

# Formation Energy Prediction of Crystalline Compounds using Deep Convolutional Network Learning on Voxel Image Representation

Ali Davariashitani<sup>1</sup> and Sara Kadkhodaei<sup>1,2</sup>

<sup>1</sup>*Department of Civil, Materials, and Environmental Engineering,  
2095 Engineering Research Facility, 842 W. Taylor St., Chicago, IL 60607*

<sup>2</sup>*To whom correspondence should be addressed; Email: sarakad@uic.edu*

## Abstract

Emerging machine-learned models have enabled efficient and accurate prediction of compound formation energy, with the most prevalent models relying on graph structures for representing crystalline materials. Here, we introduce an alternative approach based on sparse voxel images of crystals. By developing a sophisticated network architecture, we showcase the ability to learn the underlying features of structural and chemical arrangements in inorganic compounds from visual image representations, subsequently correlating these features with the compounds’ formation energy. Our model achieves accurate formation energy prediction by utilizing skip connections in a deep convolutional network and incorporating augmentation of rotated crystal samples during training, performing on par with state-of-the-art methods. By adopting visual images as an alternative representation for crystal compounds and harnessing the capabilities of deep convolutional networks, this study extends the frontier of machine learning for accelerated materials discovery and optimization. In a comprehensive evaluation, we analyse the predicted convex hulls for 3,115 binary systems and introduce error metrics beyond formation energy error. This evaluation offers valuable insights into the impact of formation energy error on the performance of the predicted convex hulls.

## I. INTRODUCTION

Machine learning has emerged as an effective approach for developing predictive models for high-throughput screening of materials [1–8]. For example, machine-learned models for formation energy prediction can construct a convex hull for a rapid assessment of the thermodynamic stability of compounds at a fraction of the computation cost and time needed for density functional theory (DFT)-calculated convex hulls with reasonable accuracy [9]. In materials research, a machine learning model can be characterized by two aspects; the representation of the material as a readable entity (or input) to the learning algorithm and the learning algorithm itself. Several machine learning approaches have investigated a variety of representations as simple as a pool of physicochemical attributes (e.g., atomic number, cohesive energy, band gap, and heat of melting), and composition vectors [10–17] up to more advanced graph representations of composition and structure of crystal compounds [18–25]. The use of image representations for machine learning, however, has been less explored in the materials research community. Image representation can be especially useful because of the significant advancements that have been made in pattern recognition (or representation learning) of visual images in the field of computer vision (a field of computer science that deals with processing and understanding visual data like images or videos). These advancements are largely because of the evolution towards more sophisticated architectures of convolutional neural networks (e.g., Residual Neural Network (ResNet)[26], EfficientNet[27], U-Net [28]) which has enabled adopting increasingly deeper networks. Inspired by this untapped opportunity for materials representa-

tion learning, we develop a sparse voxel image representation of crystalline materials that is input into a very deep convolutional neural network (CNN) with a sophisticated architecture inspired by ResNet.

We use the formation energy prediction of crystalline compounds as a platform for demonstrating the performance of our deep-learning model on voxel images of crystals. Formation energy is an ideal platform because large databases of DFT-calculated formation energies are available (e.g., Materials Project [29] and AFLOW (Automatic Flow)[30]), which provide the large amount of data needed for training our deep CNN. Additionally, there are several available machine learning approaches for formation energy prediction with which we compare the performance of our model. We show that our model’s formation energy predictive performance is comparable to the state-of-the-art machine learning models’ prediction. We present a thorough comparison of 3,115 binary convex hulls constructed from our model’s formation energy against DFT-calculated binary convex hulls in the Materials Project database. By introducing multiple error metrics for assessing binary convex hulls, we showcase how the error in the formation energy prediction is projected into the performance of a predicted convex hull.

Among machine learning methods for formation energy prediction of crystal compounds, graph neural networks have shown promising performance because the graph data structure can efficiently capture the physical, compositional, and structural information of crystal compounds [18–25]. In their pioneering work, Xie and Grossman developed the crystal graph convolutional neural network (CGCNN) for an accurate, efficient formation energy prediction [18]. CGCNN uses a graph representation of crystal structures combined with physical

attributes of chemical species, where atoms in the periodic crystal structure constitute the nodes (with each node containing the physical attributes of the atom as a vector) and connections between atoms constitute the edges of the graph. In a more advanced representation, the Atomistic Line Graph Neural Network (ALIGNN) [23] captures the bond angular information of crystal compounds by utilizing the concept of line graphs [31]. ALIGNN creates a graph on top of a regular atomistic bond graph by considering bonds as nodes and bond pairs with a shared atom as edges. It has been shown that ALIGNN outperforms other machine learning approaches on several benchmark datasets for predicting material properties [32].

While some studies have explored image-inspired representations of crystalline materials [33–38], they have not fully harnessed sophisticated components, such as skip connections, in the architecture of the convolutional network. Skip connections have been shown as necessary components for making convolutional networks deeper (with many convolutional layers), with deep CNNs being particularly adept at learning features from image representations. As a result, these studies generally underperform compared to graph-based machine learning approaches. For instance, Kaundinya *et al.* [35] focusing exclusively on cubic structures, employ a transformed variant of an image-like representation, inputted to a Gaussian process regression and a neural network, achieving mean absolute errors (MAEs) of 0.329 eV per atom and 0.350 eV per atom, respectively (over ten times higher than ALIGNN’s MAE). Moreover, the image representations employed in these studies are not directly input to the machine learning model; they undergo transformation into alternative domains before being fed into the learning algorithms. For example, Ref. [35] utilizes a 2-point spatial correlation function of ionization energy, Pauling electronegativity, and heat of fusion over a voxelized domain of the crystal structure. These spatial correlations are subsequently transformed into low-dimensional representations of the material’s internal structure using principal component analysis (PCA) to serve as the input to the learning algorithm. Ref. [33] also transforms image representation of crystals. This study exclusively focuses on the Bi-Se system and uses an autoencoder to transform separate voxel images of the lattice vectors and basis points into 2-dimensional (2D) crystal graphs in the latent space. These graphs serve as descriptors for the formation energy predictive model. An earlier study by Noh *et al.* [38] also used a 3-dimensional (3D) grid-based image representation of vanadium oxide crystal structures as input to an autoencoder, where its latent space forms a vector which serves as the input to a second-step variational autoencoder. Like Ref. [33], the crystal representation in Ref. [38] is decomposed into a unit cell image (length of the cell edges and angles between them) and basis image (atomic positions within a unit cell).

Previous studies that have used image-like represen-

tations of materials as direct input for learning models have typically relied on continuous density representations rather than the sparse images used in this study. For instance, Hoffman *et al.* [34] employ density representations of atomic positions as 3D pixel images, necessitating the utilization of a U-net [28], an advanced autoencoder, for the segmentation of density fields into atoms. In another study [37], Kajita *et al.* introduced a generic 3D voxel descriptor that compacts any field quantities, including the electron density field. They examined a model that input the 3D voxel descriptor into a CNN to predict the Hartree energy and exchange-correlation energy functionals in 680 oxides.

Our study stands apart from previous work by utilizing a sparse voxel image representation of crystals that is directly fed into a sophisticated deep convolutional neural network with skip connections for feature (or representation) learning. Unlike earlier studies, our voxel image representation solely focuses on the visual depiction of the crystal structure itself, devoid of any physicochemical attributes. Within a voxelized domain of the crystal structure, we utilize normalized atomic number, periodic group, and row numbers to color the voxels occupied by atoms, representing them as channels in an RGB (red, green and blue) voxel image. This straightforward color-coding scheme enables us to differentiate between different chemical compositions of the same crystal structure without incorporating additional attributes, resulting in a direct visualization of the crystal compound. We directly input these sparse voxel images into a deep CNN. Utilizing skip connections allows us to design a deep 15-layer network that fully harness the power of convolutional layers to autonomously uncover the underlying physical, chemical, and structural features that connect the crystal structure and chemistry of the material to its formation energy. Apart from harnessing the full potential of convolutional layers, the use of an unprocessed image representation for materials holds significant technical importance, particularly because its ease of invertibility makes it a suitable representation for generative machine learning models. The materials research community is experiencing a paradigm shift from predictive models for high-throughput screening to generative models for discovery [33, 39, 40]. Given this emerging shift, the image representation employed in this study can potentially offer distinct advantages for generative models. The sparse voxel image representation introduced here for predictive modeling, if generated via a generative machine learning model, directly corresponds to a crystal structure, thereby eliminating the need for any transformation, interpretation, or intervention.

In this study, we present new advancements in image-based machine learning models for material property prediction. The introduced model is not restricted to any specific type of crystal structure or chemical space, making it generally applicable to any crystal structure or chemical composition. By utilizing a deep convolutional neural network enabled by skip connections, the model

achieves a significant improvement in formation energy prediction. Additionally, the model’s consistency and rotational invariance are improved through the employment of rotational sampling on crystal structure data. It is worth noting that there is ample room for enhancing the predictive performance of deep CNN models on voxel crystal images through the design of more advanced and efficient architectures, as this area has received comparatively less investigation. This study lays the foundation for exploring a new domain of learning methods for materials prediction and discovery.

