ADVANCED SCIENCE Open Access
www.advancedscience.com

# A Cluster-Based Deep Learning Model Perceiving Series Correlation for Accurate Prediction of Phonon Spectrum

*Chao Liang, Yilimiranmu Rouzhahong, Shunwei Yao, Junhao Liang, Chunlin Yu, Biao Wang,* and Huashan Li**

**The spectral properties are the most prevalent continuous representation for characterizing transport phenomena and excitation responses, yet their accurate predictions remain a challenge due to the inability to perceive series correlations by existing machine learning (ML) models. Herein, a ML model named cluster-based series graph networks (CSGN) is developed based on the dynamical theory of crystal lattices to predict phonon density of states (PDOS) spectrum for crystal materials. The multiple atomic cluster representation is constructed to capture the diverse vibration modes, while the mixture Gaussian process and dynamic time warping mechanism are compiled to project from clusters to PDOS spectrum. Accurate predictions of complicated spectra with multiple or overlapping peaks are achieved. The high performance of CSGN model can be attributed to the pertinent feature extraction and the appropriate similarity evaluation, which enable the natural perception of structure-property relation and intrinsic series correlations as confirmed in the predictive results. The transferable and interpretable CSGN model advances ML predictions of spectral properties and reveals the potential of designing ML methods based on physical mechanisms.**

## 1. Introduction

Spectral properties serve as the most prevalent continuous representation, influenced by both intrinsic physical-chemical features and external activating conditions.[1] It can reflect the diverse condensed states, transport phenomena, and excitation responses in physical, chemical, and biomaterial systems.[2] The large-scale prediction of spectral properties based on machine-learning (ML) model would noticeably accelerate the material design and phenomenon interpretation.[3] Compared to the prediction of single-value properties that focuses on the precise mapping to target properties, spectrum prediction is a much more challenging problem owing to the additional complexity of representing a continuous target function and capturing the intrinsic correlation between different parts of the spectrum.[4]

Recently, the predictions of fragmentation and molecular spectra from the mass spectrum have been accomplished by various ML algorithms.[5] Considering that each fragment/molecule is related to a single peak within the input sequence, such prediction of the discrete spectrum is essentially a single-value prediction task. The progress in predicting spectral-type properties is currently restricted in material science, wherein only a few ML models were developed to predict the phononic and electronic density of states (DOS). The Euclidean neural network (E(3)NN) model applied graph-based representation to predict PDOS, which straightly transferred the single-value predictive model into the spectral property prediction.[6] The Mat2Spec

C. Liang, S. Yao, J. Liang, C. Yu, H. Li
School of Physics
Sun Yat-Sen University
Guangzhou 510275, China
E-mail: lihsh25@mail.sysu.edu.cn
Y. Rouzhahong, B. Wang
School of Materials Science and Engineering
Dongguan University of Technology
Dongguan 523808, China
E-mail: wangbiao@mail.sysu.edu.cn
H. Li
Guangdong Provincial Key Laboratory of Magnetoelectric Physics and Devices
School of Physics
Sun Yat-sen University
Guangzhou 510275, China
H. Li
Center for Neutron Science and Technology
School of Physics
Sun Yat-sen University
Guangzhou 510275, China

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/advs.202406183

© 2024 The Author(s). Advanced Science published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

**DOI: 10.1002/advs.202406183**

model improved the PDOS prediction by exploring different loss functions (MAE, KL divergence, and Wasserstein distance) and combining them with contrastive learning.[7] The physically informed ML model promoted the electron DOS prediction by designing the mixed loss function based on spectral geometric properties, statistical features, and physical features.[8]

Despite that the upgraded loss functions have substantially improved the continuity of target functions, the overall prediction quality of spectral properties is far from satisfactory, with the prediction error sharply increasing with system complexity.[9] This can primarily be ascribed to the failure to perceive the intrinsic correlation within a spectrum and evaluate the similarity between different spectra. Specifically, the intrinsic correlation between excitations at various energetic scopes within a PDOS spectrum originates from the distinct vibration modes associated with the same vibrational cluster.[10] Each phonon spectrum is contributed by multi-scale vibrational clusters and thus possesses complicated energetic correlations that can be partially transferred to other materials with similar vibrational clusters.[11] Such correlations are impossible to perceive by applying general deep neural networks on the crystal graph.[12] Besides, peak features are the critical information of a spectrum, yet the current loss functions based on Lp norm distances emphasize the accumulative single-point errors rather than the deviation of peak features, leading to an inaccurate gradient for the prediction of spectra with multiple or overlap peaks.[13] To overcome the above challenges, it is essential to design a spectrum-oriented ML model based on the continuous representation and the underlying structure-property relationship.

Herein, we developed a spectrum prediction model named cluster-based series graph networks (CSGN) to perceive series correlations and to identify spectral patterns of PDOS. The model comprises a phonon representation module describing the contributions of multi-scale atomic clusters, as well as a spectral predictive module based on mixture Gaussian process (MGP) and dynamic time warping (DTW) mechanism,[13,14] which is an advanced statistical method used primarily for measuring similarity between two temporal series with complex evolutionary features. The results verified that the series correlations have been successfully learned by the CSGN model by projecting each peak of PDOS on cluster vibrations with our continuous spectral representation method. Taking advantage of the prior physical information, the CSGN model achieves state-of-the-art (SOTA) prediction performance and interpretability comparable to the atomic-scale simulations, which provides a potential architecture for exploring complicated spectral properties in multidisciplinary science and discovering outstanding thermal conductive materials.

## 2. Results

### 2.1. Development of CSGN to Perceive Series Correlation

Typical ML models in material science for predicting single-valued property rely on the construction of nonlinear mapping $y = f(\sum_{n=1}^{N} x_n)$ between the microscopic structure and target property $y$, wherein $x_n$ is the chemical environment of the $n$th atom in the primitive cell (**Figure 1a**). Such models cannot be employed to appropriately predict PDOS spectrum given their inappropriate perception of series patterns and intrinsic correlations in continuous energetic space. Considering that a PDOS spectrum $y_{1\to t}$ is composed of various vibrational excitations induced by multi-scale atomic clusters, several peaks in different energy scopes may arise from the same cluster, while similar clusters in different materials would lead to analogous peaks in spectrum (Figure 1b).

