




StrainTensorNet: Predicting crystal structure elastic properties using SE(3)-equivariant graph neural networks

Teerachote Pakornchote , Annop Ektarawong , and Thiparat Chotibut ^{*}
 Chula Intelligent and Complex Systems, Department of Physics, Faculty of Science,
 Chulalongkorn University, Bangkok 10330, Thailand



(Received 27 June 2023; accepted 5 November 2023; published 4 December 2023)

Accurately predicting the elastic properties of crystalline solids is vital for computational materials science. However, traditional atomistic-scale *ab initio* approaches are computationally intensive, especially for studying complex materials with a large number of atoms in a unit cell. We introduce a data-driven approach to efficiently predict the elastic properties of crystal structures using SE(3)-equivariant graph neural networks (GNNs). This approach yields important scalar elastic moduli with an accuracy comparable to that of recent data-driven studies. Importantly, our symmetry-aware GNN model also enables the prediction of the strain energy density (SED) and the associated elastic constants, the fundamental tensorial quantities that are significantly influenced by a material's crystallographic group. The model consistently distinguishes independent elements of SED tensors, in accordance with the symmetry of the crystal structures. Finally, our deep learning model possesses meaningful latent features, offering an interpretable prediction of the elastic properties.

DOI: [10.1103/PhysRevResearch.5.043198](https://doi.org/10.1103/PhysRevResearch.5.043198)

I. INTRODUCTION

The elastic properties of crystalline solids such as elastic constants, the bulk modulus, and the shear modulus are important macroscopic quantities that determine materials' mechanical characteristics. Computational study of these elastic properties can provide theoretical guidelines to understand various phenomena in solid materials, e.g., the mechanical stability of crystal structures [1,2], the pressure-driven deformation and phase transition of materials [3,4], the response of materials to sound wave propagation [5,6], and the hardness of materials [7–9], to name a few. From an atomistic-scale description, an *ab initio* approach based on density functional theory has been employed to investigate the macroscopic elastic properties of materials, yielding, for example, elastic constants and elastic moduli that are in agreement with experimental results [10–13].

Three atomistic-scale computational methods are usually adopted to calculate the elastic constants of crystal structures: an energy-based method [14], a stress-strain method [10,15], and a calculation from group velocity [16–18]. The energy-based and stress-strain methods are more standard and derive the elastic constants using an energy tensor and a stress tensor, respectively, acquired from *ab initio* calculations. Although both methods yield a comparable prediction of elastic constants, the energy-based method is

computationally less efficient, involving a larger plane-wave basis set and a larger number of k -point meshes so that the predicted elastic constants converge to reliable values [19]. Indeed, the more efficient stress-strain method is commonly employed to determine elastic constants using established software such as the Vienna *ab initio* simulation package (VASP), QUANTUM ESPRESSO, and the Cambridge serial total energy package (CASTEP) [20–22]. The stress-strain method is also utilized to create a database for elastic properties of inorganic compounds [23,24].

However, atomistic-scale simulation can be computationally prohibitive, especially when the number of atoms in a unit cell grows large. Such a constraint limits an *ab initio* method's capability to investigate more complex crystalline solids. On the other hand, advances in machine learning bring about alternative approaches to computational materials science. This data-driven paradigm can reasonably predict the elastic properties of crystal structures, provided sufficient training data from an *ab initio* method [25,26]. Even when the number of atoms in a unit cell is large, such an approach can efficiently predict the bulk and shear moduli of complex materials such as alloys [27,28]. Applying machine learning together with *ab initio* calculations is also potentially useful in searching for novel superhard materials [29,30].

Machine learning models based on graph neural networks (GNNs) have received increasing attention in studying solid materials and molecules. With GNNs, it is natural to encode atomistic-scale descriptions of solids into a computational framework; atomic locations and the associated atomic attributes can be directly embedded into node attributes of a graph, whereas pairwise interactions among atoms can be encoded into the edge attributes. Efficient GNN training procedures have also been proposed [31–33], enabling GNNs to learn the representation of a complex relationship be-

^{*}Corresponding author: thiparatc@gmail.com

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

tween the input and its associated prediction [34,35]. These neural networks can also be endowed with the translation-rotation-equivariant property, so that input atomic locations of molecules (or point clouds) in \mathbb{R}^3 that differ only in their orientation or their centroid can be identified. Enforcing an SE(3)-equivariant property helps the networks to extract a more compact (translation- and rotation-independent) relationship between the inputs and their predictions. Due to these appealing features, variations of SE(3)-equivariant GNNs have been developed to study materials science [36].

In this paper, we adopt a data-driven approach using GNNs to predict the elastic properties of crystal structures. Accounting for the symmetry of crystalline solids, our SE(3)-equivariant GNNs take as an input atomistic descriptions of strained materials, such as atomic locations and their associated atomic attributes, and predict the strain energy density (SED) of the materials. The prediction of the SED, which is the energy stored in a crystal structure when distorted by either tensile or compressive strains in particular directions, can be obtained efficiently and relatively accurately, given sufficient training data. The model thus provides an alternative approach to the standard *ab initio* energy-based prediction method.

After the SED is computed, we can then calculate the elastic constants and construct the elastic tensor with a simple analytical expression; see Sec. II A. Other macroscopic (scalar) elastic properties including the bulk modulus, shear modulus, Young's modulus, and Poisson's ratio immediately follow from the elastic constants. Section IV reports our model prediction results of these elastic properties. The prediction performances of the scalar elastic properties are comparable to those of recent data-driven work [25,26]. Importantly, however, the data-driven prediction of the SED and the associated elastic constants, which are fundamental tensorial quantities that depend on the crystallographic group, are reported here. The trained model consistently reveals the independent components of strain energy tensors dictated by the symmetry of the crystal structures, as evident from the degeneracy structure of the distance metrics computed in Sec. IV B. Our symmetry-aware GNN architecture and the data set used to train the model are provided in Sec. III. Also, as opposed to a black-box deep learning prediction, we show in Sec. IV C that the latent features (data representations) that the model learned are physically meaningful, rendering our GNNs explainable. We conclude this work in Sec. V and provide supplementary results and necessary derivations in the Appendixes.

II. LINEAR ELASTICITY BACKGROUND

A. The strain energy tensor and the elastic tensor

Assuming linear elastic deformation and no external stress, the strain energy density (SED) of isotropic materials can be expressed as [37]

$$U(\epsilon_1, \epsilon_2, \dots, \epsilon_6) = \frac{1}{2} \sum_{i,j=1}^6 C_{ij} \epsilon_i \epsilon_j, \quad (1)$$

where C_{ij} is an *elastic constant* and ϵ_i is a strain component $i \in \{1, 2, \dots, 6\}$ in Voigt notation. With the goal of obtaining

the elastic constant, it suffices to focus on the SED tensor of rank 2, which we now show. Denote $U_{ij} \equiv U(\epsilon_i, \epsilon_j)$ for $i \neq j$ and $U_{ii} \equiv U(\epsilon_i)$, which are the energy stored in a distorted crystal structure when the lattice is strained by at most two strain components (two Voigt indices). In this paper, we consider U_{ij} in units of eV per atom. Since the elastic constant tensor is symmetric, the SED tensor is also symmetric and can be expressed as

$$U_{ij} = C_{ij} \epsilon_i \epsilon_j + \frac{1}{2} (C_{ii} \epsilon_i^2 + C_{jj} \epsilon_j^2) \quad \text{for } i \neq j, U_{ii} = \frac{1}{2} C_{ii} \epsilon_i^2. \quad (2)$$

Thus the elastic constants can then be analytically obtained as a function of the SED:

$$C_{ij} = \frac{1}{\epsilon_i \epsilon_j} [(1 + \delta_{ij}) U_{ij} - (1 - \delta_{ij})(U_{ii} + U_{jj})], \quad (3)$$

where δ_{ij} is the Kronecker delta.

Because the SED tensor of interest and the elastic tensor are symmetric tensors of rank 2 with 6×6 components, there are $(6 + 1)6/2 = 21$ independent components. In the matrix form, we denote the SED tensor as

$$\mathbf{U} = \begin{bmatrix} U_{11} & U_{12} & U_{13} & U_{14} & U_{15} & U_{16} \\ & U_{22} & U_{23} & U_{24} & U_{25} & U_{26} \\ & & U_{33} & U_{34} & U_{35} & U_{36} \\ & & & U_{44} & U_{45} & U_{46} \\ & & & & U_{55} & U_{56} \\ & & & & & U_{66} \end{bmatrix}, \quad (4)$$

where only the upper triangular elements are shown (the lower diagonal elements are left blank for brevity). For crystalline solids, however, the upper triangular elements are not completely independent, depending on the symmetry of the crystal structure. For instance, the SED tensor of a cubic lattice is

$$\mathbf{U}_{\text{cubic}} = \begin{bmatrix} U_{11} & U_{12} & U_{12} & U_{14} & U_{14} & U_{14} \\ & U_{11} & U_{12} & U_{14} & U_{14} & U_{14} \\ & & U_{11} & U_{14} & U_{14} & U_{14} \\ & & & U_{44} & U_{45} & U_{45} \\ & & & & U_{44} & U_{45} \\ & & & & & U_{44} \end{bmatrix}. \quad (5)$$

Note that some components of the elastic constants C_{ij} can be zero in many crystal structures, e.g., a cubic lattice has 9 zeros out of 21 independent components. However, due to the property of SED from Eq. (2), U_{ij} is never 0. For the purpose of machine learning regression, working with SED helps avoid zero-inflation problems in the training data set. Figure 1 shows the distributions of C_{ij} and U_{ij} , which cover different ranges depending on the indices. Importantly, the distribution of elastic constants can concentrate around zero, but this is not the case for the SED. As an illustrative example of how the SED can avoid a zero-inflation problem, consider the elastic constants of a diamond cubic crystal structure. For diamond, C_{11} , C_{12} , C_{14} , C_{44} , and C_{45} are 1054, 126, 0, 562, and 0 GPa, respectively. Then, Eq. (2) gives that U_{11} , U_{12} , U_{14} , U_{44} , and U_{45} subjected to 2% strain are 7.506, 16.807, 11.509, 4.002, and 8.005 meV/atom, respectively. Note that, in this paper, we consider U_{ij} in the units of eV per atom instead of per volume. The magnitude of U_{12} is the largest as it is the sum of C_{12} and C_{11} , while U_{14} and U_{45} are smaller (but nonzero) because C_{14} and C_{45} are zero.

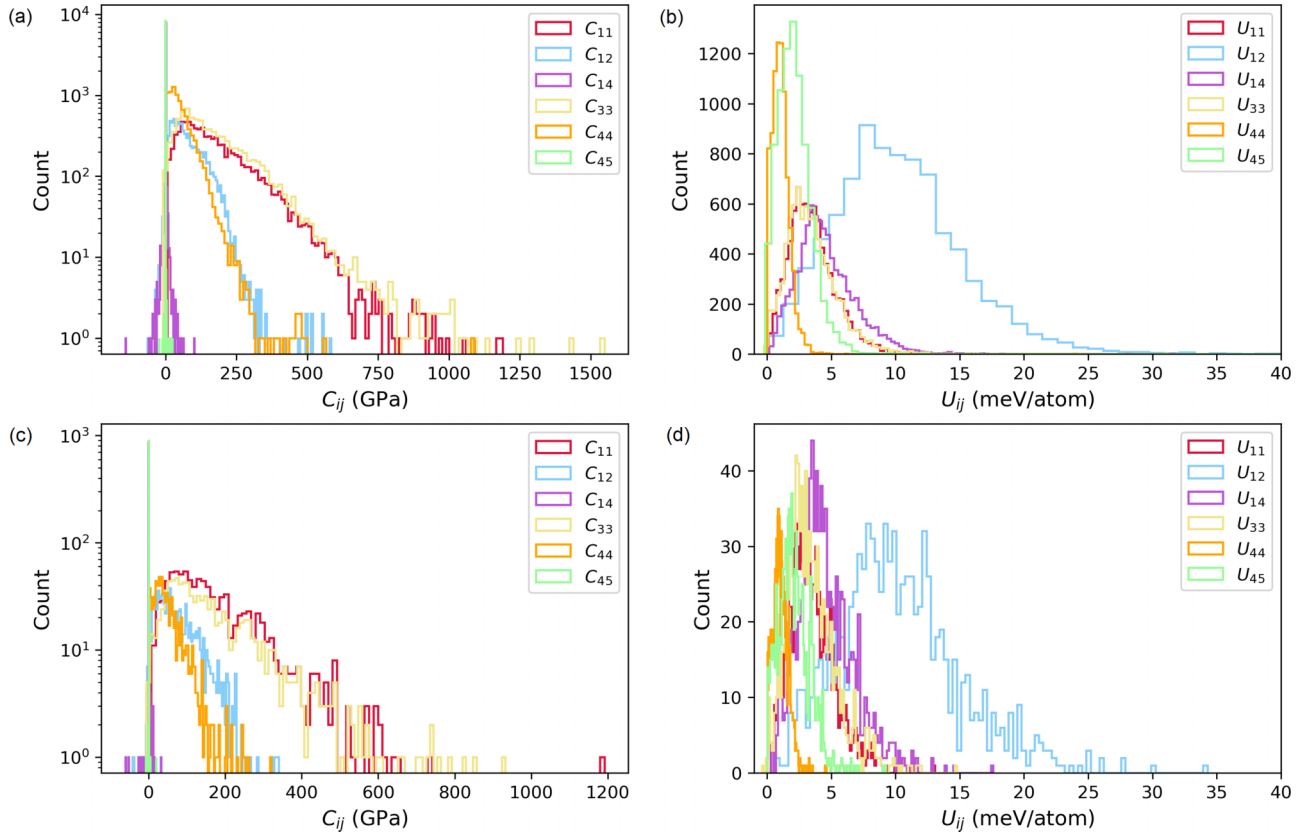


FIG. 1. Histograms of (a) C_{ij} and (b) U_{ij} in the training set and of (c) C_{ij} and (d) U_{ij} in the test set. For visualization purposes, only some ij components, specifically 11, 12, 14, 33, 44, and 45, are presented. For each ij component, U_{ij} (C_{ij}) are distributed over different ranges of energy (modulus). For $i \neq j$, U_{ij} are distributed over a broader range than U_{ii} are. For $ij = 45$, the elastic constant is concentrated around 0 GPa. Importantly, this concentration-around-zero issue can be ameliorated when the elastic constants are converted to the strain energy density. This conversion to SED reduces the number of zero values in the data set, which significantly improves model training and model predictive performances.

B. Strain operations

The unit cell of a crystal can be compactly parametrized by a lattice matrix $\mathbf{L} = [\mathbf{a} \ \mathbf{b} \ \mathbf{c}]$, where \mathbf{a} , \mathbf{b} , and \mathbf{c} are vectors of lattice parameters in Cartesian coordinates. If the system is applied by strain, the lattice matrix will be deformed by

$$\mathbf{L}'(\epsilon_1, \epsilon_2, \dots, \epsilon_6) = \boldsymbol{\epsilon}_I \mathbf{L}, \quad (6)$$

where $\boldsymbol{\epsilon}_I$ is the strain matrix

$$\boldsymbol{\epsilon}_I = \begin{bmatrix} \epsilon_1 & \frac{\epsilon_6}{2} & \frac{\epsilon_5}{2} \\ \frac{\epsilon_6}{2} & \epsilon_2 & \frac{\epsilon_4}{2} \\ \frac{\epsilon_5}{2} & \frac{\epsilon_4}{2} & \epsilon_3 \end{bmatrix} + \mathbf{I}. \quad (7)$$

In this paper, we assume the crystal structures will be deformed by at most two strain components, so that $\mathbf{L}' = \mathbf{L}'(\epsilon_i, \epsilon_j)$.