## II. RESULTS

### A. Machine Learning Approach

*Material Representation.* To generate sparse voxel images of crystals, we employ a series of steps. Initially, we construct a cubic box with a fixed side length of 17 Å and position the crystal unit cell at its center. By applying 3D rigid-body rotation to the unit cell and replicating it throughout the box, a point cloud representation is formed. This point cloud, representing the atoms, is then represented as a sparse voxel image using a regular voxel grid. The voxel images adopt a color-coding scheme akin to an RGB image, with the three channels representing the normalized values of the atomic number, group, and period for voxels containing atoms. Voxels that are unoccupied by atoms are assigned zero values. To rotate each crystal unit cell efficiently, we have developed a computational algorithm, the details of which can be found in Section IV and Supplementary Figure S1.

Our voxel crystal representation differs from the image representations used in references [33, 34] as it adopts a sparse approach. These references employ a 3D density field representation around the atomic positions. For instance, in Ref. [34], a Gaussian density field centered at the atomic coordinates is defined to determine the voxel values. In the density field approach, voxels neighboring the atomic coordinates contain density field values that reflect the atom sizes, while we do not assign any values to neighboring voxels. From a technical standpoint, the sparse voxel image provides a discrete input to the convolutional layer, whereas the density field image offers a continuous input. As depicted in Figure 1, our CNN architecture, by applying multiple convolutions in the early layers without pooling (which we call the delayed pooling approach), automatically forms a field around the atomic coordinates. This means that our model’s architecture discovers the volumetric density fields without relying on predefined functions such as the Gaussian function. We postpone the pooling operation until after the 5th convolutional layer to ensure that the density fields around input voxels are sufficiently large for meaningful interactions to occur.

*Convolutional Neural Network Design.* Advanced deep CNN architectures, developed in the field of computer

vision, incorporate skip connections to enhance model performance and enable the construction of substantially deeper networks by mitigating optimization challenges associated with increased network depth. One notable example is ResNet [26], which utilizes residual blocks comprising convolution layers and activation functions like traditional CNNs, but with the addition of short-cut highways that connect the beginning and end of each block (referred to as identity mapping skip connections). These skip connections enable the transfer of lower-level information from earlier layers to deeper layers, providing better conditioning for the optimization problem and facilitating easier learning [26].

In our approach, we adopt the architecture of residual blocks to construct a 15-layer CNN with 7 skip connections. The overall architecture, as depicted in Figure 1, consists of a deep CNN followed by a fully connected neural network for the prediction of formation energy using sparse voxel images of crystals. The deep CNN part of the architecture is employed for feature learning of voxel crystal images. These learned features are then flattened and passed as input to the fully connected neural network, which performs the final prediction of the formation energy. In our network design, we deliberately delay the introduction of pooling layers in our CNN. The first pooling layer is introduced only after the fifth convolutional kernel, with subsequent pooling layers added after the eleventh and fifteenth kernels, respectively. A detailed description of our CNN architecture can be found in section IV. In the context of materials representation learning, the use of skip connections in our CNN allows for the bypassing of local atomic features discovered in the shallower layers, while progressively learning more global features of crystal compounds across the layers of the deep network. This hierarchical learning approach facilitates the extraction of relevant abstractions, enabling the model to capture both local and global features within the crystal structures.

Our CNN, inspired by the ResNet architecture described in Ref. [26], incorporates slight modifications to better suit our specific task. In contrast to the original design, we choose not to adopt the batch normalization technique in our residual blocks. This decision is based on the observation that batch normalization hampers the training of our CNN, likely due to the intrinsic differences between sparse crystal images and natural images (such as those in ImageNet [41]). Consequently, the batch normalization process may not yield the intended benefits for our crystal image representation. Furthermore, we adjust the way in which we handle the number of channels within our network. Instead of doubling the number of channels after each convolution layer, as outlined in the original ResNet design, we increase the number of channels, after each pooling, by concatenating the side skip connections with the output of the convolution layer. This alternative approach allows for a more effective utilization of information from both the skip connections and the convolutional layers, promoting better

feature representation within our network. By tailoring the ResNet-inspired architecture to the characteristics of our crystal images, we optimize the training process and enhance the performance of our CNN for the specific task of crystal compound formation energy prediction.

**Data sets.** We obtained a data set of 139,367 crystal structures along with their corresponding DFT-calculated formation energies (the target variables) from Materials Project (v2021.05.13) [29]. From this, 15,354 structures are excluded because they either require a high resolution or a large image (more details in section IV). To train our model, we split the data into train (60%), validation (20%), and test (20%) sets. During the data pre-processing stage, we removed 9,175 crystal structures from the train set that either contain two atoms occupying the same voxel or have a unit cell that does not fit in the 17-Å cubic box, as described in detail in section IV. During training, we employ data augmentation by randomly rotating each crystal image before feeding it into the model at each epoch (see Supplementary Figure S1). This technique helps alleviate overfitting (see Supplementary Figure S4) and enhances the predictive performance of our model. Data augmentation is particularly beneficial as it effectively increases the size of the train data and implicitly enforces the rotation-invariance of crystal compounds with respect to their formation energy, as explained further in this section. To monitor the training process and prevent overfitting, we use predictions on the validation data. Once the model is trained, we evaluate its overall performance using the test data, as outlined below. In section III, we delve into the significance of data augmentation and skip connections in our CNN architecture, highlighting their role in improving the model’s performance.

## B. Formation Energy Prediction Assessment

In this section, we examine the performance of our model’s prediction. As detailed in section IV, we employ an ensemble averaging technique for predicting the formation energy. Figure 2 (a) shows the parity plot of the formation energy prediction of our model against the DFT-calculated formation energies on both the train and test sets. The results indicate an MAE of 0.042 eV per atom and 0.046 eV per atom on the train and test sets, respectively. Over 89% of the samples in the test set exhibit absolute errors below 0.1 eV per atom, and only about 2% of the samples have absolute errors exceeding 0.2 eV per atom (see Supplementary Figure S2(b)). The formation energy prediction error (i.e., predicted formation energy - DFT formation energy) shows a slightly positive skew normal distribution with a median and mean value of 0.003 eV per atom and -0.003 eV per atom on the test set (see Supplementary Figure S2(b)). As shown in Figure 2(b) and (c), our model tends to exhibit higher errors for crystal compounds with more positive and larger formation energies. This trend has also

been observed in other studies [16, 42]. To exemplify this trend, we analyze four equally populated subsets of our test set sorted by the formation energy with respective formation energy ranges of [-4.47,-2.39), [-2.39,-1.47), [-1.47,-0.46), and [-0.46,5.33) eV per atom with calculated MAEs of 0.037, 0.039, 0.046, and 0.064 eV per atom, respectively. The relatively diminished prediction performance observed for larger, positive-value ranges of formation energy can be attributed to an inherent bias in the existing dataset. The data available in the Materials Project predominantly comprises chemically stable structures characterized by negative formation energies. In contrast, the occurrence of chemically unstable crystal structures with positive formation energies remains a minority within this dataset. Notably, less than 10% of all samples possess positive formation energy (see Supplementary Figure S2(a)). Pandey *et al.* [43] have elucidated how this disparity in data distribution impacts the model’s predictive capabilities.

We conducted a comparative analysis of our model’s predictive performance with state-of-the-art machine learning models, including ElemNet [16] and Roost (Representation Learning from Stoichiometry) [17] as the best models based on compositional features, and ALIGNN [23] and CGCNN [18] as the top-performing graph-based models. Table I presents a comparison of the formation energy MAEs between different models, including two architectures of our CNN: a 3-layer CNN without skip connections (shallow CNN), which was utilized in our previous work for predicting the synthesizability of crystalline compounds [36], and the 15-layer CNN with skip connections (deep CNN). The deep CNN model in this study outperforms Roost and ElemNet, and performs on par with CGCNN, albeit slightly underperforms ALIGNN. It is worth mentioning that optimizing the architecture of a CNN is an empirical process, and as there are limited regression studies using deep CNNs on visual images, there is potential for improvement by modifying the CNN design. The significant improvement of the deep CNN compared to the shallow CNN in this study (MAE of 0.046 eV per atom versus 0.337 eV per atom on the test set) highlights the importance of network depth and skip connections in enhancing the predictive performance. As shown in Supplementary Figure S3, we compare the learning curves of our model and CGCNN with respect to the size of the training dataset. Notably, our image-based model demands a larger training dataset to attain an equivalent level of accuracy compared to CGCNN. This difference is expected considering the substantial depth of our CNN architecture, comprising a significant number of trainable parameters (2,678,641 parameters), thereby necessitating large volume of training data for effective learning. It is important to note that ALIGNN and CGCNN incorporate physical attributes such as electronegativity, group number, covalent radius, valence electrons, first ionization energy, electron affinity, and atomic volume as node features in their graph representation, thereby incorporating additional information that is more challenging to



capture from the visual image representation employed in our work.

### C. Approximate Rotational Invariance

We mitigate overfitting in our deep learning model by applying augmentation to the training data through random rigid-body rotations. As shown in Supplementary Figure S4, this simple technique effectively addresses overfitting. This data augmentation method, when combined with an ensemble averaging approach, also confers approximate rotation invariance to the formation energy prediction. In our ensemble averaging method, the formation energy is predicted by averaging the results of 50 randomly rotated instances of the given crystal structure (see section IV for more details). To assess the degree of rotation invariance in our predictive model, we randomly select 20 crystal samples from the test set and subject each sample to 500 random rotations. Figure 3 illustrates the range of predicted formation energies for each crystal sample across different rotations. The interquartile range (IQR) for these 20 crystal structures exhibits an average and maximum values of 0.009 and 0.018 eV per atom, respectively. Among the distributions of predicted formation energies for these 20 samples, the average and maximum standard deviations are 0.007 and 0.013 eV per atom, respectively. Notably, these values are nearly an order of magnitude smaller than the MAE of the test set (0.046 eV per atom as shown in Figure 2(a)). Since the test set MAE measures the precision of our model relative to DFT, the approximately tenfold reduction in prediction span for rotated samples indicates the approximate rotation invariance of the formation energy.

