, our "element-only" encoding model provides close agreement on the predicted vacancy formation enthalpies because it relies only on the host crystal structure as input. This means that no new DFT is required to screen different databases (e.g., Materials Project) than the models were trained on (e.g., NRELMatDb) as predictions on nominally identical materials provides the same quantitative and qualitative outlook for vacancy formation enthalpies. While our training database consists of 15 cation elements, we tested an element-wise CV strategy to gauge model performance when predicting vacancies in compounds whose elements were missing from the training set. Finally, since the complexity of the DFT defect relaxations limits the size of the training data that can be collected, we have shown that the model error is still expected to decrease significantly as more data is collected in the future.

We demonstrated the model's significant utility for novel materials discovery in an exercise of identifying promising candidate oxides in the context of various clean energy applications: solar thermochemical water splitting and energy storage, CO_2 conversion, and SOFC cathodes and electrolytes. We screened structures drawn from a different database (Materials Project) than the source of the training structures (NRELMatDB) using the simplest, "Element-only" graph encoding strategy (i.e., requiring only the host crystal structure as input). Narrowing down the $\sim 35,000$ oxides initially queried to as few as ~ 10 depending on the stringency of down-selection criteria, we identify candidates exhibiting the greatest potential based on predicted vacancy defect enthalpy and host oxide stability criteria, which also "rediscovers" known materials from previous experimental literature. This ML strategy therefore efficiently reveals a handful of top candidates from an intractably large space for brute-force DFT or experiments, and will help facilitate the discovery of optimal materials in the future (along with significant potential for more chemistry and structural diversity in the training data). Even stricter down-selection can now additionally be explored based on more detailed properties from first principles calculations that are only tractable across a small number of materials. The final critical contribution of this study is the prediction of defect formation enthalpies across all sites to rapidly estimate defect densities at finite temperatures. By accounting for configurational entropy in high-throughput, we can assess material performance at finite temperatures, rather than relying purely on a zero kelvin picture from individual defect predictions across just a handful of materials.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This material is based upon work supported by the U.S. Department of Energy (DOE), Office of Energy Efficiency and Renewable Energy (EERE), specifically the Hydrogen and Fuel Cell Technologies Office. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. Part of the work was performed under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under contract No. DE- AC52-07NA27344. The National Renewable Energy Laboratory (NREL) is operated by the Alliance for Sustainable Energy, LLC, for the DOE under Contract No. DE-AC36-08GO28308. This work used High-Performance Computing resources at NREL, sponsored by DOE-EERE. The views expressed in this article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Data availability statement

The datasets generated during and/or analysed during the current study are available in the Zenodo repository: "A database of vacancy formation enthalpies for materials discovery" at `https://zenodo.org/record/5999073` (DOI: 10.5281/zenodo.5999073).[49] All data generated or analysed during this study are included in this published article (and its supplementary information files).

Code availability statement

The open-source CGCNN code for training models that directly predict vacancy formation enthalpies has been distributed on the "Paper1" branch at the following GitHub repository (`https://github.com/mwitman1/cgcnndefect/tree/Paper1`), which is a modified fork of the original open-source CGCNN code (`https://github.com/txie-93/cgcnn`).

Notes and references

1 C. L. Muhich, B. D. Ehrhart, I. Al-Shankiti, B. J. Ward, C. B. Musgrave and A. W. Weimer, Wiley Interdiscip. Rev. Energy Environ., 2016, **5**, 261–287.

2 C. F. Dickens and J. K. Nørskov, J. Phys. Chem. C, 2017, **121**, 18516–18524.

3 X. Qian, J. He, E. Mastronardo, B. Baldassarri, C. Wolverton and S. M. Haile, Chem. Mater., 2020, **32**, 9335–9346.

4 H. P. Komsa, J. Kotakoski, S. Kurasch, O. Lehtinen, U. Kaiser and A. V. Krasheninnikov, Phys. Rev. Lett., 2012, **109**, 1–5.

5 J. Lu, A. Carvalho, X. K. Chan, H. Liu, B. Liu, E. S. Tok, K. P. Loh, A. H. Castro Neto and C. H. Sow, Nano Lett., 2015, **15**, 3524–3532.

