nature machine intelligence

Supplementary information

https://doi.org/10.1038/s42256-026-01326-5

Efficient equivariant framework for universal charge density prediction

In the format provided by the authors and unedited

Contents

Supplementary Notes **2**
Simplified Equivariant Message for Scalar Fields 2

Supplementary Discussions **5**
1. Performance on NMC Dataset 5
2. Construction of Spherical Bessel Radial Basis Functions for Ablation Study 5
3. Validate EdenGNN on LCAO-basis DFT 5
4. Workflow to Generate the Superposition of Atomic Charge Density 6
5. Validate the Hyperparameter Settings of EdenGNN 6
6. The Optimization Strategies for Training Models 7

Supplementary Tables **8**

Supplementary Figures **9**

References **38**

Supplementary Notes

Simplified Equivariant Message for Scalar Fields

In a graph neural network (GNN), a crystal structure is represented by a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X}, \mathbf{C})$, where $\mathcal{V}$ and $\mathcal{E}$ denote the sets of node (atom) and edge (bond) features, respectively, while $\mathbf{X}$ and $\mathbf{C}$ represent atom positions and cell parameters. The message-passing in the equivariant GNN—essentially a graph convolution by aggregating the high-dimensional features $V_{c'l_im_i}$ from neighbor atoms—employs spherical tensor kernels composed of trainable radial filters $\mathcal{F}_{cc'}^{l_ol_il_f}(r)$ and spherical harmonics $Y_{m_f}^{l_f}(\hat{\mathbf{r}})$ [1, 2]:

$$M_{cl_om_o}^{a} = \sum_{b\in\mathcal{N}(a)} \sum_{l_il_fm_im_fc'} C_{l_im_il_fm_f}^{l_om_o} V_{c'l_im_i}^{b} \left[\mathcal{F}_{cc'}^{l_ol_il_f}(r_{ab}) Y_{m_f}^{l_f}(\hat{\mathbf{r}}_{ab})\right] \tag{1}$$

where $C_{l_im_il_fm_f}^{l_om_o}$ are Clebsch-Gordan (CG) coefficients which transform the direct product of IRREPs into direct sums, preventing the exponential growth of feature dimensions after successive convolution layers. Parity equivariance is enforced by restricting the allowed product channels $l_i \otimes l_f \to l_o$ in Eq. (1) to those satisfying $p_ip_f = p_o$.

Predicting the PS charge density $\tilde{n}(\mathbf{r})$ is challenging due to the overwhelming number of grid points. By exploiting the atomic structure, $\tilde{n}(\mathbf{r})$ has a natural formulation via spherical harmonics expansion with basis functions centered at atoms $\{a\}$:

$$\tilde{n}(\mathbf{r}) = \sum_a \sum_{clm} h_{clm}^{a} R_{cl}^{a}(|\mathbf{r}-\mathbf{r}_a|) Y_m^l(\widehat{|\mathbf{r}-\mathbf{r}_a|}) \tag{2}$$

where spherical harmonics $Y_m^l(\hat{\mathbf{r}})$ and radial basis $R_{nl}(r)$ together form the basis. Using Eq. (2), the basis-based methods indirectly fit the expansion coefficients $h_{nlm}(i)$. For convenience the contracted Gaussian basis is typically employed [3], at the cost of accuracy due to their limited completeness. The grid-based methods, on the other hand, directly predict the charge density on grids using message-passing from atoms to grids [4–6]. The efficiency-accuracy trade-off between the two methods can be understood from Eq. (1). Basis-based methods perform a single atom-to-grid message-passing step with parametric analytic radial basis functions, where the tensor product scales as $\mathcal{O}(L^2)$ owing to the sparsity of CG coefficients coupling to $l_o = 0$. In contrast, grid-based methods employ multiple atom-to-grid message-passing layers with learned radial basis functions, resulting in a prohibitive $\mathcal{O}(L^6)$ scaling cost of tensor product for the hundreds of thousands of edges per atom.

The state-of-the-art grid-based methods incorporate equivariant tensor products for atom-to-grid message passing. A single layer of equivariant tensor product for scalar property reduces to a simple form of tensor contraction of $\mathcal{O}(L^2)$ scaling. While in [6] a two-layer atom-to-grid equivariant message passing is employed at the cost of $\mathcal{O}(L^6)$ scaling, deeper layers have not been reported yet in the literature due to their prohibitive computational time. In the following, we prove that the two-layer case can be reformulated to $\mathcal{O}(L^2)$ scaling, which is the theoretical minimum complexity of spherical tensor operations outputing scalars. For notational simplicity, the constant coefficients are often absorbed into the radial filters throughout the derivation. The derivation is inspired by the linearity of high order equivariant operations and the fact that the charge density is a scalar field. We start from the first equivariant message passing layer $\mathbf{p}_{cl_om_o}^{1}(\mathbf{r})$:

$$\mathbf{p}_{cl_om_o}^{1}(\mathbf{r}) = \mathrm{Gate}_{cl_o}(\mathbf{r}) \left[\sum_{l_il_fm_im_f} \sum_{c'} C_{l_im_il_fm_f}^{l_om_o} \mathcal{F}_{cc'}^{l_ol_il_f}(r) \mathbf{h}_{c'l_im_i} Y_{m_f}^{l_f}(\hat{\mathbf{r}}) + \Phi_{cl_om_o}\right] \tag{3}$$

$$\begin{aligned} \mathrm{Gate}_{cl}(\mathbf{r}) &= \phi\left[\sum_{l_il_fm_im_f} \sum_{c'} C_{l_im_il_fm_f}^{00} \tilde{\mathcal{F}}_{\substack{0l_il_f\\cl,c'}}^{1}(r) \mathbf{h}_{c'l_im_i} Y_{m_f}^{l_f}(\hat{\mathbf{r}}) + \Phi_{cl}^{s}\right] \\ &= \phi\left[\sum_{c'l'm'} \tilde{\mathcal{F}}_{cl,c'l'}^{1}(r) \mathbf{h}_{c'l'm'} Y_{-m'}^{l'}(\hat{\mathbf{r}}) + \Phi_{cl}^{s}\right] \end{aligned} \tag{4}$$

where $\text{Gate}_{cl}(\mathbf{r})$ is the non-linearity gate function acting on different spherical tensor channels, the input of which is the scalar part of a tensor product plus a scalar node attribute term Φ^s_{cl}. When coupling to scalar, the conservation of angular momentum requires that the CG coefficients $C^{00}_{l_1m_1l_2m_2}$ be non-zero only for $l_1 = l_2, m_1 + m_2 = 0$, hence $O(L^6)$ tensor product reduces to $O(L^2)$ scaling tensor contraction. The basis-method is a special case of equivariant message passing with a single layer equivariant message passing and no gate non-linearity.

The second equivariant message layer starts from the result of the first layer $\mathbf{p}^1_{cl_om_o}(\mathbf{r})$ and couples it with spherical harmonics to scalar:

$$\begin{aligned}
p^2_c(\mathbf{r}) &= \sum_{l_im_il_fm_f}\sum_{c'} C^{00}_{l_im_il_fm_f}\underset{cc'}{\mathcal{F}^2_{0l_il_f}}(r)\left[\mathbf{p}^1_{c'l_im_i}(\mathbf{r}) + \Phi_{c'l_im_i}\right]Y^{l_f}_{m_f}(\hat{\mathbf{r}}) \\
&= \sum_{c'l'm'}\mathcal{F}^2_{c,c'l'}(r)\Phi_{c'l'm'}Y^{l'}_{-m'}(\hat{\mathbf{r}}) + \sum_{c'l'm'}\mathcal{F}^2_{c,c'l'}(r)Y^{l'}_{-m'}(\hat{\mathbf{r}})\mathbf{p}^1_{c'l'm'}(\mathbf{r}) \\
&= \sum_{c'l'm'}\mathcal{F}^2_{c,c'l'}(r)\Phi_{c'l'm'}Y^{l'}_{-m'}(\hat{\mathbf{r}}) + \sum_{c'l'm'}\mathcal{F}^2_{c,c'l'}(r)Y^{l'}_{-m'}(\hat{\mathbf{r}})\text{Gate}_{c'l'}(\mathbf{r})\Phi_{c'l'm'} \\
&\quad + \sum_{c'l'm'}\text{Gate}_{c'l'}(\mathbf{r})\sum_{l_il_fm_im_f}\sum_{\tilde{c}}\mathcal{F}^2_{c,c'l'}(r)\underset{c'\tilde{c}}{\mathcal{F}^1_{l'l_il_f}}(r)C^{l'm'}_{l_im_il_fm_f}\mathbf{h}_{\tilde{c}l_im_i}Y^{l'}_{-m'}(\hat{\mathbf{r}})Y^{l_f}_{m_f}(\hat{\mathbf{r}})
\end{aligned} \tag{5}$$

We have neglected here the gate function for simplicity, but it should not be an issue since it is merely a repetition of Eq. (4) to be multiplied by $p^2_c(\mathbf{r})$. The third term in the R.H.S of Eq.(5) is computationally expensive for incorporating the direct product of two spherical harmonics, which can be decomposed into a direct sum using CG coefficients:

$$Y^{l'}_{-m'}(\hat{\mathbf{r}})Y^{l_f}_{m_f}(\hat{\mathbf{r}}) = \sum_{l=|l'-l_f|}^{l'+l_f}\sum_{m=-l}^{l}\alpha(l',l_f,l)C^{lm}_{l'-m'l_fm_f}Y^l_m(\hat{\mathbf{r}})$$

where $\alpha(l',l_f,l)$ are coefficients of coupling channels. Thus it is further written as a contraction between CG coefficients:

$$\begin{aligned}
&\sum_{c'l'm'}\text{Gate}_{c'l'}(\mathbf{r})\sum_{l_il_fm_im_f}\sum_{\tilde{c}}\mathcal{F}^2_{c,c'l'}(r)\underset{c'\tilde{c}}{\mathcal{F}^1_{l'l_il_f}}(r)C^{l'm'}_{l_im_il_fm_f}\mathbf{h}_{\tilde{c}l_im_i}Y^{l'}_{-m'}(\hat{\mathbf{r}})Y^{l_f}_{m_f}(\hat{\mathbf{r}}) \\
=&\sum_{c'l'}\sum_{\tilde{c}}\text{Gate}_{c'l'}(\mathbf{r})\sum_{ll_il_f}\sum_{mm_i}\mathcal{F}^2_{c,c'l'}(r)\underset{c'\tilde{c}}{\mathcal{F}^1_{l'l_il_f}}(r)\mathbf{h}_{\tilde{c}l_im_i}\alpha(l',l_f,l)\sum_{m'm_f}C^{l'm'}_{l_im_il_fm_f}C^{lm}_{l'-m'l_fm_f}Y^l_m(\hat{\mathbf{r}})
\end{aligned} \tag{6}$$

The contraction between CG coefficients results in δ functions[1]:

$$\sum_{m'm_f}C^{l'm'}_{l_im_il_fm_f}C^{lm}_{l'-m'l_fm_f} = (-1)^{l_f+m-m_i}\frac{\sqrt{(2l+1)(2l'+1)}}{2l_i+1}\delta_{l_i,l}\delta_{m_i,-m}$$

with the third term in the R.H.S of Eq. (5) reduced to:

$$\begin{aligned}
&\sum_{c'l'm'}\text{Gate}_{c'l'}(\mathbf{r})\sum_{l_il_fm_im_f}\sum_{\tilde{c}}\mathcal{F}^2_{c,c'l'}(r)\underset{c'\tilde{c}}{\mathcal{F}^1_{l'l_il_f}}(r)C^{l'm'}_{l_im_il_fm_f}\mathbf{h}_{\tilde{c}l_im_i}Y^{l'}_{-m'}(\hat{\mathbf{r}})Y^{l_f}_{m_f}(\hat{\mathbf{r}}) \\
=&\sum_{\tilde{c}l_im_i}\mathbf{h}_{\tilde{c}l_im_i}Y^{l_i}_{-m_i}(\hat{\mathbf{r}})\sum_{c'l'l_f}\text{Gate}_{c'l'}(\mathbf{r})\mathcal{F}^2_{c,c'l'}(r)\underset{c'\tilde{c}}{\mathcal{F}^1_{l'l_i,l_f}}(r)\alpha(l',l_f,l_i)(-1)^{l_f}\frac{\sqrt{(2l'+1)}}{\sqrt{(2l_i+1)}}
\end{aligned} \tag{7}$$

With some redefinition of trainable parameters, the two-layer atom-to-grid equivariant message passing is reformulated into an $\mathcal{O}(L^2)$ scaling form:

$$\begin{aligned}
p^2_c(\mathbf{r}) &= \sum_{c'l'm'}\Phi_{c'l'm'}\mathcal{F}^2_{c,c'l'}(r)Y^{l'}_{-m'}(\hat{\mathbf{r}}) + \sum_{c'l'm'}\text{Gate}_{c'l'}(\mathbf{r})\Phi_{c'l'm'}\mathcal{F}^2_{c,c'l'}(r)Y^{l'}_{-m'}(\hat{\mathbf{r}}) \\
&\quad + \sum_{\tilde{c}l_im_i}\mathbf{h}_{\tilde{c}l_im_i}Y^{l_i}_{-m_i}(\hat{\mathbf{r}})\sum_{c'l'}\text{Gate}_{c'l'}(\mathbf{r})\underset{cc'\tilde{c}}{\bar{\mathcal{F}}^2_{l'l_i}}(r)
\end{aligned} \tag{8}$$

[1] See Eq. (8) in p. 259 of [7]

Though $\mathcal{O}(L^2)$ scaling, Eq. (8) incorporates a large number of grids and channels, hence it is memory prohibitive. The nontrivial parts are the fully trainable radial filters $\mathcal{F}$ and the non-linearity gate functions. If we give up the non-linearity functions, we could greatly simplify the grid-based methods to a basis-based form with fully trainable radial basis functions.

Supplementary Discussions

1. Performance on NMC Dataset

The NMC dataset contains 2,000 $Li_xNi_yMn_zCo_{(1-y-z)}O_2$ battery cathode materials with varying levels of Li content [5], split into 1,450 training, 50 validation, and 500 test structures. The original NMC dataset is calculated with VASP but lacks the specific pseudopotential information. For applicability of downstream tasks such as electronic structure calculations, we recalculated the pseudo charge density as well as the augmentation occupancies for the original structures, the exact version and hashes of the pseudopotentials are reported in our Zenodo data repository.

The model was trained to predict both the pseudo-charge densities and augmentation occupancies simultaneously. The radial cutoff for RBFs $\text{cutoff}_{\text{probe}}$ is 3Å and the number of channels N_c is 3, other model hyperparameters are identical to those used for training the MD trajectory datasets. The rationale for these hyperparameters is justified in Supplementary Discussions 5. On the test set, EdenGNN achieved a pseudo-charge density MAE of 0.096% and an augmentation occupancy MAE of 0.00085. While this density error is marginally larger than that reported by ChargE3Net (0.061%), it is fully adequate for downstream applications such as predicting electronic structures. This is illustrated by a representative band structure comparison for one of the test structures shown in Supplementary Fig. 11. Across the 500 test structures, the band energy MAE is 0.045 eV. Crucially, EdenGNN maintains this level of accuracy while achieving a 150-fold inference speedup compared to ChargE3Net, as demonstrated in Fig. 3(a) of the main text.

2. Construction of Spherical Bessel Radial Basis Functions for Ablation Study

To isolate the effect of the radial basis construction, we replace only the trainable radial basis in EdenGNN with a fixed spherical Bessel basis, while keeping all other settings unchanged, including $L_{\max}$, the number of channels, the cutoff radii, and the network architecture. In particular, the comparison is performed under the same controlled setting as in Supplementary Discussion 5, so that the observed performance difference can be attributed to the basis choice itself.

For the fixed basis, we constructed the l-dependent radial functions from spherical Bessel functions with a cosine cutoff envelope. Let r_c denote the cutoff radius and let α_{nl} be the n-th positive root of the spherical Bessel function $j_l(\alpha)$, i.e.,

$$j_l(\alpha_{nl}) = 0, \qquad n = 1, \ldots, N_c, \quad l = 0, \ldots, L_{\max}. \tag{9}$$

The corresponding radial basis is written as

$$F_{cl}(r) = f_{\text{cut}}(r)\, j_l\left(\alpha_{nl}\frac{r}{r_c}\right), \tag{10}$$

where the cutoff envelope is

$$f_{\text{cutoff}}(r) = \frac{1}{2}\left[\cos\left(\pi\frac{r}{r_c}\right) + 1\right]\Theta(r_c - r), \tag{11}$$

and $\Theta(\cdot)$ is the Heaviside step function. This construction enforces smooth decay to zero at the cutoff boundary, while the roots α_{nl} provide a complete fixed radial expansion for each angular momentum channel. In our implementation, the same set of $L_{\max}$ and N_c values is used for both the trainable and fixed-basis models; therefore, any change in accuracy directly reflects the advantage of learning the radial filters rather than using a predetermined Bessel expansion.

3. Validate EdenGNN on LCAO-basis DFT

In this section, validation the applicability of EdenGNN to LCAO-basis DFT by training a foundation model using ABACUS [8]. The dataset consists of structural data for non-magnetic materials sampled from the MP database [9]. This dataset differs from the one used to train EdenGNN-Uni, as the norm-conserving pseudopotentials employed here provide less comprehensive support for f-block elements compared to the projector augmented-wave (PAW) method. The dataset was randomly partitioned into 50,000, 2,000, and

2,000 samples for training, validation, and testing, respectively. The model was trained on a single NVIDIA A800-SXM4-80GB GPU for 168 GPU hours, completing 698,000 steps with a batch size of 1. The hyperparameters are set as follows: $L_{\max} = 6$, $N_c = 4$, $\text{cutoff}_{\text{atom}} = 4.0$ Å, $\text{cutoff}_{\text{probe}} = 3.0$ Å. The meaning of these parameters and the justification for their selection are discussed in Supplementary Discussion 5.