To demonstrate the impact of ensemble averaging on improving the performance and robustness of our model, we compare the achieved approximate rotation invariance between ensemble-averaged predictions (ensemble size of 50) and predictions without ensemble averaging. Supplementary Figure S6 showcases the range of formation energy predictions for the same crystal structures as shown in Figure 3, but without employing ensemble averaging. By comparing Supplementary Figure S6 and Figure 3, we observe that the variation in formation energy predictions for different rotations increases approximately 6 to 7 times when ensemble averaging is not employed. Apart from the approximate rotation invariance, the ensemble averaging approach also provides a valuable metric - variance of the predictions - that can be used to assess the predictive uncertainty of our model, enabling us to evaluate the reliability of our model effectively.

To gain further insights into the overall effect of ensemble averaging on the model’s performance, Supplementary Figure S7 displays the range of formation energy MAE (prediction error relative to DFT-calculated formation energy) for different rotated instances of the test data. In the case without ensemble averaging, the MAE calculated over 400 instances of the test data ex-

hibits an IQR value of  $3.1 \times 10^{-4}$  eV per atom and a median of 0.05949 eV per atom. In contrast, for the ensemble averaging case with 50 instances of the test set, the IQR and median values are  $5.8 \times 10^{-5}$  eV per atom and 0.04649 eV per atom, respectively. The comparison between the two scenarios, as depicted in Supplementary Figures S6 and S7, demonstrates that ensemble averaging significantly reduces the variation in formation energy predictions for different rotations and leads to a lower MAE overall. These results highlight the effectiveness of ensemble averaging in enhancing the performance and robustness of our model.

Ensemble averaging enhances the reliability of predictions and diminishes the MAE. However, it concurrently amplifies the computational expenses involved in the prediction process, potentially rendering it impractical for exhaustive explorations of extensive chemical spaces or integration into generative models. The ensemble size for averaging serves as an adjustable parameter in our model, enabling users to strike a balance between computational efficiency and predictive reliability. In the context of broader investigations, a multi-tiered screening approach can be employed: a preliminary, low-level exploration utilizing a reduced ensemble size for high-throughput screening, followed by a comprehensive, high-level investigation involving a larger ensemble size to ensure precise predictions within a limited chemical space.

### D. Binary Convex Hull Prediction Assessment

Formation energy convex hulls are commonly used for rapid stability assessment of chemical compounds based on the energy above the hull [9]. A convex hull represents compounds with the lowest formation energy at any composition within a given chemical space, with reference to the pure end members. Figure 4 illustrates a comparison between the predicted convex hulls generated by the deep CNN model in this study and the convex hulls constructed from DFT-calculated formation energies for selected binary systems (chosen from a set of 3,115 predicted binary convex hulls). In this section, it is important to mention that the predicted convex hulls have been constructed solely on the basis of crystal structures for which DFT formation energies are available. This means that no new crystal structures for any binary chemical space outside of the Materials Project have been taken into consideration. As depicted in Figure 4, the formation energy MAE exhibits significant variation among different systems. For instance, the MAE for Tm-Pt is 0.042 eV per atom, while for Tb-N it is 0.115 eV per atom. However, relying solely on the MAE as a measure is insufficient to assess the deviation of the predicted convex hull from the DFT (or true) convex hull. Notably, the predicted convex hull for Tm-Pt and Tb-N demonstrates good agreement with the DFT, despite their considerably different MAEs. Tm-Pt exhibits one of the lowest MAEs, while Tb-N has one of the largest MAEs among

the 3,115 predicted binary convex hulls. To gain a more comprehensive understanding of how errors in formation energy impact convex hull predictions, we introduce additional metrics such as depth error and hull accuracy, in addition to the formation energy MAE (for more details, refer to section IV).

The depth error is an evaluation metric that quantifies the deviation in the depth of the predicted convex hull compared to the true convex hull. It is computed as the difference between the confined areas of the predicted and true convex hulls, normalized by the true area (refer to the definition in section IV). The confined area is defined by the zero formation energy line, which connects the pure end members of the binary system. The depth error takes values ranging from -1 to large positive values, where the extremes indicate an extremely shallower (or flat line) or deeper convex hull prediction, respectively, compared to the DFT convex hull. In addition to the depth error, the hull accuracy metric measures the number of correct predictions of crystal samples that form the convex hull. As explained in more detail below, metrics like the depth error are crucial for assessing the accuracy of stability assessment based on the energy above the hull.

In Figure 4, we present several examples of binary convex hulls, illustrating different combinations of MAE, depth error, and hull accuracy. These examples serve to emphasize the importance of considering these metrics together for a comprehensive evaluation of the predicted convex hull. As an extreme case, consider the Nd-Y binary system, which exhibits an MAE of 0.027 eV per atom (among the lowest values in the set of 3,115 convex hull MAEs). However, despite the low MAE, the predicted convex hull deviates significantly from the DFT convex hull, with a depth error of almost 7,000 percent and a hull accuracy of only 33 percent. This discrepancy arises from the extremely shallow DFT convex hull, where even small errors in the formation energy can lead to substantial deviations in the predicted convex hull. In contrast, the Tb-N binary system displays a large MAE of 0.115 eV per atom, but the corresponding depth error is only -0.04 (or -4 percent). This discrepancy is attributed to the N-rich samples, as evident in Figure 4(c) and Supplementary Table S5, where our formation energy model exhibits poor performance (for details on other binary systems, please refer to Supplementary Tables S3-S8). However, due to the convex hull comprising only three samples and having a relatively deep structure (i.e., a large range of formation energy), the large MAE does not result in a significant change in the depth of the convex hull. Similarly, Figures 4(d) and (f) depict a similar scenario for the Mo-Se and V-Se systems, where despite large MAEs for the Se-rich samples, the depth error remains relatively low. The Tm-Pt binary system represents a combination of low MAE, low depth error, and high hull accuracy. The Na-Cl system demonstrates moderate values for MAE, depth error, and hull accuracy. Notably, the Tm-Pt and V-Se systems achieve the

highest hull prediction accuracy (refer to Figures 4(a) and (d)), despite having different ranges of formation energy MAEs. This observation can be attributed to the larger number of compounds forming the DFT hull, which leads to higher accuracy values even if a few samples are misplaced on the hull in the prediction. These examples highlight that a high MAE in predicting formation energy does not necessarily result in poor convex hull prediction, and conversely, a low MAE in predicting formation energy does not guarantee a good agreement between the predicted and true convex hulls. It is crucial to consider all these metrics together to obtain a comprehensive assessment of the predictive performance of the convex hulls.

To conduct a comprehensive evaluation of binary convex hull prediction, we calculate the MAE and depth error for all 3,115 binary systems across the entire dataset (train, validation, and test). Figure 5 (a) illustrates the relationship between the MAE and depth error of these binary convex hulls. Interestingly, the majority of convex hulls exhibit low depth errors regardless of their MAE, encompassing both low and high MAE ranges. There is no clear correlation observed between the MAE and depth error. Notably, the largest depth errors tend to occur in the lower ranges of MAE, aligning with our observation in the Nd-Y example presented in Figure 4(e). This observation leads us to conclude that factors such as the range of formation energy (or the depth of the convex hull) play a more crucial role in determining the depth error than the MAE alone. Hence, in Figure 5 (b), we depict the relationship between the depth error and the formation energy range, represented by  $\Delta E$  (i.e.,  $\Delta E = \max(\text{DFT } E_f) - \min(\text{DFT } E_f)$ ) across all binary systems. This plot illustrates that a majority of poor predictions of binary convex hull depths (large positive errors or close to -1) appear at low ranges of  $\Delta E$ .

To gain further insights into the binary convex hull predictions made by our model, we group the 3,115 binary systems into five material classes: ceramics (combinations of a non-metal and a metal), semiconductors (combinations of two metalloids), metals (combinations of two metals), semimetals (combinations of a metalloid and a metal), and nonmetals (combinations of two non-metal elements). Figure 6 (a) displays the density distribution of the formation energy MAE for each class of binary convex hulls. Interestingly, the MAE distribution exhibits a distinct shift in its peak across different material classes. Metallic convex hulls demonstrate the lowest MAE, while nonmetallic convex hulls exhibit the highest. The MAE progressively increases as we transition from metallic to ionic materials, with semimetals, semiconductors, and ceramics falling between the lowest and highest MAE values, respectively. This analysis of convex hull MAE for different material classes sheds light on the limitations of our predictive model. The reason for the difference in MAE between metallic and nonmetallic compounds can be attributed to the fact that there is a wider range of formation energies for nonmetallic compounds and a

significantly lower number of non-metallic compounds in the data set (see Supplementary Figure S8 for more details). Due to these factors, the performance of our model is comparatively lower for non-metallic compounds. In Figure 6 (b), we observe that unlike the MAE, the depth error distribution for binary convex hulls does not show a clear distinction among different material classes. The depth error range depicted in Figure 6 (b) is limited to -1 to 1. Semi-metallic and ceramic materials exhibit the best performance, with their distributions peaking close to zero. Metallic and semiconductor materials show almost uniform distributions without a clear peak. Non-metallic materials display a significantly negative skewness in the depth error distribution, indicating that the predicted convex hulls are shallower than the true convex hulls (i.e., underprediction of the convex hull). Semiconductors also exhibit positive skewness, albeit to a lesser extent. Supplementary Table S1 provides further statistics on binary convex hull predictions for different material classes.