6 H. Jeen, W. S. Choi, M. D. Biegalski, C. M. Folkman, I.-C. Tung, D. D. Fong, J. W. Freeland, D. Shin, H. Ohta, M. F. Chisholm and H. N. Lee, Nat. Mater., 2013, **12**, 1057–1063.

7 S. Zhang and G. Galli, npj Computational Materials, 2020, **6**, 1–9.

8 C. Menéndez, D. Chu and C. Cazorla, NPJ Computational Materials, 2020, **6**, 1–9.

9 S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson and G. Ceder, Comput. Mater. Sci., 2013, **68**, 314–319.

10 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, APL Mater., 2013, **1**, 011002.

11 J. E. Saal, S. Kirklin, M. Aykol, B. Meredig and C. Wolverton, JOM, 2013, **65**, 1501–1509.

12 S. Lany, Phys. Rev. B, 2013, **87**, 085112.

13 S. Lany, J. Phys. Condens. Matter, 2015, **27**, 283203.

14 V. Stevanović, S. Lany, X. Zhang and A. Zunger, Physical Review B, 2012, **85**, 115104.

15 K. Choudhary, K. F. Garrity, A. C. E. Reid, B. DeCost, A. J. Biacchi, A. R. Hight Walker, Z. Trautt, J. Hattrick-Simpers, A. G. Kusne, A. Centrone, A. Davydov, J. Jiang, R. Pachter, G. Cheon, E. Reed, A. Agrawal, X. Qian, V. Sharma, H. Zhuang, S. V. Kalinin, B. G. Sumpter, G. Pilania, P. Acar, S. Mandal, K. Haule, D. Vanderbilt, K. Rabe and F. Tavazza, npj Comput. Mater., 2020, **6**, 173.

16 F. Bertoldo, S. Ali, S. Manti and K. S. Thygesen, npj Comput. Mater., 2022, **8**, 56.

17 A. M. Deml, A. M. Holder, R. P. O'Hayre, C. B. Musgrave and V. Stevanović, J. Phys. Chem. Lett., 2015, **6**, 1948–1953.

18 J. B. Varley, A. Samanta and V. Lordi, The journal of physical chemistry letters, 2017, **8**, 5059–5063.

19 A. Goyal, P. Gorai, S. Anand, E. S. Toberer, G. J. Snyder and V. Stevanovic, Chemistry of Materials, 2020, **32**, 4467–4480.

20 Z. Wan, Q. D. Wang, D. Liu and J. Liang, Phys. Chem. Chem. Phys., 2021, **23**, 15675–15684.

21 R. B. Wexler, G. S. Gautam, E. B. Stechel and E. A. Carter, J. Am. Chem. Soc., 2021, **143**, 13212–13227.

22 A. Mannodi-Kanakkithodi, M. Y. Toriyama, F. G. Sen, M. J. Davis, R. F. Klie and M. K. Chan, npj Computational Materials, 2020, **6**, 1–14.

23 A. Mannodi-Kanakkithodi, X. Xiang, L. Jacoby, R. Biegaj, S. T. Dunham, D. R. Gamelin and M. K. Chan, Patterns, 2022, 100450.

24 N. C. Frey, D. Akinwande, D. Jariwala and V. B. Shenoy, ACS Nano, 2020, **14**, 13406–13417.

25 G. Cheon, L. Yang, K. McCloskey, E. J. Reed and E. D. Cubuk, arXiv, 2020, arXiv:2012.02920.

26 L. Cian, G. Lancioni, L. Zhang, M. Ianese, N. Novelli, G. Serra and F. Maresca, arXiv, 2021, arXiv:2109.14012.

27 V. Fung, J. Zhang, E. Juarez and B. G. Sumpter, npj Comput. Mater., 2021, **7**, 1–8.

28 M. Karamad, R. Magar, Y. Shi, S. Siahrostami, I. D. Gates and A. Barati Farimani, Phys. Rev. Mater., 2020, **4**, 1–6.