On the test set, the model achieves a density error of 0.61%, which error is comparable to the error of EdenGNN-Uni on PAW datasets (0.67%), indicating the transferability and robustness of EdenGNN across different DFT methods. To evaluate out-of-distribution performance, we further calculate the band structures for 100 randomly sampled structures from the GNoME database, ensuring the chemical systems excluded from the training set. The resulting KS eigenenergy MAE $\text{MAE}_{\text{eigen}}$ is 0.097 eV. The corresponding parity plot is presented in Supplementary Fig. 12.

For the DFT calculations in this section, we employed the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional and a k-point grid with a density of 0.15 Å^{-1}. The self-consistent field (SCF) convergence criterion `scf_thr` is set to 10^{-6}. The energy cutoff (`ecutwfc`) is set to 100 Rydberg; we refer readers to the ABACUS documentation for further details regarding this precision parameter. The trained model weights, dataset splits, and the exact pseudopotentials and basis sets (along with their hashes) are publicly available in our Zenodo data repository.

4. Workflow to Generate the Superposition of Atomic Charge Density

The superposition of atomic charge density (SACD) prior is generated directly from the native output of the electronic-structure code, rather than from any separate atomic-density reconstruction. For the VASP datasets used in this work, we obtain the SACD prior by setting `ICHARG = 12` (and `NELM = 0` to disable self-consistent calculations) with a strictly consistent pseudopotential set and the same calculation parameters used throughout dataset generation and charge density prediction. This workflow ensures that the atomic-center charge density is produced in a standardized and reproducible manner, without introducing additional interpolation artifacts or post-processing noise.

The same principle applies to other supported DFT packages in EdenGNN. In these cases, we also rely on the built-in SACD functionality provided by the code itself, together with a unified pseudopotential setup, so that the resulting SACD prior is defined on the same physical baseline across all samples. Because SACD depends on the pseudopotential choice, we keep the corresponding pseudopotentials fixed for each dataset to guarantee internal consistency and to avoid mixing priors generated under different atomic reference states.

5. Validate the Hyperparameter Settings of EdenGNN

To validate the hyperparameter choices used in the main text, we perform controlled scans of the maximum angular momentum $L_{\max}$, the number of radial channels N_c, the atomic cutoff radius, the probe cutoff radius, and the density loss fraction, while keeping all other settings fixed at the reference configuration ($L_{\max} = 6$, $N_c = 4$, $\text{cutoff}_{\text{atom}} = 4.0$ Å, $\text{cutoff}_{\text{probe}} = 4.0$ Å, and loss density fraction = 0.9). The corresponding results are shown in Supplementary Fig. 13a–13f.

The density prediction error decreases rapidly as $L_{\max}$ increases from 3 to 4, and then becomes nearly saturated for $L_{\max} \geq 5$. A similar trend is observed for the radial channels: the error is reduced from 0.118% at $N_c = 2$ to 0.087% at $N_c = 3$, and remains at a comparable level for $N_c \geq 4$. These results indicate that $N_c = 3$ already provides a good balance between expressivity and efficiency. The cutoff-radius scans show that the model is more sensitive to the probe cutoff than to the atomic cutoff when the cutoff is too small, and the inverse is observed when the cutoff is larger. This suggests that we can further improve the accuracy by increasing the interaction ranges of atoms. The density loss has already saturated at $\text{cutoff}_{\text{probe}} = 3.0$ Å, and the choice of $\text{cutoff}_{\text{probe}} = 4.0$ Å is to provide a safety margin. On the other hand, using $\text{cutoff}_{\text{probe}} = 3.0$ Å instead of 4.0 Å would increase the training speed. Finally, as expected, varying the density loss fraction from 0.5 to 0.9 would increase the density accuracy while decreasing the augmentation occupancy accuracy, because the two optimization targets are competing with each other. Hence we choose a dual network strategy in training EdenGNN-Uni to maximize the performance. Overall, these scans support the hyperparameter setting adopted in the main text as a compromise between accuracy and computational cost.

6. The Optimization Strategies for Training Models

Throughout the work we used the AdamW optimizer with a ReduceLROnPlateau learning rate scheduler to train all models, including the ChargE3Net and EAC-Net models. The ReduceLROnPlateau scheduler is implemented by monitoring the validation metrics and reducing the learning rate by a factor of 0.5 if the validation metrics do not improve for `patience` evaluation times. As a robust, feedback-driven scheduler, ReduceLROnPlateau automatically adjusts the learning rate based on validation metrics, thereby reducing the difficulty of engineering hyperparameters. The detailed optimization settings including batch size and learning rate scheduler are summarized in Table 2.

Supplementary Tables

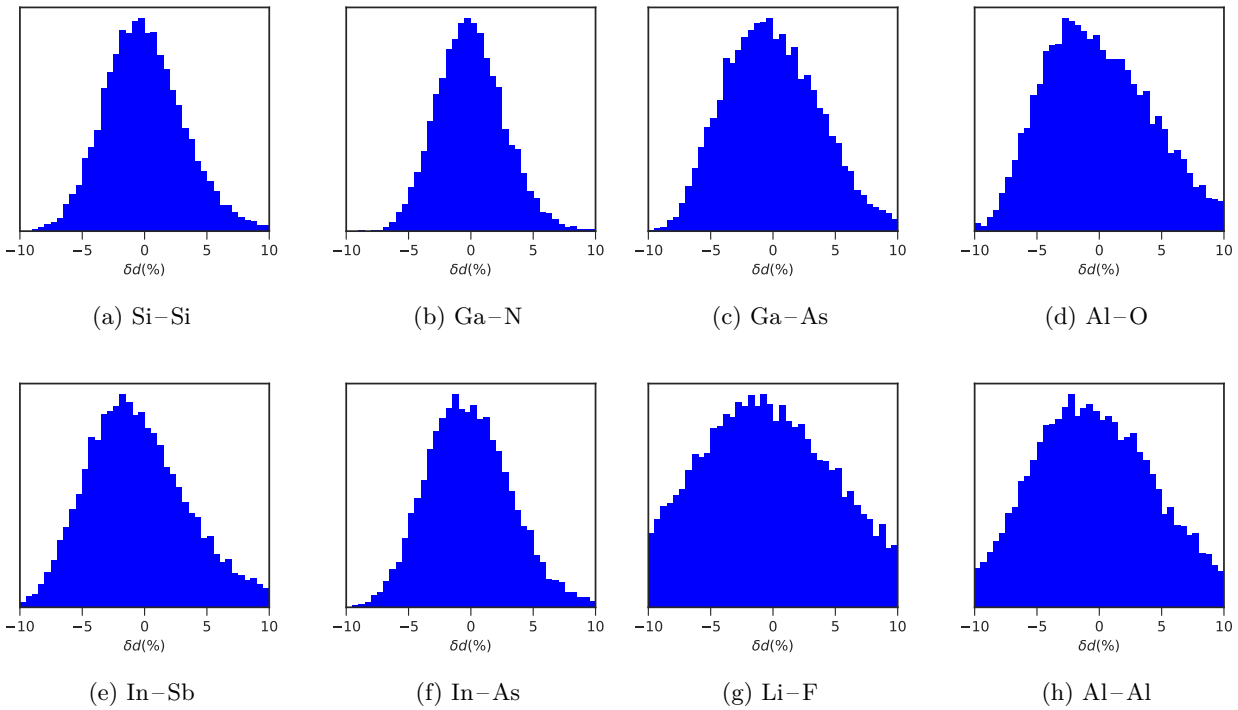
Supplementary Table 1: Overlap statistics and generalizability of EdenGNN-Uni on disjoint test sets. Comp, Re-Form, Chem-Sys and Proto denote composition, reduced formula, chemical system and AFLOW-style protostructure [10], respectively. The overlap percentages indicate the proportion of test samples that share the same descriptor with the MP training set. The Disjoint subset is constructed by excluding test structures whose Comp, Re-Form, Chem-Sys or Proto labels appear in the MP training set.

Test Subset	Size (N)	Overlap	$\varepsilon_{\tilde{n}}$ (%)	MAE_{aug}	$\text{MAE}_{\text{eigen}}$ (eV)
MP Test Set	5,000	9.6% (Comp) 17.7% (Re-Form) 54.9% (Chem-Sys) 2.9% (Proto)	0.67	0.0085	0.102
MP Comp-Disjoint	4,519	0% (Comp)	0.66	0.0084	0.100
MP Re-Form-Disjoint	4,107	0% (Re-Form)	0.69	0.0086	0.102
MP Chem-Sys-Disjoint	2,254	0% (Chem-Sys)	0.68	0.0087	0.106
MP Proto-Disjoint	4,853	0% (Proto)	0.67	0.0083	0.102
GNoME Test Set	1,000	0% (Comp) 0% (Re-Form) 0% (Chem-Sys) 3.3% (Proto)	0.66	0.0105	0.099
GNoME Proto-Disjoint	967	0% (Proto)	0.66	0.0105	0.099

Supplementary Table 2: Optimization settings for models trained in this work. Abbreviations: MD-traj (Molecular Dynamics trajectories), MP (Materials Project), LR (learning rate). For the MP dataset, patience is measured in evaluation steps rather than epochs. For EAC-Net, the batch size is 40 probes, because at each training step EAC-Net optimizes a fixed number of probes which may belong to 1 or more structures, while other models optimize a fixed number of structures at each step.

Model	Dataset	Initial LR	Patience	Batch Size
EdenGNN	MD-traj	5×10^{-3}	5 epochs	1
EdenGNN	MoS_2	1×10^{-2}	5 epochs	1
EdenGNN-Aug	MP	1×10^{-2}	5×1000 steps	8
EdenGNN-Den	MP	5×10^{-3}	5×3000 steps	1
ChargE3Net	MD-traj	1×10^{-2}	50 epochs	8
EAC-Net	MD-traj	1×10^{-3}	50 epochs	40 probes

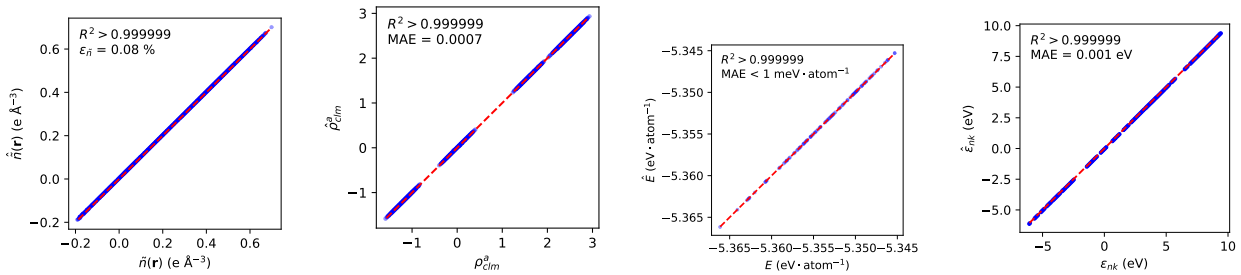
Supplementary Figures



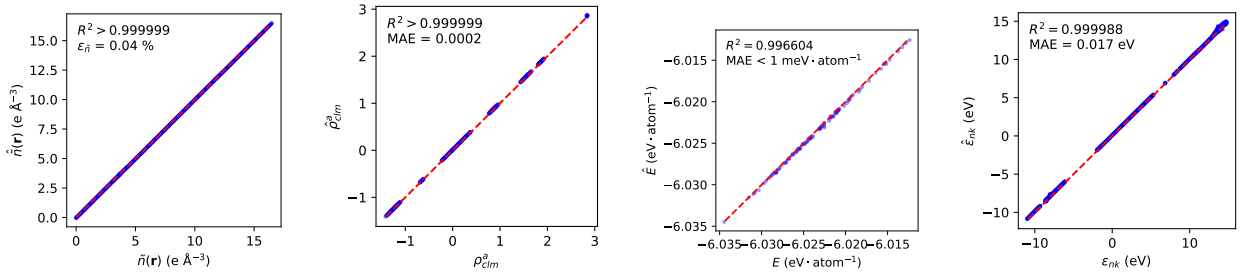
(a) Si−Si (b) Ga−N (c) Ga−As (d) Al−O

(e) In−Sb (f) In−As (g) Li−F (h) Al−Al

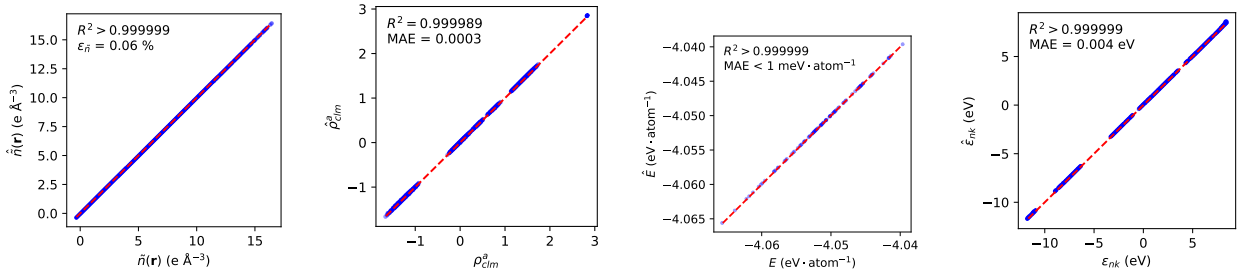
Supplementary Fig. 1: Relative bond length distributions of structures in the test sets of the molecular dynamics trajectory datasets. The relative bond length is defined as $\delta d = (d - \bar{d})/\bar{d}$, where $\bar{d}$ is the average bond length.


$R^2 > 0.999999$
$\varepsilon_{\tilde{n}} = 0.08$ %
$\tilde{n}(\mathbf{r})$ (e Å^{-3})
$\hat{\tilde{n}}(\mathbf{r})$ (e Å^{-3})
$R^2 > 0.999999$
MAE = 0.0007
ρ^a_{clm}
$\hat{\rho}^a_{clm}$
$R^2 > 0.999999$
MAE < 1 meV · atom^{-1}
E (eV · atom^{-1})
$\hat{E}$ (eV · atom^{-1})
$R^2 > 0.999999$
MAE = 0.001 eV
ε_{nk} (eV)
$\hat{\varepsilon}_{nk}$ (eV)

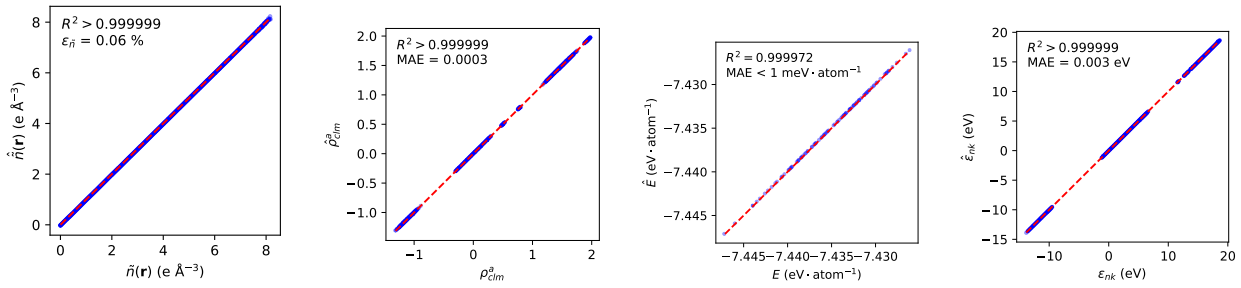
(a)


$R^2 > 0.999999$
$\varepsilon_{\tilde{n}} = 0.04$ %
$\tilde{n}(\mathbf{r})$ (e Å^{-3})
$\hat{\tilde{n}}(\mathbf{r})$ (e Å^{-3})
$R^2 > 0.999999$
MAE = 0.0002
ρ^a_{clm}
$\hat{\rho}^a_{clm}$
$R^2 = 0.996604$
MAE < 1 meV · atom^{-1}
E (eV · atom^{-1})
$\hat{E}$ (eV · atom^{-1})
$R^2 = 0.999988$
MAE = 0.017 eV
ε_{nk} (eV)
$\hat{\varepsilon}_{nk}$ (eV)

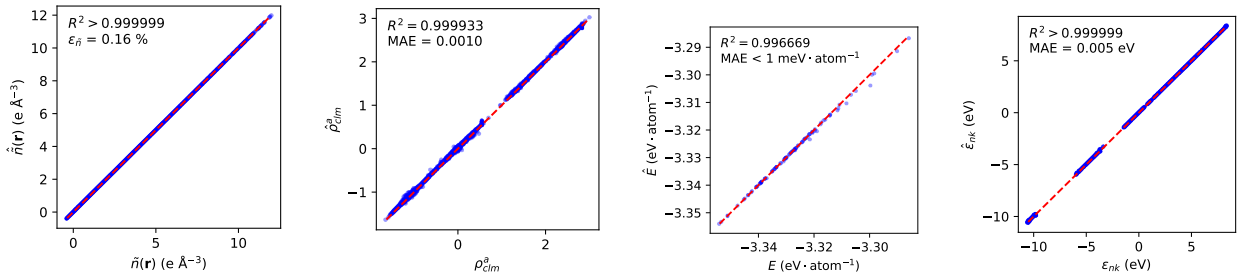
(b)


$R^2 > 0.999999$
$\varepsilon_{\tilde{n}} = 0.06$ %
$\tilde{n}(\mathbf{r})$ (e Å^{-3})
$\hat{\tilde{n}}(\mathbf{r})$ (e Å^{-3})
$R^2 = 0.999989$
MAE = 0.0003
ρ^a_{clm}
$\hat{\rho}^a_{clm}$
$R^2 > 0.999999$
MAE < 1 meV · atom^{-1}
E (eV · atom^{-1})
$\hat{E}$ (eV · atom^{-1})
$R^2 > 0.999999$
MAE = 0.004 eV
ε_{nk} (eV)
$\hat{\varepsilon}_{nk}$ (eV)