To gain further insights into the binary convex hull predictions, we provide Supplementary Figure S9, which showcases the formation energy MAE for different pairs of elements in our binary systems. It is evident that certain elements, such as C and F, consistently exhibit high MAEs for the predicted convex hull, regardless of the second element they are paired with. For instance, the formation energy MAEs of carbides, nitrides, and fluorides in our study are relatively large, with respective values of 0.101, 0.096, and 0.073 eV per atom (see Supplementary Table S2). The average MAEs over distinct binary systems containing C, N, and F are amongst the highest, with respective values of 0.178, 0.118, and 0.178 eV per atom (see Supplementary Table S2 for details). To ensure that any observed bias in our model is not due to an uneven distribution of chemical elements in our training data, we analyze the frequency of chemical elements in our training data, as depicted in Supplementary Figure S10 over the periodic table. Notably, elements such as C, N, and F are among the more frequent ones. Oxygen, which is the most common element in our train data set, ranks among the top 20 in terms of binary convex hull prediction MAEs for systems involving oxygen (see Supplementary Table S2). Supplementary Table S2 provides a comprehensive list of MAEs, average MAEs, and median depth errors for pairs of elements grouped by their positions in the periodic table. Additionally, Figure 7 displays the MAEs and depth errors for different pairs of elements, specifically selected based on their frequent appearance in the binary systems that formed the most convex hulls within the analyzed binary dataset. The heatmaps presented in Supplementary Figures S9 and 7 reveal that nonmetals and halogens, including C, F, H, S, N, and Cl, exhibit elevated average MAEs. The depth error heatmap in Figure 7 further demonstrates that there is no one-to-one correspondence between the MAE and depth error.

Aside from depth error and hull accuracy, we define two

other error metrics; the positional distance and the adjacency distance. These metrics enable us to evaluate the performance of our model in predicting the order of structures at a given composition. Given two permutations (i.e., predicted and DFT) of the list of crystal samples at a fixed composition, the positional distance measures the number of elements needed to be swapped to turn one list to the other and the adjacency distance measures the minimum number of adjacent transpositions needed to transform one permutation into another. For example, for two permutations  $P1=[1, 2, 3]$  and  $P2=[2, 1, 3]$ , the positional and adjacency distances are 2 and 1, respectively. Supplementary Tables S3-S8 report the positional and adjacency distances, the MAE, and the hull match (whether the predicted crystal structure on the hull matches DFT) at different compositions of the example binary systems of Figure 4. For example, as shown in Supplementary Table S5, on the  $N_2$  end-member of N-Tb, the predicted sample on the hull disagrees with the DFT-calculated sample. The predicted list of 14 crystal samples has a positional and adjacency distance of 8 and 10. As shown in Supplementary Table S8, the Mo-side of Mo-Se shows a positional and adjacency distance of 8 and 4 for a list of 7 crystal samples with the predicted sample on the hull matching the DFT-calculated sample.

### III. DISCUSSION

This work introduces the utilization of a sparse voxel image representation of crystal compounds in combination with a deep 15-layer CNN as a learning algorithm for material property regression. It provides valuable insights into the optimal design of deep CNNs as predictive models for material properties. While deep CNNs involve intricate architectures with numerous (hyper)parameters, we focus on two crucial design aspects that have significantly improved the predictive performance of our model: data augmentation and skip connections. We discuss each aspect briefly below.

Data augmentation plays a pivotal role in our training process, where we employ the augmentation of rotated crystal images in our train set. This technique effectively reduces overfitting, as demonstrated by the narrowing gap between the validation and training errors in Supplementary Figure S4. In the absence of data augmentation, the validation error plateaus at a fraction of the total epochs while the training error continues to decrease, indicating overfitting (as shown in Supplementary Figure S4). By implementing data augmentation, the training and validation errors converge, indicating a more balanced model performance. Additionally, data augmentation enables the deep CNN to identify general underlying features by implicitly enforcing rotation invariance in crystals' formation energy. Rather than "memorizing" patterns from arbitrarily oriented crystal structures, the network "learns" general features from multiple randomly rotated orientations of the same crystal

structure (as many rotations as the number of epochs). This approach mitigates overfitting to specific crystal orientations and facilitates the identification of rotation-invariant features embedded in crystal images.

Instead of relying on augmentation of rotated samples in the training data, an alternative approach is to employ neural networks that explicitly enforce rotation equivariance. Euclidean neural networks (e.g., E(3) equivariant neural networks or E(3)NN)[44–47] are an example of such networks that utilize sophisticated filters, such as radial functions and spherical harmonics, to achieve equivariance to 3D Euclidean transformations, including rotation. While E(3)NN networks offer an explicit solution for rotation equivariance, in our experience, we have found the data augmentation approach to be more feasible due to the relative ease of optimization of conventional CNNs compared to equivariant CNNs. Nonetheless, the application of E(3)NN networks with graph representations has shown promise in previous studies. For example, successful utilization of E(3)NN networks with graph representation of crystalline materials has been reported in the literature[48–51]. Although data augmentation and E(3)NN differ in their technical approaches, they both serve as regularization methods to achieve rotation equivariance and alleviate overfitting in the network. E(3)NN achieves this through the implementation of sophisticated filters that effectively reduce the number of parameters in the network, leading to a more compact and regularized model. On the other hand, data augmentation addresses overfitting by expanding the training data size, compensating for the large number of trainable parameters in the network. While E(3)NN focuses on parameter reduction to enforce regularization, data augmentation increases the diversity and variability of the training data. Both techniques contribute to enhancing the network’s generalization capabilities and improving its performance on unseen data.

The incorporation of skip connections into the 15-layer deep convolutional network leads to a significant improvement in the prediction of formation energy, as demonstrated in Supplementary Figure S5. This enhancement can be attributed to the ability of skip connections to bypass local atomic features discovered in the shallower layers of the network. Simultaneously, the deep architecture of the network facilitates the exploration and discovery of more global features inherent in crystalline materials. Deep neural networks often suffer from the degradation problem, where the performance deteriorates as the network becomes deeper. This occurs because randomly initialized weights tend to approach zero as the number of layers increases, causing the optimizer to behave chaotically [52]. To address this issue, architectures like ResNets [53, 54] employ skip connections. Traditionally, skip connections are recognized for their role in alleviating optimization challenges by producing smoother loss functions, facilitating easier training [52]. However, our work sheds light on an additional aspect of skip connections beyond their optimization benefits. We

demonstrate that skip connections serve as a mechanism to capture the essential physicochemical information at different levels. By allowing the outputs of different layers (both shallow and deep) to bypass through identity mapping, skip connections enable the network to leverage local atomic fingerprints from shallower layers while simultaneously learning abstract, generalized features from deeper layers. In this way, skip connections facilitate the integration of both local and global information, leading to improved performance in formation energy prediction.

## IV. METHODS

*Data Collection and Voxel Image Preparation.* We gather crystal structure information in CIF format and the corresponding DFT-calculated formation energies from the Materials Project database (v2021.05.13)[29]. To extract the structural information, we utilize the Atomic Simulation Environment (ASE) package [55]. Our in-house Python code is then employed to generate sparse voxel images of the crystals. In the voxelization process, we repeat the crystal unit cell (cubic or non-cubic) in space to fill a cubic box with an edge size of 17 Å. We eliminate a crystal structure if its unit cell does not fit in the cubic box. The box is then voxelized using a  $32 \times 32 \times 32$  grid, resulting in images with dimensions of  $32 \times 32 \times 32$  voxels. To ensure that each voxel contains at most one atom, we set the minimum interatomic distance to be greater than the diagonal of a voxel,  $d_v$ , calculated as  $d_v = (17/32) \times \sqrt{3} = 0.92$  Å. Consequently, crystal structures with minimum interatomic distances larger than 0.92 Å are filtered out. The 3D sparse voxel images of crystals are color-coded using three channels, similar to an RGB image. These channels represent the normalized atomic number, group number, and period number. For lanthanides and actinides, we assign a group number of 3.5. During training, to introduce variability and enhance generalization, we apply a random rotation to each crystal image at each epoch. Rather than applying a direct rotation to the unit cell and subsequently executing the computationally intensive task of filling the 17 Å box - a method which becomes intractably repetitive - we initially construct a larger ‘encompassing’ box with an edge equal to the diagonal of the 17- Å cubic box. During the data pre-processing stage, we fill the larger box by replicating the crystal unit cell in all directions only once. Consequently, whenever an instance of a crystal structure input is requested, either for training or prediction, we perform a random rigid-body rotation to the larger box, while the 17- Å box remains unchanged and consistently populated after each rotation. Thereafter, we perform the voxelization of the 17- Å box to generate the final sparse voxel images. Supplementary Figure S1 visually details the rotation methodology.