Due to an applied strain, the atomic coordinates \mathbf{r} must also be transformed accordingly as

$$\mathbf{r}' = \boldsymbol{\epsilon}_I \mathbf{L} \mathbf{r}_f, \quad (8)$$

where \mathbf{r}_f is a fractional coordinate of an atom in an unstrained lattice. We note that \mathbf{r}_f typically shifts from its equilibrium value when strain is applied (say, in density functional theory or in experiments); however, the atomic relaxation is neglected for simplicity [38].

III. MACHINE LEARNING WITH SE(3)-EQUIVARIANT GRAPH NEURAL NETWORKS

A. Crystal graphs

A crystal structure can be represented as a multigraph whose nodes and edges encode atoms and their pairwise connections, respectively. A pair of nodes describing an atom of type m and an atom of type n can be connected by multiple edges, encapsulating the interactions between the atom of type m in a unit cell and atoms of type n in the unit cell as well as in other periodic cells of consideration; see Fig. 2. An atom of type m can interact with an atom of the same type in periodic cells if the interaction range of consideration is large enough, represented by self-loops in the node m of the crystal graph. Each atom's location and its atomic features (e.g., atomic mass, electronegativity, polarizability, atomic radius, etc.) in a unit cell are accounted for by the attributes of a single node.

More formally, a crystal graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a set of nodes (vertices) \mathcal{V} and a set of edges \mathcal{E} defined as

$$\begin{aligned} \mathcal{V} &= \{(\mathbf{f}_n, \mathbf{r}_n) \mid \mathbf{f}_n \in \mathbb{R}^M, \mathbf{r}_n = \mathbf{L} \mathbf{r}_{f_n} \in \mathbb{R}^3\}, \\ \mathcal{E} &= \{\Delta \mathbf{r}_{mn}^{(\mathbf{T})} \mid \Delta \mathbf{r}_{mn}^{(\mathbf{T})} = \mathbf{r}_m - \mathbf{r}_n + \mathbf{T}; \mathbf{r}_m, \mathbf{r}_n \in \mathbb{R}^3\}, \end{aligned}$$

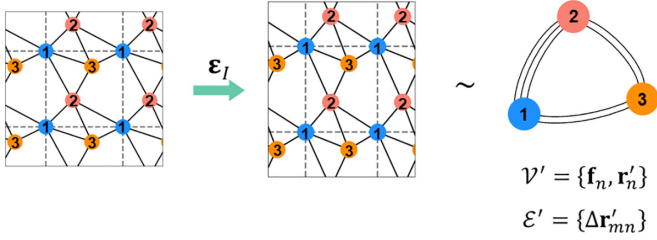


FIG. 2. A two-dimensional lattice and its *strained* crystal graph embedding. Gray dashed lines are periodic boundaries. The whole structure can be spanned by the three atoms in three different bases in the unit cell, represented by three different nodes. Black lines are (undirected) edges connecting two neighbor nodes. The strained lattice is represented as a multigraph on the right, which is the crystal graph input into our SE(3)-equivariant GNN building block. Note that to describe the crystal graph for the unit cell and the atoms on its boundary, only three nodes are required. The multiedges of each pair of nodes are distinguished by different crystallographic directions.

where m and n are indices of the two atoms in the unit cell, \mathbf{f}_n is a vectorized atomic feature of an atom of type n , M is the number of atomic features, \mathbf{r}_{f_n} is the fractional coordinate of an atom of type n , and \mathbf{T} is a translation vector within the space of interest. Each node embeds two kinds of information: atomic attributes \mathbf{f} and atomic positions \mathbf{r} in the unit cell. From this definition, the unit cell and its periodic images are compactly represented as a multigraph; only a finite number of nodes describing the atoms in the unit cell is required, and the number of edges joining the two nodes grows linearly with the number of periodic images of the atoms in the structure of interest (see Fig. 2).

A crystal graph describing the whole crystal structure will require information about infinitely many edges. In practice, one typically embeds only partial information of the structure using the description of the unit cell and a finite number of edges encoding interactions with neighbor atoms [39]. One method to construct the edges is to connect a target atom to other atoms within a finite radius; however, still, this method generates an excessive number of edges, which is computationally infeasible for GNN machine learning. Alternatively, some algorithms determine the connection between atoms using, e.g., the Voronoi tessellation, generating a moderate number of edges [40,41]. However, for layered materials, using Voronoi tessellation causes the edges connecting atoms between layers to be absent. These interlayer connections are crucial in differentiating the structure strained in the out-of-plane direction from structures strained in other directions. In this paper, we define an edge between a pair of atoms $\Delta\mathbf{r}$ (dropping the translation vector superscript and the atom indices subscript for brevity) to be nonzero only if each component Δr_β , where $\beta \in \{x, y, z\}$, satisfies the following spatial cutoff criterion: $|\Delta r_\beta| \leq \min(|a_\beta| + |b_\beta| + |c_\beta|, s) + \delta$, where a_β , b_β , and c_β are the β components of the corresponding lattice parameters constituting the lattice matrix \mathbf{L} , s is a cutoff distance, and δ is a small cutoff extension of our choice. This can reduce the number of edges from the hard radial cutoff criterion and ensures the connections of atoms between layers for layered materials by an appropriate choice of s and δ .

B. Rotational and translational equivariance

If two crystal graph inputs constructed from an identical crystal graph strained by two different sets of strained components (ϵ_i, ϵ_j) and (ϵ_k, ϵ_l) yielding the new vertices $\mathcal{V}' = \{\mathbf{f}_n, \mathbf{r}'_n\}$ and $\mathcal{V}'' = \{\mathbf{f}_n, \mathbf{r}''_n\}$, respectively, are equivalent up to a three-dimensional (3D) rotation and a translation, then we expect our machine learning model to predict the same SED. This symmetry-aware property can be implemented with geometric deep learning models [42]. We will equip a latent representation of our model with this symmetry-aware (equivariant) property, which is defined as follows.

A latent representation $\phi : \mathcal{V} \rightarrow \mathcal{Y}$ is *equivariant* if for each transformation $T_g : \mathcal{V} \rightarrow \mathcal{V}$, with g being an element of an abstract group G , there exists a transformation $S_g : \mathcal{Y} \rightarrow \mathcal{Y}$ such that the following condition holds:

$$S_g[\phi(v)] = \phi(T_g[v]),$$

for all $g \in G$, $v \in \mathcal{V}$. In our case, the group G of interest is SE(3), providing the latent representation with a 3D rotation- and translation-equivariant property. This latent feature ϕ can be achieved with SE(3)-equivariant GNNs, known as tensor field networks (TFNs) [43], and with a TFN with an appropriate attention mechanism, known as SE(3)-Transformers [44]; see more detailed recipes of these two models in Appendix A.

The key idea that enables the SE(3)-equivariant property of these two GNNs is that its message-passing kernel is constructed from translational-invariant spherical harmonic bases, with the structure-preserving transformation S_g given by the Wigner D matrices (see Appendix A). Under a 3D rotation, each spherical harmonic J basis function of the kernel transforms according to

$$Y_J \left(\mathbf{R}_g^{-1} \frac{\Delta\mathbf{r}}{\|\Delta\mathbf{r}\|} \right) = \mathbf{D}_J^*(g) Y_J \left(\frac{\Delta\mathbf{r}}{\|\Delta\mathbf{r}\|} \right),$$

where $\mathbf{D}_J(g)$ is the J th Wigner D matrix and \mathbf{R}_g is a rotation matrix associated with $g \in \text{SO}(3)$, making the learned latent representation ϕ equivariant [43]. The multihead attention mechanism in SE(3)-Transformers also induces an equivariant property on the product rule between each key and value pair, rendering the whole message-passing algorithm equivariant under SE(3) [44]. In this paper, we use SE(3)-Transformers to first build a compressed representation of the orientations of a strained crystal structure; then such learned latent representation will be passed to other networks with a high expressive power to perform the prediction (regression) task of the SED. The following section contains the entire framework for the SED prediction.

C. Model architecture and training procedure

We now describe our GNN-based model that predicts the SED tensor. The architecture of our model, which we term *StrainTensorNet*, is illustrated in Fig. 3. The model is designed to receive two types of input: a crystal graph representing a *strained* crystal structure and a one-hot vector of dimension $6(6+1)/2 = 21$ [45] indicating the *strained components*, with a value 1 in the ij component that the strain operation is applied on and with a value 0 otherwise.

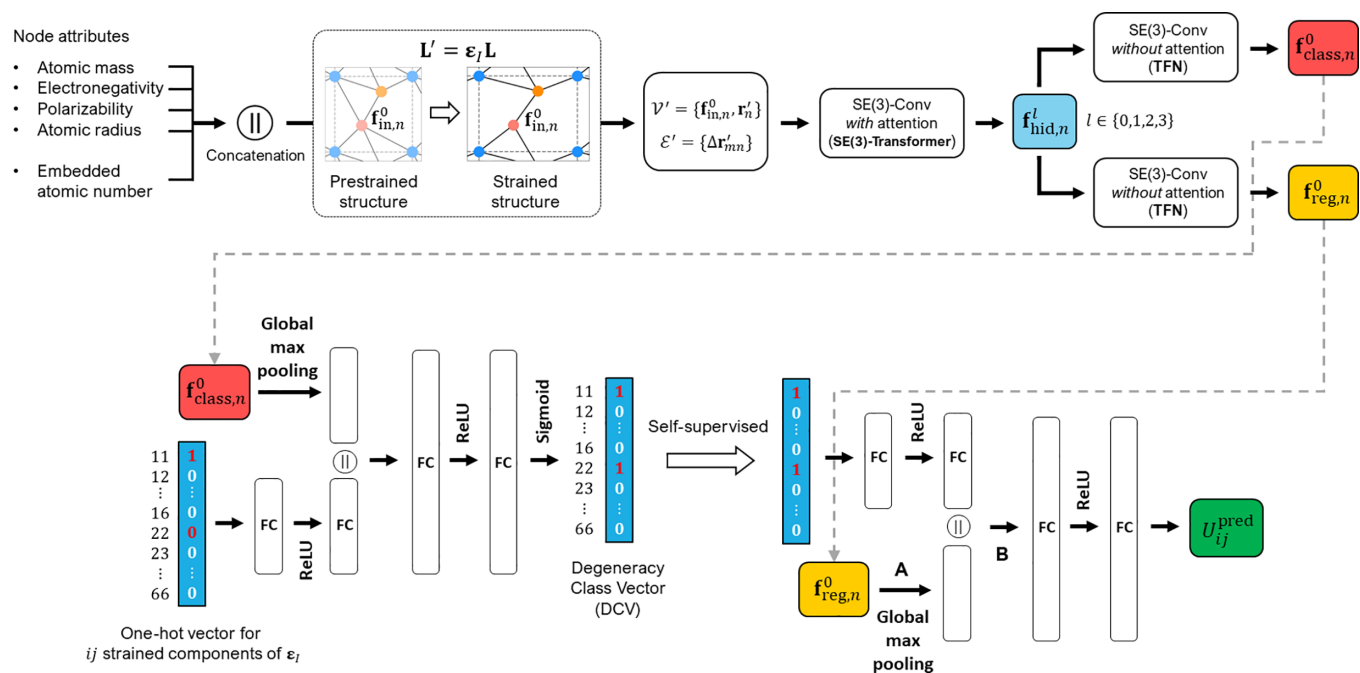


FIG. 3. With a strained crystal graph and the strained component one-hot vector as the input, StrainTensorNet employs a self-supervised approach combining both classification and regression networks to predict both the degeneracy class vector and the U_{ij} . The $\mathbf{f}_{in,n}^0 \equiv \mathbf{f}_n$ denotes the input feature of the n th node, where a superscript, 0, indicates the input into the $l = 0$ channel of the spherical harmonics in the SE(3)-Transformer, which is a channel whose feature is invariant under rotation (transforms like a scalar quantity). $\Delta \mathbf{r}'_{mn}$ is the edge connecting a target node m and a neighbor node n . The degeneracy class vector from the classification network will be concatenated with the latent representation of the strained crystal graph ($\mathbf{f}_{reg,n}^0$) for the final regression network to predict the SED. FC, fully connected layer; ReLU, rectified linear unit.

The crystal graph of a strained crystal structure is an input feature of the SE(3)-equivariant GNN building block, which generates *two* latent features. The first latent feature ($\mathbf{f}_{class,n}^0$ in Fig. 3) is fed into a *classification* neural network that predicts a vector of dimension 21, identifying all the upper triangular elements of the SED tensor whose values are exactly identical (by symmetry) to the component indicated by the input one-hot vector. This inherent classification informs the model to discern the degenerate structure of the SED tensor that depends on the symmetry of an input crystal graph (see Fig. 4). For example, for a cubic lattice strained in the direction 11, the input one-hot vector $(1, 0, \dots, 0)^T$ together with the latent feature $\mathbf{f}_{class,n}^0$ gives a prediction of a vector that has a value close to 1 in the indices $ij = 11, 22, 33$ and a value close to 0 in the other 18 indices. Finally, the predicted *degeneracy class vector* (DCV) together with the second latent feature ($\mathbf{f}_{reg,n}^0$ in Fig. 3) will be fed into the final neural network that predicts (regresses on) the SED.

To generate an SE(3)-equivariant latent representation of an input crystal graph, as alluded to in the previous section, we first used the SE(3)-Transformer that receives the strained crystal graph input. The material structure from the database is first transformed according to Eq. (6) by a strain of a fixed magnitude, which is then converted into the strained crystal graph input. We note that in the *ab initio* calculation the atomic coordinates in the strained lattice are optimized to be in the equilibrium positions under an applied strain but in this paper the fractional coordinates are kept as in the prestrain

condition. The strained crystal graph input is given by

$$\mathcal{V}' = \{(\mathbf{f}_n, \mathbf{r}'_n) \mid \mathbf{f}_n \in \mathbb{R}^M, \mathbf{r}'_n = \epsilon_l \mathbf{L} \mathbf{r}_{f_n} \in \mathbb{R}^3\},$$

$$\mathcal{E}' = \{\Delta \mathbf{r}'_{mn}(\mathbf{T}') \mid \Delta \mathbf{r}'_{mn}(\mathbf{T}') = \mathbf{r}'_m - \mathbf{r}'_n + \mathbf{T}'; \mathbf{r}'_m, \mathbf{r}'_n \in \mathbb{R}^3\},$$

where the prime notation indicates that the atomic positions and the translation vector are of the strained lattice.

The training data of $\{U_{ij}\}$ are computed, via Eq. (2), from $\{C_{ij}\}$ extracted from the Materials Project database [24] (see Appendix E). For molecules and materials, if the number of training data is large enough, only atomic numbers are sufficient for the node (atomic) attributes [46]. Nevertheless, in this paper, features such as atomic mass, electronegativity, polarizability, atomic radius, atomic number, and ij indices are used as node attributes. Polarizability data are obtained from Ref. [47], whereas other attributes are obtained from the Materials Project database. The atomic number is vectorized through an embedding layer of dimension 512, which is then concatenated with the other four node attributes (see Fig. 3). Note that the elastic constants are influenced by the atomic mass since the phonon frequencies in sound waves, particularly the acoustic modes, are inversely proportional to the square root of the atomic mass. Additionally, the resistance to changes in bond length correlates with both electronegativity and polarizability. These three atomic properties are thus incorporated into node attributes, and they significantly improve the model prediction accuracy.