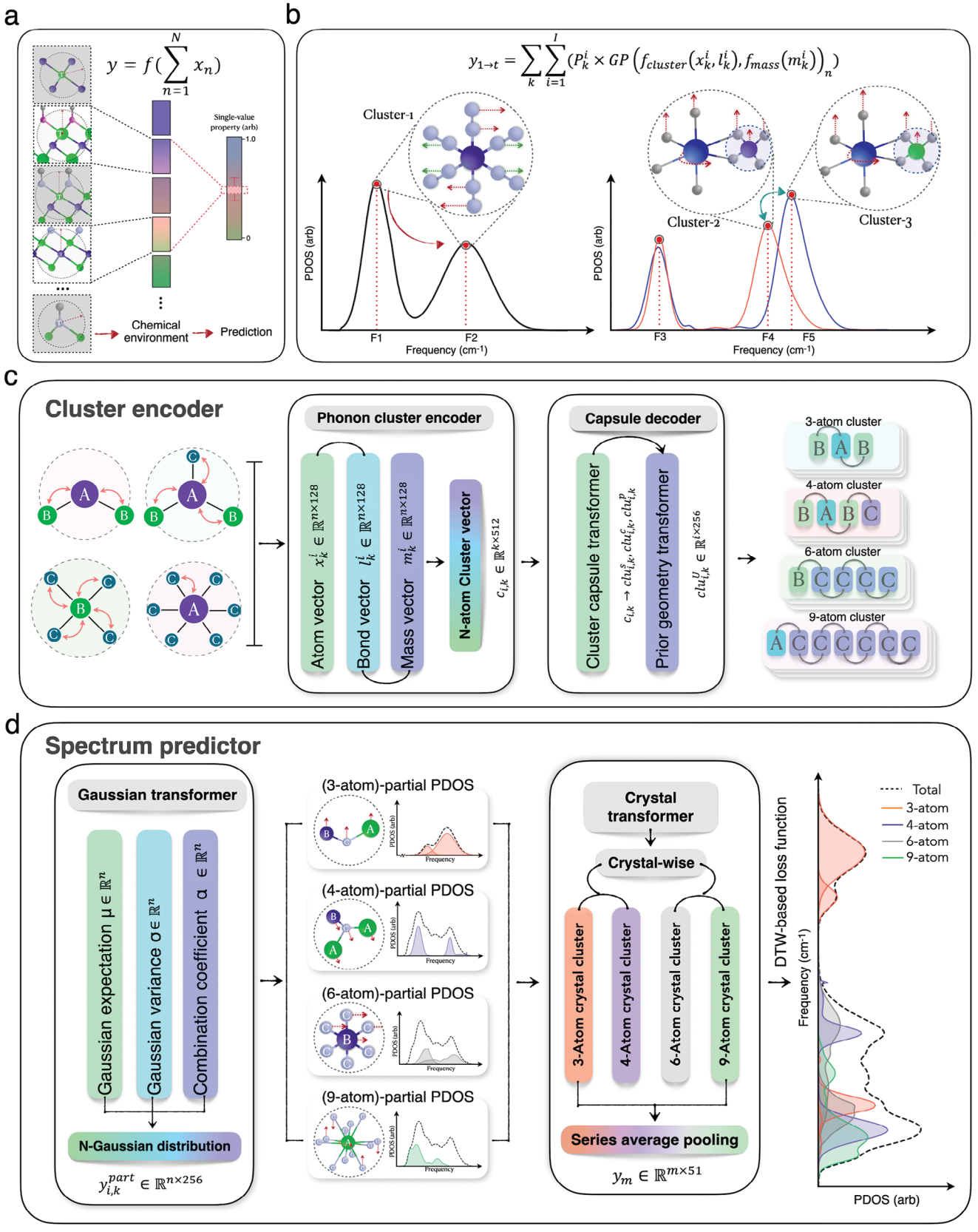
Accordingly, we modified the prediction model for PDOS as the weighted summation of cluster contributions, via the projection of cluster graph to mixture Gaussian function.

$$F(y_{1\to t}|x,\ l,m,n) = \sum_{k}\sum_{i=1}^{I}\left(P_k^i \times GP\left(f_{cap}\left(f_{cluster}\left(x_k^i, l_k^i\right),\ f_{mass}\left(m_k^i\right)\right)\right)_n\right) \tag{1}$$

wherein $x_k^i$ denotes the $k$-scale vibrational clusters with the $i$th atom serving as center atom (each cluster containing $k$-1atoms), $l_k^i$ and $m_k^i$ indicates the relevant bond configuration vector, and mass vector, and $P_k^i$ is the superposition weight of clusters after n-order Gaussian process. The nonlinear function $f_{cluster}$ generates a cluster graph based on structural features, $f_{mass}$ projects the mass information on cluster graph to reflect the dependence of vibrational frequency on atom mass, the capsule transformer $f_{cap}$ refines the cluster representation by perceiving symmetry information. The $GP(\cdot)_n$ of MGP[14] generates a continuous spectrum in energy space by linearly superimposing $n$ independent Gaussian distributions to represent the multiple vibrational modes of each cluster. The hypermeter $n = 10$ is set to exceed the maximum number of vibrational modes that can be generated by each cluster, with the unnecessary peaks automatically eliminated during training.

The spectrum prediction model was concretized by establishing the cluster encoder (CE) block (Figure 1c) and the spectrum predictor (SP) block (Figure 1d). The objective of CE block is to represent the multi-scale vibrational clusters and to encode the physical characteristics relevant to vibrational excitation. For each cluster, the cluster graph was built based on the standard atomic chemical environments $x_k^i$ via representing the interaction between surrounding atoms and bonds within a cluster. Additional features including chemical bond configuration $l_k^i$ and atom mass distribution $m_k^i$ were then incorporated via concatenation operation. The cluster graphs for all kinds of clusters with 3, 4, 6, and 9 atoms iterating over the atoms in primitive cell were concatenated to form a material graph (the cluster sizes were designed to encompass diverse configurations with symmetry and to reach desired prediction performance as shown in Figure S1, Supporting Information). The double counting issue for vibrational modes regarding clusters with various sizes and center atoms was automatically addressed through training the $P_k^i$ parameter of each cluster, wherein the cluster weights can be appropriately assigned as validated in Section S1 (Supporting Information). The cluster representation $c_{i,k}$ of the $i$th atom with $k$-1 neighbor atoms is composed of three feature vectors $clu_{i,k}^s$, $clu_{i,k}^c$, and $clu_{i,k}^p$, which are spatial transformation, chemical environment, and presence weight of each cluster capsule representation. The capsule transformer $f_{cap}$ was implemented to encode the geometric transformations into the cluster representation, which enables the identification of equivalent clusters stemming from


© 2024 The Author(s). Advanced Science published by Wiley-VCH GmbH





© 2024 The Author(s). Advanced Science published by Wiley-VCH GmbH

crystal symmetry and the appropriate evaluation of similarity between different clusters.[15] During the network training, the capsule transformer intelligently found the spatial transformations that maximize the similarity of chemical environments among different clusters and updated the chemical environments based on cluster similarity.

In order to predict diverse phonon spectra in continuous energy space, we constructed the SP block. The MGP was implemented on individual cluster graphs to generate the partial PDOS $y_{i,k}^{part}$ for the $i$-center atom cluster with $k$-1 atom via the Gaussian transformer module of the SP block. The partial PDOS of all clusters within a material was then merged, wherein the $y_{i,k}^{part}$ was multiplied by the related cluster weight $P_k^i$ (obtained from the $clu_{i,k}^p$ after training). The PDOS $y_{1\to t}$ of the target material was eventually obtained by linear superposition of all partial PDOS through series average pooling layer. The DTW function based on statistical mechanism was employed to evaluate the distinction between predicted and ground truth spectra by quantifying the similarity of the geometric features of two sequences.[13] The loss function was set as the summation of DTW discrepancy and mean absolute errors (MAE). This shape-aware loss function based on DTW discrepancy is capable of perceiving complex shift or dilatation of multiple and overlap peaks, and thus can precisely assess the deviation of peak distribution and peak shape in the predicted spectrum compared to its ground truth counterpart.

The performance of CSGN was examined by the database composed of 1523 crystal materials, which was obtained from the Materials Project (MP) database and predicted by DFT calculations.[16] The crystal database was divided into the training, validation, and testing datasets at a ratio of 8:1:1. The CSGN models were developed using the Python 3.8 and TensorFlow 2.15 framework.[17] Detailed algorithm architecture, data structure, mixed Gaussian processes, loss function selection, equivariance perception of cluster encoder block, and training process are documented in Sections S1 and S2 (Supporting Information).

### 2.2. Microscopic Vibrational Origin of Phonon Spectrum Learned by CSGN

Since the ML framework was developed based on the dynamical theory of crystal lattices, it would be expected that our CSGN model not only offers accurate prediction of PDOS spectrum, but also appropriately identifies the origin of each peak to vibrational clusters.[18] To verify such a premise, we compared our predicted PDOS spectra and the extracted cluster contributions to the ground truth data obtained by DFT simulations in previous studies (PDOS from the MP database, and vibration modes from the Phonon website). The comprehensive analyses of various materials demonstrated the efficacy of CSGN model, and the detailed results for the $BaPdF_4$ crystal will be discussed in this section (**Figure 2**). The $BaPdF_4$ crystal was selected as a representative material owing to its relatively complex spectrum, diverse vibrational clusters, and high crystal symmetry that facilitates the interpretation of structural-property mapping (Figure 2a).