*Convolutional Neural Network.* We develop a 15-convolutional-layer network consisting of 7 residual blocks and 3 average pooling layers, followed by a fully



connected neural network (see Supplementary Figure S11). Each residual block consists of two convolutional layers, each followed by a rectified linear unit (ReLU) activation layer and a skip connection that connects the beginning of the block to its end. In each convolutional layer, we use a kernel of size 3 and padding of type SAME with stride 1 to ensure that the filter is applied to all the voxels of the input. To merge a skip connection (i.e., side stream) with the mainstream coming from the convolutional layer, we either use addition or concatenation. We use the concatenation of outputs only before a pooling layer in order to double the number of channels while reducing the image size during pooling. The addition of outputs is used elsewhere as the method of merging in the residual blocks.

The deep convolutional network consists of three distinct segments, each containing a different image size and ending with a pooling layer. The first segment consists of a single convolutional layer, followed by an activation layer. In this layer, we increase the number of channels from 3 to 32. This single layer is followed by two residual blocks, each consisting of two convolutional and activation layers, outputting 32 channels. We utilize concatenation to combine the outputs of the mainstream and skip connection, rendering the number of channels of the output of this segment equal to 64. This segment ends with an average pooling layer, reducing the image size by half ( $16 \times 16 \times 16$ ). The second segment consists of three residual blocks, followed by an average pooling. The images passing through this segment have 64 channels, and at the end of the segment, their size is reduced by half ( $8 \times 8 \times 8$ ) and their channels are doubled (128). The last segment consists of two residual blocks and an average pooling layer, but in this case, the last block uses addition instead of concatenation, keeping the channels as 128 and reducing the size to  $4 \times 4 \times 4$ . A detailed schematic of the network is shown in Supplementary Figure S11.

The last pooling layer is flattened to a vector of size ( $4 \times 4 \times 4 \times 128 = 8192$ ) and is connected to a fully connected network with a node architecture of 16-16-1 with linear activation functions. The Keras package [56] is used to build and train this network. The 3D images of the train set are randomly rotated in 3D space and input to the network for 500 epochs in batches of size 32. The mean squared error (MSE) is used as the loss function. To train the network, we use the Adam optimizer with a learning rate of 0.001, the exponential decay rates of 0.9 and 0.999 for the first and second moment estimates, respectively, and a machine precision threshold (or  $\epsilon$ ) of  $1e-07$ .

*Rotational Ensemble Averaging* Once the model is trained, we employ an ensemble averaging method for prediction of the formation energy. Once a crystal sample is input into the trained model, an ensemble of 50 randomly rotated instances of the sample is generated and the formation energy prediction is averaged over the ensemble. The ensemble averaging method improves the prediction accuracy and robustness of our model, as de-

tailed in section II, Figure 3, and Supplementary Figures S6 and S7.

*Error Metrics.* The evaluation of the formation energy prediction and the constructed convex hull is performed using the following error metrics:

Formation Energy Mean Absolute Error (MAE): The MAE is calculated using the formula:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

where  $y_i$  represents the true formation energy of sample  $i$  (DFT-calculated formation energy obtained from the Materials Project database),  $\hat{y}_i$  corresponds to the model’s prediction of the formation energy for sample  $i$ , and the sum runs over total of  $n$  samples. When computing the MAE for a binary convex hull prediction, only crystal compounds (or samples) from that specific binary system are included.

Depth error for Convex Hull: The depth error for the convex hull measures the difference in the confined area between the predicted and true convex hulls, and is defined as:

$$\text{Depth error} = \frac{A_{\text{predicted}} - A_{\text{true}}}{A_{\text{true}}} \quad (2)$$

where  $A_{\text{predicted}}$  and  $A_{\text{true}}$  represent the areas enclosed by the predicted and true (or DFT-calculated) convex hulls, respectively.

Accuracy of Convex Hull Prediction: The accuracy of the convex hull prediction is calculated as the percentage of correctly predicted crystal samples on the hull with respect to the crystal samples on the DFT-calculated hull. In other words, the hull accuracy measures the percentage of predictions on the hull that matches the DFT-calculated samples on the hull. Accordingly, if our model mistakenly predicts a crystal sample to be on the hull while the DFT-calculated sample is above the hull, the hull accuracy measure will not be affected (e.g., see Figure 4(a)).

## DATA AVAILABILITY

The data developed or used by this study is available on our GitHub repository at: [https://github.com/kadkhodaei-research-group/XIE-SPP/tree/main/training/formation-energy/data\\_sets](https://github.com/kadkhodaei-research-group/XIE-SPP/tree/main/training/formation-energy/data_sets)

## CODE AVAILABILITY

The codes developed or utilized in this study are openly accessible to support transparency and facilitate further research. They can be found in our GitHub repository at: <https://github.com/kadkhodaei-research-group/XIE-SPP>

## AUTHOR CONTRIBUTION STATEMENT

A.D. and S.K. conceptualized the presented idea. A. D. designed the networks and the computational framework. Both A.D. and S.K. analyzed the data and wrote the manuscript. S.K. supervised the project.

## ACKNOWLEDGEMENTS

This research is based upon work supported by the National Science Foundation (NSF) under Award Numbers DMR-2119308. We used resources at the Electronic Visualization Laboratory (EVL) at UIC available through the NSF Award CNS-1828265. Additionally, we would like to thank Zahra Kadhodaie for helpful suggestions regarding the design of the convolutional network.

## COMPETING INTERESTS

The authors declare no financial or non-financial competing interests.

## REFERENCES

- [1] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, "Machine learning for molecular and materials science," *Nature*, vol. 559, no. 7715, pp. 547–555, 2018. [Online]. Available: <https://doi.org/10.1038/s41586-018-0337-2>
- [2] G. Pilania, "Machine learning in materials science: From explainable predictions to autonomous design," *Computational Materials Science*, vol. 193, p. 110360, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0927025621000859>
- [3] J. F. Rodrigues, L. Florea, M. C. F. de Oliveira, D. Diamond, and O. N. Oliveira, "Big data and machine learning for materials science," *Discover Materials*, vol. 1, no. 1, p. 12, 2021. [Online]. Available: <https://doi.org/10.1007/s43939-021-00012-0>
- [4] J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques, "Recent advances and applications of machine learning in solid-state materials science," *npj Computational Materials*, vol. 5, no. 1, p. 83, 2019. [Online]. Available: <https://doi.org/10.1038/s41524-019-0221-0>
- [5] V. Stanev, K. Choudhary, A. G. Kusne, J. Paglione, and I. Takeuchi, "Artificial intelligence for search and discovery of quantum materials," *Communications Materials*, vol. 2, no. 1, p. 105, Oct 2021. [Online]. Available: <https://doi.org/10.1038/s43246-021-00209-z>
- [6] J. Wei, X. Chu, X.-Y. Sun, K. Xu, H.-X. Deng, J. Chen, Z. Wei, and M. Lei, "Machine learning in materials science," *InfoMat*, vol. 1, no. 3, pp. 338–358, 2019. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/inf2.12028>
- [7] D. Morgan and R. Jacobs, "Opportunities and challenges for machine learning in materials science," *Annual Review of Materials Research*, vol. 50, no. 1, pp. 71–103, 2020. [Online]. Available: <https://doi.org/10.1146/annurev-matsci-070218-010015>
- [8] G. H. Gu, J. Noh, I. Kim, and Y. Jung, "Machine learning for renewable energy materials," *J. Mater. Chem. A*, vol. 7, pp. 17 096–17 117, 2019. [Online]. Available: <http://dx.doi.org/10.1039/C9TA02356A>
- [9] C. J. Bartel, A. Trewartha, Q. Wang, A. Dunn, A. Jain, and G. Ceder, "A critical examination of compound stability predictions from machine-learned formation energies," *npj Computational Materials*, vol. 6, no. 1, p. 97, 2020. [Online]. Available: <https://doi.org/10.1038/s41524-020-00362-y>
- [10] F. Faber, A. Lindmaa, O. A. von Lilienfeld, and R. Armiento, "Crystal structure representations for machine learning models of formation energies," *International Journal of Quantum Chemistry*, vol. 115, no. 16, pp. 1094–1101, 2015. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.24917>
- [11] L. Ward, R. Liu, A. Krishna, V. I. Hegde, A. Agrawal, A. Choudhary, and C. Wolverton, "Including crystal structure attributes in machine learning models of formation energies via voronoi tessellations," *Phys. Rev. B*, vol. 96, p. 024104, Jul 2017. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.96.024104>
- [12] B. Meredig, A. Agrawal, S. Kirklin, J. E. Saal, J. W. Doak, A. Thompson, K. Zhang, A. Choudhary, and C. Wolverton, "Combinatorial screening for new materials in unconstrained composition space with machine learning," *Phys. Rev. B*, vol. 89, p. 094104, Mar 2014. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.89.094104>
- [13] L. Ward, A. Agrawal, A. Choudhary, and C. Wolverton, "A general-purpose machine learning framework for predicting properties of inorganic materials," *npj Computational Materials*, vol. 2, no. 1, p. 16028, 2016. [Online]. Available: <https://doi.org/10.1038/npjcompumats.2016.28>
- [14] A. Dunn, Q. Wang, A. Ganose, D. Dopp, and A. Jain, "Benchmarking materials property prediction methods: the matbench test set and automatminer reference algorithm," *npj Computational Materials*, vol. 6, no. 1, p. 138, 2020. [Online]. Available: <https://doi.org/10.1038/s41524-020-00406-3>
- [15] G. G. C. Peterson and J. Brgoch, "Materials discovery through machine learning formation energy," *Journal of Physics: Energy*, vol. 3, no. 2, p. 022002, mar 2021. [Online]. Available: <https://dx.doi.org/10.1088/2515-7655/abe425>
- [16] D. Jha, L. Ward, A. Paul, W.-k. Liao, A. Choudhary, C. Wolverton, and A. Agrawal, "Elemnet: Deep learning the chemistry of materials from only elemental composition," *Scientific Reports*, vol. 8, no. 1, p. 17593, 2018. [Online]. Available: <https://doi.org/10.1038/s41598-018-35934-y>
- [17] R. E. A. Goodall and A. A. Lee, "Predicting materials properties without crystal structure: deep representation learning from stoichiometry," *Nature Communications*, vol. 11, no. 1, p. 6280, 2020. [Online]. Available: <https://doi.org/10.1038/s41467-020-19964-7>
- [18] T. Xie and J. C. Grossman, "Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties," *Phys. Rev. Lett.*, vol. 120, p. 145301, Apr 2018. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.120.145301>