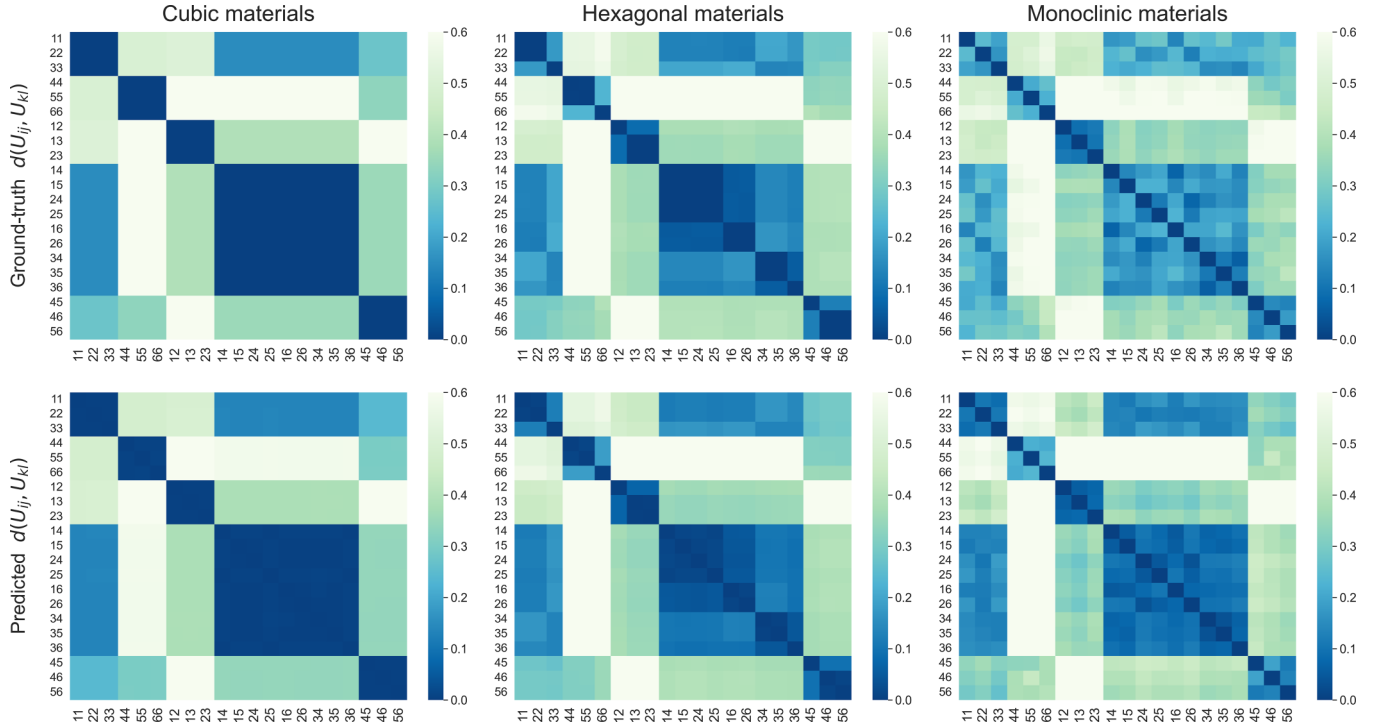


FIG. 4. The distance metrics $d(U_{ij}, U_{kl})$ computed from the test set of the ground-truth SED (top row) and the predicted SED (bottom row). The columns organize the crystal structure: The left, middle, and right columns are for cubic, hexagonal, and monoclinic materials, respectively. The model prediction shows an excellent agreement with the ground truth for crystals with high symmetry (cubic) and a good agreement with the ground truths for crystals with lower symmetry (hexagonal and monoclinic).

Since the atomic features are scalar attributes, the input feature vector will be fed into the $l = 0$ channel of the transformer network since such a channel transforms like a scalar (see Appendix A). We will denote each node's input feature vector as $\mathbf{f}_{in,n}^0 = \mathbf{f}_n$. For computational feasibility, we will restrict the output feature vector of the SE(3)-Transformers to consist of four channels $\mathbf{f}_{hid,n}^l$ with $l \in \{0, 1, 2, 3\}$. These latent features are fed into two different SE(3)-equivariant GNNs *without* attention (TFNs), outputting a nodewise classification feature vector $\mathbf{f}_{class,n}^0$ and the nodewise regression feature vector $\mathbf{f}_{reg,n}^0$; see Fig. 3. Note that the attention mechanism is not deployed in the last GNN layer as it yields a better prediction accuracy.

To train the model to discern different orientations of a strained crystal graph and classify its appropriate degeneracy class vector, in each epoch, we draw a new realization of a crystal graph in a different orientation and of a new one-hot vector in the same class of vector associated with the original strained component. Specifically, a new crystal graph input is sampled from a uniformly random orientation (uniformly random Euler angles) around the center of mass in the unit cell of the original strained crystal graph. A new one-hot vector of the strained components is chosen such that the element 1 is drawn uniformly from one of the nonzero components of the DCV. For example, if the original strained component is 11 for a cubic material, a new one-hot vector in each epoch is drawn from the situations where only one of the strained components 11, 22, or 33 is 1, while the other components are 0.

Our self-supervised approach combines both the regression errors and the classification errors in the global loss function

$$\mathcal{L} = \frac{1}{N} \sum_{n=1}^N \left| U_{ij}^{(n)} - U_{ij}^{(n),\text{pred}} \right| + \lambda \mathcal{L}_{\text{class}}, \quad (9)$$

where the SED prediction $U_{ij}^{(n),\text{pred}}$ is the output from the regression network, $\mathcal{L}_{\text{class}}$ is the binary cross-entropy loss function for multilabel classification, λ specifies the relative significance between the classification and the regression errors, and N is the total number of training samples. We stop the training when the gradient of the loss is relatively small over multiple epochs. Training was performed on NVIDIA A100 GPUs.

D. Choosing a subset of strained crystal structures for training data

By virtue of the rotational equivariance, we expect Strain-TensorNet to efficiently predict the 21 components of \mathbf{U} using training data consisting of only nondegenerate components. We have selected the number of nondegenerate components for training that depends on crystal systems, i.e., cubic, tetragonal, hexagonal, trigonal, monoclinic, and triclinic. For cubic materials, only six components are used for training (the number of nondegenerate components for other crystal systems can be found in Table I). Recall that, as discussed earlier, the cubic lattice has five distinct U_{ij} as shown in Eq. (5);

TABLE I. The minimal set of ij strained components of each crystal system used to generate an initial crystal graph input.

Crystal system	Number of training indices	Training indices
Cubic	6	11, 12, 14, 15, 44, 45
Hexagonal	13	11, 12, 13, 14, 15, 16, 33, 34, 36, 44, 45, 46, 66
Tetragonal	14	11, 12, 13, 14, 15, 16, 26, 33, 34, 36, 44, 45, 46, 66
Trigonal	16	11, 12, 13, 14, 15, 16, 24, 25, 33, 34, 36, 44, 45, 46, 66
Orthorhombic	21	all
Monoclinic	21	all
Triclinic	21	all

however, the strain tensor transforms the cubic structures into six distinct structures (according to the symmetry in Laue class $m\bar{3}m$) as follows:

$$\begin{bmatrix} \mathcal{T}_1 & \mathcal{T}_2 & \mathcal{T}_2 & \mathcal{O}_2 & \mathcal{M}_2 & \mathcal{M}_2 \\ & \mathcal{T}_1 & \mathcal{T}_2 & \mathcal{M}_2 & \mathcal{O}_2 & \mathcal{M}_2 \\ & & \mathcal{T}_1 & \mathcal{M}_2 & \mathcal{O}_2 & \\ & & & \mathcal{O}_1 & \mathcal{M}_1 & \mathcal{M}_1 \\ & & & & \mathcal{O}_1 & \mathcal{M}_1 \\ & & & & & \mathcal{O}_1 \end{bmatrix},$$

where each element in the matrix stands for a crystal system of $\mathbf{L}'(\epsilon_i, \epsilon_j)$; \mathcal{T} , \mathcal{O} , and \mathcal{M} are tetragonal, orthorhombic, and monoclinic lattices, respectively; and a different number labeling the subscript indicates a different structure. Despite the fact that U_{14} (C_{14}) is equal to U_{15} (C_{15}) by the cubic symmetry, $\mathbf{L}'(\epsilon_1, \epsilon_4)$ and $\mathbf{L}'(\epsilon_1, \epsilon_5)$ possess different lattice symmetries which are orthorhombic and monoclinic lattices, respectively. This is because ϵ_1 strains the structure in the x direction, while ϵ_4 and ϵ_5 strain the structure in both the y and z directions, and both the x and y directions, respectively. Since $\mathbf{L}'(\epsilon_1, \epsilon_4)$ and $\mathbf{L}'(\epsilon_1, \epsilon_5)$ are not equivalent up to a rotation, the SE(3) kernel regards the two inputs as different inputs. By exploiting the SE(3) kernel to identify inputs that are equivalent up to a rotation, for cubic materials, it suffices to use the training data consisting of the distinct input structures, strained only in six directions, i.e., 11, 12, 14, 15, 44, and 45.

E. The distance between SED components

To evaluate our model capability to predict the symmetry-dependent degeneracy pattern of the SED tensor, we use the Canberra distance between two components of the SED tensor as a metric:

$$d(U_{ij}, U_{kl}) \equiv \frac{1}{N} \sum_{n=1}^N \frac{|U_{ij}^{(n)} - U_{kl}^{(n)}|}{|U_{ij}^{(n)}| + |U_{kl}^{(n)}|}, \quad (10)$$

where N represents the number of samples and ij and kl are Voigt indices. With this metric, the distance between U_{ij} and U_{kl} is zero if they belong to the same degeneracy class (their values are identical). The top row of Fig. 4 shows the ground-truth distance pattern for cubic, hexagonal, and monoclinic materials. Dark blue and bright green colors indicate $d(U_{ij}, U_{kl})$ values that are closest to zero and 0.6, respectively. It is important to note that the degeneracy pattern of U_{ij} depends on the crystal symmetry, which results in a unique Canberra distance pattern for each crystal system. For instance, in cubic materials, U_{11} , U_{22} , and U_{33} are identical (in

the same degeneracy class); so the distance pattern of cubic materials displays a dark blue box corresponding to indices 11, 22, and 33. The distance patterns in Figs. 4 and 5 are averaged over the samples with the same crystal system in the data set. Hence the distance pattern of low-symmetry crystals, such as monoclinic and triclinic, will strongly depend on the data set.

IV. DISCUSSION AND RESULTS

A. Rationale for the StrainTensorNet model architecture

The underlying principle for using an equivariant network as a core component is to construct a compact representation of the SED tensor, such that the crystal structures strained by different sets of ϵ_i and ϵ_j that are equivalent up to rotation (and translation) possess the same SED. Such a symmetry-dependent SED tensor of specific crystal structures is demonstrated in the ground-truth distance patterns $d(U_{ij}, U_{kl})$ in the top row of Fig. 4, where StrainTensorNet can well approximate such symmetry-dependent patterns; see the bottom row of Fig. 4.

Figure 6 shows our earlier model development with the goal of compactly representing the SED tensor, from the simplest but less expressive model to StrainTensorNet, which can compactly represent the SED tensor rather accurately. All these trial GNN-based models were trained on crystal structures with fewer than 500 edges. The simplest model [Fig. 6(a)] takes as an input only the crystal graph with node attributes, i.e., atomic mass, atomic radius, electronegativity, polarizability, and embedded atomic number. This minimal equivariant model does not yield a sufficiently accurate representation of U_{ij} ; we found that the model is more biased towards predicting the SED tensor whose distance matrix pattern is more akin to that of the cubic crystals. This could be because the model is not sufficiently expressive and thus attempts to fit the majority of the training data set comprising more higher-symmetry structures (see Table IV in Appendix E for the crystal system statistics of our curated data set).

To improve the minimal model's expressiveness, we introduce the degeneracy class vector (DCV) as an additional input, to inform the model about the symmetry-dependent class of the SED tensor. The DCV is first embedded by fully connected layers and then concatenated with global max-pooled features or node attributes, as shown in Figs. 6(b) and 6(c), respectively. These symmetry-informed models yield lower mean absolute error (MAE) and root-mean-square error (RMSE) of the SED tensor compared with the minimal model,

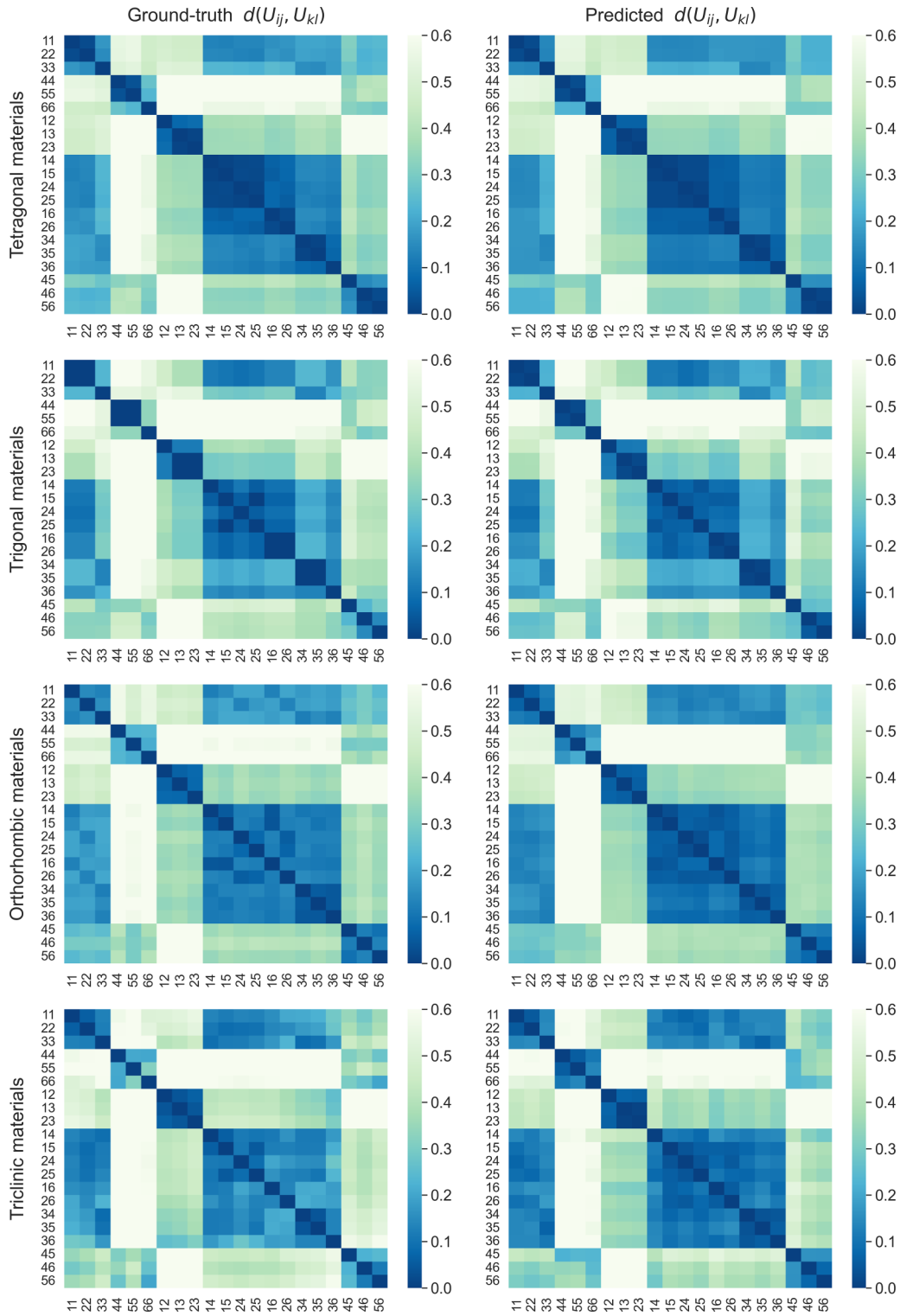


FIG. 5. The distance metric (computed from the test set) of tetragonal, trigonal, orthorhombic, and triclinic materials (top to bottom, respectively) for the ground truth (left column) and StrainTensorNet’s prediction (right column).