The PDOS spectrum of $BaPdF_4$ crystal contains four main peaks (as denoted by arrows in Figure 2e), which represent the varying activation of vibration modes under different energy excitations. The predicted spectrum precisely matches the ground truth spectrum as shown in Figure 2d. To investigate the contributions of vibrational clusters on each peak, we extracted the predicted partial PDOS spectra related to the $k$-atom ($k$ = 3, 4, 6, and 9) clusters with different types of atoms serving as cluster centers (there is only one inequivalent atom in each element type). The normalized 2-sample Kolmogorov–Smirnov (2-KS) was then calculated between the partial PDOS spectra and the four spectral peaks (Figure 2b). The null hypothesis is that the two distributions are identical, and the alternative hypothesis is that they are not identical, which is justified by the low p-value. The detailed calculated 2-KS results are documented in Table S1 (Supporting Information). The results suggest that the P4 pattern with high excitation energy arises from the contributions of 3, 4-atom Pd-center clusters and 3-atom F-center clusters (Figure 2d,e), consistent with the truth phonon modes in the same energy scope obtained from the Phonon website (Figure 2c). In addition, the 9-atom vibrational clusters contribute only to the P1 and P2 patterns (Figure 2g), which are in accordance with the true energetic distribution of 9-atom phonon modes (Figure 2c) and can be explained by the excitation of cluster with a large mass in low energy range. In all, the decomposition of PDOS predicted by the CSGN model (Figure 2d–g) aligns well with the multi-scale cluster vibrational modes (Figure 2c).

The above spectrum decomposition also justifies the underlying mechanism of the CSGN model to perceive intrinsic correlations within a spectrum. In particular, the 3-atom F-center cluster possesses just one configuration in $BaPdF_4$ crystal, yet contributes to three PDOS peaks with distinct excitation energies (Figure 2d). This can be ascribed to the existence of three different vibration modes (P1, P3, and P4) regarding the same cluster as revealed by the ground truth data (Figure 2c). The projections of a single cluster on multiple peaks are also observed and examined for the Pd-center and Ba-center clusters (Figure 2f), which indicate that the complex energetic correlations have been perceived by the CSGN model. Such intrinsic correlation among different peaks learned from the training set can be naturally transferred to the predicted spectra of new materials containing similar vibrational clusters.

### 2.3. Global Trends of PDOS Spectrum Distribution Reflected by CSGN

The appropriate perceptions of the structural similarity and the consequential spectral correlation among diverse crystal materials are essential to the accurate prediction of PDOS spectrum. We comprehensively analyzed the global trends of the predicted

**Figure 1.** Development of the CSGN (cluster-based series graph networks) model to percept series correlation and to predict PDOS (phonon density of states). a) Schematic illustration of single-value property prediction. b) Schematic illustration of spectral pattern identification via multi-atom cluster representation, where $F_1$ to $F_5$ are frequency values for the main peaks induced by different clusters. c) Schematic framework of the cluster encoder block to generate cluster representation relevant to multi-scale vibrational excitation, where $x_k^i$, $l_k^i$, and $m_k^i$ are atom, bond, and mass vectors for the $i$-center atom, $c_{i,k}$ is the $i$-center atom cluster graph with $k$ scale. d) Schematic framework of the spectrum predictor block to project cluster representation on mixture Gaussian spectra.


© 2024 The Author(s). Advanced Science published by Wiley-VCH GmbH

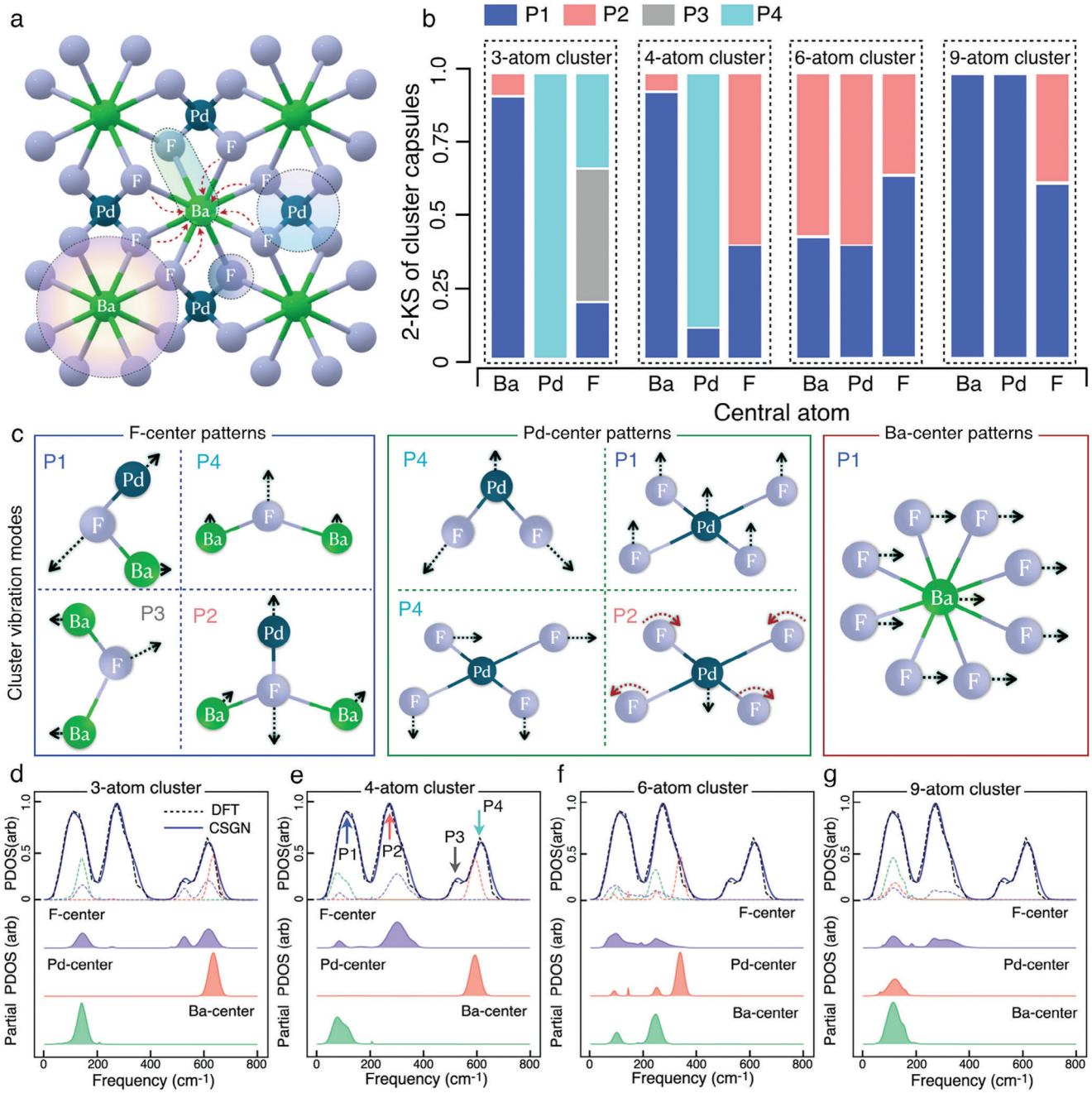



**Figure 2.** Microscopic interpretability analysis for PDOS prediction via the CSGN model. a) Schematic illustration of the multi-scale clusters with different central atoms, wherein the cluster domains and atomic interactions are denoted by shaded regions and red arrows, respectively. b) Projected contributions of different clusters to the four PDOS peaks (P1, P2, P3, and P4 are denoted by arrows in blue, orange, gray, and green respectively in panel e) in representative $BaPdF_4$ crystal from the CSGN model. The lengths of color bars display the normalized 2-sample Kolmogorov–Smirnov (2-KS) calculated between the partial spectra of specific clusters and the four spectral peaks (P1 to P4). c) Schematic illustration of representative phonon modes extracted from the four peaks (P1 to P4) of the $BaPdF_4$ phonon structures, with the arrows indicating the vibrational directions. d–g) Decomposition of the predicted PDOS spectrum for $BaPdF_4$ crystal on individual clusters with various atom numbers (3, 4, 6, and 9) and central atom types. The blue solid and black dashed lines present the predicted and ground truth spectra, respectively. The purple, red, and green shaded regions denote the contributions from clusters with F, Pd, and Ba centers.