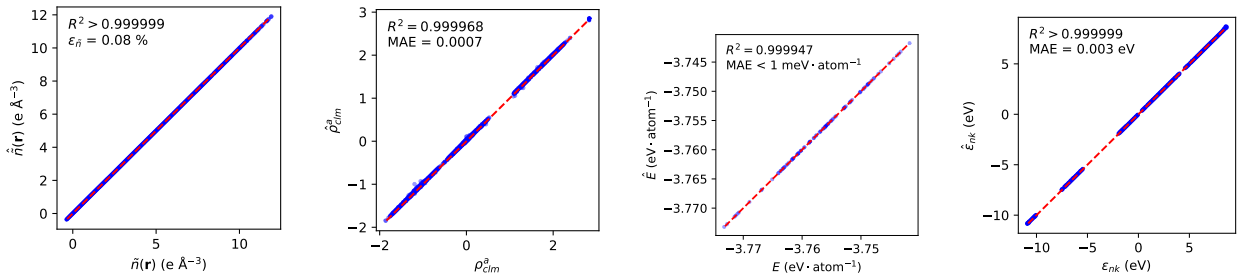
(c)


$R^2 > 0.999999$
$\varepsilon_{\tilde{n}} = 0.06$ %
$\tilde{n}(\mathbf{r})$ (e Å^{-3})
$\hat{\tilde{n}}(\mathbf{r})$ (e Å^{-3})
$R^2 > 0.999999$
MAE = 0.0003
ρ^a_{clm}
$\hat{\rho}^a_{clm}$
$R^2 = 0.999972$
MAE < 1 meV · atom^{-1}
E (eV · atom^{-1})
$\hat{E}$ (eV · atom^{-1})
$R^2 > 0.999999$
MAE = 0.003 eV
ε_{nk} (eV)
$\hat{\varepsilon}_{nk}$ (eV)

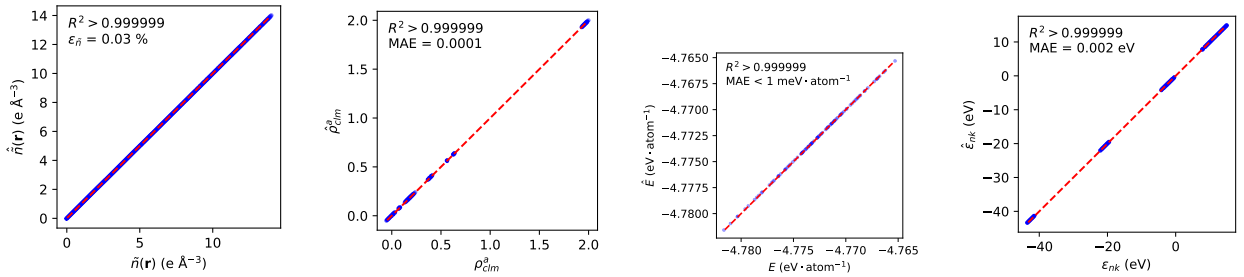
(d)



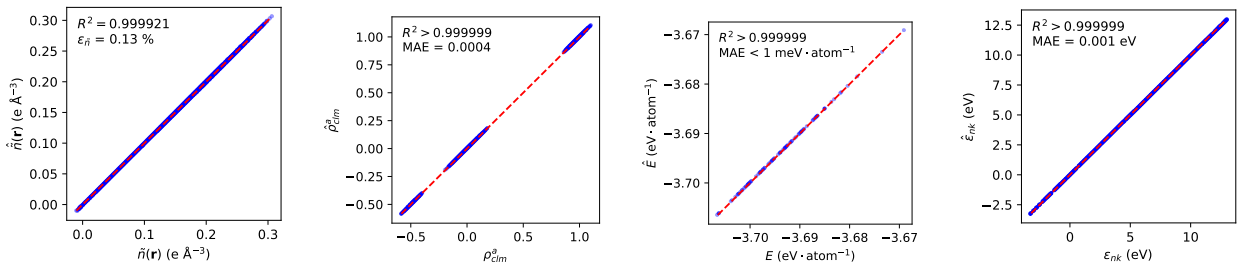
(e)



(f)

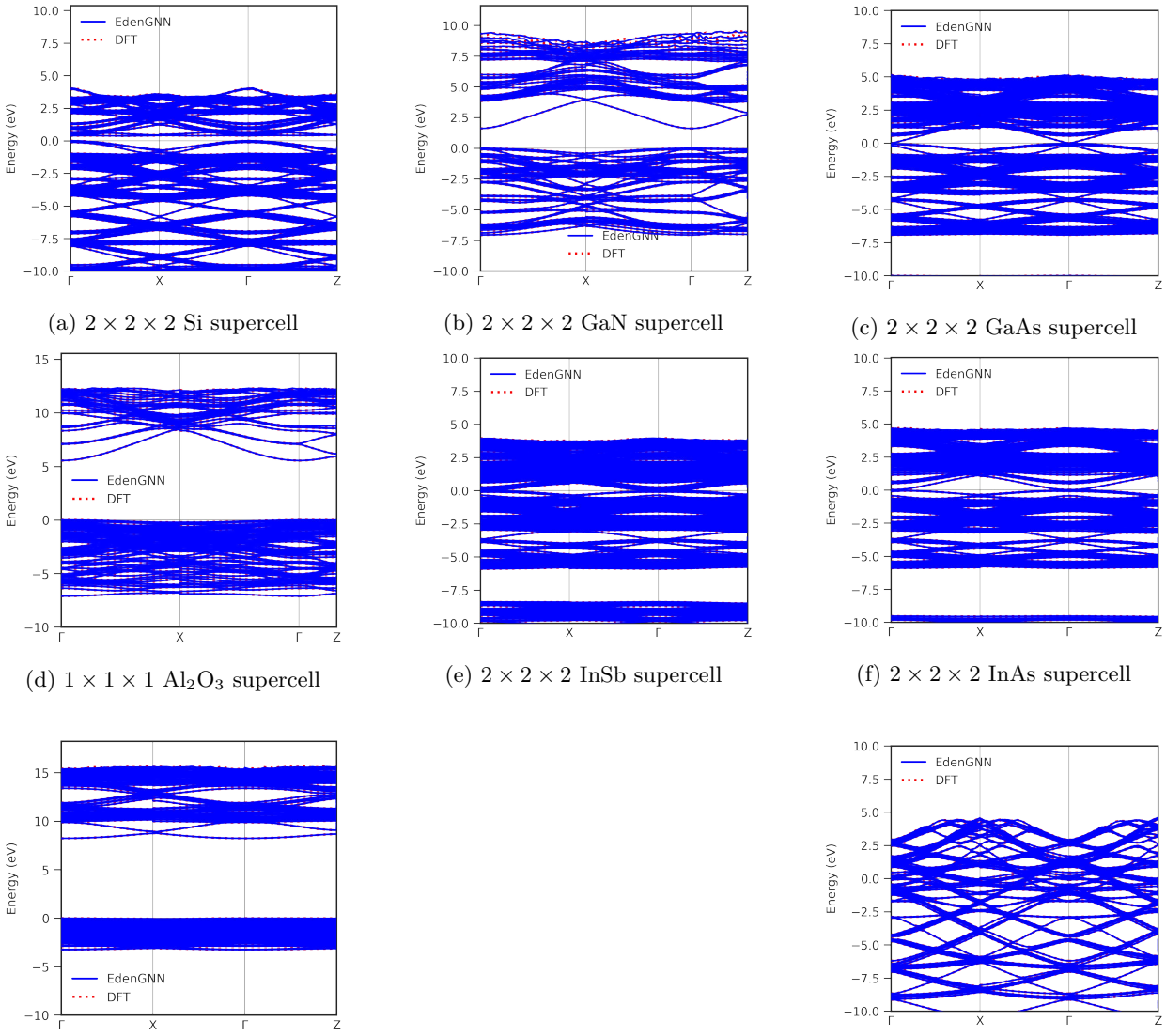


(g)



(h)

Supplementary Fig. 2: Parity plots of model predictions versus DFT benchmarks for systems from the molecular dynamics trajectories datasets. The systems are: (a) Si, (b) GaN, (c) GaAs, (d) Al_2O_3, (e) InSb, (f) InAs, (g) LiF and (h) Al. In each subfigure, the panels from left to right illustrate the pseudo density $\tilde{n}(\mathbf{r})$, the augmentation occupancy ρ^a_{clm}, the total energy E, and the eigenenergy ε_{nk} respectively. The coefficients of determination R^2 and the average errors are shown in each panel.



(a) $2 \times 2 \times 2$ Si supercell

(b) $2 \times 2 \times 2$ GaN supercell

(c) $2 \times 2 \times 2$ GaAs supercell

(d) $1 \times 1 \times 1$ Al_2O_3 supercell

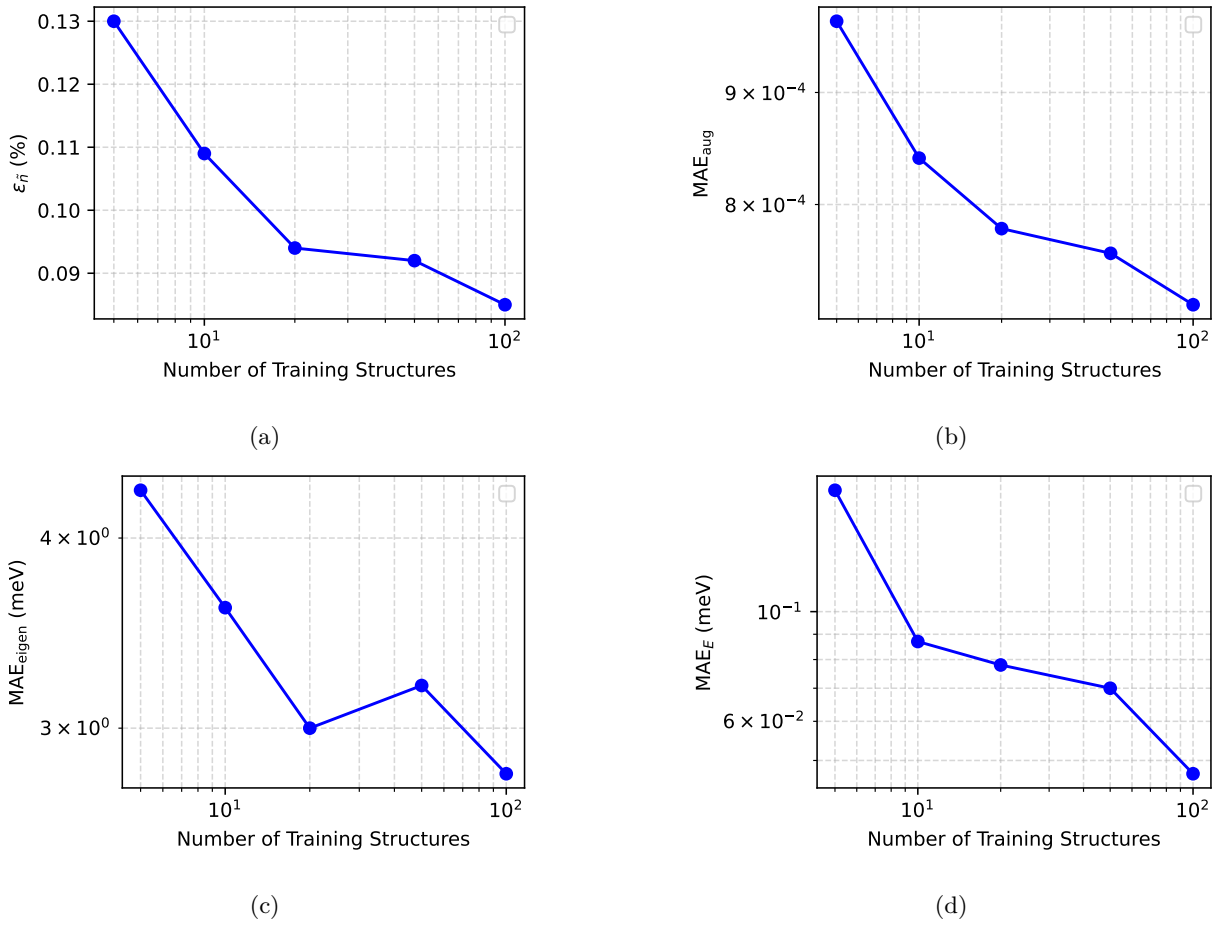
(e) $2 \times 2 \times 2$ InSb supercell

(f) $2 \times 2 \times 2$ InAs supercell

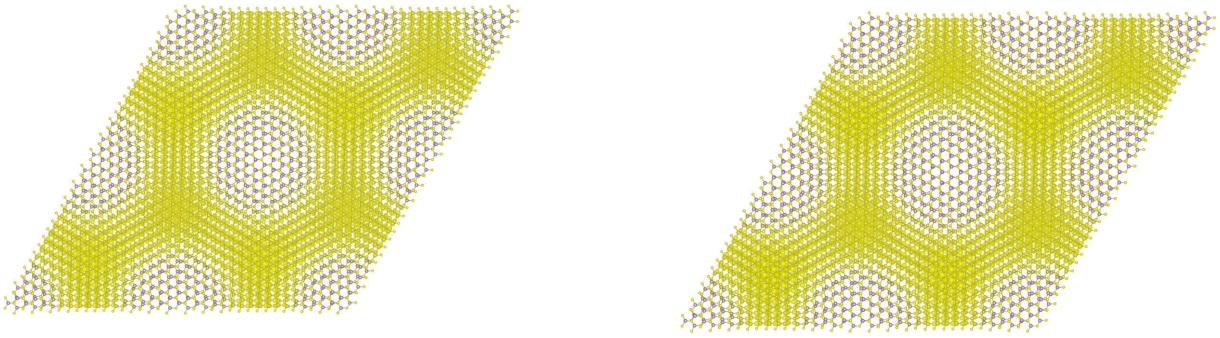
(g) $2 \times 2 \times 2$ LiF supercell

(h) $2 \times 2 \times 2$ Al supercell

Supplementary Fig. 3: Comparison of band structures predicted by EdenGNN and calculated by DFT for materials in the molecular dynamics trajectories dataset. The structure is chosen as the first frame in the trajectories of the test sets.

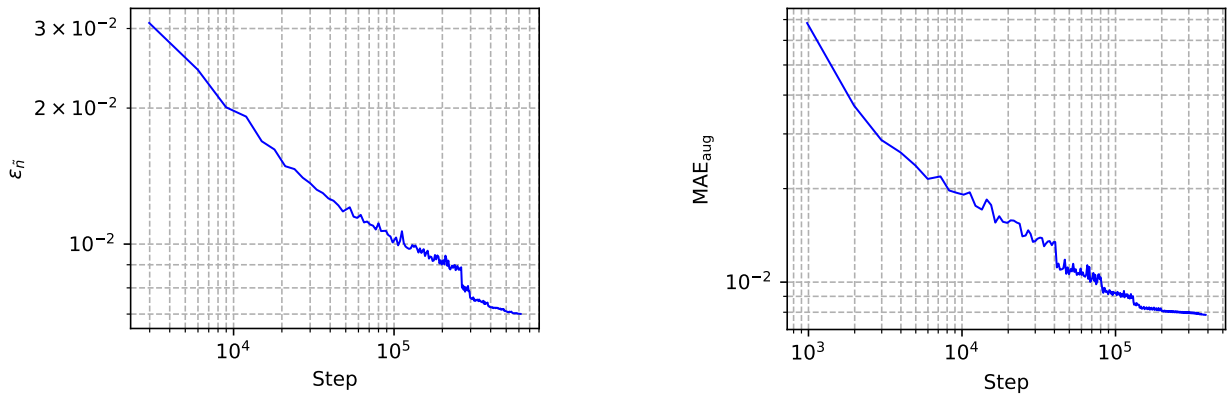


Supplementary Fig. 4: Learning curves for the InAs system. Subfigures (a-d) illustrate the variations of evaluation metrics across training set sizes of 5, 10, 20, 50, and 100: (a) normalized mean-absolute error (MAE) $\varepsilon_{\tilde{n}}$ of the pseudo charge density, (b) MAE of augmentation occupancies $\mathrm{MAE}_{\mathrm{aug}}$, (c) MAE of eigenenergies $\mathrm{MAE}_{\mathrm{eigen}}$, and (d) MAE of total energies MAE_E.



(a) (b)

Supplementary Fig. 5: Predicted charge density of the valence band maximum (VBM) state for 3.5° twisted MoS_2 superlattice. (a) The VBM charge density predicted by EdenGNN. (b) The VBM charge density calculated by DFT. Both figures are plotted using the same isovalue of 4.3×10^{-5} e/Å^3.