while the model in Fig. 6(b) performs slightly better than the model in Fig. 6(c) (see Table VI in Appendix G). Moreover, the predicted distance patterns are less biased towards those of the cubic lattice. Specifically, for crystal structures with

lower symmetry such as hexagonal, tetragonal, trigonal, and orthorhombic, the predicted distance patterns are relatively similar to their ground-truth distance patterns. As expected, informing the model about the DCV of the input helps force

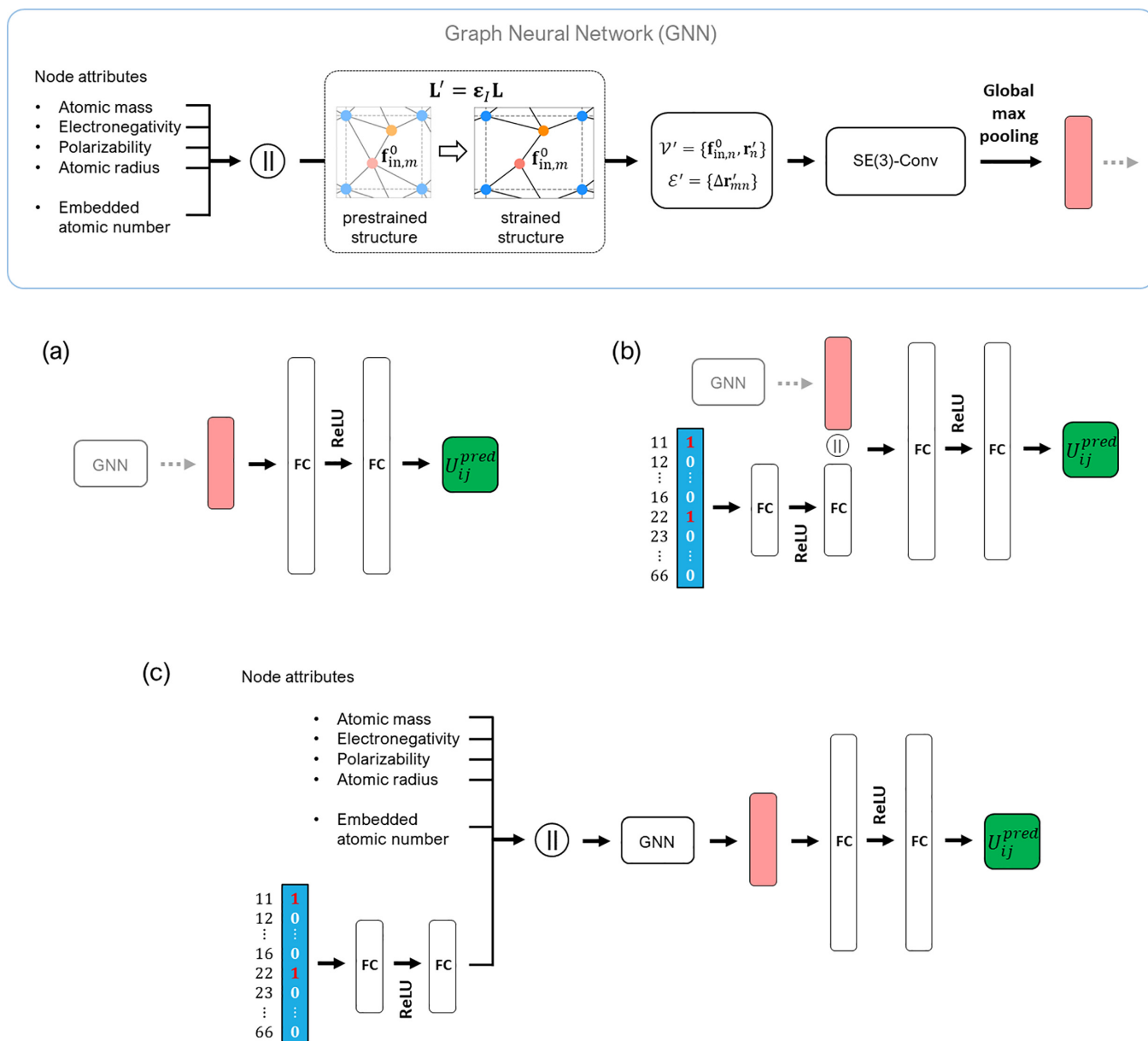


FIG. 6. Earlier trial models with SE(3)-equivariant GNNs as a fundamental building block. Top: The building block consists of a strained crystal graph input that is fed into SE(3)-equivariant GNNs [SE(3)-Transformer followed by a TFN, similar to that of the building block in StrainTensorNet], whose output is the global max-pooled feature from the TFN. (a) A minimal equivariant GNN model for SED regression. (b) An SED regression network that takes both the known degeneracy class vector and the minimal equivariant GNN’s representation as an input. (c) The same network as (a), but the input node attributes also incorporate the degeneracy class vector.

the distance between U_{ij} and U_{kl} belonging to the same degeneracy class to be closer to zero.

While these models with a DCV included as an input give good predictions of the SED tensor, they require knowledge of the prestrained crystal symmetry *a priori* (to properly assign the value of the DCV). To overcome this limitation, we use the self-supervised method schematically shown in Fig. 3 to modify the model to predict, rather than to require, the DCV. The classification neural network that predicts the DCV takes as an input a one-hot vector of the strained ij components, representing the ϵ_i and ϵ_j components in (7). This self-supervised technique enables StrainTensorNet to predict U_{ij} without any

prior knowledge of the prestrained crystal symmetry. Importantly, StrainTensorNet can express the degeneracy class of the SED tensor relatively well; see Figs. 4 and 5.

B. StrainTensorNet’s prediction results

Table II summarizes StrainTensorNet’s prediction errors on the test set. StrainTensorNet provides data-driven prediction of U_{ij} with a reasonable accuracy. Using these predicted SED tensors together with Eq. (3), elastic tensors of materials can also be calculated. Our model’s prediction accuracy significantly depends on the elemental composition of the compounds. Figure 7 depicts the MAE of U_{ij} and C_{ij} , aver-

TABLE II. Statistics of the data set and the model prediction. Bold terms emphasize the enhanced accuracy of StrainTensorNet over previous studies.

Properties	This work			Mazhnik <i>et al.</i> [25]			Zhao <i>et al.</i> [26]
	Average	MAE	RMSE	Average	MAE	RMSE	RMSE
U_{ij} (meV/atom)	2.652	0.655	1.288				
C_{ij} (GPa)	42.92	10.37	17.69				
B_V (GPa)	107.07	11.48	19.73	111.83	11.11	19.54	16.530
G_V (GPa)	50.98	9.61	17.10	54.81	8.24	11.43	15.780
Y	129.62	22.29	38.63	138.95	19.15	26.23	
ν	0.401	0.037	0.133	0.286	0.041	0.105	

aged over 21 ij components and categorized by elements. It is interesting to note that compounds containing a period 2 element (Be, B, C, N, O, or F), any of the period 6 transition metals Ta, W, Re, or Os, or any of the actinides Th, Pa, or Pu, which have high bulk and shear moduli on average, exhibit a higher MAE for U_{ij} compared with the overall MAE of the entire test set. As a result, their MAEs for C_{ij} are also significantly higher than the overall MAE of the test set. In contrast, compounds containing lanthanides, which exhibit medium bulk and shear moduli on average, except for Gd, demonstrate a considerably lower MAE than the overall MAE of the test set, for both U_{ij} and C_{ij} .

To further investigate the prediction results of the elastic constants, we plot comparisons between the prediction [obtained from the predicted SED tensors together with Eq. (3)] and the ground truth of the elastic constants in Fig. 8. Figure 8(a) shows the prediction results of every C_{ij} component of all crystal structures, whereas Figs. 8(b)–8(f) show the prediction results of C_{22} , C_{23} , C_{26} , C_{55} , and C_{56} of cubic materials. For cubic materials, the model is trained only on a nondegenerate subset of 21 U_{ij} components, specifically components 11, 12, 14, 15, 44, and 45. Notably, the model is able to predict other unseen components reasonably well. However, the prediction of the elastic constant components that are exactly zero is challenging. For example, the ground-truth values of C_{26} and C_{56} of cubic materials are exactly 0 GPa, while the means and standard deviations of these components from our model predictions are 0.69 and 4.45 GPa, and -0.43 and 4.85 GPa, respectively (see Fig. 9 for the histograms of the prediction results). Although the standard deviations are relatively large, the means are less than 1 GPa. It is worth noting that the first-principles calculations also make some errors in these values, but these are usually not computed since they are known to be exactly 0 GPa.

To see how well StrainTensorNet discerns the geometric relationship between different strained crystal graph inputs, we plot the Canberra distance metric of the predicted and ground-truth SED tensors for comparison in Fig. 4. The predicted distance metric of cubic materials, as shown in the left column of Fig. 4, closely resembles the ground-truth distance metric, although there are minor discrepancies in the values that belong to the same degeneracy class. For instance, the distances between the predicted U_{11} , U_{22} , and U_{33} are 4.8–6.2 $\mu\text{eV}/\text{atom}$, resulting in percentage errors of 1.83–2.43% [computed from $\frac{1}{N} \sum_{n=1}^N |U_{ij}^{(n)} - U_{kl}^{(n)}| / U_{av}^{(n)} \times 100\%$, where $U_{av}^{(n)} = (|U_{ij}^{(n)}| + |U_{kl}^{(n)}|) / 2$]. For hexagonal

materials, U_{11} are identical to U_{22} , but not to U_{33} . The predicted distance metric of hexagonal materials accordingly forms a dark blue box in the 11 and 22 components. The percentage error between the predicted U_{11} and U_{22} is 2.22%. For monoclinic materials, the values of each U_{ij} are not necessarily identical to one another, and the distance metric is averaged over the monoclinic materials data. The resulting distance pattern may differ significantly depending on the data set. Notably, these agreements between the predicted and ground-truth distance metrics in various crystal structures reveal that StrainTensorNet gives an excellent prediction of the degenerate structure of SED tensors of strained crystalline materials.

Lastly, the model can also predict other elastic properties, including Voigt's bulk modulus B_V , Voigt's shear modulus G_V , Young's modulus Y , and Poisson's ratio ν , using the predicted C_{ij} together with Eqs. (B1)–(B4). Table II summarizes the MAE and RMSE of our predicted elastic properties compared with previous works. The MAE and RMSE of B_V are 12.59 and 20.83 GPa, respectively, which are comparable to those of Mazhnik *et al.* [25]. The MAE and RMSE of G_V are 10.49 and 18.95 GPa, respectively, with the RMSE of G_V being comparable to that reported by Zhao *et al.* [26]. Our work also yields smaller errors for ν than were reported by Mazhnik *et al.* [25]. Note that our data set (see Appendix E) differs from the data set in the relevant work of Refs. [25,26]. Figure 10 presents the MAE of B_V , G_V , Y , and ν , categorized by elements. The model produces high MAE for compounds containing period 2 elements, mainly due to their larger MAEs in U_{ij} and C_{ij} (see Fig. 7). On the other hand, the model can produce smaller MAE for lanthanide compounds, as their MAEs in U_{ij} and C_{ij} (see Fig. 7) are smaller.

C. Interpretability of StrainTensorNet's latent features

To investigate how latent features facilitate successful prediction of the SED tensor, we employed the diffusion maps to faithfully visualize the three-dimensional data representation of the high-dimensional latent features of the entire test set [48,49]. Figures 11 and 12 reveal the dimensionality reduction of the latent features in the diffusion coordinates (φ_1 , φ_2 , φ_3). Two different latent features were considered: the global max-pooling layer from the GNN representation of the input crystal graph (denoted as **A** in Fig. 3) and the concatenation between **A** and the embedded DCV (denoted as **B** in Fig. 3). These low-dimensional representations in Figs. 11 and 12 are colored by

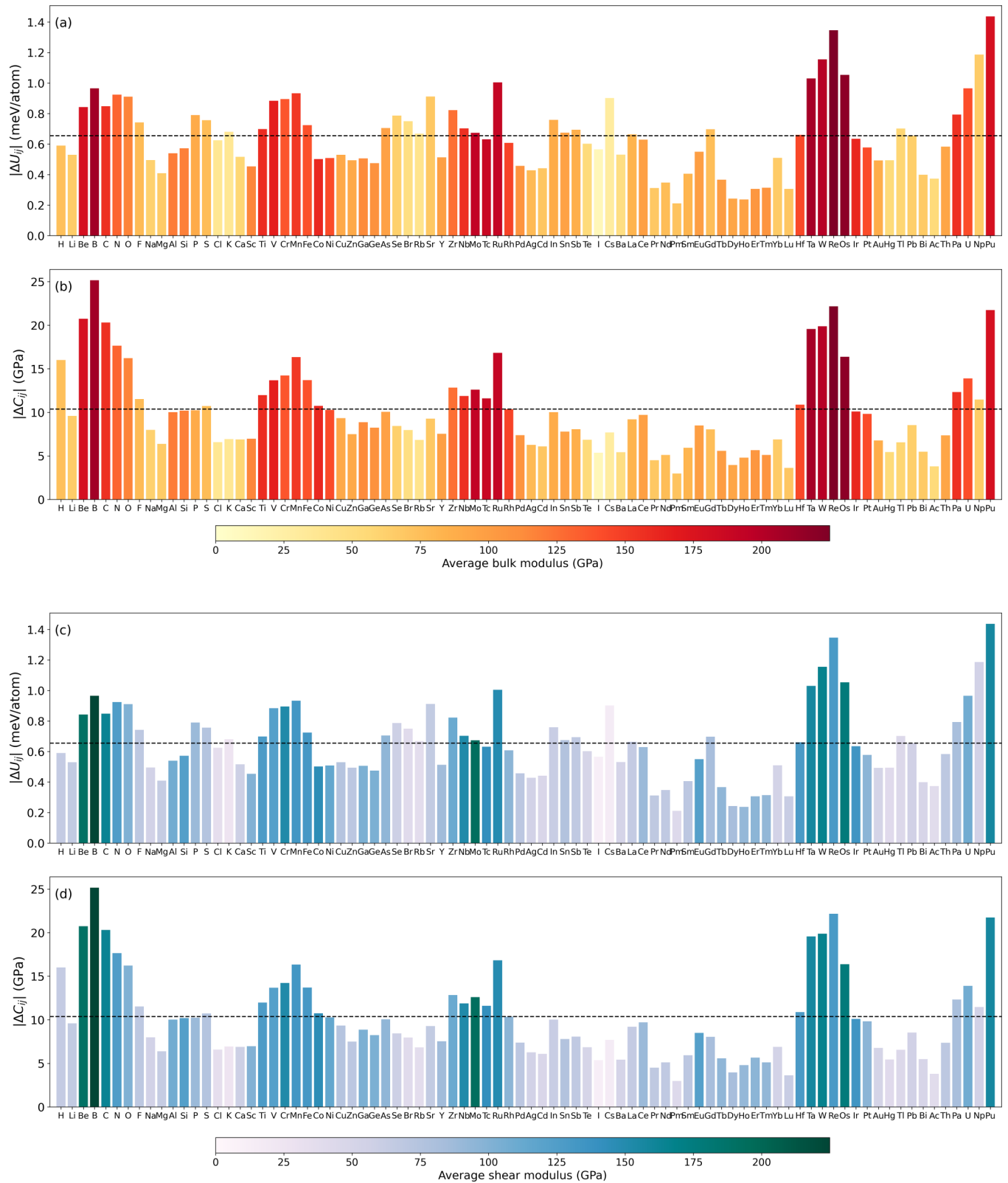


FIG. 7. Bar charts displaying the MAE of the predicted U_{ij} and C_{ij} components, averaged over 21 ij components. The charts are categorized based on the elements present in the compounds in the test set and are colored according to the average bulk modulus for (a) and (b) and according to the average shear modulus for (c) and (d), as indicated by the color bar below each set of bar charts. The dashed lines represent the MAE of each property, averaged over the data in the test set. It can be seen that most elements whose compounds have high bulk and shear moduli exhibit errors of U_{ij} and C_{ij} that exceed the data set’s MAE.