29 P. R. Kaundinya, K. Choudhary and S. R. Kalidindi, arXiv, 2022, arXiv:2201.08348.

30 K. T. Schütt, O. T. Unke and M. Gastegger, arXiv, 2021, arXiv:2102.03150.

31 T. Xie and J. C. Grossman, Phys. Rev. Lett., 2018, **120**, 145301.

32 Y. Zuo, M. Qin, C. Chen, W. Ye, X. Li, J. Luo and S. P. Ong, Mater. Today, 2021, **51**, 126–135.

33 T. Hsu, N. Keilbart, S. Weitzner, J. Chapman, P. Xiao, T. A. Pham, S. R. Qiu, X. Chen and B. C. Wood, arXiv, 2021, arXiv:2109.11576.

34 J. P. Mailoa, M. Kornbluth, S. Batzner, G. Samsonidze, S. T. Lam, J. Vandermause, C. Ablitt, N. Molinari and B. Kozinsky, Nat. Mach. Intell., 2019, **1**, 471–479.

35 O. T. Unke, S. Chmiela, M. Gastegger, K. T. Schütt, H. E. Sauceda and K.-R. Müller, Nat. Commun., 2021, **12**, 7273.

36 S. Pandey, J. Qu, V. Stevanović, P. St. John and P. Gorai, Patterns, 2021, **2**, 100361.

37 A. H. McDaniel, E. C. Miller, D. Arifin, A. Ambrosini, E. N. Coker, R. O'Hayre, W. C. Chueh and J. Tong, Energy Environ. Sci., 2013, **6**, 2424.