© 2024 The Author(s). Advanced Science published by Wiley-VCH GmbH

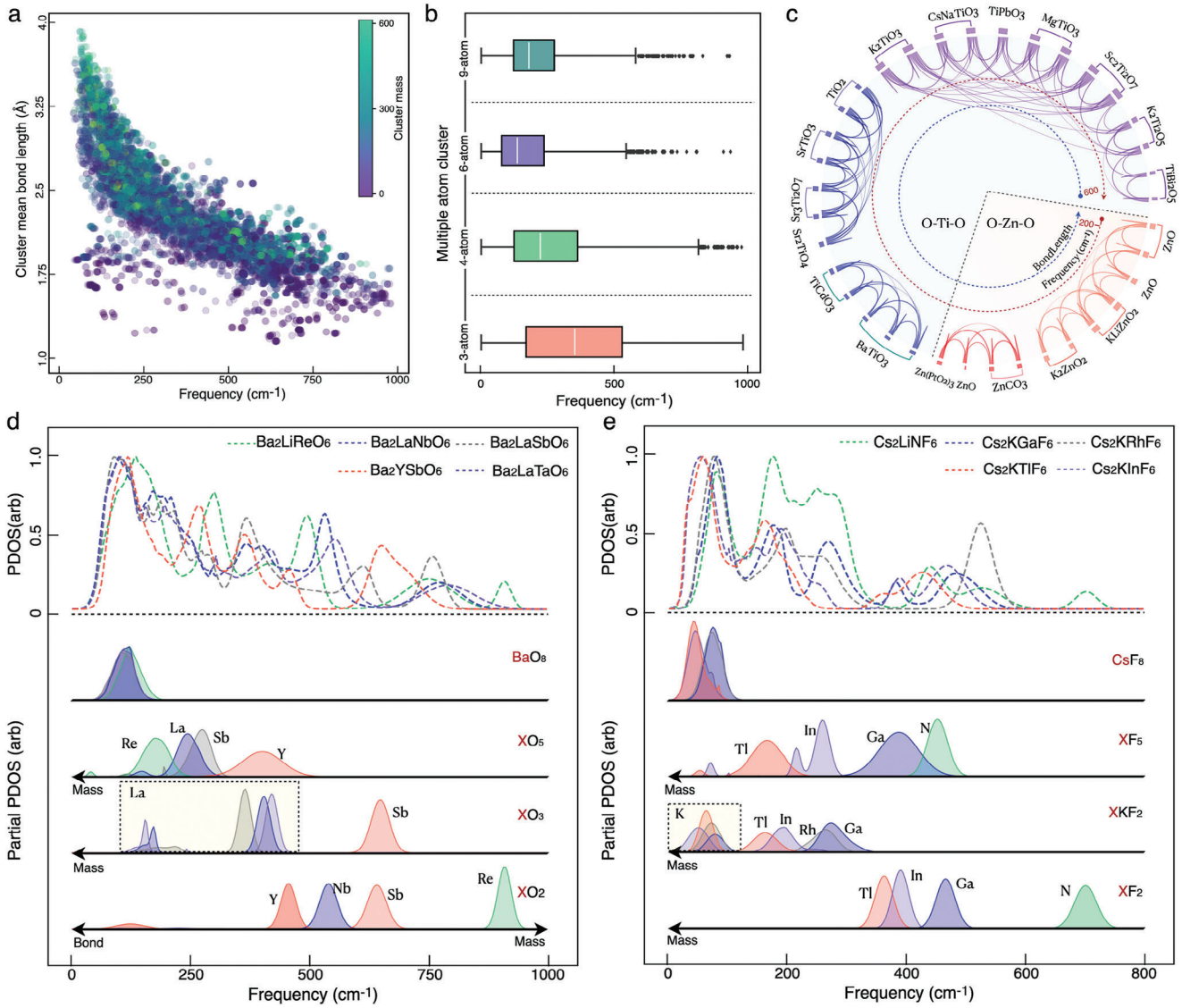



**Figure 3.** Global analysis of the physical trends in PDOS (phonon density of states) spectrum distribution predicted by the CSGN (cluster-based series graph networks) model. a) Correlation analysis among cluster mean bond length, cluster mass, and major peak frequency of all clusters in the phonon datasets, with the cluster mass indicated by colors of nodes. b) Dependence of major peak frequency on cluster scale via the box plot. The 3-, 4-, 6-, and 9-atom cluster groups are presented in dark green, purple, light green, and orange, respectively. The box plots show maximum and minimum values (whiskers), upper and lower quartiles (box boundaries), and median values (horizontal lines). c) Cluster-based correlation analysis of partial PDOS for 36 materials via the chord plot. Each node in the outermost ring represents a material containing either O–Ti–O or O–Zn–O clusters. The color and thickness of chords denote the similarity of partial PDOS generated by the same clusters in different materials. The inner circles in red and blue dashed lines illustrate the variation trends of vibrational frequency and bond length, respectively. d,e) Comparative analysis of phonon spectra in two representative material groups, including Ba-X-Y-O ($Ba_2LiReO_6$, $Ba_2LaNbO_6$, $Ba_2LaSbO_6$, $Ba_2YSbO_6$, and $Ba_2LaTaO_6$) and Cs-X-Y-F ($Cs_2LiNF_6$, $Cs_2KGaF_6$, $Cs_2KRhF_6$, $Cs_2KTlF_6$, and $Cs_2KInF_6$). The PDOS of each material is presented at the top with dashed line, and the partial PDOS of relevant clusters is shown at the bottom. The colors of dashed lines and shaded regions are consistent and denote the corresponding material. The arrows indicate the increasing directions of cluster mass or bond length.

PDOS spectra and the relevant material features to justify the efficacy of our CSGN model based on cluster representation. For individual clusters in all materials, the frequencies of the main peak in the predicted partial PDOS spectra were extracted and collectively presented in **Figure** 3a. The results imply that our cluster representation is capable of generating diverse PDOS spectra to continuously cover the entire energy range of relevant phonon modes, with the entire data documented in Section S4 (Supporting Information). The overall trend that the peak frequency increases with decreasing cluster mass and average bond length can be found in the peak distribution, which is consistent with the vibrational origin of phonon spectra. To quantify the strain effect in our model, we have conducted two series of additional calculations for predicting PDOS with material datasets in different strain conditions as shown in Figures S5 and S6 (Supporting Information). The test results verified the robustness of


© 2024 The Author(s). Advanced Science published by Wiley-VCH GmbH

CSGN model to strain effect and indicated that the CSGN model successfully learned the underlying structure-property relation. From the projection of peak distribution, the clusters with different sizes contribute to distinct energy ranges, with the average frequency declining with cluster scales (Figure 3b).

To unveil the impact of crucial material features (including element type, bond configuration, and condensed phase) on evaluating cluster similarity and predicting partial PDOS, we sampled the O–Ti–O and O–Zn–O clusters in 36 representative materials to calculate the correlation of partial PDOS via the chord diagram (Figure 3c). These two 3-atom clusters were chosen for their symmetric structure and rich embedding crystal materials. In the chord diagram, the nodes at the edge represent different clusters from distinct materials, and the connecting lines indicate the similarity between the partial PDOS of different cluster nodes. The Pearson coefficients are presented by the line thickness, and each group of clusters with coefficients over 0.8 is highlighted by the same color. The O–Ti–O and O–Zn–O clusters generate three and two groups of similar partial PDOS in the relatively high and low energy ranges respectively, wherein each group contains various types of crystal materials. The materials with analogous chemical components present in the same group, and those with identical stoichiometry but different phases possess the largest correlations, which confirm that the similarity of clusters among materials has been appropriately learned by the CSGN model. In addition, the clusters within the same group exhibit analogous bond lengths, while noticeable differences in bond lengths are observed between different groups, suggesting that the impact of varying bond lengths on partial PDOS sufficiently reflects those of the global chemical environment. The diverse partial PDOS associated with the simple O–Ti/Zn–O clusters can thus be attributed to the appropriate perception of delicate change in cluster microscopic structure and the subsequent mapping on Gaussian distributions.