- [19] C. W. Park and C. Wolverton, "Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery," *Phys. Rev. Mater.*, vol. 4, p. 063801, Jun 2020. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevMaterials.4.063801>
- [20] H. Zhan, X. Zhu, Z. Qiao, and J. Hu, "Graph neural tree: A novel and interpretable deep learning-based framework for accurate molecular property predictions," *Analytica Chimica Acta*, p. 340558, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0003267022011291>
- [21] P. R. Kaundinya, K. Choudhary, and S. R. Kalidindi, "Prediction of the electron density of states for crystalline compounds with atomistic line graph neural networks (alignn)," *JOM*, vol. 74, no. 4, pp. 1395–1405, 2022. [Online]. Available: <https://doi.org/10.1007/s11837-022-05199-y>
- [22] P. P. Meyer, C. Bonatti, T. Tancogne-Dejean, and D. Mohr, "Graph-based metamaterials: Deep learning of structure-property relations," *Materials & Design*, vol. 223, p. 111175, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0264127522007973>
- [23] K. Choudhary and B. DeCost, "Atomistic line graph neural network for improved materials property predictions," *npj Computational Materials*, vol. 7, no. 1, p. 185, 2021. [Online]. Available: <https://doi.org/10.1038/s41524-021-00650-1>
- [24] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, "SchNet – a deep learning architecture for molecules and materials," *The Journal of Chemical Physics*, vol. 148, no. 24, p. 241722, 2018. [Online]. Available: <https://doi.org/10.1063/1.5019779>
- [25] C. Chen, W. Ye, Y. Zuo, C. Zheng, and S. P. Ong, "Graph networks as a universal machine learning framework for molecules and crystals," *Chemistry of Materials*, vol. 31, no. 9, pp. 3564–3572, 2019. [Online]. Available: <https://doi.org/10.1021/acs.chemmater.9b01294>
- [26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: <http://arxiv.org/abs/1512.03385>
- [27] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *CoRR*, vol. abs/1905.11946, 2019. [Online]. Available: <http://arxiv.org/abs/1905.11946>
- [28] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241. [Online]. Available: [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- [29] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, "Commentary: The materials project: A materials genome approach to accelerating materials innovation," *APL Materials*, vol. 1, no. 1, p. 011002, 2013. [Online]. Available: <https://doi.org/10.1063/1.4812323>
- [30] S. Curtarolo, W. Setyawan, G. L. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M. J. Mehl, H. T. Stokes, D. O. Demchenko, and D. Morgan, "Aflow: An automatic framework for high-throughput materials discovery," *Computational Materials Science*, vol. 58, pp. 218–226, 2012. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0927025612000717>
- [31] Z. Chen, X. Li, and J. Bruna, "Supervised community detection with line graph neural networks," 2017. [Online]. Available: <https://arxiv.org/abs/1705.08415>
- [32] K. Choudhary, "The atomistic line graph neural network," <https://github.com/usnistgov/alignn.git>, 2021.
- [33] T. Long, N. M. Fortunato, I. Opahle, Y. Zhang, I. Samathrakakis, C. Shen, O. Gutfleisch, and H. Zhang, "Constrained crystals deep convolutional generative adversarial network for the inverse design of crystal structures," *npj Computational Materials*, vol. 7, no. 1, p. 66, 2021. [Online]. Available: <https://doi.org/10.1038/s41524-021-00526-4>
- [34] J. Hoffmann, L. Maestrati, Y. Sawada, J. Tang, J. M. Sellier, and Y. Bengio, "Data-driven approach to encoding and decoding 3-d crystal structures," 2019. [Online]. Available: <https://arxiv.org/abs/1909.00949>
- [35] P. R. Kaundinya, K. Choudhary, and S. R. Kalidindi, "Machine learning approaches for feature engineering of the crystal structure: Application to the prediction of the formation energy of cubic compounds," *Phys. Rev. Mater.*, vol. 5, p. 063802, Jun 2021. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevMaterials.5.063802>
- [36] A. Davariashtiyani, Z. Kadkhodaie, and S. Kadkhodaie, "Predicting synthesizability of crystalline materials via deep learning," *Communications Materials*, vol. 2, no. 1, p. 115, 2021. [Online]. Available: <https://doi.org/10.1038/s43246-021-00219-x>
- [37] S. Kajita, N. Ohba, R. Jinnouchi, and R. Asahi, "A Universal 3D Voxel Descriptor for Solid-State Material Informatics with Deep Convolutional Neural Networks," *Scientific Reports*, vol. 7, no. 1, p. 16991, Dec. 2017. [Online]. Available: <https://doi.org/10.1038/s41598-017-17299-w>
- [38] J. Noh, J. Kim, H. S. Stein, B. Sanchez-Lengeling, J. M. Gregoire, A. Aspuru-Guzik, and Y. Jung, "Inverse design of solid-state materials via a continuous representation," *Matter*, vol. 1, no. 5, pp. 1370–1384, 2019. [Online]. Available: <https://doi.org/10.1016/j.matt.2019.08.017>
- [39] S. Kim, J. Noh, G. H. Gu, A. Aspuru-Guzik, and Y. Jung, "Generative adversarial networks for crystal structure prediction," *ACS Central Science*, vol. 6, no. 8, pp. 1412–1420, 2020, pMID: 32875082. [Online]. Available: <https://doi.org/10.1021/acscentsci.0c00426>
- [40] B. Sanchez-Lengeling and A. Aspuru-Guzik, "Inverse molecular design using machine learning: Generative models for matter engineering," *Science*, vol. 361, no. 6400, pp. 360–365, 2018. [Online]. Available: <https://www.science.org/doi/abs/10.1126/science.aat2663>
- [41] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255. [Online]. Available: <https://doi.org/10.1109/CVPR.2009.5206848>
- [42] Y. Jiang, D. Chen, X. Chen, T. Li, G.-W. Wei, and F. Pan, "Topological representations of crystalline compounds for the machine-learning prediction of materials properties," *npj Computational Materials*,

- vol. 7, no. 1, p. 28, 2021. [Online]. Available: <https://doi.org/10.1038/s41524-021-00493-w>
- [43] S. Pandey, J. Qu, V. Stevanović, P. St. John, and P. Gorai, “Predicting energy and stability of known and hypothetical crystals using graph neural network,” *Patterns*, vol. 2, no. 11, p. 100361, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2666389921002233>
- [44] T. S. Cohen and M. Welling, “Group equivariant convolutional networks,” *CoRR*, vol. abs/1602.07576, 2016. [Online]. Available: <http://arxiv.org/abs/1602.07576>
- [45] N. Thomas, T. E. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff, and P. Riley, “Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds,” *CoRR*, vol. abs/1802.08219, 2018. [Online]. Available: <http://arxiv.org/abs/1802.08219>
- [46] M. Geiger and T. Smidt, “e3nn: Euclidean neural networks,” 2022. [Online]. Available: <https://arxiv.org/abs/2207.09453>
- [47] T. E. Smidt, M. Geiger, and B. K. Miller, “Finding symmetry breaking order parameters with euclidean neural networks,” *Physical Review Research*, vol. 3, no. 1, jan 2021. [Online]. Available: <https://doi.org/10.1103/PhysRevResearch.3.012002>
- [48] Z. Chen, N. Andrejevic, N. C. Drucker, T. Nguyen, R. P. Xian, T. Smidt, Y. Wang, R. Ernstorfer, D. A. Tennant, M. Chan, and M. Li, “Machine learning on neutron and x-ray scattering and spectroscopies,” *Chemical Physics Reviews*, vol. 2, no. 3, p. 031301, 2021. [Online]. Available: <https://doi.org/10.1063/5.0049111>
- [49] Y. Cheng, G. Wu, D. M. Pajerowski, M. B. Stone, A. T. Savici, M. Li, and A. J. Ramirez-Cuesta, “Direct prediction of inelastic neutron scattering spectra from the crystal structure\*,” *Machine Learning: Science and Technology*, vol. 4, no. 1, p. 015010, feb 2023. [Online]. Available: <https://dx.doi.org/10.1088/2632-2153/acb315>
- [50] R. Okabe, A. Chotrattanapituk, A. Boonkird, N. Andrejevic, X. Fu, T. S. Jaakkola, Q. Song, T. Nguyen, N. Drucker, S. Mu, B. Liao, Y. Cheng, and M. Li, “Virtual node graph neural network for full phonon prediction,” 2023. [Online]. Available: <https://arxiv.org/abs/2301.02197>
- [51] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, “E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials,” *Nature Communications*, vol. 13, no. 1, may 2022. [Online]. Available: <https://doi.org/10.1038/s41467-022-29939-5>
- [52] H. Li, Z. Xu, G. Taylor, and T. Goldstein, “Visualizing the loss landscape of neural nets,” *CoRR*, vol. abs/1712.09913, 2017. [Online]. Available: <http://arxiv.org/abs/1712.09913>
- [53] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” vol. abs/1512.03385, 2015. [Online]. Available: <http://arxiv.org/abs/1512.03385>
- [54] K. He, X. Zhang, S. Ren, and J. Sun., “Identity mappings in deep residual networks,” *CoRR*, vol. abs/1603.05027, 2016. [Online]. Available: <http://arxiv.org/abs/1603.05027>
- [55] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dulak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, “The atomic simulation environment—a python library for working with atoms,” *Journal of Physics: Condensed Matter*, vol. 29, no. 27, p. 273002, 2017. [Online]. Available: <http://stacks.iop.org/0953-8984/29/i=27/a=273002>
- [56] F. Chollet *et al.* (2015) Keras. [Online]. Available: <https://github.com/fchollet/keras>