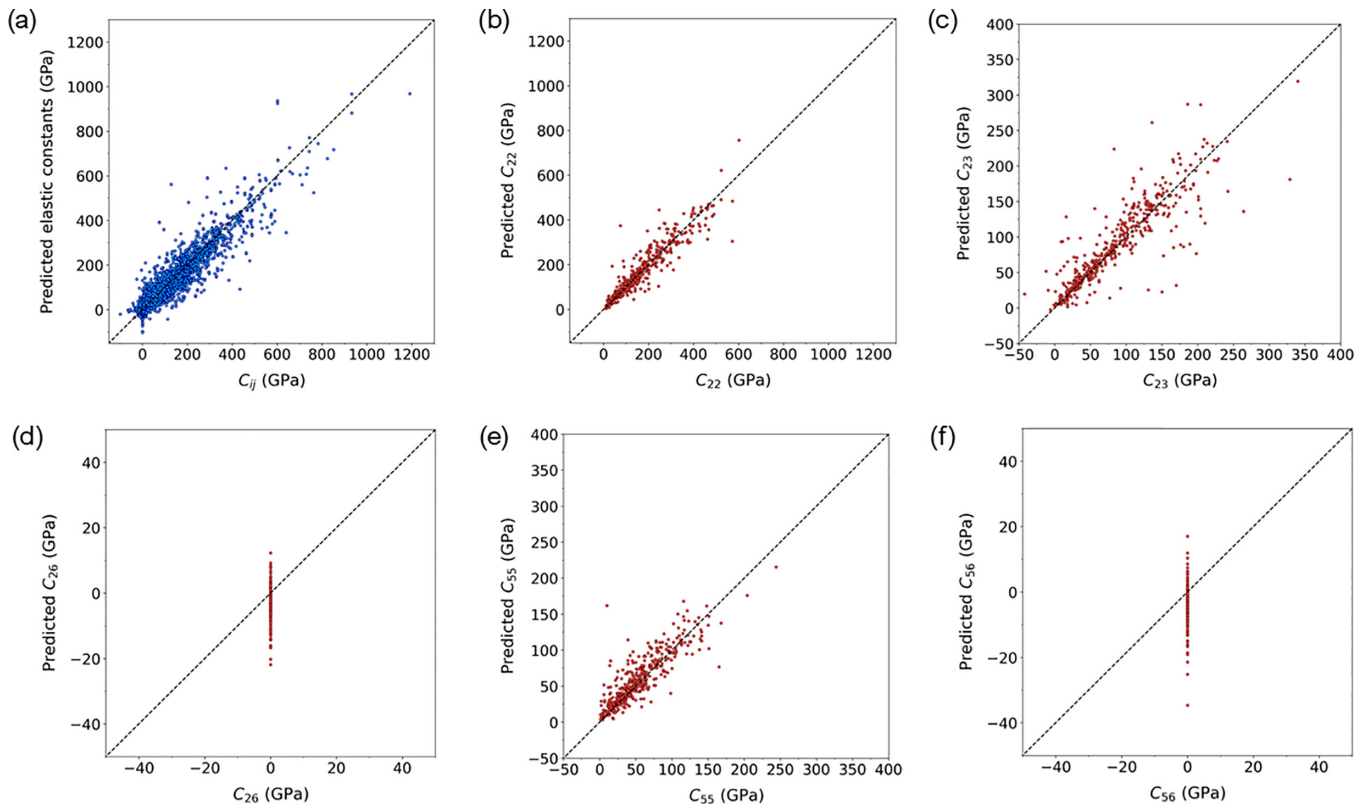


FIG. 8. The predicted values (vertical axes) vs the ground-truth values (horizontal axes) of the elastic constants. (a) All the components of the elastic constants of the whole test set of all crystal systems. (b)–(f) The results for only *cubic materials*, C_{22} (b), C_{23} (c), C_{26} (d), C_{55} (e), and C_{56} (f), where we note that the 22, 23, 26, 55, and 56 indices *did not* appear in the input of the training set for cubic materials.

the energy scale [Figs. 11(a) and 12(a)] and by the strained component ij [Figs. 11(b), 11(c), 12(b), and 12(c)].

What does the max-pooled feature of the crystal graph input \mathbf{A} represent? Figure 12(c) shows that the crystal graph latent representations of the *same material* that are strained in different ij components are *almost identical*, in fact sharing greater resemblance than the representations of different materials that belong to the same symmetry class. This is consistent with the result of Bronstein *et al.* [42], which shows that the latent representation in SE(3)-equivariant GNNs of the original input is very similar to those of the mildly spatially distorted inputs and, interestingly, is less similar to those of inputs spatially translated further away. Hence the latent feature \mathbf{A} seems to uniquely encode the information of the input material, rather than the symmetry of the input. Thus \mathbf{A} alone does not suffice to meaningfully represent the symmetry-dependent SED tensor [Figs. 12(a) and 12(b)].

On the other hand, with the embedded DCV in the latent feature \mathbf{B} , the network can differentiate different strained components within the same material class. When strained in 21 different components, the latent features \mathbf{A} of three example materials (trigonal LuTlTe₂ phase, tetragonal YMnSi phase, and cubic Ta₃Ru phase) that were not differentiable within the same material class [Fig. 12(c)] are now clearly segregated in the φ_2 variable through the help of the embedded DCV; see Fig. 11(c). In fact, the latent feature \mathbf{B} , which combines

both the input crystal graph representation and the embedded DCV, offers an interpretable latent representation of the SED (Fig. 11), which we now discuss.

The latent feature \mathbf{B} is organized such that the (φ_1, φ_2) coordinate encodes the strain energy density that varies from a smaller value in the $(+, +)$ quadrant to a larger value in the $(-, -)$ quadrant [see Fig. 11(a)]. Since U_{ij} stored by the tensile strain ($i, j \in \{1, 2, 3\}$) is typically higher than U_{ij} stored by the shear strain ($i, j \in \{4, 5, 6\}$), Figs. 11(a) and 11(b) consistently reveal that the latent features of the materials strained by the 11, 22, or 33 component have a negative φ_2 value, whereas those of the materials strained by the 44, 55, or 66 components have a positive φ_2 value. In addition, since U_{ij} for $i \neq j$ is computed from the sum of C_{ii} , C_{jj} , and C_{ij} , it is typically larger than U_{ii} and U_{jj} . Figures 11(a) and 11(b) also consistently reveal that the latent features of the materials strained by the 44, 45, 11, and 12 components are respectively organized in the φ_2 coordinate from a more positive to a more negative value. Additionally, Figs. 11(a) and 11(c) show that materials with higher (lower) average SED will be represented by a larger negative (larger positive) φ_1 variable. In summary, the max-pooled feature from the graph neural networks \mathbf{A} encodes the material information, and, together with the information of the DCV, the concatenated latent feature \mathbf{B} effectively encapsulates both the material information and its strained component-dependent SED.

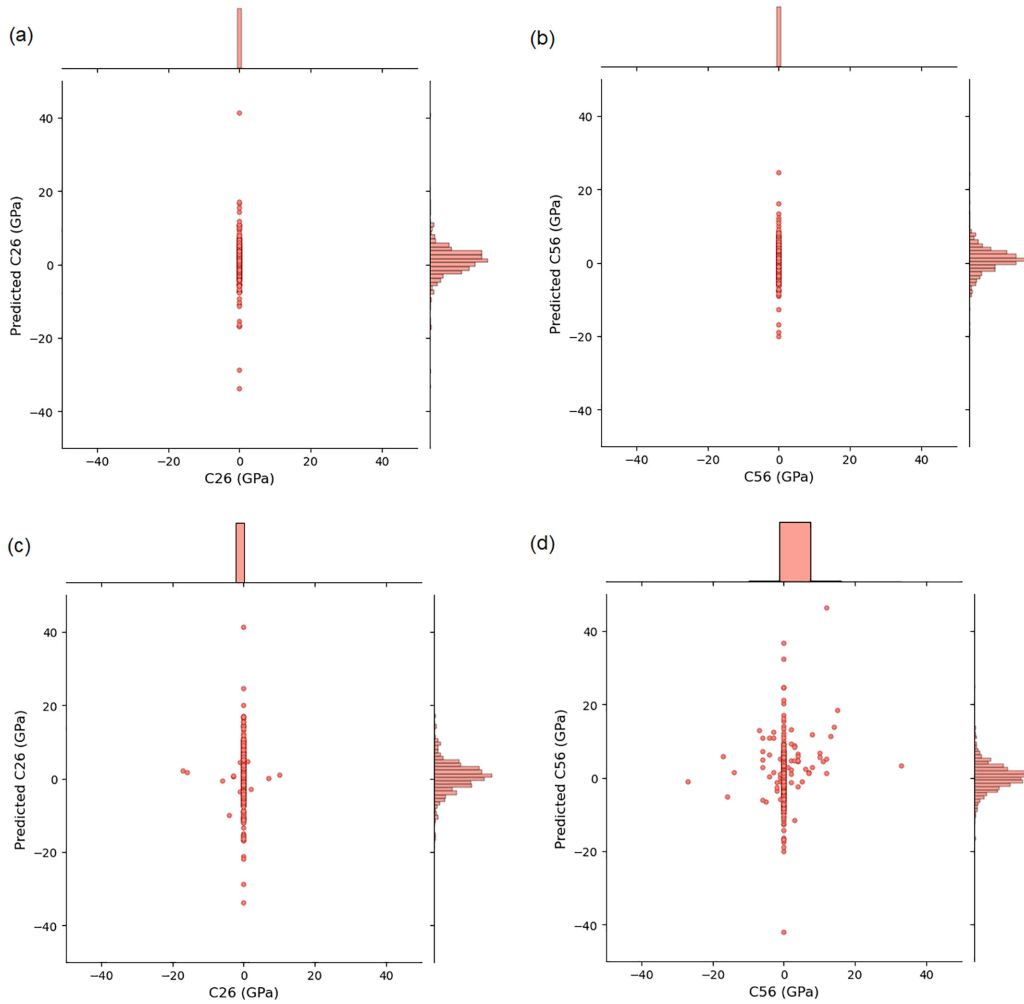


FIG. 9. Statistics of the prediction results (vertical axes) vs the ground truths (horizontal axes) of the elastic constant components that are concentrated near 0 GPa. (a) The predicted C_{26} of a cubic system has a mean and a standard deviation of 0.69 and 4.45 GPa, respectively. (b) The predicted C_{56} of a cubic system has a mean and a standard deviation of -0.43 and 4.85 GPa, respectively. (c) The predicted C_{26} of all the crystal systems has a mean and a standard deviation of 0.99 and 6.08 GPa, respectively. (d) The predicted C_{56} of all the crystal systems has a mean and a standard deviation of -0.20 and 9.22 GPa, respectively. The ground-truth distributions of C_{26} and C_{56} are centered around 0 GPa with vanishing widths. Although some of the predicted C_{26} and C_{56} are not exactly 0 GPa (for cubic system), most of the predicted values are near 0 GPa.

V. CONCLUSIONS

We have demonstrated a data-driven framework for predicting the elastic properties of crystal structures using SE(3)-equivariant graph neural networks (GNNs). By leveraging SE(3)-equivariant GNNs as building blocks, our self-supervised deep learning model accurately predicts values of the Voigt's bulk modulus, Voigt's shear modulus, Young's modulus, and Poisson's ratio that are comparable to those of other nonequivariant deep learning studies [25,26].

A key contribution is the prediction of the strain energy density (SED) and its associated elastic constants, which are tensorial quantities that depend on a material's crystallographic group. The similarity between the distance metrics of the SED components between the ground truths and the predictions demonstrates our model's capability to identify different symmetry groups of strained crystal structures. Requiring only a strained crystal graph and the strained

component as the input, our approach offers an efficient alternative to the standard *ab initio* method for designing new materials with tailored elastic properties.

The interpretability of the model is also a notable feature. The learned latent representations taking into account the degeneracy class vector are organized in a physically meaningful structure for predicting the SED tensors. This interpretability aspect enhances the transparency of model prediction, enabling the justification of whether the prediction is physically relevant.