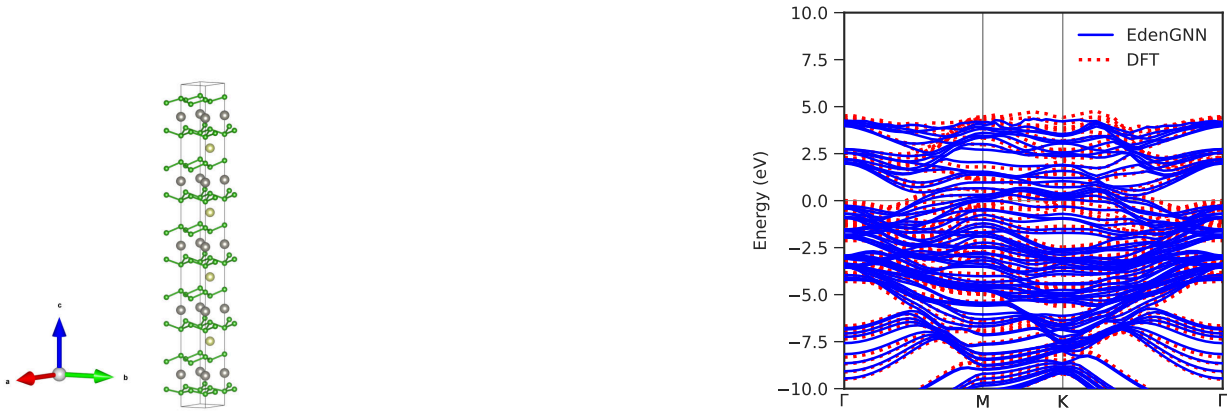


(a) (b)

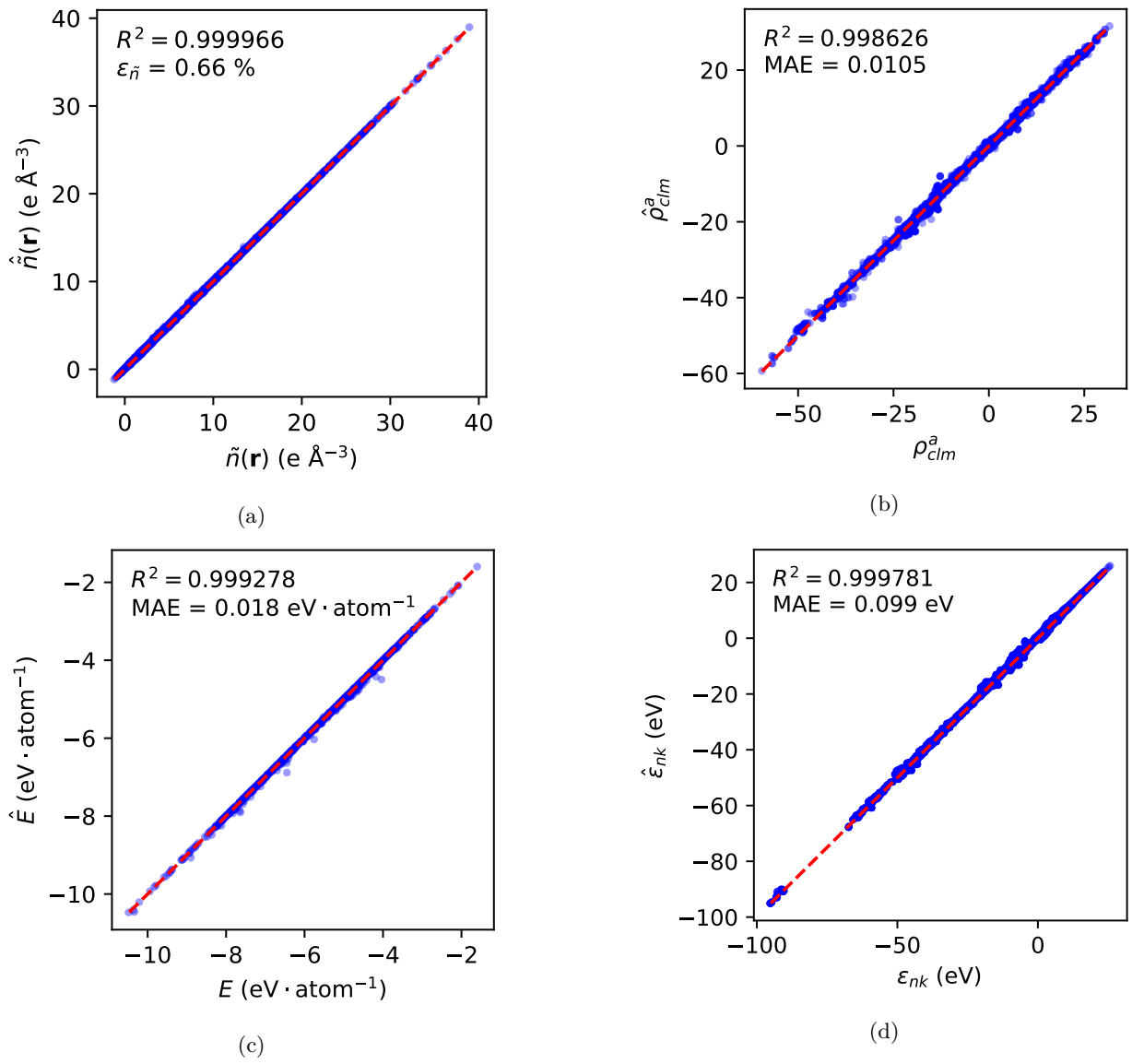
Supplementary Fig. 6: Validation loss curves for EdenGNN-Uni in the training stage. (a) The loss curve for the pseudo charge density prediction network. (b) The loss curve for the augmentation occupancy prediction network. $\varepsilon_{\tilde{n}}$ is the normalized pseudo charge density mean-absolute error (MAE), while $\mathrm{MAE}_{\mathrm{aug}}$ is the MAE of augmentation occupancies.

Note: Due to its large size, the images and detailed caption for this figure are placed at the end of this Supplementary Information document on page 37.

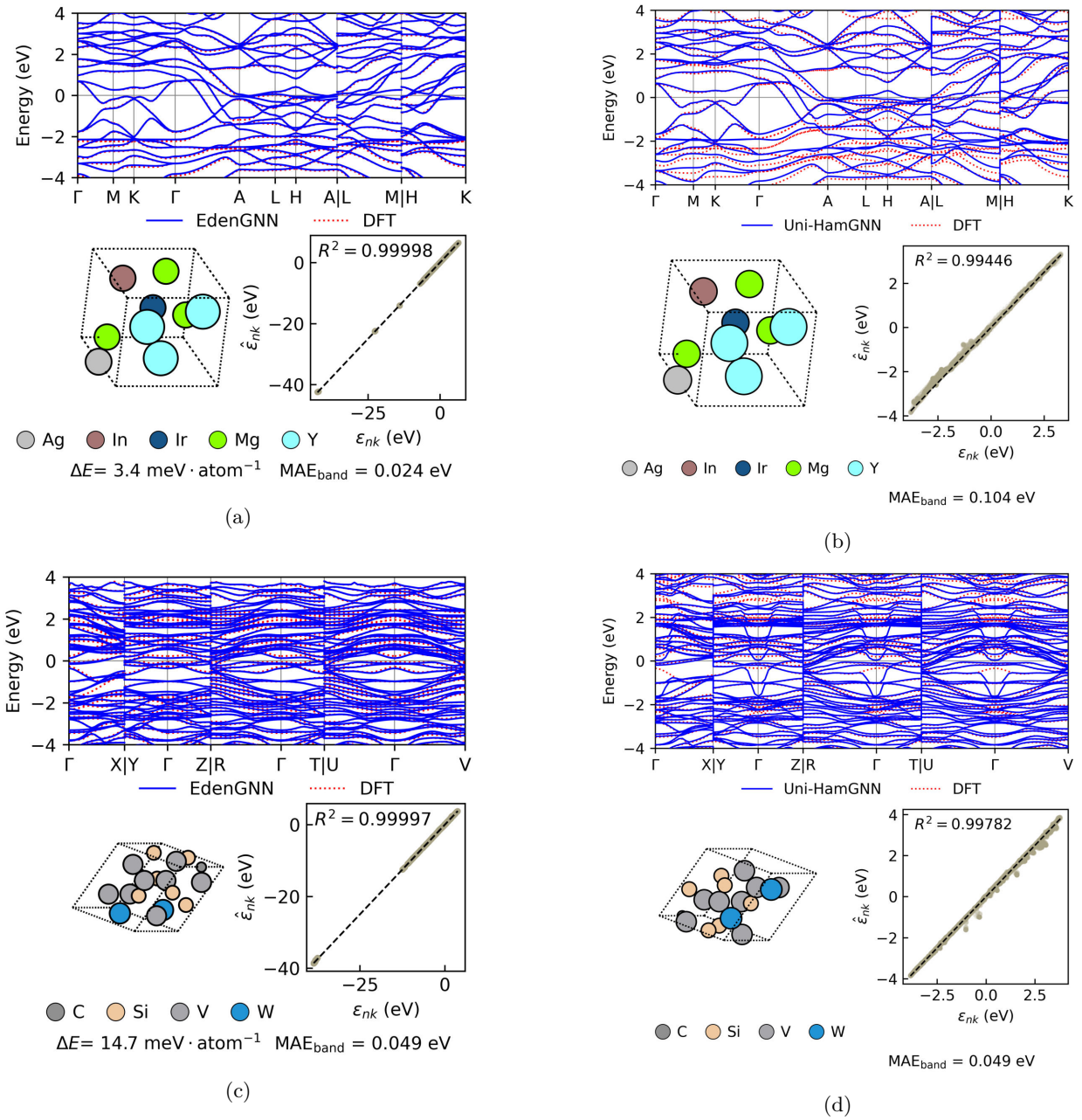
Supplementary Fig. 7: Band structure and density of state comparisons for 100 randomly sampled structures from the Materials Project database.



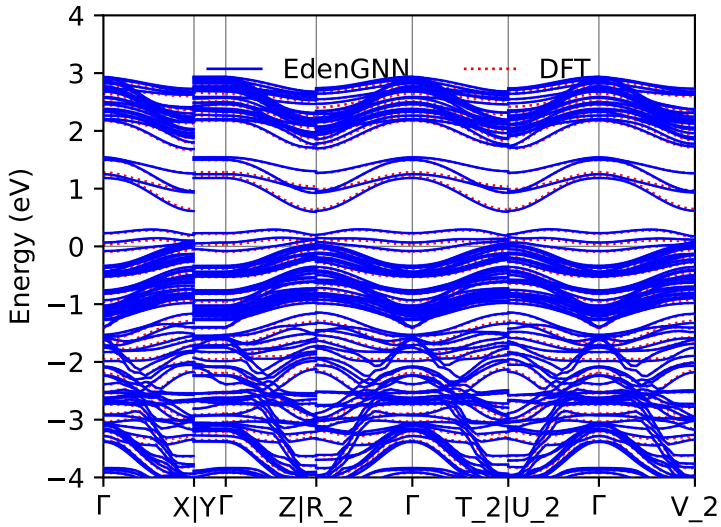
Supplementary Fig. 8: Representative band structure comparison for an outlier structure at the 95th percentile of the band energy error distribution. The plot shows the band structure of a material with Materials Project ID mp-1228750 [9], which has a mean absolute error $\mathrm{MAE}_{\mathrm{band}}$ of 0.232 eV.



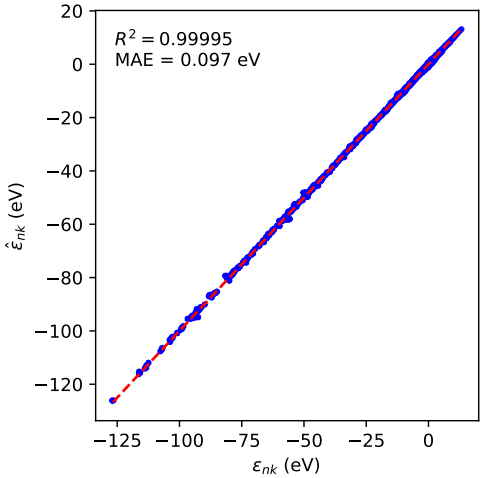
Supplementary Fig. 9: Parity plots of EdenGNN-Uni predictions versus DFT benchmarks for 1,000 novel materials randomly sampled from the GNoME database [11]. the pseudo density $\tilde{n}(\mathbf{r})$, the augmentation occupancy ρ^a_{clm}, the total energy E and the eigenenergy ε_{nk} respectively. The coefficients of determination R^2 and the average errors are shown in each panels.



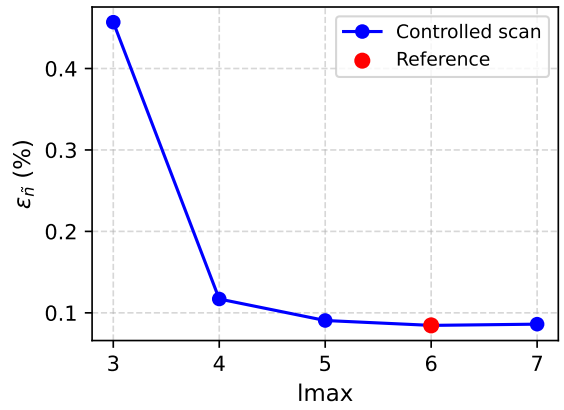
Supplementary Fig. 10: Comparison of predicted band structures by EdenGNN-Uni and Uni-HamGNN for materials in the GNoME database. For each subfigure, the upper panel illustrates the band structure comparison plot along high-symmetry lines; the lower-left panel displays the atomic structure; and the lower-right panel shows the parity plot of the eigenenergy ε_{nk} with the coefficient of determination (R^2) indicated. At the bottom, the elements of the compound, energy error ΔE and band energy mean absolute error MAE$_{\text{band}}$ are listed. For the result of EdenGNN-Uni, MAE$_{\text{band}}$ is evaluated only for states above -4 eV relative to the Fermi level E_f, whereas for Uni-HamGNN, MAE$_{\text{band}}$ is evaluated for states within [-4, 4] eV relative to the Fermi level. The representative compounds include: (a, b) $AgInIrMg_3Y_3$ and (c, d) $CSi_6V_8W_2$; for the latter, Uni-HamGNN exhibits ghost state artifacts near Fermi level.



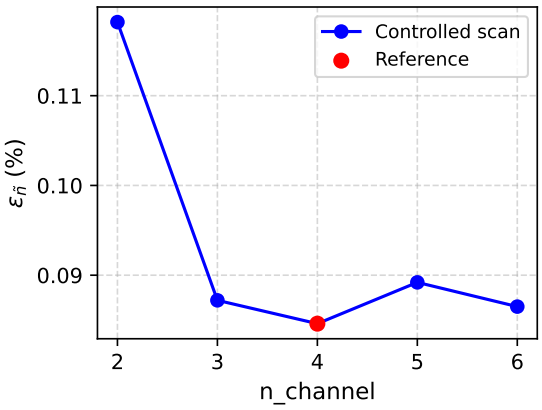
Supplementary Fig. 11: Representative band structure comparison for a $Li_{0.42}Ni_{0.17}Co_{0.83}O_2$ test structure in the NMC dataset.



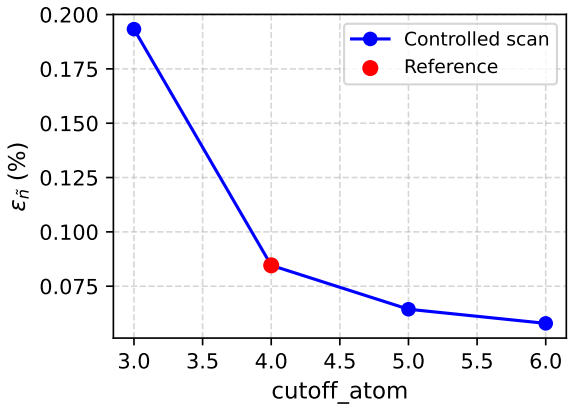
Supplementary Fig. 12: Parity plot of KS eigenenergies on 100 test structures from the GNoME database. The model is trained and validated on LCAO-basis DFT data.



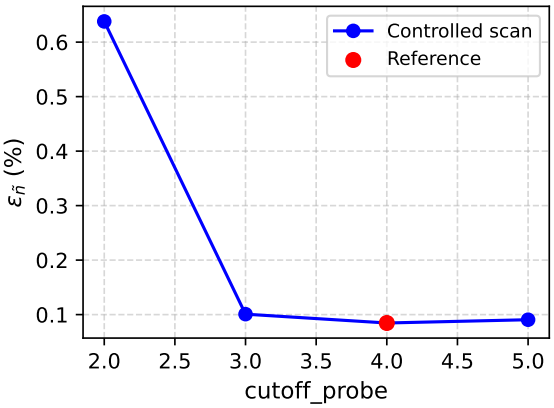
(a)



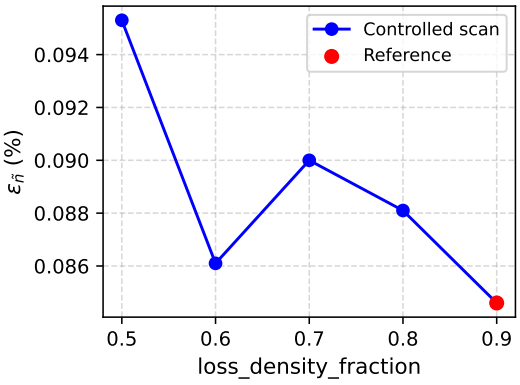
(b)



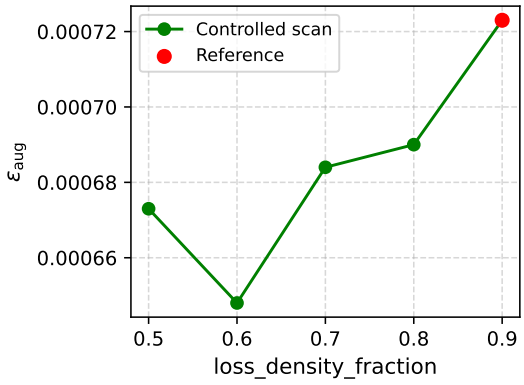
(c)



(d)