	Shallow CNN (this work)	Deep CNN (this work)	ALIGNN	CGCNN	ElemNet	Roost
MAE (eV per atom)	0.337	0.046	0.022	0.039	0.1	0.06

TABLE I. **Benchmarking model performance against existing models.** Comparison of the formation energy MAE of different models (ElemNet [16], Roost [17], ALIGNN [23], and CGCNN [18] and this work.

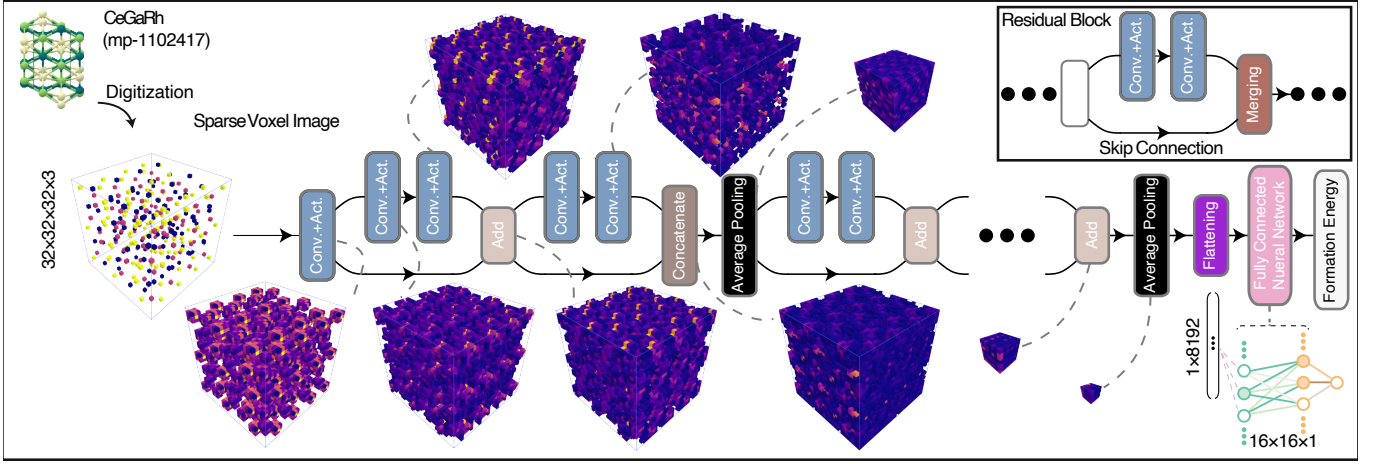


FIG. 1. **The overall design of the deep convolutional neural network and the fully connected neural network of this study.** The crystal structures are digitized into 3D colored sparse voxel images which are input to a deep convolutional neural network. The network consists of 7 residual blocks arranged in sequence in combination with merging and pooling layers. The architecture of each residual block is shown in the inset, which consists of a skip connection used to bypass the output of the previous block to the next. The latent features learned by the convolutional neural network are flattened and input into a fully connected neural network which performs the final prediction of the formation energy.

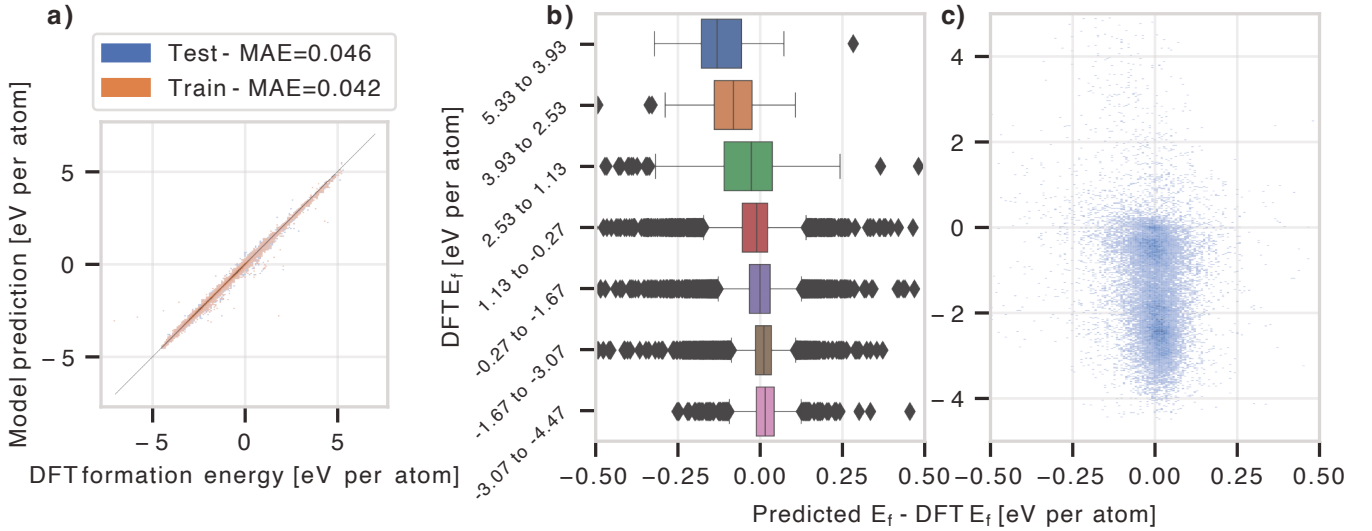
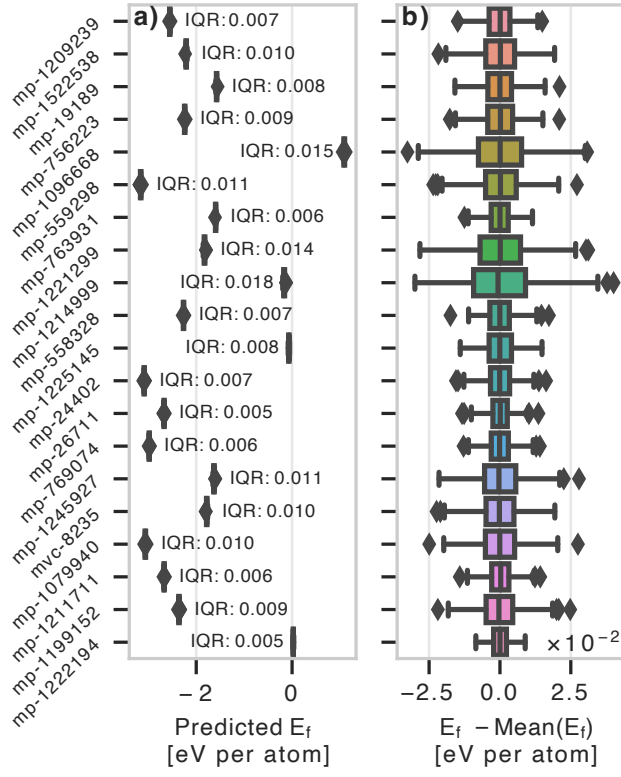


FIG. 2. **Formation energy prediction evaluation.** a) The parity plot for samples in the train and test sets. The MAE of formation energy prediction for the test and train data is reported in the legend. b,c) Distribution of the prediction error of test data over different ranges of formation energy. b) Box and whisker representation of prediction error (i.e., predicted  $E_f$  - DFT  $E_f$ ) for different intervals of DFT formation energy. The left side, middle line, and right side of each box show respectively the first quartile, median, and third quartile of the error. The whisker line shows the minimum and maximum of the error. c) The scatter plot of samples in the test set showing the DFT formation energy versus prediction error.



**FIG. 3. Assessing the approximate rotation invariance of the model's formation energy prediction.** a) The formation energy prediction span over 500 randomly rotated instances of each crystal sample identified by its Materials Project's number. The interquartile range (IQR) associated with the length of each box is reported next to each crystal sample and represents the spread of the data from the 25th to the 75th percentile. b) Box and whisker representation of the formation energy prediction spread (i.e., predicted  $E_f$  - mean (predicted  $E_f$ )) among 500 random rotations of each crystal sample.