The combination of interpretability and the consideration of crystallographic groups sets our model apart from recent data-driven methods for predicting elastic properties of materials. We hope this work is a stepping stone toward an efficient data-driven approach for materials discovery and design and opens up avenues for approaching more challenging tensor prediction tasks, e.g., predicting dielectric and piezoelectric

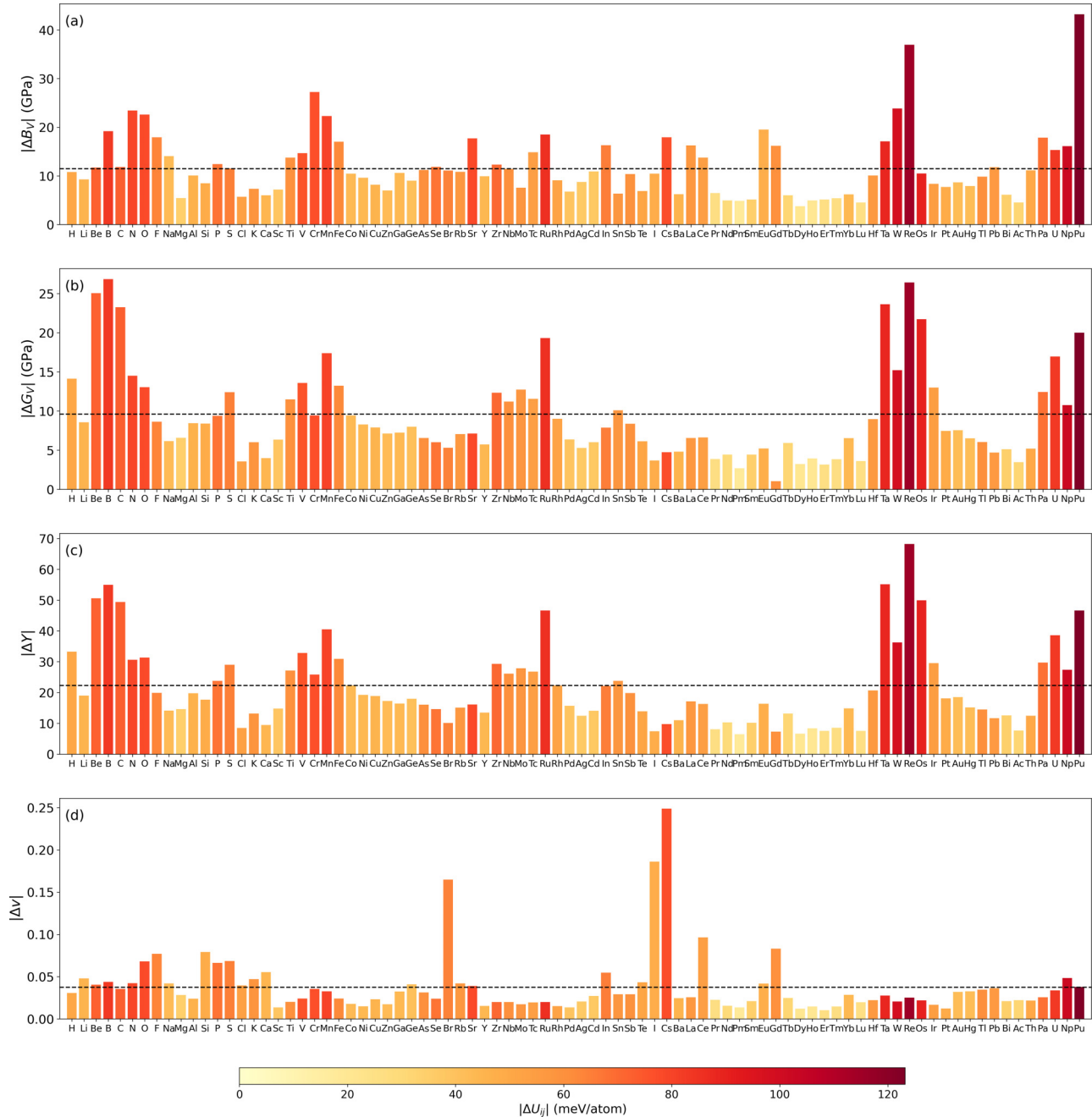


FIG. 10. Bar charts displaying the MAE of the predicted elastic properties, i.e., (a) B_V , (b) G_V , (c) Y , and (d) ν . The charts are categorized based on the elements present in the compounds in the test set and colored according to the average bulk modulus, as indicated in the color bar below. The dashed lines represent the MAE of each property, averaged over the data in the test set. It can be seen that most elements whose compounds have a high bulk modulus exhibit errors of B_V , G_V , and Y that exceed the data set’s MAE.

tensors, which are second-rank and third-rank tensors, respectively.

The data set used in this paper and a detailed numerical implementation of these models have been made available [50].

ACKNOWLEDGMENTS

This research is supported by the Program Management Unit for Human Resources and Institutional Development,

Research and Innovation (Grant No. B05F650024) and by Thailand Science Research and Innovation Fund Chulalongkorn University [CU_FRB65_ind (5)_110_23_40]. The authors acknowledge high-performance computing resources including NVIDIA A100 GPUs from Chula Intelligent and Complex Systems Lab, Faculty of Science, and from the Center for AI in Medicine (CU-AIM), Faculty of Medicine, Chulalongkorn University, Thailand. The authors also acknowledge the National Science and Technology Development Agency, National e-Science Infrastructure Consortium, Chulalongkorn University and the Chulalongkorn

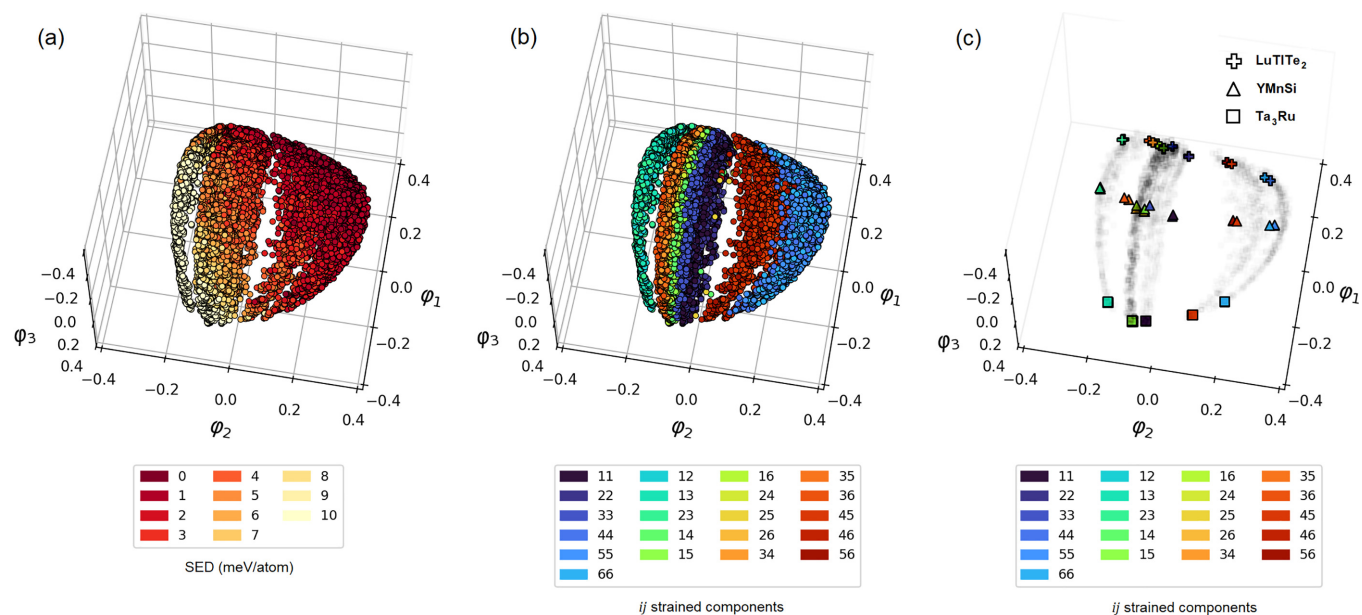


FIG. 11. Three-dimensional diffusion map representation of the concatenated latent features that are fed into the SED tensor prediction layer (the latent feature **B** in Fig. 3). (a) is colored by the SED scale, and (b) is colored by the strained component ij . (c) highlights the latent features of three example materials colored by the strained component ij as in (b); the features of other materials are shown in transparent gray. The three materials are the trigonal LuTlTe₂ phase (pluses), the tetragonal YMnSi phase (triangles), and the cubic Ta₃Ru phase (squares), which have low, medium, and high elastic constants, respectively. These are all the latent features in the test data set. A more elaborate discussion can be found in Sec. IV C.

Academic Advancement into Its 2nd Century Project, NSRF via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation [Grants No. B05F650021, No. B37G660013] (Thailand) for providing computing infrastructure.

APPENDIX A: INGREDIENTS OF SE(3)-TRANSFORMERS

Here we summarize the computational building block of SE(3)-equivariant GNNs with a built-in attention mechanism [SE(3)-Transformers]. A more detailed discussion of

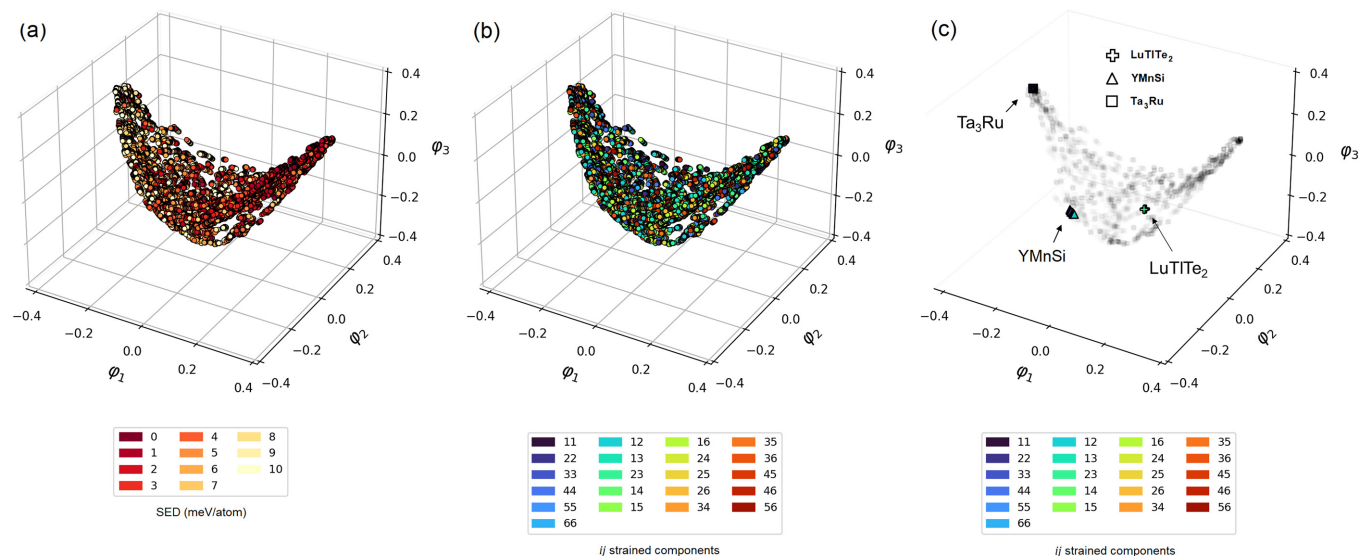


FIG. 12. Three-dimensional diffusion map representation of the global max-pooled latent features of the graph neural networks (i.e., the latent feature **A** in Fig. 3). (a) is colored by the SED scale, and (b) is colored by the strained component ij . (c) highlights the latent features of three example materials colored by the strained component ij as in (b); the features of other materials are shown in transparent gray. The three materials are the trigonal LuTlTe₂ phase (pluses), the tetragonal YMnSi phase (triangles), and the cubic Ta₃Ru phase (squares), which have low, medium, and high elastic constants, respectively. See the discussion and the interpretation of these plots in Sec. IV C.

this deep learning architecture can be found in the original proposal [44].

In an SE(3)-equivariant GNN *without* attention, also known as a tensor field network (TFN) [43], given a multi-channel input feature of the j th node, $\mathbf{f}_{\text{in},j}^k$, of a crystal graph, the GNN will update each crystal graph input feature by the following message-passing convolution:

$$\mathbf{f}_{\text{out},i}^l = \mathbf{W}^{ll} \mathbf{f}_{\text{in},i}^l + \sum_{k \geq 0} \sum_{j \neq i} \mathbf{W}^{lk} (\mathbf{r}_j - \mathbf{r}_i) \mathbf{f}_{\text{in},j}^k, \quad (\text{A1})$$

where i and j are the indices of a target node and its neighboring nodes, respectively; l and k specify type- l and type- k feature vectors which transform under SO(3) rotation according to the $(2l+1) \times (2l+1)$ and $(2k+1) \times (2k+1)$ Wigner D matrices, respectively [type-0 vector is invariant under SO(3) rotation]; \mathbf{W}^{ll} is the diagonal matrix whose diagonal is identically parametrized by a trainable parameter w^{ll} describing a self-interaction message passing; and \mathbf{W}^{lk} is a convolution kernel that propagates the channel- k feature from neighboring nodes to the channel- l feature of the node of interest such that

$$\mathbf{W}^{lk}(\mathbf{r}_j - \mathbf{r}_i) = \sum_{J=|k-l|}^{k+l} \varphi_J^{lk}(\|\mathbf{r}_j - \mathbf{r}_i\|) \mathbf{W}_J^{lk}(\mathbf{r}_j - \mathbf{r}_i), \quad (\text{A2})$$

with

$$\mathbf{W}_J^{lk}(\mathbf{r}_j - \mathbf{r}_i) = \sum_{m=-J}^J Y_{Jm} \left(\frac{\mathbf{r}_j - \mathbf{r}_i}{\|\mathbf{r}_j - \mathbf{r}_i\|} \right) \mathbf{Q}_{Jm}^{lk}, \quad (\text{A3})$$

where $\varphi_J^{lk}(\|\mathbf{r}\|)$ is a learnable radial function, Y_{Jm} is the m th component (m is the magnetic quantum number in quantum mechanics) of the spherical harmonic function $Y_J: \mathbb{R}^3 \rightarrow \mathbb{R}^{2J+1}$, and \mathbf{Q}_{Jm}^{lk} is a Clebsch-Gordan matrix with dimension $(2l+1) \times (2k+1)$. The SO(3)-equivariant basis kernel \mathbf{W}_J^{lk} imposes a fixed functional constraint on the convolution kernel along the angular direction, whereas the radial direction is still trainable via φ_J^{lk} .

The SE(3)-Transformer introduces the attention α_{ij} to each edge of a TFN, so that the message-passing rule of Eq. (A1) is weighted by the attentions. Without isolating out the self-interaction term, for brevity, this weighted message-passing rule can be compactly expressed as

$$\mathbf{f}_{\text{out},i}^l = \sum_j \alpha_{ij} \mathbf{v}_{ij}^l, \quad (\text{A4})$$

where the channel- l value message, \mathbf{v}_{ij}^l , is the same as that of the above TFN, defined as

$$\mathbf{v}_{ij}^l = \sum_{k \geq 0} \mathbf{W}_V^{lk}(\mathbf{r}_j - \mathbf{r}_i) \mathbf{f}_{\text{in},j}^k, \quad (\text{A5})$$

and the attention on the edge ij (which is invariant under SO(3) [44]) is

$$\alpha_{ij} = \frac{\exp(\mathbf{q}_i^\top \mathbf{k}_{ij})}{\sum_{j'} \exp(\mathbf{q}_i^\top \mathbf{k}_{ij'})}, \quad (\text{A6})$$

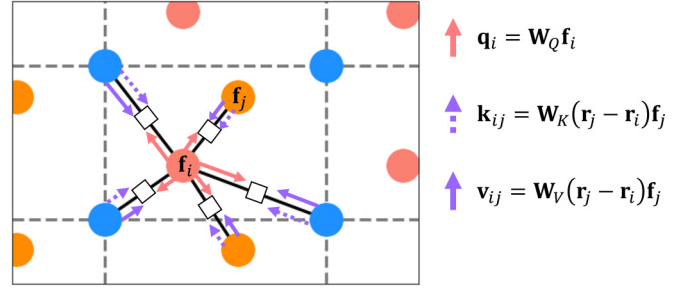


FIG. 13. A schematic of a GNN message-passing rule that aggregates features from neighboring nodes of a crystal graph with an attention mechanism. A white square on each edge depicts an attention head. A target node (pink circle) sends a query vector of node i (\mathbf{q}_i , pink arrows) to match with key vectors (\mathbf{k}_{ij} , dotted purple arrows) and the value message (\mathbf{v}_{ij} , solid purple arrows) from neighboring nodes (blue and orange circles). The angular momentum channels are not illustrated in the visualization.

with a query vector on a node i , \mathbf{q}_i , and a key vector on the edge ij , \mathbf{k}_{ij} , given by

$$\mathbf{q}_i = \bigoplus_{l \geq 0} \sum_{k \geq 0} \mathbf{W}_Q^{lk} \mathbf{f}_{\text{in},i}^k, \quad (\text{A7})$$

$$\mathbf{k}_{ij} = \bigoplus_{l \geq 0} \sum_{k \geq 0} \mathbf{W}_K^{lk} (\mathbf{r}_j - \mathbf{r}_i) \mathbf{f}_{\text{in},j}^k, \quad (\text{A8})$$

respectively. $\bigoplus_{l \geq 0}$ denotes the direct sum over the l channels. Figure 13 depicts the message passing with the attention mechanism of the SE(3)-Transformer model.