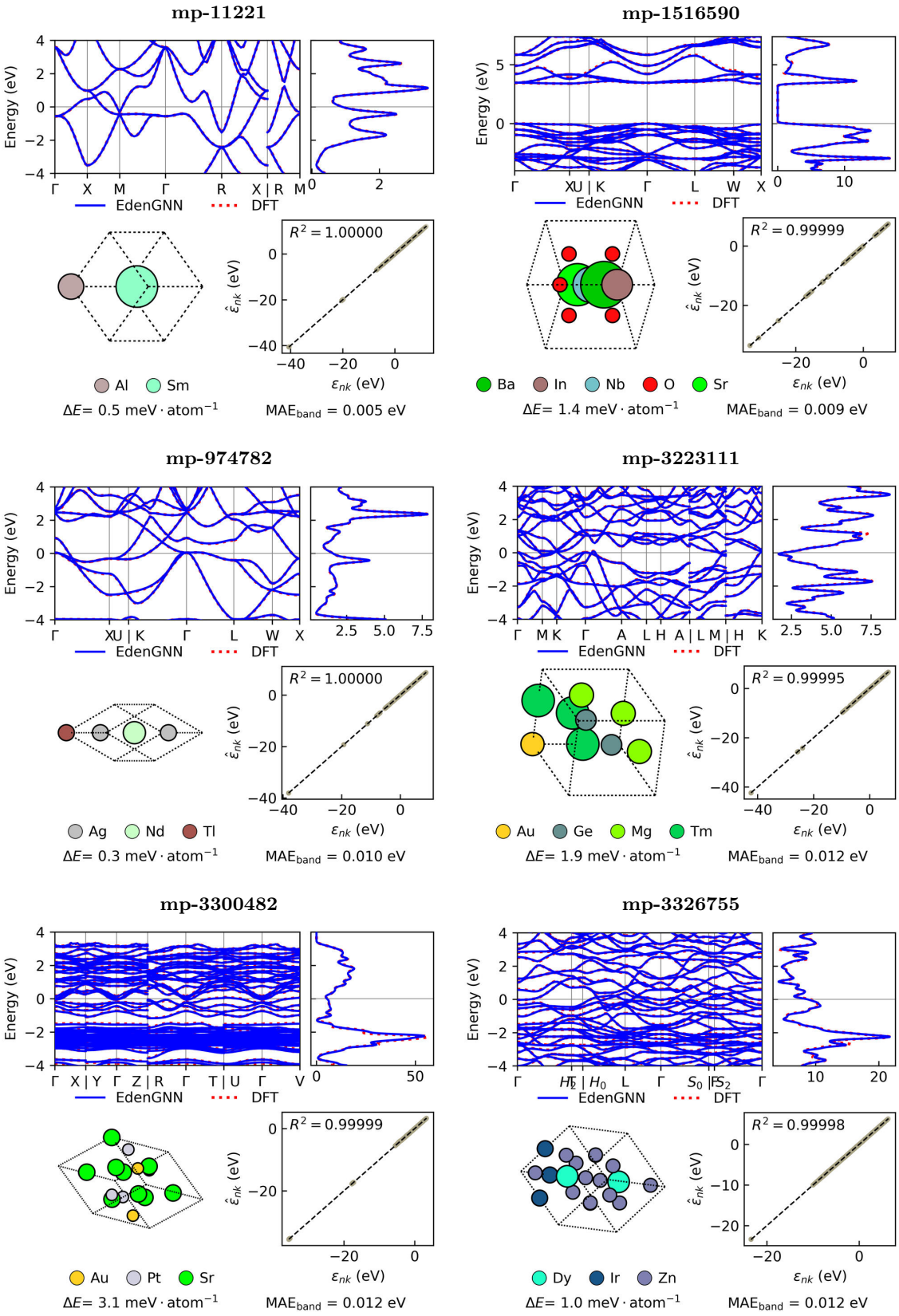


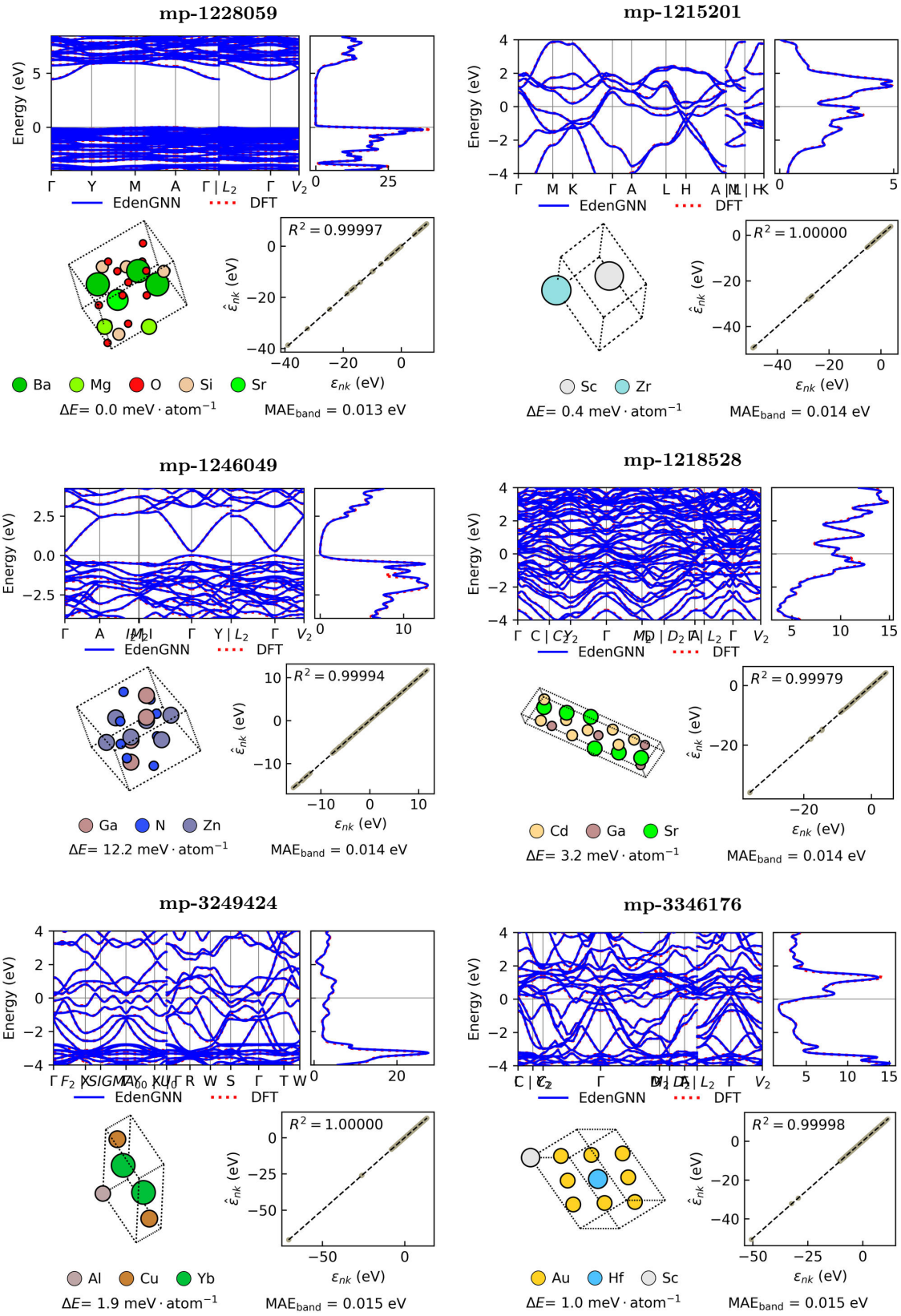
(e)

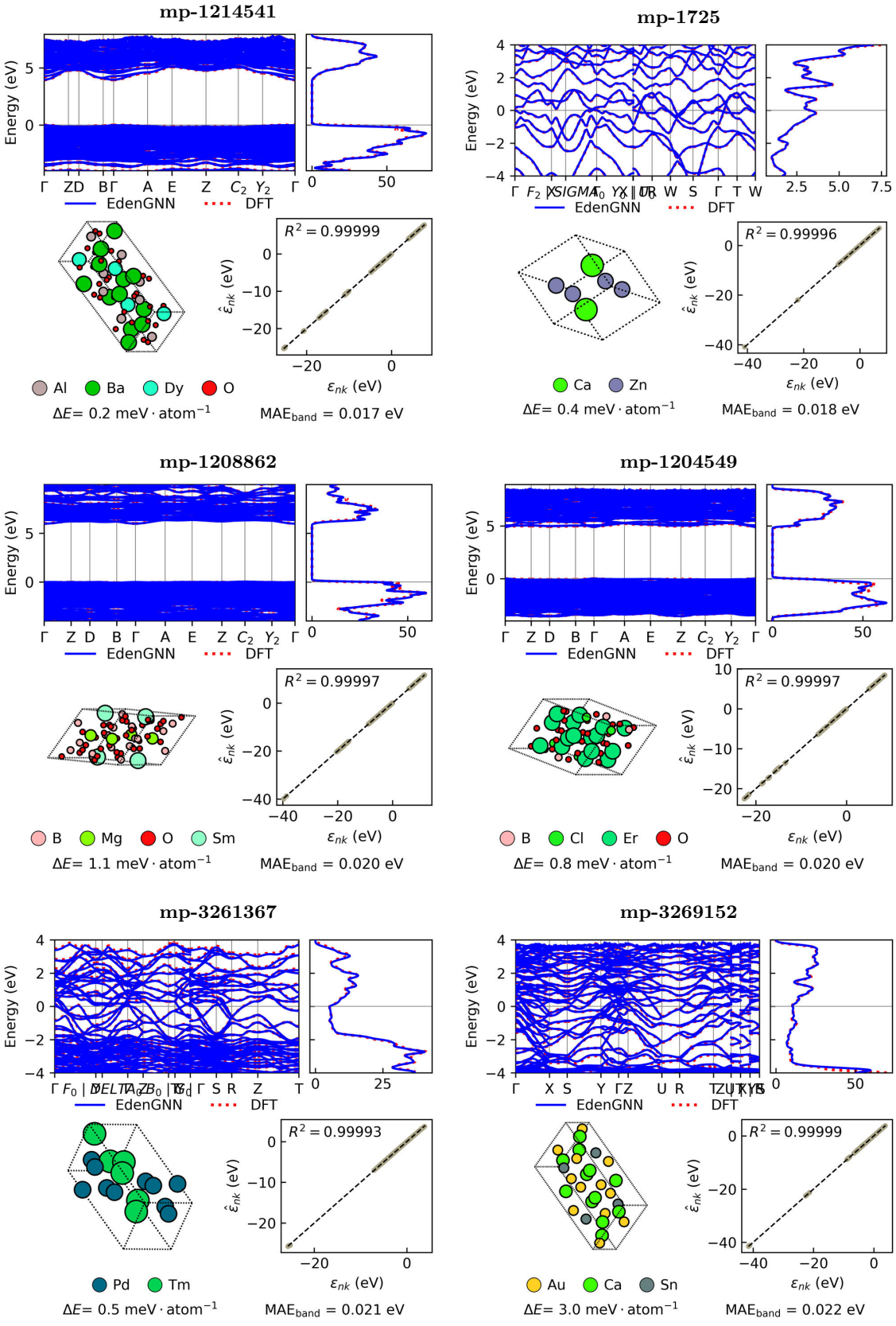


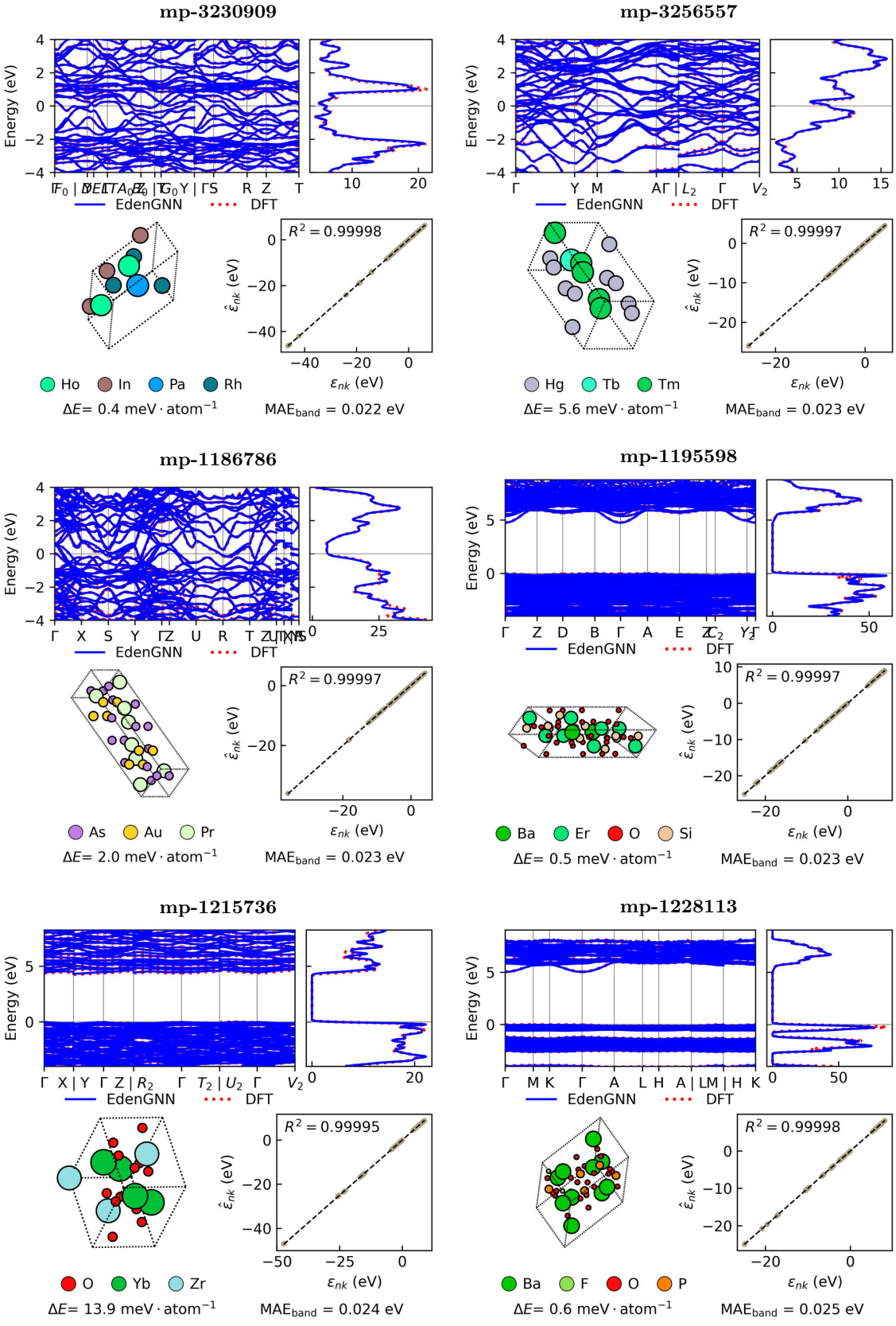
(f)

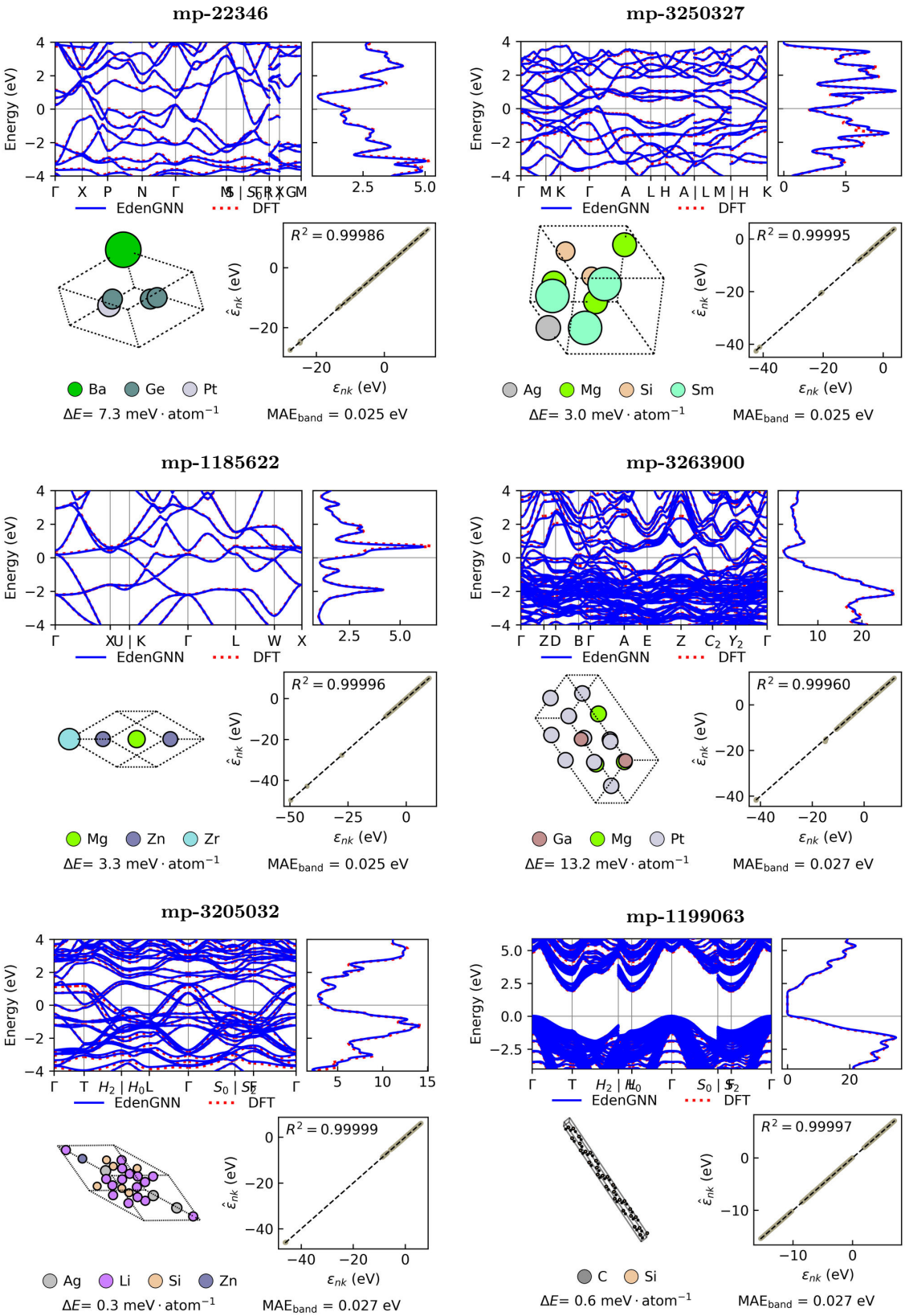
Supplementary Fig. 13: Hyperparameter validation for EdenGNN on the InAs system. All scans are performed with one parameter varied at a time while keeping the reference configuration fixed.