38 J. E. Miller, A. H. McDaniel and M. D. Allendorf, Adv. Energy Mater., 2014, **4**, 1300469.

39 B. T. Gorman, M. Lanzarini-Lopes, N. G. Johnson, J. E. Miller

and E. B. Stechel, Front. Energy Res., 2021, **9**, 1–18.
40 S. Lany, The Journal of Chemical Physics, 2018, **148**, 071101.
41 J. E. Park, Z. J. L. Bare, R. J. Morelock, M. A. Rodriguez, A. Ambrosini, C. B. Musgrave, A. H. McDaniel and E. N. Coker, Front. Energy Res., 2021, **9**, 1–8.
42 G. Sai Gautam, E. B. Stechel and E. A. Carter, Chem. Mater., 2020, **32**, 9964–9982.
43 G. Sai Gautam, E. B. Stechel and E. A. Carter, Advanced Theory and Simulations, 2020, **3**, 2000112.
44 S. L. Millican, J. M. Clary, C. B. Musgrave and S. Lany, Chem. Mater., 2022, **34**, 519–528.
45 Y. Kumagai, N. Tsunoda, A. Takahashi and F. Oba, Phys. Rev. Mater., 2021, **5**, 1–12.
46 J. Vieten, B. Bulfin, P. Huck, M. Horton, D. Guban, L. Zhu, Y. Lu, K. A. Persson, M. Roeb and C. Sattler, Energy Environ. Sci., 2019, **12**, 1369–1384.
47 D. Maiti, B. J. Hare, Y. A. Daza, A. E. Ramos, J. N. Kuhn and V. R. Bhethanabotla, Energy Environ. Sci., 2018, **11**, 648–659.
48 I. Tezsevin, M. C. M. van de Sanden and S. Er, J. Phys. Chem. Lett., 2021, **12**, 4160–4165.
49 M. Witman, A. Goyal, T. Ogitsu, A. H. McDaniel and S. Lany, Zenodo, 2022, 10.5281/zenodo.5999073.
50 A. Belsky, M. Hellenbrandt, V. L. Karen and P. Luksch, Acta Crystallographica Section B: Structural Science, 2002, **58**, 364–369.
51 A. Goyal, P. Gorai, H. Peng, S. Lany and V. Stevanović, Computational Materials Science, 2017, **130**, 1–9.
52 P. E. Blöchl, Physical Review B, 1994, **50**, 17953–17979.
53 G. Kresse, Physical Review B, 1999, **59**, 1758–1775.
54 G. Kresse and J. Furthmüller, Computational Materials Science, 1996, **6**, 15–50.
55 J. P. Perdew, K. Burke and M. Ernzerhof, Phys. Rev. Lett., 1996, **77**, 3865–3868.
56 S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys and A. P. Sutton, Physical Review B, 1998, **57**, 1505–1509.
57 A. Sharan and S. Lany, The Journal of Chemical Physics, 2021, **154**, 234706.
58 M. Pavone, A. B. Munoz-Garcia, A. M. Ritzmann and E. A. Carter, The Journal of Physical Chemistry C, 2014, **118**, 13346–13356.
59 E. Ertekin, L. K. Wagner and J. C. Grossman, Physical Review B, 2013, **87**, 155210.
60 J. A. Santana, J. T. Krogel, P. R. Kent and F. A. Reboredo, The Journal of Chemical Physics, 2017, **147**, 034701.
61 H. Peng, D. O. Scanlon, V. Stevanovic, J. Vidal, G. W. Watson and S. Lany, Physical Review B - Condensed Matter and Materials Physics, 2013, **88**, 1–7.
62 S. Lany, Physical Review B, 2008, **78**, 245207.
63 I. E. Castelli, D. D. Landis, K. S. Thygesen, S. Dahl, I. Chorkendorff, T. F. Jaramillo and K. W. Jacobsen, Energy & Environmental Science, 2012, **5**, 9034–9043.
64 A. A. Emery, J. E. Saal, S. Kirklin, V. I. Hegde and C. Wolverton, Chemistry of Materials, 2016, **28**, 5621–5634.
65 R. J. Carrillo and J. R. Scheffe, Solar Energy, 2017, **156**, 3–20.
66 J. T. Hancock and T. M. Khoshgoftaar, J. Big Data, 2020, **7**, 28.
67 J. Behler and M. Parrinello, Phys. Rev. Lett., 2007, **98**, 1–4.
68 A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai and S. Chintala, Advances in Neural Information Processing Systems 32, Curran Associates, Inc., 2019, pp. 8024–8035.
69 C. Zhang, S. Bengio, M. Hardt, B. Recht and O. Vinyals, arXiv, 2016, arXiv:1611.03530.
70 Y. Cheng, L. Zhu, G. Wang, J. Zhou, S. R. Elliott and Z. Sun, Comput. Mater. Sci., 2020, **183**, 109803.
71 K. Choudhary and B. G. Sumpter, arXiv, 2022, arXiv:2205.08366.
72 M. D. Witman, A. Goyal, T. Ogitsu, A. H. McDaniel and S. Lany, ChemRxiv, 2022, 10.26434/chemrxiv–2022–frcns.
73 B. Meredig and C. Wolverton, Phys. Rev. B, 2009, **80**, 245119.
74 Y. Jung and J. Hu, J. Nonparametr. Stat., 2015, **27**, 167–179.
75 Y. Bengio and Y. Grandvalet, Advances in Neural Information Processing Systems, 2003.
76 D. R. Barcellos, M. D. Sanders, J. Tong, A. H. McDaniel and R. P. O'Hayre, Energy Environ. Sci., 2018, **11**, 3256–3265.
77 A. M. Bergeson-Keller, M. D. Sanders and R. P. O'Hayre, Energy Technology, 2022, **10**, 2100515.
78 A. Zunger, S.-H. Wei, L. G. Ferreira and J. E. Bernard, Phys. Rev. Lett., 1990, **65**, 353–356.
79 M. Aykol, S. S. Dwaraknath, W. Sun and K. A. Persson, Sci. Adv., 2018, **4**, eaaq0148.
80 Y. Titov, N. Belyavina, M. Slobodyanik, O. Nakonechna, N. Strutynska and M. Tymoshenko, Open Chem., 2020, **18**, 1294–1303.
81 P. M. Tirmali, D. K. Mishra, B. P. Benglorkar, S. M. Mane, S. L. Kadam and S. B. Kulkarni, J. Chinese Adv. Mater. Soc., 2018, **6**, 207–221.
82 C. Zhang, X. Wang, H. Yan, D. Kang, L. Li, X. Lu, D. Han, F. Yan and J. Zhu, J. Electron. Mater., 2014, **43**, 1071–1075.
83 H. Mevs and H. Müller-Buschbaum, J. Less Common Met., 1990, **158**, 147–152.
84 L. Suescun, O. Chmaissem, J. Mais, B. Dabrowski and J. D. Jorgensen, J. Solid State Chem., 2007, **180**, 1698–1707.
85 N. A. Strange, J. E. Park, A. Goyal, R. T. Bell, J. A. Trindell, J. D. Sugar, K. H. Stone, E. N. Coker, S. Lany, S. Shulda et al., Inorganic Chemistry, 2022, **61**, 6128–6137.
86 F. Sibieude, M. Ducarroir, A. Tofighi and J. Ambriz, Int. J. Hydrogen Energy, 1982, **7**, 79–88.
87 M. D. Allendorf, R. B. Diver, N. P. Siegel and J. E. Miller, Energy & Fuels, 2008, **22**, 4115–4124.
88 X. D. Zou, S. Hovmöller, M. Parras, J. M. González-Calbet, M. Vallet-Regí and J. C. Grenier, Acta Crystallogr. Sect. A