APPENDIX B: MACROSCOPIC ELASTIC MODULI

Here are the formulas of elastic moduli computable from the elastic constants: Voigt's bulk modulus

$$B_V = \frac{1}{9} \sum_{i,j=1}^3 C_{ij}, \quad (\text{B1})$$

Voigt's shear modulus

$$G_V = \frac{1}{30} \sum_{i,j=1}^3 (C_{ii} - C_{ij}) + \frac{1}{5} \sum_{i=4}^6 C_{ii}, \quad (\text{B2})$$

Young's modulus

$$Y = \frac{9BG}{3B + G}, \quad (\text{B3})$$

and the Poisson's ratio

$$\nu = \frac{3B - 2G}{6B + 2G}. \quad (\text{B4})$$

APPENDIX C: SYMMETRY IN THE STRAIN ENERGY TENSOR

The strain energy density (SED) tensor \mathbf{U} is derived from the elastic tensor \mathbf{C} . The symmetry in \mathbf{U} follows the symmetry in \mathbf{C} , which depends on the crystal system.

1. Hexagonal and tetragonal I (for Laue class 4/mmm)

For hexagonal and tetragonal I lattices (Laue class 4/mmm) the elastic tensor and SED tensor are as follows:

$$\mathbf{C}_{\text{hex/tet}} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 \\ & C_{11} & C_{13} & 0 & 0 & 0 \\ & & C_{33} & 0 & 0 & 0 \\ & & & C_{44} & 0 & 0 \\ & & & & C_{44} & 0 \\ & & & & & C_{66} \end{bmatrix}, \quad (C1)$$

$$\mathbf{U}_{\text{hex/tet}} = \begin{bmatrix} U_{11} & U_{12} & U_{13} & U_{14} & U_{14} & U_{16} \\ & U_{11} & U_{13} & U_{14} & U_{14} & U_{16} \\ & & U_{33} & U_{34} & U_{34} & U_{36} \\ & & & U_{44} & U_{45} & U_{46} \\ & & & & U_{44} & U_{46} \\ & & & & & U_{66} \end{bmatrix}. \quad (C2)$$

2. Tetragonal II (for Laue class 4/m)

For tetragonal II lattices (Laue class 4/m) the elastic tensor and SED tensor are as follows:

$$\mathbf{C}_{\text{tetII}} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & C_{16} \\ & C_{11} & C_{13} & 0 & 0 & -C_{16} \\ & & C_{33} & 0 & 0 & 0 \\ & & & C_{44} & 0 & 0 \\ & & & & C_{44} & 0 \\ & & & & & C_{66} \end{bmatrix}, \quad (C3)$$

$$\mathbf{U}_{\text{tetII}} = \begin{bmatrix} U_{11} & U_{12} & U_{13} & U_{14} & U_{14} & U_{16} \\ & U_{11} & U_{13} & U_{14} & U_{14} & U'_{16} \\ & & U_{33} & U_{34} & U_{34} & U_{36} \\ & & & U_{44} & U_{45} & U_{46} \\ & & & & U_{44} & U_{46} \\ & & & & & U_{66} \end{bmatrix}, \quad (C4)$$

where $U'_{16} = -C_{16}\epsilon_1\epsilon_6 + \frac{1}{2}(C_{11}\epsilon_1^2 + C_{66}\epsilon_6^2)$.

3. Rhombohedral I (for Laue class $\bar{3}m$)

For rhombohedral I lattices (Laue class $\bar{3}m$) the elastic tensor and SED tensor are as follows:

$$\mathbf{C}_{\text{rhomI}} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & 0 & 0 \\ & C_{11} & C_{13} & -C_{14} & 0 & 0 \\ & & C_{33} & 0 & 0 & 0 \\ & & & C_{44} & 0 & 0 \\ & & & & C_{44} & C_{14} \\ & & & & & C_{66} \end{bmatrix}, \quad (C5)$$

$$\mathbf{U}_{\text{rhomI}} = \begin{bmatrix} U_{11} & U_{12} & U_{13} & U_{14} & U''_{14} & U_{16} \\ & U_{11} & U_{13} & U'_{14} & U''_{14} & U_{16} \\ & & U_{33} & U_{34} & U_{34} & U_{36} \\ & & & U_{44} & U_{45} & U_{46} \\ & & & & U_{44} & U'_{46} \\ & & & & & U_{66} \end{bmatrix}, \quad (C6)$$

where $U'_{14} = -C_{14}\epsilon_1\epsilon_4 + \frac{1}{2}(C_{11}\epsilon_1^2 + C_{44}\epsilon_4^2)$, $U''_{14} = \frac{1}{2}(C_{11}\epsilon_1^2 + C_{44}\epsilon_4^2)$, and $U'_{46} = C_{14}\epsilon_4\epsilon_6 + \frac{1}{2}(C_{44}\epsilon_4^2 + C_{66}\epsilon_6^2)$.

4. Rhombohedral II (for Laue class 3)

For rhombohedral II lattices (Laue class 3) the elastic tensor and SED tensor are as follows:

$$\mathbf{C}_{\text{rhomII}} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & 0 \\ & C_{11} & C_{13} & -C_{14} & -C_{15} & 0 \\ & & C_{33} & 0 & 0 & 0 \\ & & & C_{44} & 0 & -C_{15} \\ & & & & C_{44} & C_{14} \\ & & & & & C_{66} \end{bmatrix}, \quad (C7)$$

$$\mathbf{U}_{\text{rhomII}} = \begin{bmatrix} U_{11} & U_{12} & U_{13} & U_{14} & U_{15} & U_{16} \\ & U_{11} & U_{13} & U'_{14} & U'_{15} & U_{16} \\ & & U_{33} & U_{34} & U_{34} & U_{36} \\ & & & U_{44} & U_{45} & U''_{46} \\ & & & & U_{44} & U'_{46} \\ & & & & & U_{66} \end{bmatrix}, \quad (C8)$$

where $U''_{46} = -C_{15}\epsilon_4\epsilon_6 + \frac{1}{2}(C_{44}\epsilon_4^2 + C_{66}\epsilon_6^2)$.

5. Orthorhombic, monoclinic, and triclinic

The strain energy tensors of orthorhombic, monoclinic, and triclinic lattices do not possess any degenerate structure, and the strain energy is expressed by the most general tensor of the form (4) in the main text.

APPENDIX D: THE MINIMAL SET OF STRAINED COMPONENTS FOR EACH CRYSTAL SYSTEM IN A TRAINING DATA SET

The equivariant property of the SE(3)-Transformer should be able to identify strained crystal graph inputs that are equivalent up to a rotation; thus to reduce the number of training

TABLE III. Statistics of the number of edges for 12 105 materials created by the cell-radii and crystal near-neighbor (CrystalNN) methods. SD, standard deviation; 25% 50%, and 75%, 25th, 50th, and 75th percentiles.

	Cell radii	CrystalNN
Mean	2115.16	143.47
SD	4665.66	129.67
Min	6	6
25%	368	58
50%	1024	112
75%	1776	224
Max	75864	2272

data, we require one independent ij component of the SED tensor for each degeneracy class vector. The other strained components can be incorporated during the randomized update of the one-hot strained component vector within the same degeneracy class in each training epoch; see Sec. III C. The minimal set of ij indices of each crystal system can be chosen from the degenerate structure of the SED tensor for each crystal system in Appendix C. Our chosen set is presented in Table I in the main text.

APPENDIX E: DATA SET

Materials data are pooled from the Materials Project database [24]. There are 12 105 materials in the database that have the elastic tensor. However, only some materials are used in this paper due to memory restrictions. The memory usage increases with the number of edges, and so the computationally feasible data are restricted to materials with a maximum number of edges of 2000 constructed by a cell-radii approach. Of the 12 105 materials from the database, there are 9296 materials with this property (about 76.79%; Table III). Figure 14 shows that the selected materials have bulk and shear moduli distributed in the range 0–600 GPa, which can be a good representation of the whole database. In addition, we have found that the validation errors do not decrease significantly even when the number of edges increases to 3000. The statistics of crystal systems curated for training and test sets of these 9296 materials is shown in Table IV. The data

TABLE IV. The number of data for training and test sets categorized by crystal system.

Crystal system	Training set		Test set	
	Count	%	Count	%
Cubic	3865	46.17	441	47.83
Hexagonal	1369	16.35	146	15.84
Tetragonal	1580	18.87	163	17.68
Trigonal	530	6.33	60	6.51
Orthorhombic	731	8.73	79	8.57
Monoclinic	253	3.02	27	2.93
Triclinic	43	0.51	6	0.65

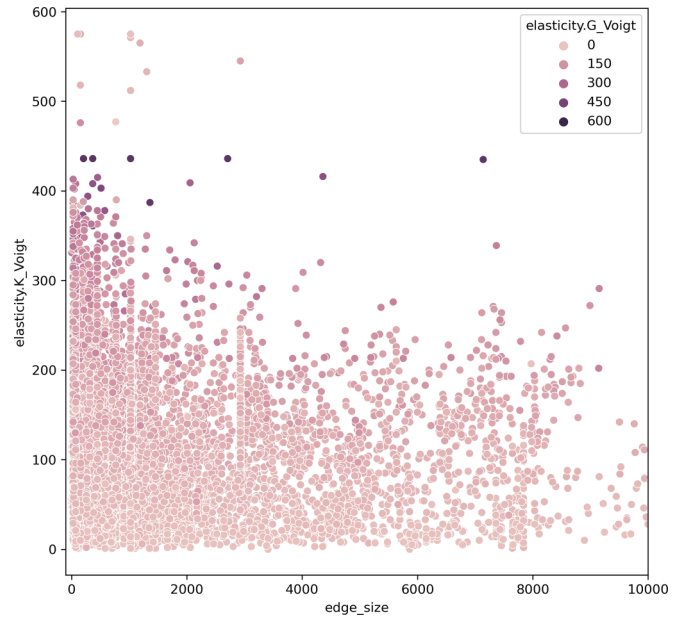


FIG. 14. The bulk modulus of materials plotted against the number of edges created by the cell-radii method, with the hue indicating the shear modulus. The data below a hard cutoff of 2000 edges can already cover the range of values of bulk and shear moduli of the whole data set.

set is partitioned into a training set and a validation set using the stratified method (in SCIKIT-LEARN) with a random seed number of 0.

APPENDIX F: MODEL PARAMETERS

The hyperparameters that significantly affect the model performance are the number of convolution layers, the number of l channels in the SE(3) kernel, and the number of hidden units of the learnable radial function in the kernel ϕ . The number of convolution layers is 4, where the model performance only slightly improves with more layers. The learnable radial function ϕ is a feedforward network, consisting of an input layer that is fed into a fully connected hidden layer with a ReLU activation, which is forwarded to another fully connected layer that outputs a hidden feature vector. We use 128 hidden units in the hidden layer of ϕ to avoid oversquashing, and the performance only marginally improves expanding beyond 128 hidden units. The dimension of the hidden feature vectors $\mathbf{f}_{\text{hid},n}^l$ is 32, whereas the dimension of $\mathbf{f}_{\text{reg},n}^l$ and $\mathbf{f}_{\text{class},n}^l$ is 128. The dimension of the input atomic number embedding is 512. The maximum degree of l is 4, where, of course, the model performance also improves if the degree is higher especially for low-symmetry crystal structures; however, increasing the maximum degree of l beyond 4 becomes infeasible on our computational infrastructure, with other hyperparameters fixed. The number of transformer heads is 2. We use the adaptive moment estimation (ADAM) optimizer and cosine annealing method for a scheduler with an initial learning rate of 10^{-4} .

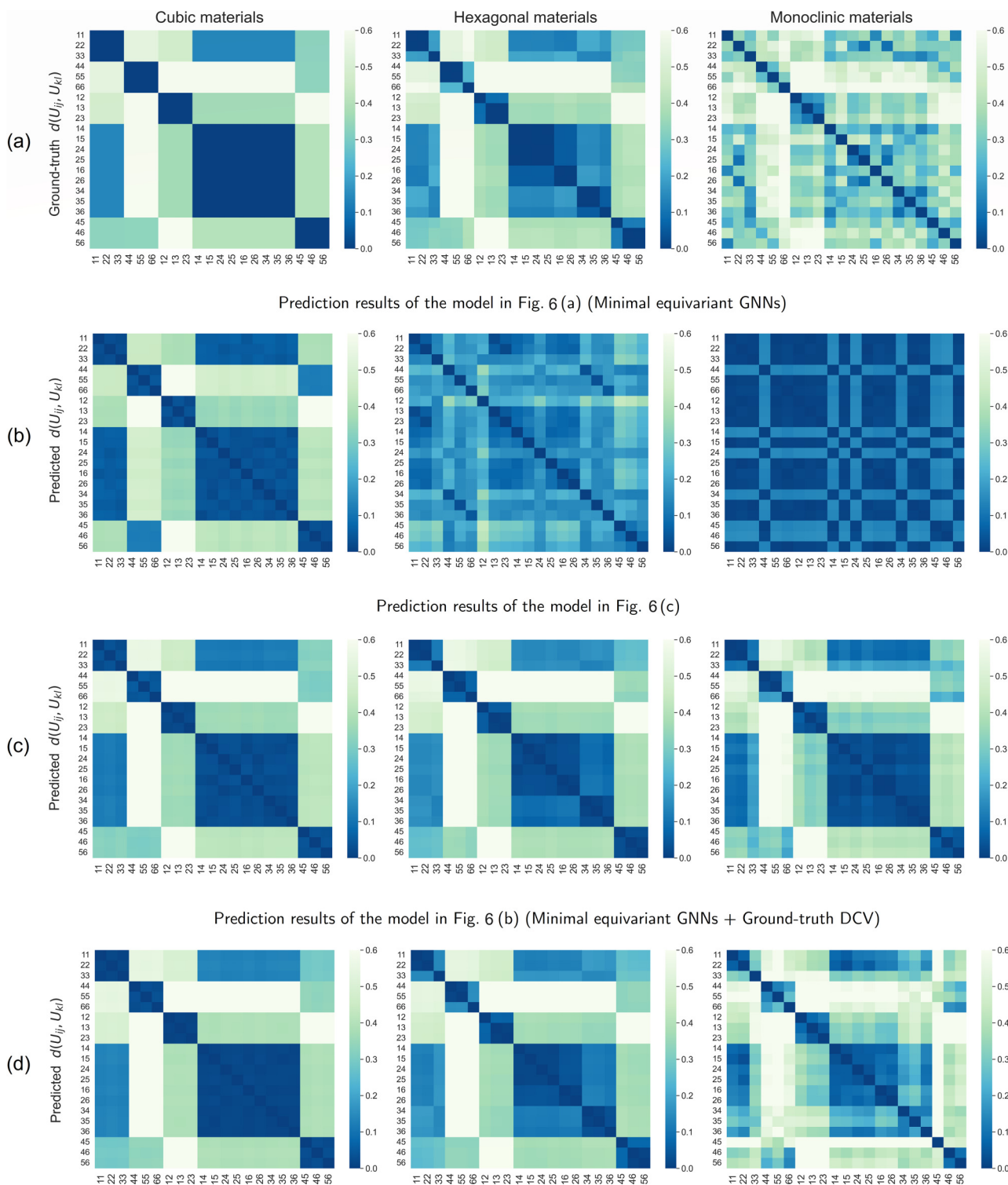


FIG. 15. The distance metrics (of the materials whose numbers of edges are less than 500) of (a) the ground truths, (b) the model in Fig. 6(a), (c) the model Fig. 6(c), and (d) the model in Fig. 6(b). The distance metrics in (d) suggest that the concatenation of the strained crystal graph latent representation and the degeneracy class vector significantly increases the expressiveness of the SED regression network, motivating the self-supervised network architecture in the main text.