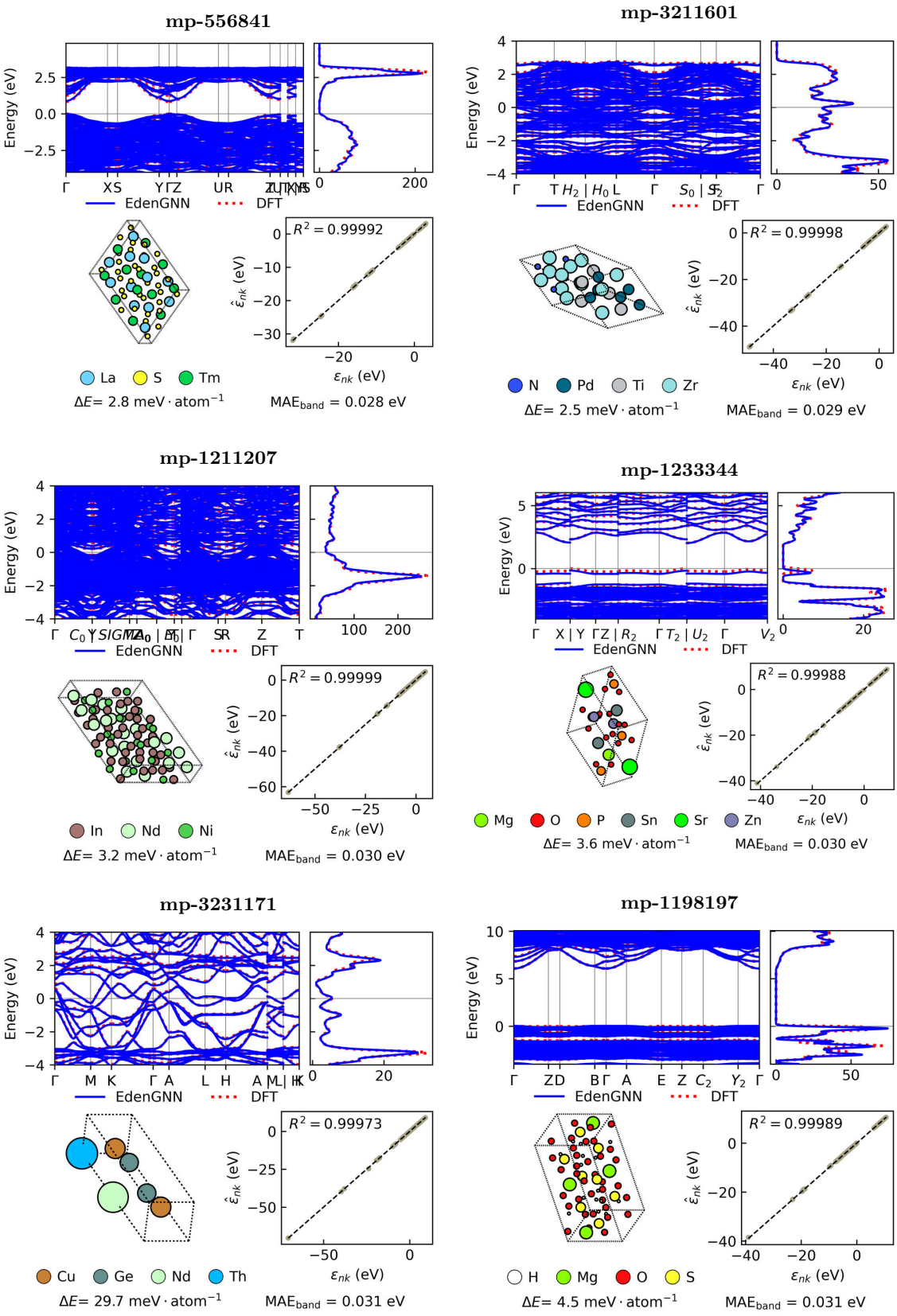

mp-11221
Energy (eV)
Γ X M Γ R X | R M
EdenGNN
DFT
$R^2 = 1.00000$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Al Sm
ΔE= 0.5 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.005 eV
mp-1516590
Γ XU | K Γ L W X
$R^2 = 0.99999$
Ba In Nb O Sr
ΔE= 1.4 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.009 eV
mp-974782
Γ XU | K Γ L W X
$R^2 = 1.00000$
Ag Nd Tl
ΔE= 0.3 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.010 eV
mp-3223111
Γ M K Γ A L H A | L M | H K
$R^2 = 0.99995$
Au Ge Mg Tm
ΔE= 1.9 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.012 eV
mp-3300482
Γ X | Y Γ Z | R Γ T | U Γ V
$R^2 = 0.99999$
Au Pt Sr
ΔE= 3.1 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.012 eV
mp-3326755
Γ H_2 | H_0 L Γ S_0 | S_2 Γ
$R^2 = 0.99998$
Dy Ir Zn
ΔE= 1.0 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.012 eV

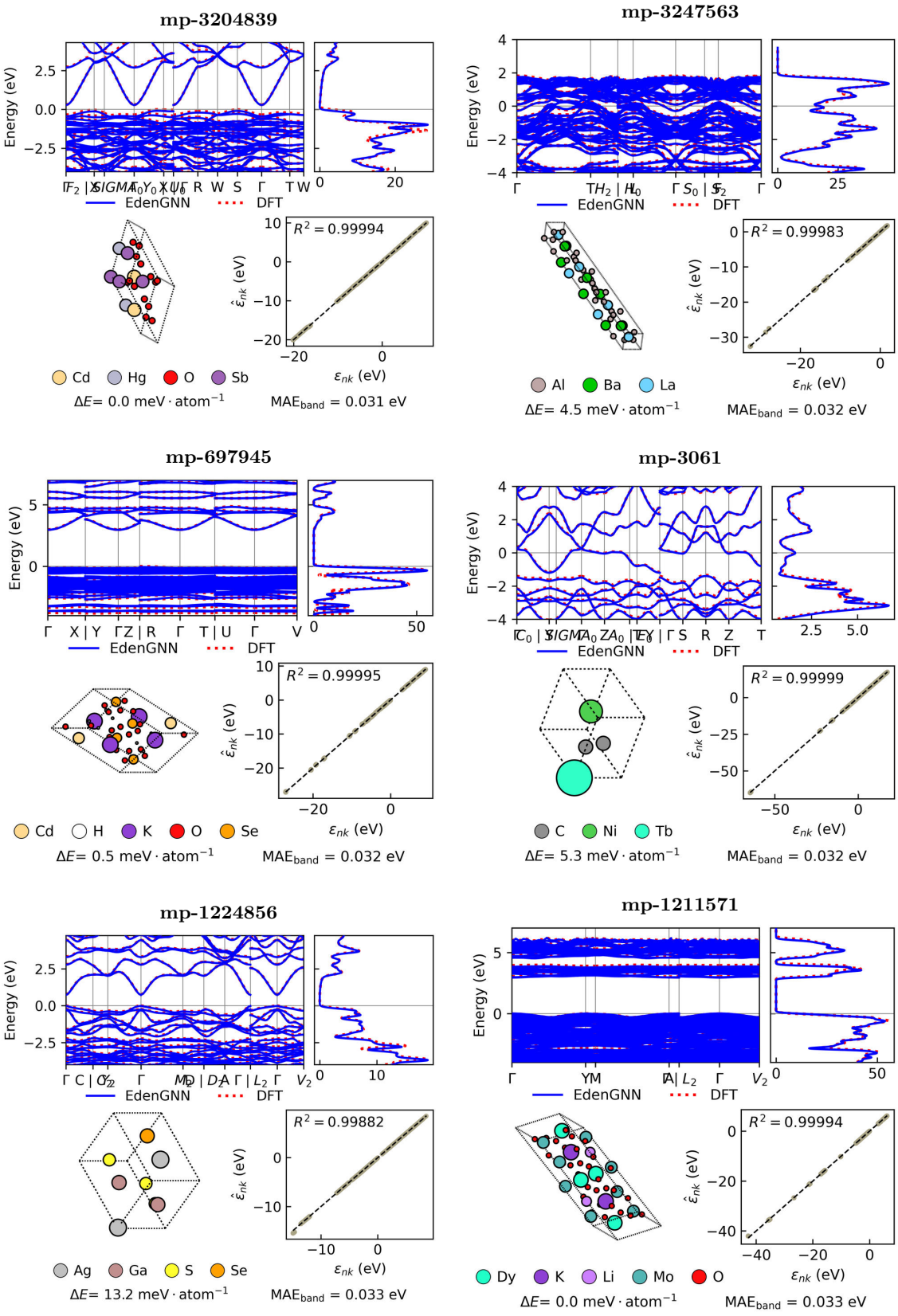

mp-1228059
Energy (eV)
Γ Y M A Γ | L_2 Γ V_2
EdenGNN DFT
$R^2 = 0.99997$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Ba Mg O Si Sr
ΔE= 0.0 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.013 eV
mp-1215201
Γ M K Γ A L H A | L H K
$R^2 = 1.00000$
Sc Zr
ΔE= 0.4 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.014 eV
mp-1246049
Γ A | L_2 Γ Y | L_2 Γ V_2
$R^2 = 0.99994$
Ga N Zn
ΔE= 12.2 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.014 eV
mp-1218528
Γ C | C_2 Γ M_2 | D_2 A | L_2 Γ V_2
$R^2 = 0.99979$
Cd Ga Sr
ΔE= 3.2 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.014 eV
mp-3249424
Γ F_2 | X SIGMA Y Y_0 | U Γ R W S Γ T W
$R^2 = 1.00000$
Al Cu Yb
ΔE= 1.9 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.015 eV
mp-3346176
C | C_2 Γ M_2 | D_2 A | L_2 Γ V_2
$R^2 = 0.99998$
Au Hf Sc
ΔE= 1.0 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.015 eV

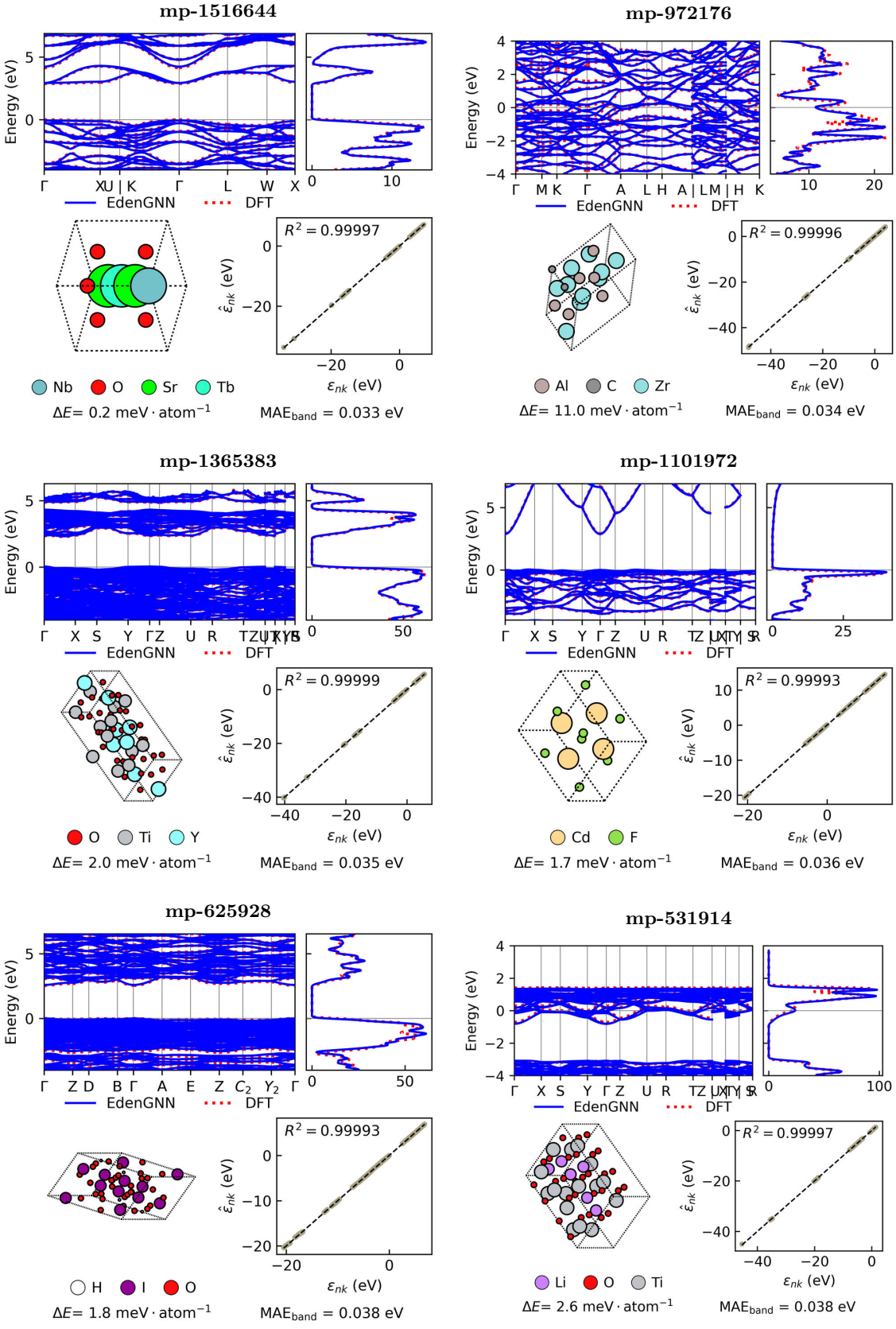

mp-1214541
Energy (eV)
Γ ZD BΓ A E Z C_2 Y_2 Γ
EdenGNN DFT
$R^2 = 0.99999$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Al Ba Dy O
ΔE= 0.2 meV · atom^{-1}
MAE$_{band}$ = 0.017 eV
mp-1725
Energy (eV)
EdenGNN DFT
$R^2 = 0.99996$
Ca Zn
ΔE= 0.4 meV · atom^{-1}
MAE$_{band}$ = 0.018 eV
mp-1208862
Energy (eV)
Γ Z D B Γ A E Z C_2 Y_2 Γ
EdenGNN DFT
$R^2 = 0.99997$
B Mg O Sm
ΔE= 1.1 meV · atom^{-1}
MAE$_{band}$ = 0.020 eV
mp-1204549
Energy (eV)
Γ Z D B Γ A E Z C_2 Y_2 Γ
EdenGNN DFT
$R^2 = 0.99997$
B Cl Er O
ΔE= 0.8 meV · atom^{-1}
MAE$_{band}$ = 0.020 eV
mp-3261367
Energy (eV)
EdenGNN DFT
$R^2 = 0.99993$
Pd Tm
ΔE= 0.5 meV · atom^{-1}
MAE$_{band}$ = 0.021 eV
mp-3269152
Energy (eV)
Γ X S Y ΓZ U R
EdenGNN DFT
$R^2 = 0.99999$
Au Ca Sn
ΔE= 3.0 meV · atom^{-1}
MAE$_{band}$ = 0.022 eV

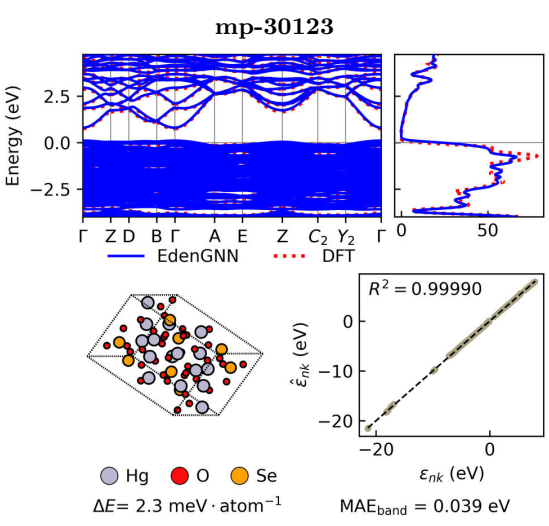

mp-3230909
Energy (eV)
EdenGNN
DFT
$R^2 = 0.99998$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Ho In Pa Rh
ΔE= 0.4 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.022 eV
mp-3256557
$R^2 = 0.99997$
Hg Tb Tm
ΔE= 5.6 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.023 eV
mp-1186786
$R^2 = 0.99997$
As Au Pr
ΔE= 2.0 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.023 eV
mp-1195598
$R^2 = 0.99997$
Ba Er O Si
ΔE= 0.5 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.023 eV
mp-1215736
$R^2 = 0.99995$
O Yb Zr
ΔE= 13.9 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.024 eV
mp-1228113
$R^2 = 0.99998$
Ba F O P
ΔE= 0.6 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.025 eV

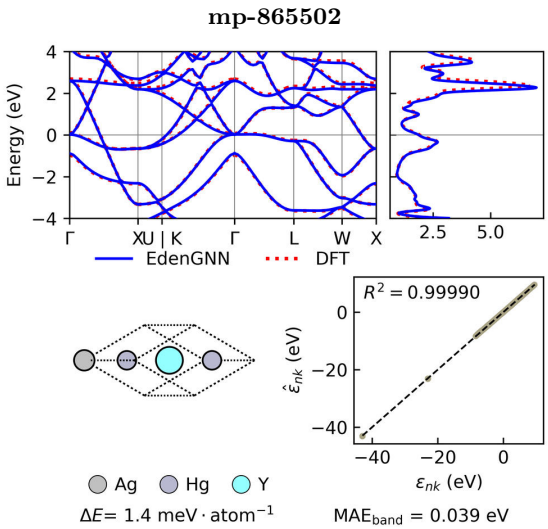

mp-22346
Energy (eV)
Γ X P N Γ M | S R X G M
EdenGNN
DFT
$R^2 = 0.99986$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Ba Ge Pt
ΔE= 7.3 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.025 eV
mp-3250327
Γ M K Γ A L H A | L M | H K
$R^2 = 0.99995$
Ag Mg Si Sm
ΔE= 3.0 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.025 eV
mp-1185622
Γ X U | K Γ L W X
$R^2 = 0.99996$
Mg Zn Zr
ΔE= 3.3 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.025 eV
mp-3263900
Γ Z D B Γ A E Z C_2 Y_2 Γ
$R^2 = 0.99960$
Ga Mg Pt
ΔE= 13.2 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.027 eV
mp-3205032
Γ T H_2 | H_0 L Γ S_0 | S_2 Γ
$R^2 = 0.99999$
Ag Li Si Zn
ΔE= 0.3 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.027 eV
mp-1199063
Γ T H_2 | H_0 L Γ S_0 | S_2 Γ
$R^2 = 0.99997$
C Si
ΔE= 0.6 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.027 eV