Found. Crystallogr., 1993, **49**, 27–35.
89 M. Orfila, M. Linares, R. Molina, J. Á. Botas, J. Marugán and R. Sanz, Int. J. Hydrogen Energy, 2017, **42**, 13532–13543.
90 Y. Liang, H. Wang, J. Zhou, Y. Li, J. Wang, T. Regier and H. Dai, J. Am. Chem. Soc., 2012, **134**, 3517–3523.
91 G. Salazar-Alvarez, H. Lidbaum, A. López-Ortega, M. Estrader, K. Leifer, J. Sort, S. Suriñach, M. D. Baró and J. Nogués, J. Am. Chem. Soc., 2011, **133**, 16738–16741.
92 Y. Shilova, M. Patrakeev, E. Mitberg, I. Leonidov, V. Kozhevnikov and K. Poeppelmeier, J. Solid State Chem., 2002, **168**, 275–283.
93 X. Song, S. Le, X. Zhu, L. Qin, Y. Luo, Y. Li, K. Sun and Y. Chen, Int. J. Hydrogen Energy, 2017, **42**, 15808–15817.
94 S. A. Miller, P. Gorai, B. R. Ortiz, A. Goyal, D. Gao, S. A. Barnett, T. O. Mason, G. J. Snyder, Q. Lv, V. Stevanović and E. S. Toberer, Chem. Mater., 2017, **29**, 2494–2501.
95 M. M. Nair and S. Abanades, Sustain. Energy Fuels, 2018, **2**, 843–854.
96 D. Fuks, Y. Mastrikov, E. Kotomin and J. Maier, J. Mater. Chem. A, 2013, **1**, 14320.
97 Y.-M. Kim, J. He, M. D. Biegalski, H. Ambaye, V. Lauter, H. M. Christen, S. T. Pantelides, S. J. Pennycook, S. V. Kalinin and A. Y. Borisevich, Nature materials, 2012, **11**, 888–894.
98 Y.-L. Lee, J. Kleis, J. Rossmeisl, Y. Shao-Horn and D. Morgan, Energy & Environmental Science, 2011, **4**, 3966–3970.
99 N. Gokon, T. Yawata, S. Bellan, T. Kodama and H.-S. Cho, Energy, 2019, **171**, 971–980.
100 D. Xiang, C. Gu, H. Xu and G. Xiao, Small, 2021, **17**, 2101524.
101 B. Steele, Journal of Power Sources, 1994, **49**, 1–14.
102 S. S. Naghavi, A. A. Emery, H. A. Hansen, F. Zhou, V. Ozolins and C. Wolverton, Nature communications, 2017, **8**, 1–6.
103 W. C. Chueh, C. Falter, M. Abbott, D. Scipio, P. Furler, S. M. Haile and A. Steinfeld, Science, 2010, **330**, 1797–1801.
104 A. H. McDaniel, Current Opinion in Green and Sustainable Chemistry, 2017, **4**, 37–43.