The underlying structure-property relationship learned by the CSGN model was further probed by scrutinizing the partial PDOS of related clusters within two representative material groups (BaXYO and CsXYO). As validated in Figure S8 (Supporting Information), the predicted PDOS of all these materials accurately matches the ground truth spectra. The top panels of Figure 3d,e present the predicted total PDOS for each material, and the bottom panels illustrate the partial PDOS of the embedding clusters with similar chemical environments. Consistent with the harmonic vibration model in condensed matter physics, our results suggest that the vibrational frequency critically depends on the mass of motional atoms, bond lengths, and restoring force constant.

For clusters with identical structures but in different crystals (groups $BaO_8$, $CsF_8$, and $LaO_3$), the partial PDOS are similar, and the peak frequency slightly drops with increasing bond lengths. Given that restoring force constant is positively correlated to the second derivative of energy with respect to interatomic distance[38], the increasing bond length would diminish the restoring force constant and thus lead to lower vibrational frequency. Regarding the clusters containing different center atoms but with the same surrounding atoms, the partial PDOS spread in a much broader frequency range. In most cases (group $XO_5$, $XO_3$, $XF_5$, $XKF_2$, and $XF_2$), the peak frequency substantially decreases with increasing cluster mass (atom mass: Re> La> Sb> Y, Tl> In> Rh> Ga> N), as expected from the inverse proportion relation between mass and vibrational frequency. However, for the cluster group with relatively small variation in cluster mass, the impact of restoring force constant becomes more important, and the opposite trend with cluster mass could be observed. Within the group $XO_2$, the PDOS peaks of the $ReO_2$ and $SbO_2$ clusters with ionic bonds are distributed in the relatively higher frequency range compared to the $NbO_2$ and $YO_2$ clusters with covalence bonds. The frequency differences within each bond type ($NbO_2$> $YO_2$, $SbO_2$> $ReO_2$) can be further explained by the varying restoring force constants as reflected in their bond energies (Sb-O:200–300 kJ $mol^{-1}$. Re-O:300–400 kJ $mol^{-1}$, Nb-O:600–800 kJ $mol^{-1}$, Y-O: 300–400 kJ $mol^{-1}$). For the cluster groups possessing multiple PDOS peaks, distinct trends can be observed for different peaks. While the second PDOS peak within the group $XKF_2$ noticeably shifts with cluster mass, the frequency of the first peak remains almost unchanged, probably because the center atom is static in relevant vibration mode. Detailed calculated results, physical, and chemical information of the selected materials are documented in Section S5 (Supporting Information).

### 2.4. Performance Analysis of CSGN

The performance of our CSGN model in predicting PDOS was examined through the comparison with SOTA models including Mat2Spec and E3NN.[6,7] The CSGN model possesses a MAE of 0.061 for the normalized testing set, which is 21.8% and 41.9% lower than those of the Mat2Spec (0.078) and E3NN (0.105) models by testing the same datasets with lower computing cost. More importantly, the prediction of complicated PDOS spectra with multiple and overlap peaks remains a challenge for existing algorithms, but can be generally achieved by the CSGN model.

Six representative materials ($RbSnCl_3$, $K_2S_3$, $K_2NaInF_6$, $CuP_2$, $NbBO_4$, and $Cs_2KGaF_6$) within the testing set were chosen to illustrate the prediction performance (**Figure** 4a), wherein the first five materials are consistent with the selections in the publications of the Mat2Spec and E3NN models. For relatively simple PDOS spectra shown in the first row, both of the Mat2Spec (green line) and CSGN (red line) models appropriately predicted the frequencies and shapes of the main peaks. The satellite peaks overlapping with the main peaks in $K_2S_3$ and $K_2NaInF_6$ can be adequately distinguished by the CSGN model yet not by the Mat2Spec model. For complicated PDOS spectra with multiple peaks spreading in broad energy range as shown in the second row ($CuP_2$, $NbBO_4$, and $Cs_2KGaF_6$), the Mat2Spec model captured the overall distributions, yet failed to predict the peak numbers and peak positions. The error of the Mat2Spec model is especially large in the high-frequency range as illustrated in the spectrum of $NbBO_4$, probably due to the insufficient size of dataset. In contrast, the CSGN model can achieve transferable prediction to generate high-quality PDOS spectra with accurate peak distribution and peak strength in the full energy range, as validated by testing materials without similar components in the training set. More comparison results can be found in Figure S9 (Supporting Information).

The total number of phonon states defined as the integral of PDOS spectrum is an intrinsic material feature, which