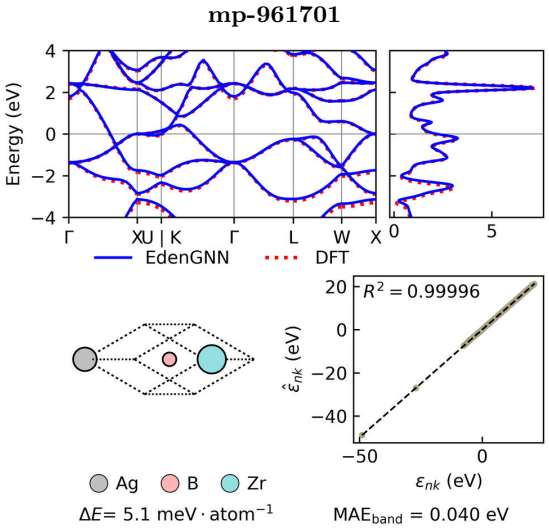

mp-556841
Energy (eV)
EdenGNN
DFT
$R^2 = 0.99992$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
La S Tm
ΔE= 2.8 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.028 eV
mp-3211601
$R^2 = 0.99998$
N Pd Ti Zr
ΔE= 2.5 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.029 eV
mp-1211207
$R^2 = 0.99999$
In Nd Ni
ΔE= 3.2 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.030 eV
mp-1233344
$R^2 = 0.99988$
Mg O P Sn Sr Zn
ΔE= 3.6 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.030 eV
mp-3231171
$R^2 = 0.99973$
Cu Ge Nd Th
ΔE= 29.7 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.031 eV
mp-1198197
$R^2 = 0.99989$
H Mg O S
ΔE= 4.5 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.031 eV

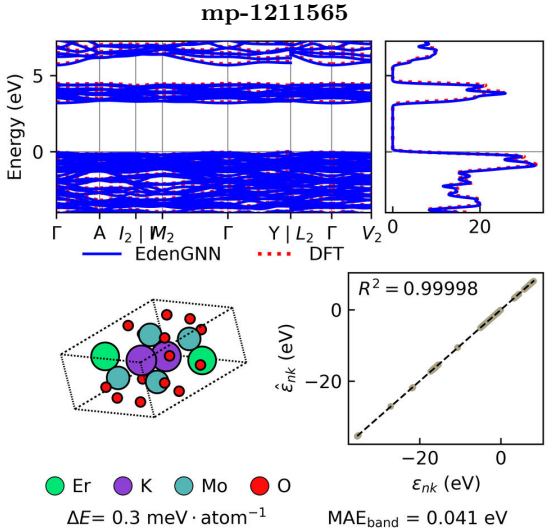

mp-3204839
Energy (eV)
EdenGNN
DFT
$R^2 = 0.99994$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Cd Hg O Sb
ΔE= 0.0 meV · atom^{-1}
MAE$_{band}$ = 0.031 eV
mp-3247563
$R^2 = 0.99983$
Al Ba La
ΔE= 4.5 meV · atom^{-1}
MAE$_{band}$ = 0.032 eV
mp-697945
$R^2 = 0.99995$
Cd H K O Se
ΔE= 0.5 meV · atom^{-1}
MAE$_{band}$ = 0.032 eV
mp-3061
$R^2 = 0.99999$
C Ni Tb
ΔE= 5.3 meV · atom^{-1}
MAE$_{band}$ = 0.032 eV
mp-1224856
$R^2 = 0.99882$
Ag Ga S Se
ΔE= 13.2 meV · atom^{-1}
MAE$_{band}$ = 0.033 eV
mp-1211571
$R^2 = 0.99994$
Dy K Li Mo O
ΔE= 0.0 meV · atom^{-1}
MAE$_{band}$ = 0.033 eV

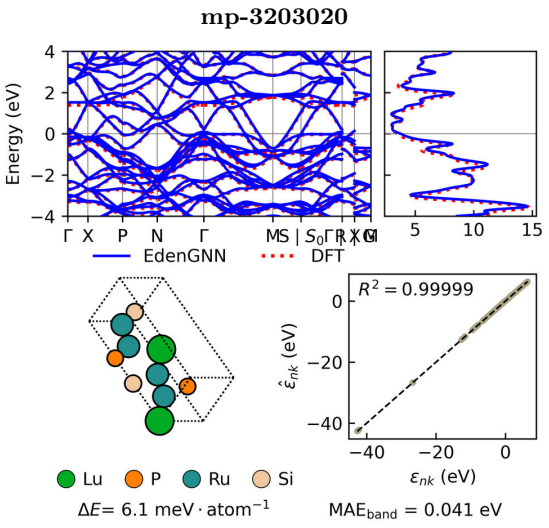

mp-1516644
Energy (eV)
Γ X U | K Γ L W X
EdenGNN DFT
$R^2 = 0.99997$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Nb O Sr Tb
ΔE= 0.2 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.033 eV
mp-972176
Γ M K Γ A L H A | L M | H K
$R^2 = 0.99996$
Al C Zr
ΔE= 11.0 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.034 eV
mp-1365383
Γ X S Y Γ Z U R T Z U | X Y | S
$R^2 = 0.99999$
O Ti Y
ΔE= 2.0 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.035 eV
mp-1101972
Γ X S Y Γ Z U R T Z | U X | T Y | S R
$R^2 = 0.99993$
Cd F
ΔE= 1.7 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.036 eV
mp-625928
Γ Z D B Γ A E Z C_2 Y_2 Γ
$R^2 = 0.99993$
H I O
ΔE= 1.8 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.038 eV
mp-531914
Γ X S Y Γ Z U R T Z | U X | T Y | S R
$R^2 = 0.99997$
Li O Ti
ΔE= 2.6 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.038 eV

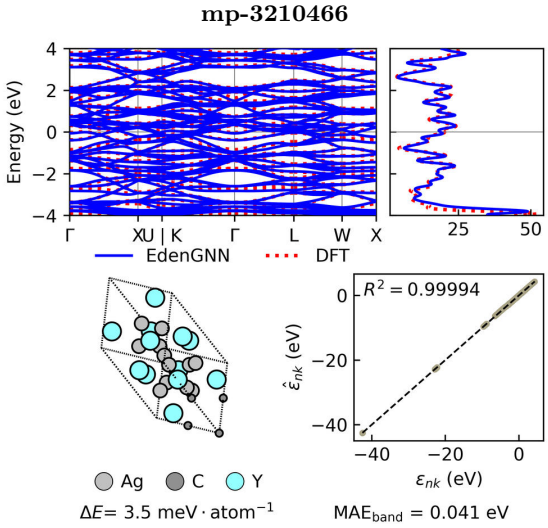

mp-30123
Energy (eV)
Γ Z D B Γ A E Z C_2 Y_2 Γ
EdenGNN
DFT
$R^2 = 0.99990$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Hg O Se
ΔE= 2.3 meV · atom^{-1}
MAE$_{band}$ = 0.039 eV

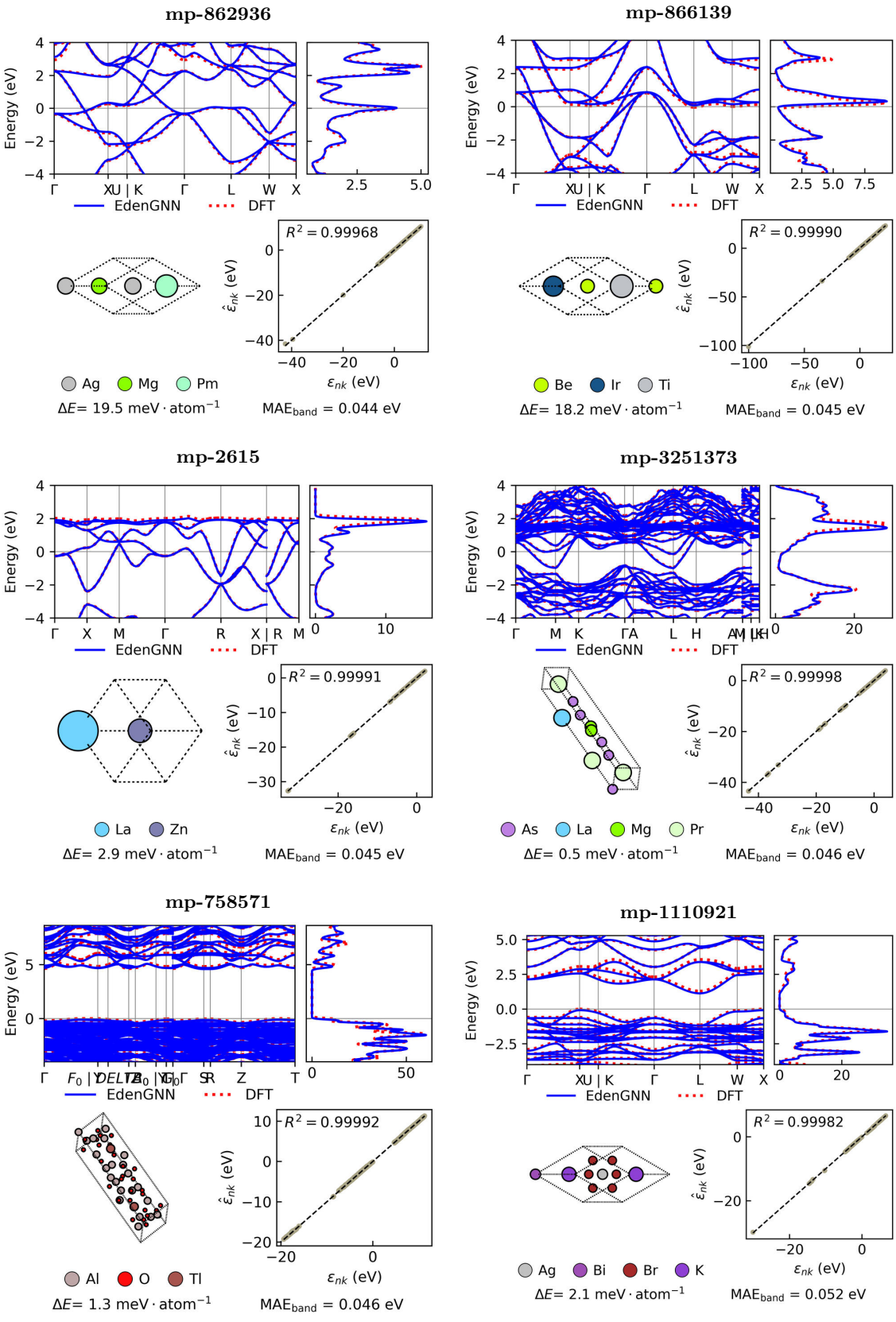

mp-865502
Energy (eV)
Γ XU | K Γ L W X
EdenGNN
DFT
$R^2 = 0.99990$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Ag Hg Y
ΔE= 1.4 meV · atom^{-1}
MAE$_{band}$ = 0.039 eV

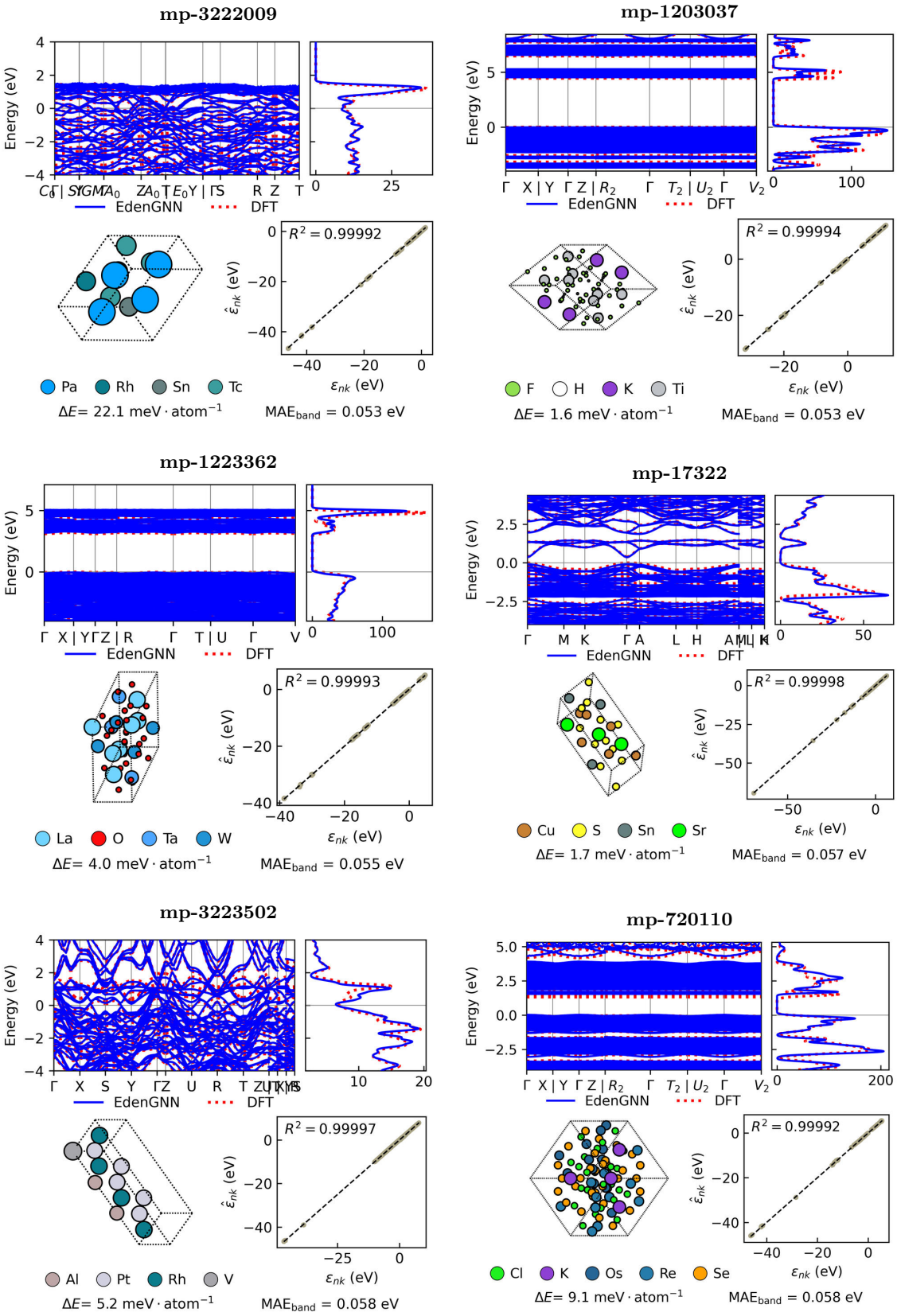

mp-961701
Energy (eV)
Γ XU | K Γ L W X
EdenGNN
DFT
$R^2 = 0.99996$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Ag B Zr
ΔE= 5.1 meV · atom^{-1}
MAE$_{band}$ = 0.040 eV

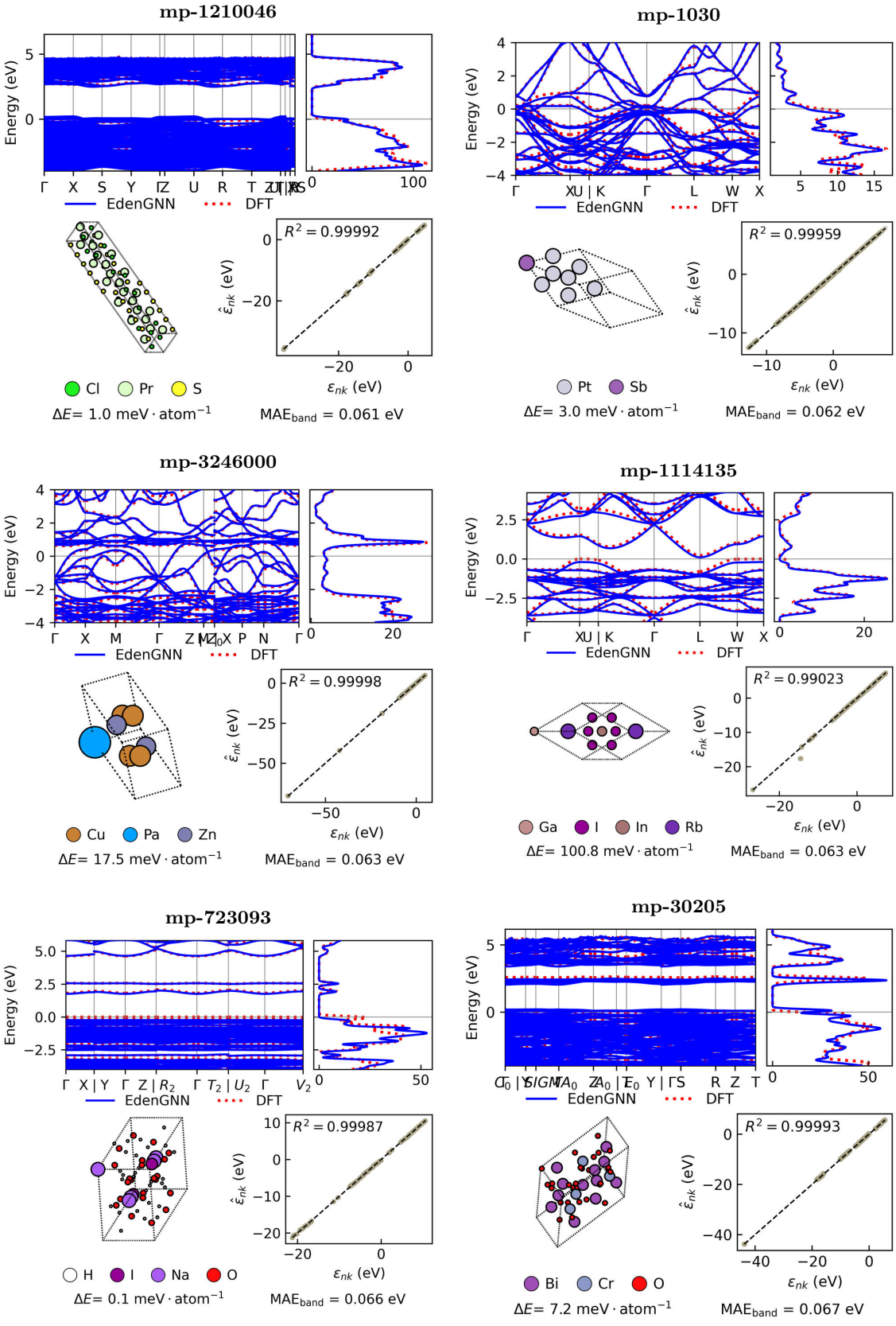

mp-1211565
Energy (eV)
Γ A I_2 | M_2 Γ Y | L_2 Γ V_2
EdenGNN
DFT
$R^2 = 0.99998$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Er K Mo O
ΔE= 0.3 meV · atom^{-1}
MAE$_{band}$ = 0.041 eV