APPENDIX G: ADDITIONAL RESULTS FOR TRIAL MODELS

This Appendix provides additional results for the earlier trial models shown in Fig. 6. Figure 15 shows the distance

metric of the ground truths of the models in Fig. 6. In Table VI the MAE and RMSE of predicted elastic properties using the model in Fig. 6(b) are provided, while Table VII provides the MAE of predicted elastic properties using the models in Figs. 6(a)–6(c).

TABLE V. The MAE and RMSE of predicted elastic properties using the model in Fig. 6(b). The data set for this model consists of materials whose number of edges is less than 2000.

Properties	Average	MAE	RMSE
U_{ij} (meV/atom)	2.652	0.661	1.258
C_{ij} (GPa)	42.92	10.98	17.86
B_V (GPa)	107.07	11.58	19.10
G_V (GPa)	50.98	9.91	16.31
Y	129.62	22.81	37.24
ν	0.401	0.035	0.081

TABLE VI. The MAE of predicted elastic properties using the model in Figs. 6(a)–6(c). The data set for these earlier models consists of materials whose number of edges is less than 500. Bold terms highlight the highest accuracy among the three models.

Properties	Average	MAE of Fig. 6(a)	MAE of Fig. 6(b)	MAE of Fig. 6(c)
U_{ij} (meV/atom)	2.749	1.628	0.842	0.866
C_{ij} (GPa)	44.10	78.90	13.29	18.66
B_V (GPa)	108.67	67.95	15.35	17.07
G_V (GPa)	53.92	79.94	12.46	14.63
Y	136.26	86.05	29.08	31.93
ν	0.394	1.758	0.044	0.061

- [1] M. Born and K. Huang, *Dynamical Theory of Crystal Lattices*, International Series of Monographs on Physics (Clarendon, Oxford, 1988).
- [2] F. Mouhat and F.-X. Coudert, Necessary and sufficient elastic stability conditions in various crystal systems, *Phys. Rev. B* **90**, 224104 (2014).
- [3] J. R. MacDonald, Review of some experimental and analytical equations of state, *Rev. Mod. Phys.* **41**, 316 (1969).
- [4] V. V. Brazhkin, High-pressure synthesized materials: Treasures and hints, *High Pressure Res.* **27**, 333 (2007).
- [5] M. Yamaguchi, T. Yagi, T. Azuhata, T. Sota, K. Suzuki, S. Chichibu, and S. Nakamura, Brillouin scattering study of gallium nitride: Elastic stiffness constants, *J. Phys.: Condens. Matter* **9**, 241 (1997).
- [6] H. J. Reichmann, S. D. Jacobsen, S. J. Mackwell, and C. A. McCammon, Sound wave velocities and elastic constants for magnesiowüstite using gigahertz interferometry, *Geophys. Res. Lett.* **27**, 799 (2000).
- [7] D. M. Teter, Computational alchemy: The search for new superhard materials, *MRS Bull.* **23**, 22 (1998).
- [8] X.-Q. Chen, H. Niu, D. Li, and Y. Li, Modeling hardness of polycrystalline materials and bulk metallic glasses, *Intermetallics* **19**, 1275 (2011).
- [9] Y. Tian, B. Xu, and Z. Zhao, Microscopic theory of hardness and design of novel superhard crystals, *Int. J. Refract. Met. Hard Mater.* **33**, 93 (2012).
- [10] O. H. Nielsen and R. M. Martin, Stresses in semiconductors: *Ab initio* calculations on Si, Ge, and GaAs, *Phys. Rev. B* **32**, 3792 (1985).
- [11] P. Labéguerie, M. Harb, I. Baraille, and M. Rérat, Structural, electronic, elastic, and piezoelectric properties of α -quartz and MXO_4 ($M = \text{Al, Ga, Fe; } X = \text{P, As}$) isomorph compounds: A DFT study, *Phys. Rev. B* **81**, 045107 (2010).
- [12] Y. Wang, J. J. Wang, H. Zhang, V. R. Manga, S. L. Shang, L.-Q. Chen, and Z.-K. Liu, A first-principles approach to finite temperature elastic constants, *J. Phys.: Condens. Matter* **22**, 225404 (2010).
- [13] A. Ektarawong, S. I. Simak, and B. Alling, Carbon-rich icosahedral boron carbides beyond B_4C and their thermodynamic stabilities at high temperature and pressure from first principles, *Phys. Rev. B* **94**, 054104 (2016).
- [14] R. Stadler, W. Wolf, R. Podloucky, G. Kresse, J. Furthmüller, and J. Hafner, *Ab initio* calculations of the cohesive, elastic, and dynamical properties of $CoSi_2$ by pseudopotential and all-electron techniques, *Phys. Rev. B* **54**, 1729 (1996).
- [15] Y. Le Page and P. Saxe, Symmetry-general least-squares extraction of elastic data for strained materials from *ab initio* calculations of stress, *Phys. Rev. B* **65**, 104104 (2002).
- [16] K. Sarasamak, S. Limpijumnong, and W. R. L. Lambrecht, Pressure-dependent elastic constants and sound velocities of wurtzite SiC , GaN , InN , ZnO , and $CdSe$, and their relation to the high-pressure phase transition: A first-principles study, *Phys. Rev. B* **82**, 035201 (2010).
- [17] L. A. Chernozatonskii, P. B. Sorokin, A. A. Kuzubov, B. P. Sorokin, A. G. Kvashnin, D. G. Kvashnin, P. V. Avramov, and B. I. Yakobson, Influence of size effect on the electronic and elastic properties of diamond films with nanometer thickness, *J. Phys. Chem. C* **115**, 132 (2011).
- [18] T. Pakornchote, A. Ektarawong, U. Pinsook, and T. Bovornratanarak, Modifying electronic and elastic properties of 2-dimensional [110] diamond by nitrogen substitution, *C* **7**, 8 (2021).

- [19] M. A. Caro, S. Schulz, and E. P. O'Reilly, Comparison of stress and total energy methods for calculation of elastic properties of semiconductors, *J. Phys.: Condens. Matter* **25**, 025803 (2013).
- [20] G. Kresse and J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, *Comput. Mater. Sci.* **6**, 15 (1996).
- [21] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos *et al.*, QUANTUM ESPRESSO: A modular and open-source software project for quantum simulations of materials, *J. Phys.: Condens. Matter* **21**, 395502 (2009).
- [22] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, and M. C. Payne, First principles methods using CASTEP, *Z. Kristallogr. - Cryst. Mater.* **220**, 567 (2005).
- [23] M. de Jong, W. Chen, T. Angsten, A. Jain, R. Notestine, A. Gamst, M. Sluiter, C. Krishna Ande, S. van der Zwaag, J. J. Plata, C. Toher, S. Curtarolo, G. Ceder, K. A. Persson, and M. Asta, Charting the complete elastic properties of inorganic crystalline compounds, *Sci. Data* **2**, 150009 (2015).
- [24] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, The Materials Project: A materials genome approach to accelerating materials innovation, *APL Mater.* **1**, 011002 (2013).
- [25] E. Mazhnik and A. R. Oganov, Application of machine learning methods for predicting new superhard materials, *J. Appl. Phys.* **128**, 075102 (2020).
- [26] Y. Zhao, K. Yuan, Y. Liu, S.-Y. Louis, M. Hu, and J. Hu, Predicting elastic properties of materials from electronic charge density using 3D deep convolutional neural networks, *J. Phys. Chem. C* **124**, 17262 (2020).
- [27] H. Levämäki, F. Tasnádi, D. G. Sangiovanni, L. J. S. Johnson, R. Armiento, and I. A. Abrikosov, Predicting elastic properties of hard-coating alloys using ab-initio and machine learning methods, *npj Comput. Mater.* **8**, 17 (2022).
- [28] N. Linton and D. S. Aidhy, A machine learning framework for elastic constants predictions in multi-principal element alloys, *APL Mach. Learn.* **1**, 016109 (2023).
- [29] P. Avery, X. Wang, C. Oses, E. Gossett, D. M. Proserpio, C. Toher, S. Curtarolo, and E. Zurek, Predicting superhard materials via a machine learning informed evolutionary structure search, *npj Comput. Mater.* **5**, 89 (2019).
- [30] W.-C. Chen, J. N. Schmidt, D. Yan, Y. K. Vohra, and C.-C. Chen, Machine learning and evolutionary prediction of superhard B-C-N compounds, *npj Comput. Mater.* **7**, 114 (2021).
- [31] S. Kearnes, K. McCloskey, M. Berndl, V. Pande, and P. Riley, Molecular graph convolutions: Moving beyond fingerprints, *J. Comput.-Aided Mol. Des.* **30**, 595 (2016).
- [32] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, Neural message passing for quantum chemistry, *Proceedings of the 34th International Conference on Machine Learning*, edited by D. Precup and Y. W. Teh, Proceedings of Machine Learning Research (PMLR, 2017), Vol. 70, pp. 1263–1272.
- [33] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, SchNet—A deep learning architecture for molecules and materials, *J. Chem. Phys.* **148**, 241722 (2018).
- [34] N. Dym and H. Maron, On the universality of rotation equivariant point cloud networks, in *Ninth International Conference on Learning Representations (ICLR, Appleton, WI, 2021)*.
- [35] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, C. Gulcehre, F. Song, A. Ballard, J. Gilmer, G. Dahl, A. Vaswani, K. Allen, C. Nash, V. Langston, C. Dyer *et al.*, Relational inductive biases, deep learning, and graph networks, [arXiv:1806.01261](https://arxiv.org/abs/1806.01261) [cs.LG].
- [36] The tensor field network (TFN), LInet, geometric message-passing neural network (GemNet), and E(n)-equivariant graph neural network (EGNN) have exploited this computational architecture to make better data-driven predictions of the properties of chemical compounds [43,51–54]. The SE(3)-Transformers model based on a TFN with an attention mechanism also demonstrates a good predictive performance on the standard Quantum Machine 9 (QM9) data set [44]. In addition, the translation-rotation equivariance concept has also been incorporated into generative modeling, in an attempt to search for novel molecules and crystal structures with desirable properties [55,56].
- [37] W. S. Slaughter, *The Linearized Theory of Elasticity* (Birkhäuser, Boston, 2002).
- [38] C. S. G. Cousins, Inner elasticity, *J. Phys. C: Solid State Phys.* **11**, 4867 (1978).
- [39] T. Xie and J. C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, *Phys. Rev. Lett.* **120**, 145301 (2018).
- [40] N. E. R. Zimmermann, M. K. Horton, A. Jain, and M. Haranczyk, Assessing local structure motifs using order parameters for motif recognition, interstitial identification, and diffusion path characterization, *Front. Mater.* **4**, 34 (2017).
- [41] H. Pan, A. M. Ganose, M. Horton, M. Aykol, K. A. Persson, N. E. R. Zimmermann, and A. Jain, Benchmarking coordination number prediction algorithms on inorganic crystal structures, *Inorg. Chem.* **60**, 1590 (2021).
- [42] M. M. Bronstein, J. Bruna, T. Cohen, and P. Velickovic, Geometric deep learning: Grids, groups, graphs, geodesics, and gauges, [arXiv:2104.13478](https://arxiv.org/abs/2104.13478) [cs.LG].
- [43] N. Thomas, T. 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, [arXiv:1802.08219](https://arxiv.org/abs/1802.08219) [cs.LG].
- [44] F. Fuchs, D. Worrall, V. Fischer, and M. Welling, SE(3)-Transformers: 3D roto-translation equivariant attention networks, in *Advances in Neural Information Processing Systems (NeurIPS 2020)*, Advances in Neural Information Processing Systems Vol. 33, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin (Curran Associates, Red Hook, NY, 2020), pp. 1970–1981.
- [45] Due to the symmetry of the SED, we identify only an upper triangular element.
- [46] V. Fung, J. Zhang, E. Juarez, and B. G. Sumpter, Benchmarking graph neural networks for materials chemistry, *npj Comput. Mater.* **7**, 84 (2021).
- [47] P. Schwerdtfeger and J. K. Nagle, 2018 Table of static dipole polarizabilities of the neutral elements in the periodic table, *Mol. Phys.* **117**, 1200 (2019).

- [48] S. Lafon, Diffusion maps and geometric harmonics, Ph.D. thesis, Yale University, 2004.
- [49] R. R. Coifman and S. Lafon, Diffusion maps, *Appl. Comput. Harmonic Anal.* **21**, 5 (2006).
- [50] https://github.com/trachote/predict_elastic_tensor.
- [51] J. Gastegger, F. Becker, and S. Günnemann, GemNet: Universal directional graph neural networks for molecules, in *Advances in Neural Information Processing Systems (NeurIPS 2020)*, Advances in Neural Information Processing Systems Vol. 34, edited by A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan (Curran Associates, Red Hook, NY, 2021), pp. 6790–6802.
- [52] V. G. Satorras, E. Hoogeboom, and M. Welling, E(n) equivariant graph neural networks, in *Proceedings of the 38th International Conference on Machine Learning*, Proceedings of Machine Learning Research Vol. 139, edited by M. Meila and T. Zhang (PMLR, Cambridge, MA, 2021), pp. 9323–9332.
- [53] B. K. Miller, M. Geiger, T. E. Smidt, and F. Noé, Relevance of rotationally equivariant convolutions for predicting molecular properties, [arXiv:2008.08461](https://arxiv.org/abs/2008.08461) [cs.LG].
- [54] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis, *Comput. Mater. Sci.* **68**, 314 (2013).
- [55] M. Xu, L. Yu, Y. Song, C. Shi, S. Ermon, and J. Tang, GeoDiff: A geometric diffusion model for molecular conformation generation, in *Tenth International Conference on Learning Representations (ICLR, Appleton, WI, 2022)*.
- [56] T. Xie, X. Fu, O.-E. Ganea, R. Barzilay, and T. S. Jaakkola, Crystal diffusion variational autoencoder for periodic material generation, in *Tenth International Conference on Learning Representations (ICLR, Appleton, WI, 2022)*.