© 2024 The Author(s). Advanced Science published by Wiley-VCH GmbH

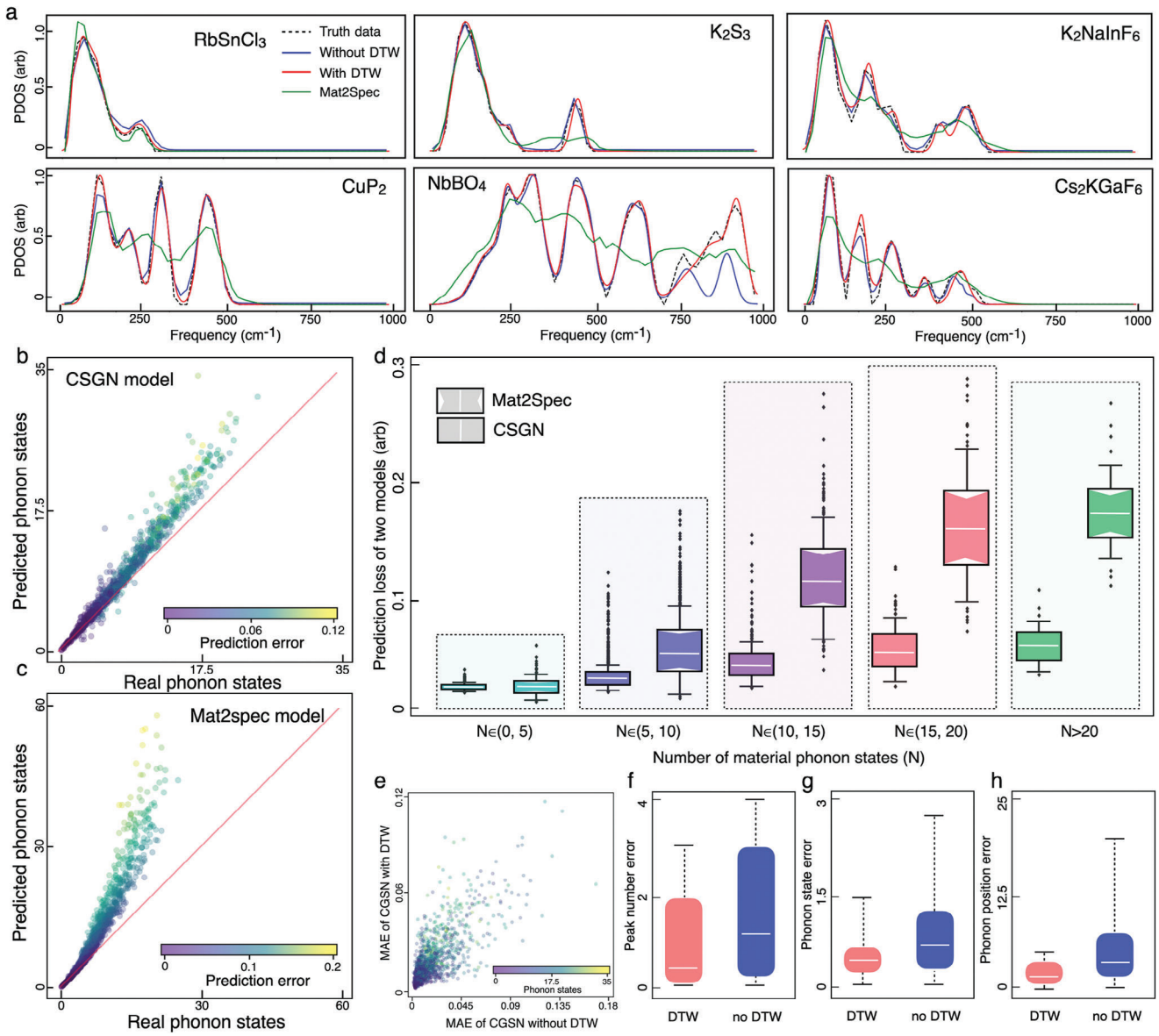



**Figure 4.** Prediction performance of PDOS (phonon density of states) spectrum via the CSGN (cluster-based series graph networks) model. a) Comparison of predictive PDOS spectra with different models for six representative materials. The spectra predicted by the CSGN model without DTW (dynamic time warping) kernel, the CSGN model with DTW, and the Mat2Spec model are shown in the blue, red, and green solid lines respectively, while the ground truth spectra are presented in the black dashed lines. b,c) Prediction performances of phonon state numbers by the CSGN and Mat2Spec models, with the mass of primitive cell illustrated by the color of node. d) Comparison of prediction loss distributions between the CSGN and Mat2Spec models for material systems with different phonon state numbers. The box plots show maximum and minimum values (whiskers), upper and lower quartiles (box boundaries), and median values (horizontal lines). e) Comparison of predictive error distributions between the CSGN models with and without DTW kernel, with the phonon state number illustrated by the color of node. f–h)The effectiveness analysis of DTW kernel for detecting prior physical features of phonon spectrum, including peak number, phonon states, and phonon position.

predictive error is technically related to the accumulative MAE of PDOS spectrum. Materials with larger phonon state numbers are likely to possess more complicated crystal structure, larger diversity of PDOS pattern, and broader range of vibration frequency. The global performances of the CSGN and Mat2Spec models in all materials were measured by comparing the predicted numbers of phonon states to the ground truth values. It turns out that the predictive error of Mat2Spec model gradually increases with the number of phonon states (Figure 4b), while the accurate prediction of phonon state number by the CSGN model is maintained in the entire material space (Figure 4c). Quantitative comparison of model performance was conducted on material groups with different phonon state numbers. From the box plots of MAE distributions (Figure 4d), the prediction performance of CSGN model exceeds the Mat2Spec model in material systems regardless of spectrum complexity, and the improvement become significant for systems with relatively large phonon state number (N >10).


© 2024 The Author(s). Advanced Science published by Wiley-VCH GmbH

The desired performance of CSGN model can be attribute to the improvement of spectrum representation (including cluster graph and Gaussian generator) and loss function. The impacts of these factors were analyzed by comparing the predictive errors of the CSGN model with and without the DTW loss function. The MAE of PDOS prediction by the CSGN model without the DTW loss function is mainly distributed in the range of (0, 0.066), which is lower than the (0, 0.078) of Mat2Spec model and higher than the (0, 0.061) of CSGN model with the DTW loss function (Figure 4e). The predictive errors of peak number (Figure 4f), phonon states (Figure 4g), and peak frequency (Figure 4h) were reduced by 45.2%, 58.7%, and 62.1% via implementing the DTW loss function. As illustrated in Figure 4a, despite that the CSGN model is capable of accurately predicting the peak positions in most materials, the DTW loss function plays an important role in refining the peak shapes especially for complicated spectra. The above results suggest that the collaboration of advanced spectrum representation and loss function is crucial for the CSGN model.

## 3. Discussion

The above comprehensive analyses demonstrated the efficacy of our CSGN model and enabled us to interpret its underlying mechanism. The key idea is to develop a transparent ML model based on the dynamical theory of crystal lattices for accurately encoding the structural/spectral diversity and similarity,[18] as well as naturally perceiving the structure-property relation and intrinsic spectral correlation, which is accomplished as follows: 1) We composed a mass decorated multiple atomic cluster representation with 40,496 cluster types to describe the rich vibration modes originating from clusters with different sizes and shapes. Such representation provides a more flexible and precise feature extraction of crystal materials compared to the traditional crystal graph comprised of spherical atomic environments with a uniform fixed radius. 2) We designed a spectral prediction method including the mapping of cluster graphs on mixture Gaussian functions and the linear superposition of predicted partial PDOS toward the predictive spectrum. This is consistent with the physical scenario that the PDOS is a superposition of all phonon contributions, and the peak frequencies are determined by the relevant vibration modes. As verified by the prediction results, the CSGN model appropriately learned the mapping from clusters to PDOS peaks and the frequency shift with changing chemical environment. The spectral correlations within a spectrum arising from the different vibration modes of the same cluster and the correlation across the materials arising from the similarity between vibrational clusters are thus adequately embedded in our predictive spectrum. The decomposition of PDOS spectrum into the contributions of atomic clusters also effectively increases the data size and address the insufficient database of PDOS spectrum. 3) We compiled a DTW-based loss function to perceive the shift or dilatation of peaks and thus to appropriately evaluate the spectral differences. This overcomes the drawback of traditional Euclidean-based function (MAE, MSE, etc.) and simple statistical (KL, WD divergence, etc.) loss functions that underestimate the prediction error due to the error cancellation for any spectrum variation with preserved integral value.

In summary, we have developed the CSGN model consisted of the cluster encoder and spectrum predictor blocks to accurately predict PDOS spectrum of crystal materials. The predictive results demonstrate that the CSGN model has successfully conducted the mapping from vibrational clusters to PDOS patterns, as well as the transformation of vibrational frequency in response to the chemical environment modification. The rational structure-property relation learned by the CSGN model guarantees the appropriate reflection of series correlations in the predictive spectra, and thus enables the accurate prediction of complicated spectra with multiple or overlap peaks. The high performance and interpretability of the CSGN model can be attributed to both the establishment of cluster-spectrum representation based on crystal lattices dynamical theory and the subsequent identification of pattern similarity with DTW loss function. This work not only provides a transferable ML model to improve the prediction of spectral properties but also reveals the opportunity to design advanced ML framework based on the fundamental physical mechanisms.

**Table 1.** Algorithm 1 Cluster encoder block.

**Input**: $x_k^{atom} = (x_k^1, \ldots, x_k^i) \in \mathbb{R}^{i\times k\times 128}$, $x_k^{bond} = (l_k^1, \ldots, l_k^i) \in \mathbb{R}^{i\times j\times 128}$, $x_k^{mass} = (m_k^1, \ldots, m_k^i) \in \mathbb{R}^{i\times k}$
**Prior modules**: $f_{cluster} = (f_a(\cdot), f_b^{atom}(), f_b^{cluster}())$, $f_{mass}(\cdot), f_{cap}(\cdot)$,
**Output**: $clu_{i,k}^s, clu_{i,k}^c, clu_{i,k}^p \in \mathbb{R}^{i\times 128}, \mathbb{R}^{i\times 128}, \mathbb{R}^{i\times 1}$
**For** $k$ iterations **do**:
**For** $i$ iterations **do**:
$\boldsymbol{c}_{i,k}^{atom} \leftarrow f_a(x_k^i)$ atom composition for each cluster
$\boldsymbol{b}_{i,k} \leftarrow f_b^{atom}(l_k^i)$ bond configuration for each cluster
$\boldsymbol{c}_{i,k}^{mass} \leftarrow f_{mass}(m_k^i)$ mass distribution for each cluster
$\boldsymbol{c}_k^A \leftarrow Softplus(c_{i,k}^{atom})$
$\boldsymbol{c}_k^B \leftarrow f_b^{cluster}(b_{i,k})$
$\boldsymbol{c}_k^M \leftarrow Softplus(c_{i,k}^{mass})$
$\boldsymbol{c}_k^{total} \leftarrow c_k^A \oplus c_k^B \oplus c_k^M$
$\boldsymbol{c}_{i,k} = \boldsymbol{clu}_{i,k}^s, \boldsymbol{clu}_{i,k}^c, \boldsymbol{clu}_{i,k}^p \leftarrow f_{cap}(\boldsymbol{c}_k^{total})$
**End for**
**End for**
**Return** $\boldsymbol{clu}_{i,k}^s = (clu_1^s, \ldots, clu_i^s)_k$
$\boldsymbol{clu}_{i,k}^c = (clu_1^c, \ldots, clu_i^c)_k$
$\boldsymbol{clu}_{i,k}^p = (clu_1^p, \ldots, clu_i^p)_k$

## 4. Experimental Section

*Framework of the CSGN*: In this work, accurate predictions of materials' spectral properties were achieved, benefited from the identification of spectrum patterns and series correlations via the CSGN model. This was accomplished through the unified training of two modules including the cluster encoder (CE) block and the spectrum predictor (SP) block.