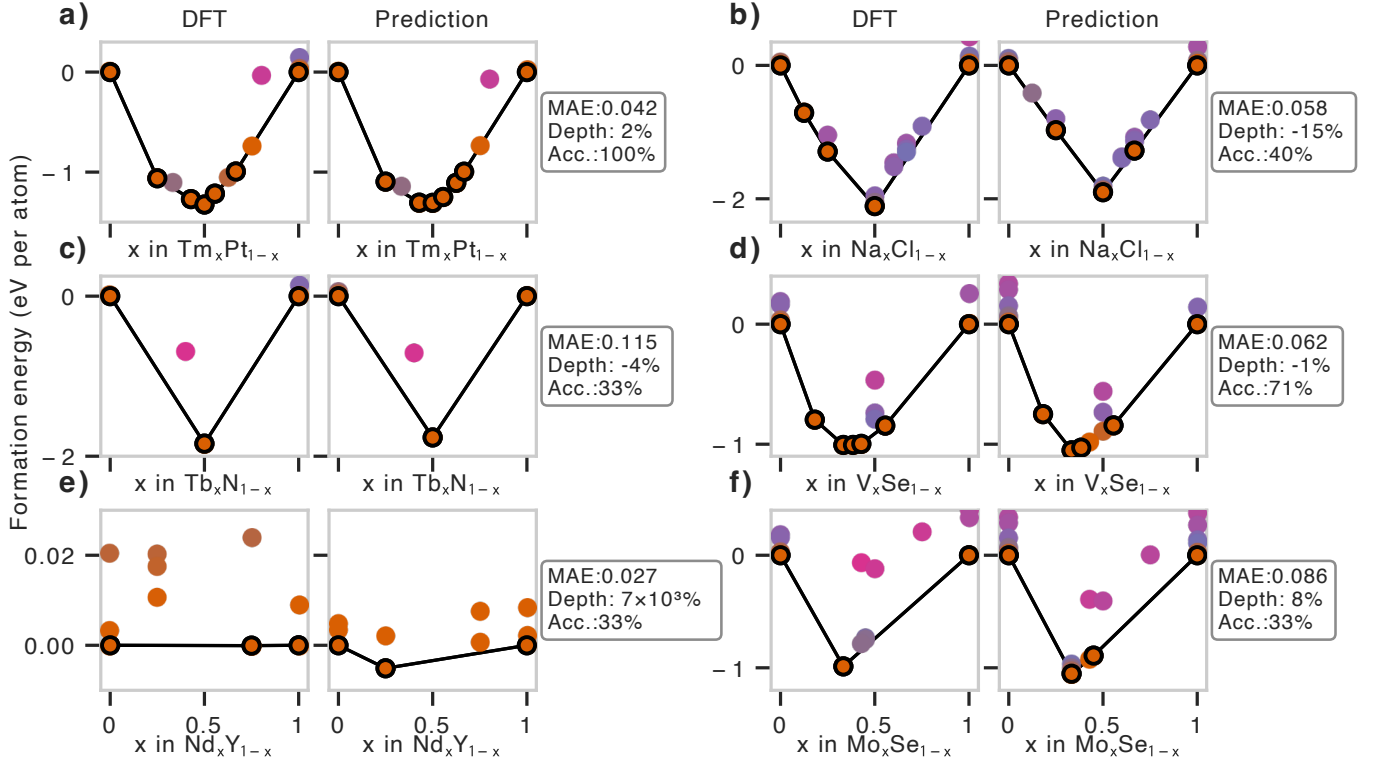


FIG. 4. **Convex hull prediction evaluation.** Comparison of the predicted binary convex hulls with DFT for a) Tm-Pt, b) Tb-N, c) Nd-Y, d) Na-Cl, e) V-Se, and f) Mo-Se. The formation energy MAE (eV per atom), convex hull depth error (%), and hull accuracy (%) values are shown in the box next to each binary system.

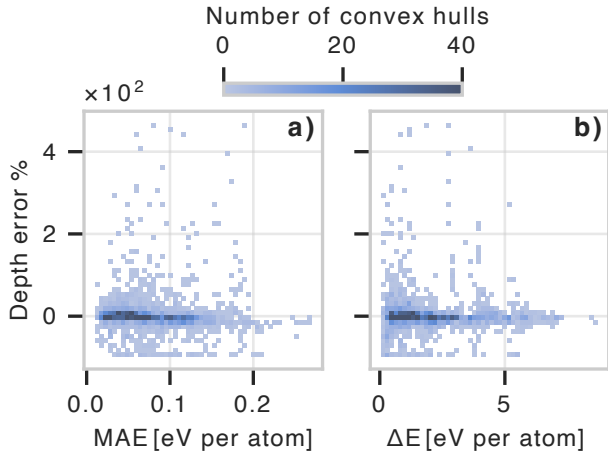


FIG. 5. **Relationship between formation energy prediction error and convex hull performance.** a) Distribution of predicted binary convex hulls over a) the formation energy MAE and depth error and b) formation energy range ( $\Delta E = \max(\text{DFT } E_f) - \min(\text{DFT } E_f)$ ) and depth error.



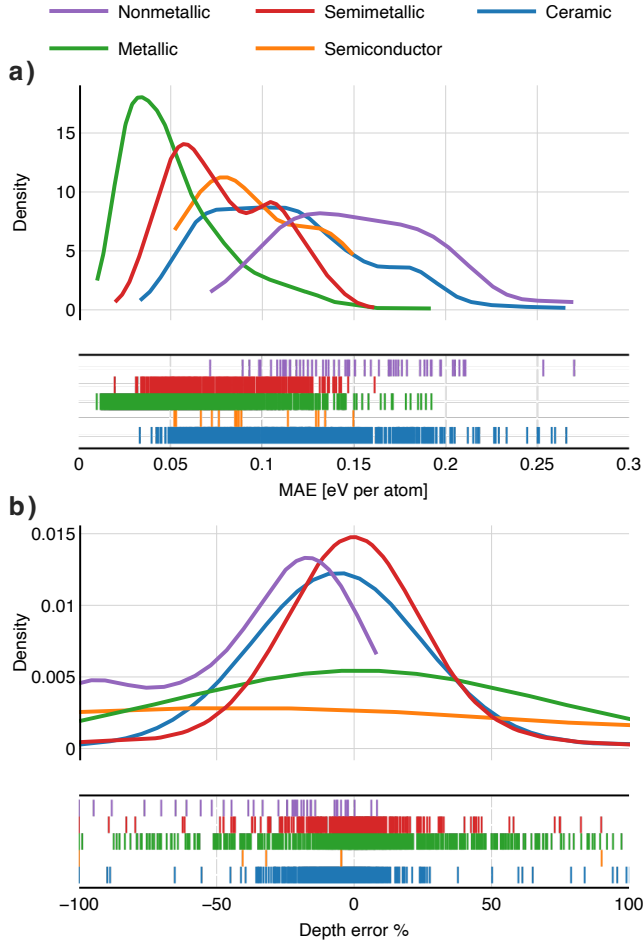


FIG. 6. **Evaluating convex hull predictions across material classes.** Kernel density estimation of the a) formation energy MAE and b) convex hull depth error for the predicted binary systems classified into distinct material classes. The rug plot positioned at the bottom of each density plot illustrates the distribution of individual binary systems within each class.

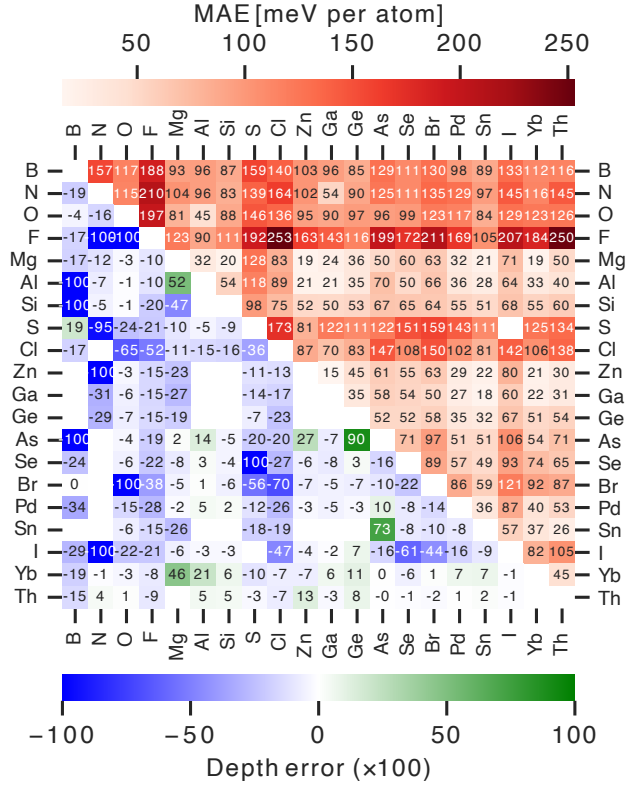


FIG. 7. **Evaluating convex hull predictions across pairs of chemical elements.** Pairwise heat map comparison illustrating the formation energy MAE and convex hull depth error for the most frequent elements in the analyzed binary systems.