The CE block perceived the input data of atoms and bonds to construct the multi-scales cluster representation of crystal materials (**Table 1**), through the hierarchical aggregation of chemical environments and the encoding of symmetric transformations. The CE block was composed of four modules including the atomic cluster module, bond distribution module, cluster mass module, and capsule-based symmetry module. The first three modules received and extracted prior features to encode the multi-scale cluster environments, and the last module transmits the full symmetry transformations of space groups to update the cluster representation.

The crystal information files (CIFs) of the MP datasets are filtered by leveraging the Pymatgen[35] interface to construct the input dataset, with the detailed information of original dataset documented in Section S1 (Supporting Information). The input dataset includes the atom


© 2024 The Author(s). Advanced Science published by Wiley-VCH GmbH

**Table 2.** Algorithm 2 Spectrum predictor block.

**Input**: cluster capsule representation
$clu^{s}_{i,k}, clu^{c}_{i,k}, clu^{p}_{i,k} \in \mathbb{R}^{1\times128}, \mathbb{R}^{1\times128}, \mathbb{R}^{1\times1}$
**Mixture Gaussian transformer**: $f_u(\cdot)$, $\Theta_g(\cdot)$, $\mathcal{N}(\cdot), f_{mol}(\cdot), f_\phi(\cdot)$
**Output**: $\gamma_m = (\gamma_1, \ldots, \gamma_m) \in \mathbb{R}^{m\times51}$
**For** *k* iterations **do**:
**For** *i* iterations **do**:
$\boldsymbol{clu^{u}_{i,k}} \leftarrow f_u(clu^{s}_{i,k}, clu^{c}_{i,k})$
$\boldsymbol{clu^{p}_{i,k}} \leftarrow Softplus(clu^{p}_{1}, \ldots, clu^{p}_{i})_k$
$\boldsymbol{\mu_{i,k}, \sigma_{i,k}, \alpha_{i,k}} \leftarrow \Theta_g(clu^{u}_{i,k}, n)$
$\boldsymbol{\gamma_{i,k}} \leftarrow \mathcal{N}(\mu_{i,k}, \sigma_{i,k}, \alpha_{i,k})$
$\boldsymbol{\gamma^{u}_{i,k}} \leftarrow clu^{p}_{i,k} \otimes \gamma_{i,k}$
**End for**
$\boldsymbol{\gamma^{u}_{k}} \leftarrow \sum_I(\gamma^{u}_{i,k})$
**End for**
$\boldsymbol{\Phi_k} \leftarrow f_\phi(\gamma^{u}_{k})$
$\boldsymbol{\gamma_m} \leftarrow \sum_k(\gamma^{u}_{k} \otimes \Phi_k)$
**Return** $\boldsymbol{\gamma_m} = (\gamma_1, \ldots, \gamma_m)$

composition $x^{atom}_{k} = (x^{1}_{k}, \ldots, x^{i}_{k})$, bond connection $x^{bond}_{k} = (l^{1}_{k}, \ldots, l^{i}_{k})$, and mass distribution $x^{mass}_{k} = (m^{1}_{k}, \ldots, m^{i}_{k})$ of all clusters with different center atoms $i \in N$ and different scales ($k$ = 3, 4, 6, and 9). For each cluster, the atom data $x^{i}_{k}$ reflects the atom number and types through the atomic embedding vectors with ($k \times 128$) dimensions, the bond data reflects the bond number, connected atom types, and bond lengths through the bond embedding vectors (with $j \times 128$ dimensions expanded by the Gaussian basis, $j$ was the bond number of cluster), and the mass data reflects the atomic mass distribution by a vector with ($k \times 1$) dimensions.

The chemical environment of each cluster was built parallelly. The atom module $f_a(\cdot)$ generates the matrix $c^{atom}_{i,k}$ with ($k \times 192$) dimensions from $x^{i}_{k}$ through the multi-layer perceptron (MLP), which encodes the atomic interactions within each cluster and modulates the similarity between different clusters. The mass module $f_{mass}(\cdot)$ builds the matrix for each cluster to highlight the dependence of vibration modes on atom mass via the MLP-based model, with the output $c^{mass}_{i,k}$ of ($k \times 128$) dimensions. The atom and mass matrices were then modulated with the *Softplus* activation function for enhancing the smoothness of material representation. The bond module consists of atom-level $f^{atom}_{b}(\cdot)$ and cluster-level $f^{cluster}_{b}(\cdot)$ sub-modules, which were both MLP-based models with outputs $b_{i,k}$ of ($j \times 192$) dimensions and $c^{B}_{k}$ of ($k \times 192$) dimensions. The former encodes the environment vector for each bond via receiving bond index and connected atom information, and the latter projects the information of all bonds on each atom to generate the bond matrix. The cluster chemical environment $c^{total}_{i,k}$ was obtained by concatenating the atom, bond, and mass matrices.

$$c^{total}_{k} = Softplus\left(f_a\left(x^{i}_{k}\right)\right) \oplus \left(f^{cluster}_{b}\left(f^{atom}_{b}\left(l^{i}_{k}\right)\right)\right) \oplus \left(Softplus\left(f_{mass}\left(m^{i}_{k}\right)\right)\right) \quad (2)$$

The cluster chemical environment was then transformed into the cluster capsule representation $c_{i,k} = \{clu^{s}_{i,k}, clu^{c}_{i,k}, clu^{p}_{i,k}\}$ through the capsule encoder $f_{cap}(\cdot)$ following the standard procedure developed in the SEN model.[15] Such transformation refines the evaluation of cluster similarity by recognizing the preservation of cluster properties under rigid geometric transformations including rotation and reflection. The geometric transformation operator $clu^{s}_{i,k}$ and the convoluted cluster chemical environment $clu^{c}_{i,k}$ are both vectors with 128 dimensions. The presence value $clu^{p}_{i,k}$ was a variable to measure the weight of corresponding cluster.

The SP block was designed to predict diverse phonon spectrums for different material systems by receiving and decoding cluster capsule representation, as illustrated in **Table 2**. The cluster partial PDOS $\gamma_{i,k}$ of *i*th atom with *k*-scale was predicted via the MGP model, which represents the activated phonon state patterns of the target cluster under continuous energy space. $f_u(\cdot)$ was a symmetry module that propagates the geometric transformations into cluster capsules by receiving symmetry operators $clu^{s}_{i,k}$, chemical environment $clu^{c}_{i,k}$, mainly including scaling, translation, rotation, inversion reflection, and mirroring reflection transformations. With the *Softplus* activation function, capsule weights of clusters $clu^{p}_{i,k}$.were obtained $\Theta_g()$ was a parametric function to calculate parameters $\mathcal{N}(\mu_{i,k}, \sigma_{i,k}, \alpha_{i,k})$ of MGP via receiving updated cluster capsules, wherein $\mu_{i,k}$, $\sigma_{i,k}$, $\alpha_{i,k}$ were respectively the internal threshold, the width of the Gaussian process, and weight parameters, and $\mathcal{N}(\cdot)$ was the MLP mathematical form. The PDOS $\gamma^{u}_{i,k}$ of each atom-center cluster successfully obtained by sampling from the corresponding MGP function with 256 dimensions. $f_{mol}(\cdot)$ was an atom-to-molecular transformer model that outputs crystal-level PDOS $\gamma^{u}_{k}$ of material under a fixed scale according to space structures with ($N \times 51$) dimensions. $f_\phi(\cdot)$ was an MLP-based model to calculate the contribution probability of crystal-level PDOS ($\Phi_k \in (0, 1)$). The total PDOS $\gamma_m$ of material can be eventually obtained by linear superposition of all partial PDOS through series average pooling.

## Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

## Acknowledgements

The authors acknowledge the support from the National Natural Science Foundation of China (52072417, 11832019, 52473244, 12404261), the NSFC original exploration project (12150001), the Guangdong Science & Technology Project (2019QN01C113), the Young Scientists Fund of the National Natural Science Foundation of China (12404261), the Natural Science Foundation of Guangdong Province (2018B030306036), the Project of Nuclear Power Technology Innovation Center of Science Technology and Industry for National Defense (HDLCXZX-2021-HD-035), and the Guangdong International Science and Technology Cooperation Program (2020A0505020005). The calculations reported were performed on resources provided by the Guangdong Provincial Key Laboratory of Magnetoelectric Physics and Devices (No. 2022B1212010008) and the National Supercomputer Center in Guangzhou.

## Conflict of Interest

The authors declare no conflict of interest.

## Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

## Keywords

cluster capsule representation, machine learning, predictive model, series correlation, spectral property

Received: June 8, 2024
Revised: September 30, 2024
Published online: October 18, 2024

[1] a) A. Boes, L. Chang, C. Langrock, M. Yu, M. Zhang, Q. Lin, M. Loncar, M. Fejer, J. Bowers, A. Mitchell, *Science* **2023**, *40*, 379; b) M. Kus´,


© 2024 The Author(s). Advanced Science published by Wiley-VCH GmbH

*Phys. Rev. Lett.* **1985**, *54*, 1343; c) X. Qi, H. Katsura, A. W. W. Ludwig, *Phys. Rev. Lett.* **2012**, *108*, 1; d) D. A. Long, M. J. Cich, C. Mathurin, A. T. Heiniger, G. C. Mathews, A. Frymire, G. B. Rieker, *Nat. Photonics* **2023**, *18*, 127.

[2] a) M. W. Schauerte, N. Pozhydaieva, J. Grawenhoff, L. M. Welp, I. Silbern, A. Wulf, F. A. Billau, T. Glatter, H. Urlaub, A. Jäschke, K. Höfer, *Nature* **2023**, *620*, 1054; b) U. J. Kim, S. Lee, H. Kim, Y. Roh, S. Han, H. Kim, Y. Park, S. Kim, M. J. Chung, H. Son, H. Choo, *Nat. Commun.* **2023**, *14*, 5262; c) Y. Li, O. Fiehn, *Nat. Methods* **2023**, *20*, 1475; d) Y. Li, T. Kind, J. Folz, A. Vaniya, S. S. Mehta, O. Fiehn, *Nat. Methods* **2021**, *18*, 1524.

[3] a) S. Xing, S. Shen, B. Xu, X. Li, T. Huan, *Nat. Methods* **2023**, *20*, 881; b) Y. Cui, K. Chen, L. Zhang, H. Wang, L. Bai, D. Elliston, W. Ren, *J. Phys. Chem. Lett.* **2023**, *14*, 7924; c) L. Cao, M. Guler, A. Tagirdzhanov, Y. Y. Lee, A. Gurevich, H. Mohimani, *Nat. Commun.* **2021**, *12*, 3718; d) V. Fung, G. Hu, P. Ganesh, B. G. Sumpter, *Nat. Commun.* **2021**, *12*, 88.

[4] a) K. Ghosh, A. Stuke, M. Todorović, P. B. Jørgensen, M. N. Schmidt, A. Vehtari, P. Rinke, *Adv. Sci.* **2019**, *6*, 1801367; b) A. Chandrasekaran, D. Kamal, R. Batra, C. Kim, L. Chen, R. Ramprasad, *npj Comput. Mater.* **2019**, *5*, 22; c) K. Venkatraman, B. D. A. Levin, K. March, P. Rez, P. A. Crozier, *Nat. Phys.* **2019**, *15*, 1237.

[5] a) J. Cox, *Nat. Biotechnol.* **2023**, *41*, 33; b) M. A. Stravs, K. Dührkop, S. Böcker, N. Zamboni, *Nat. Methods* **2022**, *19*, 865; c) S. Goldman, J. Wohlwend, M. Stražar, G. Haroush, R. J. Xavier, C. W. Coley, *Nat. Mach. Intell.* **2023**, *5*, 965.

[6] Z. Chen, N. Andrejevic, T. Smidt, Z. Ding, Q. Xu, Y. T. Chi, Q. T. Nguyen, A. Alatas, J. Kong, M. Li, *Adv. Sci.* **2021**, *8*, 2004214.

[7] S. Kong, F. Ricci, D. Guevarra, J. B. Neaton, C. P. Gomes, J. M. Gregoire, *Nat. Commun.* **2022**, *13*, 949.

[8] V. Fung, P. Ganesh, B. G. Sumpter, *Chem. Mater.* **2022**, *34*, 4848.

[9] Y. Zhou, W. Guan, C. Zhao, H. Hu, Z. He, X. Zou, X. Gong, *J. Clean. Prod.* **2023**, *418*, 138037.

[10] a) Q. D. Gibsom, T. Zhao, L. M. Daniels, H. C. Walker, R. Daou, S. Hébert, M. Zanella, M. S. Dyer, J. B. Claridge, B. Slater, M. W. Gaultois, F. Corà, J. Alaria, M. J. Rosseinsky, *Science* **2021**, *373*, 1017; b) M. Xu, D. L. Bao, A. Li, M. Gao, D. Meng, A. Li, S. Du, G. Su, S. J. Pennycook, S. T. Pantelides, W. Zhou, *Nat. Mater.* **2023**, *22*, 612.

[11] R. Senga, Y. C. Lin, S. Morishita, R. Kato, T. Yamada, M. Hasegawa, K. Suenaga, *Nature* **2022**, *603*, 68.

[12] a) T. Xie, J. C. Grossman, *Phys. Rev. Lett.* **2018**, *120*, 145301; b) H. R. Banjade, S. Hauri, S. Zhang, F. Ricci, W. Gong, G. Hautier, S. Vucetic, Q. Yan, *Sci. Adv.* **2021**, *7*, eabf1754; c) K. T. Schütt, M. Gastegger, A. Tkatchenko, K. R. Müller, R. J. Maurer, *Nat. Commun.* **2019**, *10*, 5024.

[13] a) M. Cuturi, M. Blondel, presented at 34th Int. Conf. Machine Learning, Sydney, Australia, August, **2017**, *2*, pp. 1483–1505; b) Z. Yue, Y. Wang, J. Duan, T. Yang, C. Huang, Y. Tong, B. Xu, presented at Proc. of the AAAI Conf on Artificial Intelligence, Vancouver,Canada, February, **2022**, *36*, pp. 8980–8987; c) X. Zheng, N. Xu, L. Trinh, D. Wu, T. Huang, S. Sivaranjani, Y. Liu, L. Xie, *Sci. Data* **2022**, *9*, 359; d) V. L. Guen, N. Thome, *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 1.

[14] a) A. Grisafi, D. M. Wilkins, G. Csányi, M. Ceriotti, *Phys. Rev. Lett.* **2018**, *120*, 36002; b) A. Glielmo, B. E. Husic, A. Rodriguez, C. Clementi, F. Noe, A. Laio, *Chem. Rev.* **2021**, *121*, 9722.

[15] C. Liang, Y. Rouzhahong, C. Ye, C. Li, B. Wang, H. Li, *Nat. Commun.* **2023**, *14*, 5198.

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

[17] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. A. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, presented at *12th USENIX Symp. on Operating Systems Design and Implementation, OSDI*, Savannah, GA, USA, November, **2016**, pp. 265–283.

[18] M. Born, K. Huang, M. Lax, *Am. J. Phys.* **1955**, *23*, 474.


© 2024 The Author(s). Advanced Science published by Wiley-VCH GmbH