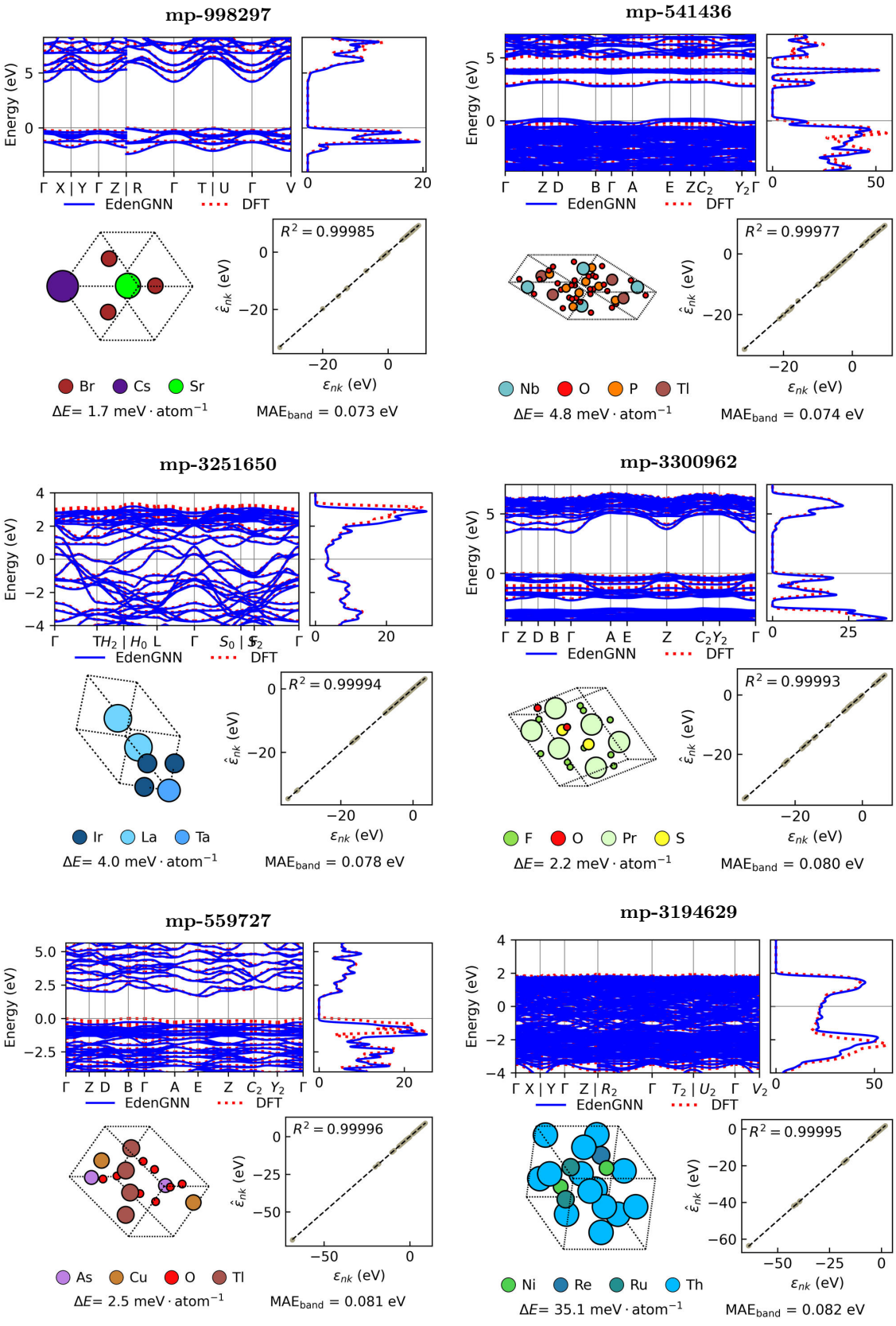

mp-3203020
Energy (eV)
Γ X P N Γ MS | S_0 Γ R X M
EdenGNN
DFT
$R^2 = 0.99999$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Lu P Ru Si
ΔE= 6.1 meV · atom^{-1}
MAE$_{band}$ = 0.041 eV

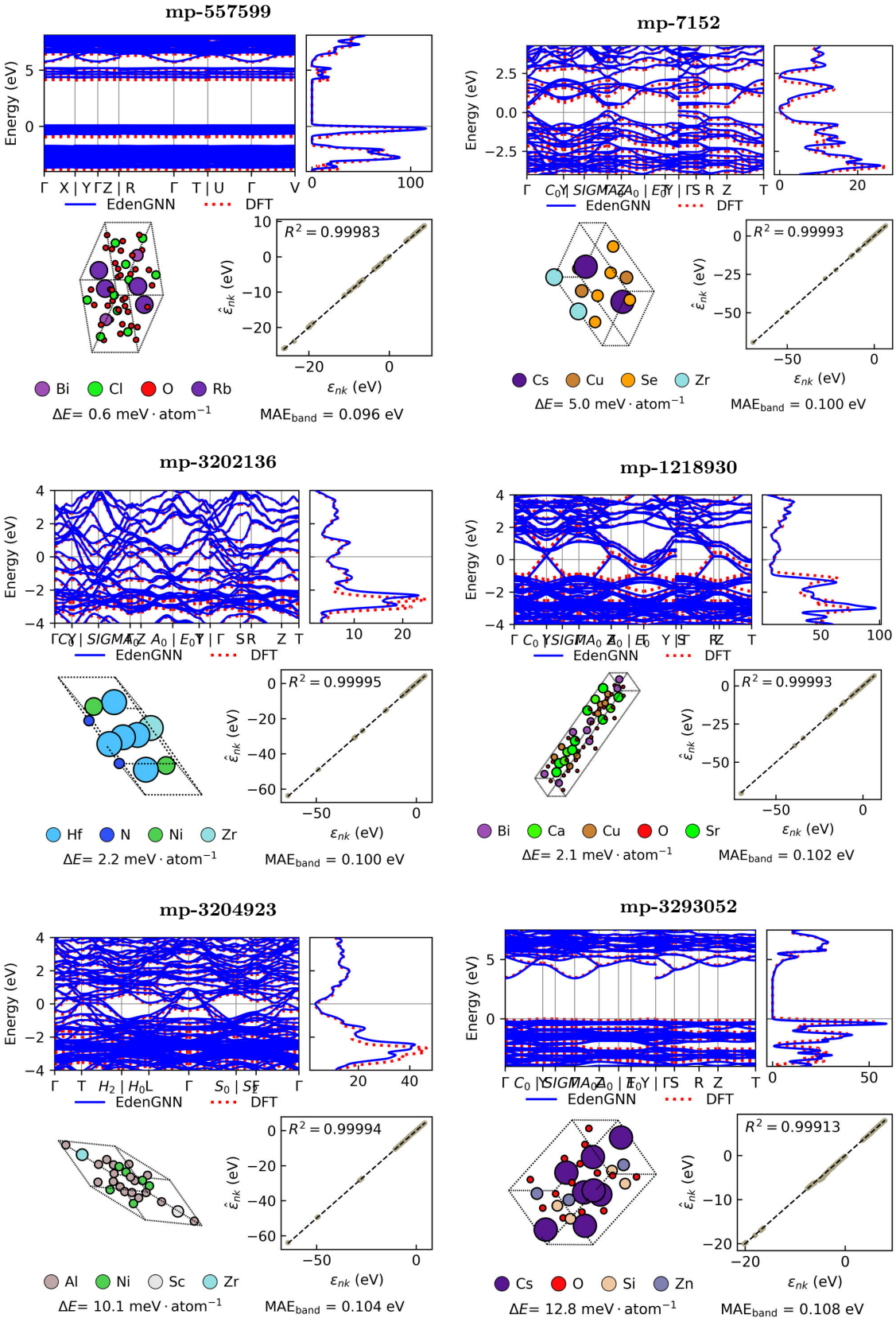

mp-3210466
Energy (eV)
Γ XU | K Γ L W X
EdenGNN
DFT
$R^2 = 0.99994$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Ag C Y
ΔE= 3.5 meV · atom^{-1}
MAE$_{band}$ = 0.041 eV

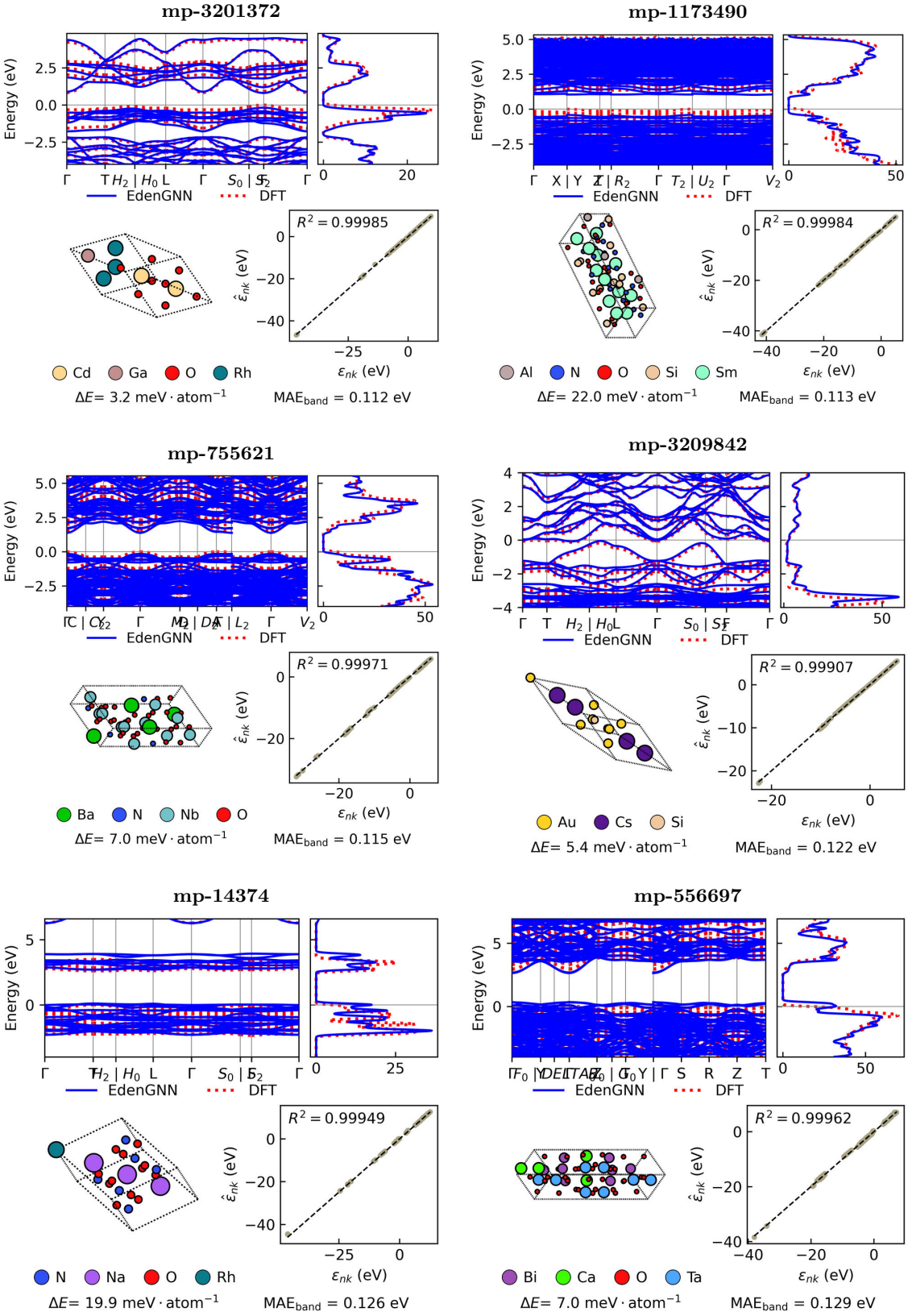

mp-862936
Energy (eV)
Γ XU | K Γ L W X
EdenGNN DFT
$R^2 = 0.99968$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Ag Mg Pm
ΔE= 19.5 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.044 eV
mp-866139
Energy (eV)
Γ XU | K Γ L W X
EdenGNN DFT
$R^2 = 0.99990$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Be Ir Ti
ΔE= 18.2 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.045 eV
mp-2615
Energy (eV)
Γ X M Γ R X | R M
EdenGNN DFT
$R^2 = 0.99991$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
La Zn
ΔE= 2.9 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.045 eV
mp-3251373
Energy (eV)
Γ M K ΓA L H A | H
EdenGNN DFT
$R^2 = 0.99998$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
As La Mg Pr
ΔE= 0.5 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.046 eV
mp-758571
Energy (eV)
Γ F_0 | Y DELTA Z_0 | Y G_0 Γ SR Z T
EdenGNN DFT
$R^2 = 0.99992$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Al O Tl
ΔE= 1.3 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.046 eV
mp-1110921
Energy (eV)
Γ XU | K Γ L W X
EdenGNN DFT
$R^2 = 0.99982$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Ag Bi Br K
ΔE= 2.1 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.052 eV

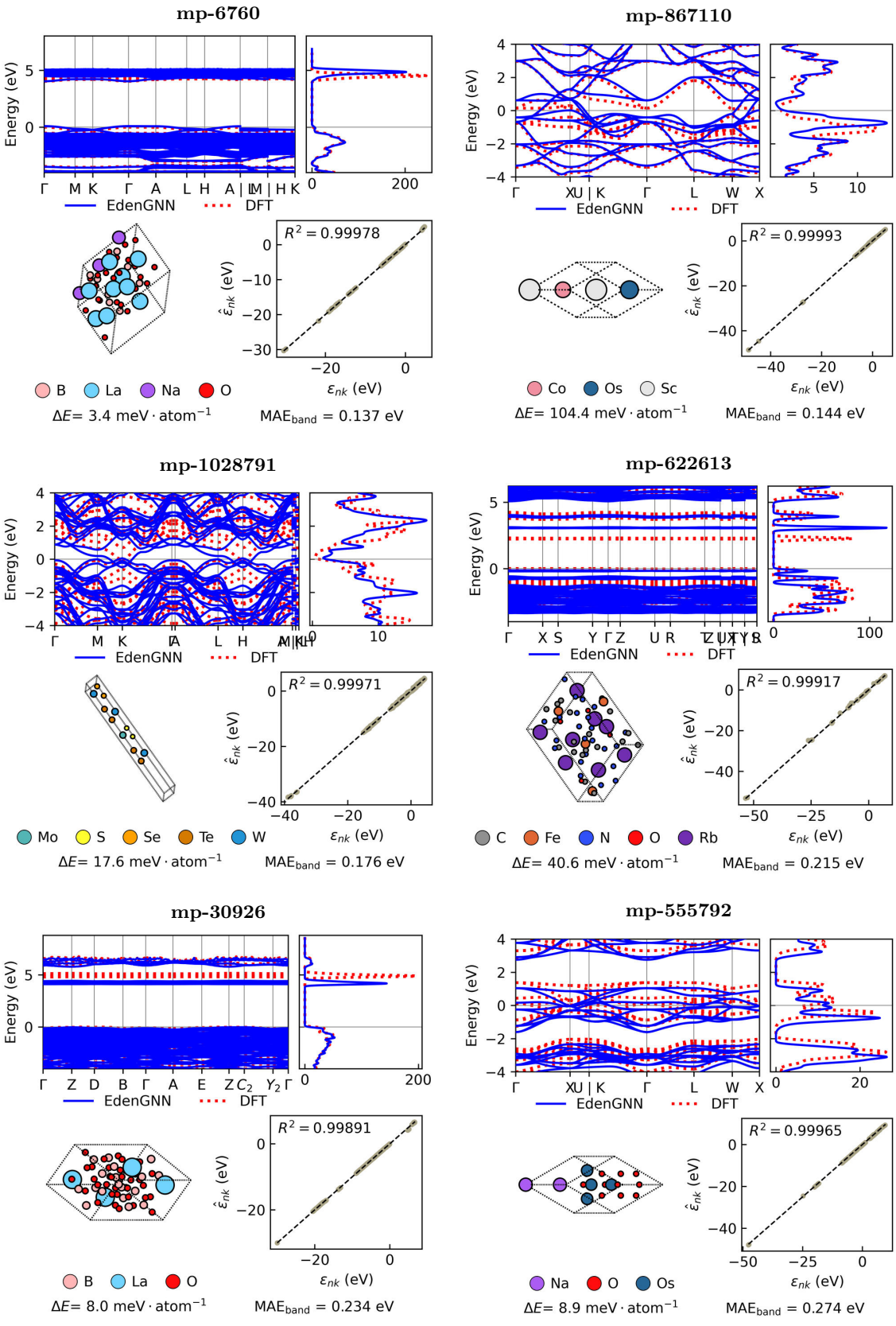

mp-3222009
Energy (eV)
EdenGNN
DFT
$R^2 = 0.99992$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Pa Rh Sn Tc
ΔE= 22.1 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.053 eV
mp-1203037
$R^2 = 0.99994$
F H K Ti
ΔE= 1.6 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.053 eV
mp-1223362
$R^2 = 0.99993$
La O Ta W
ΔE= 4.0 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.055 eV
mp-17322
$R^2 = 0.99998$
Cu S Sn Sr
ΔE= 1.7 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.057 eV
mp-3223502
$R^2 = 0.99997$
Al Pt Rh V
ΔE= 5.2 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.058 eV
mp-720110
$R^2 = 0.99992$
Cl K Os Re Se
ΔE= 9.1 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.058 eV

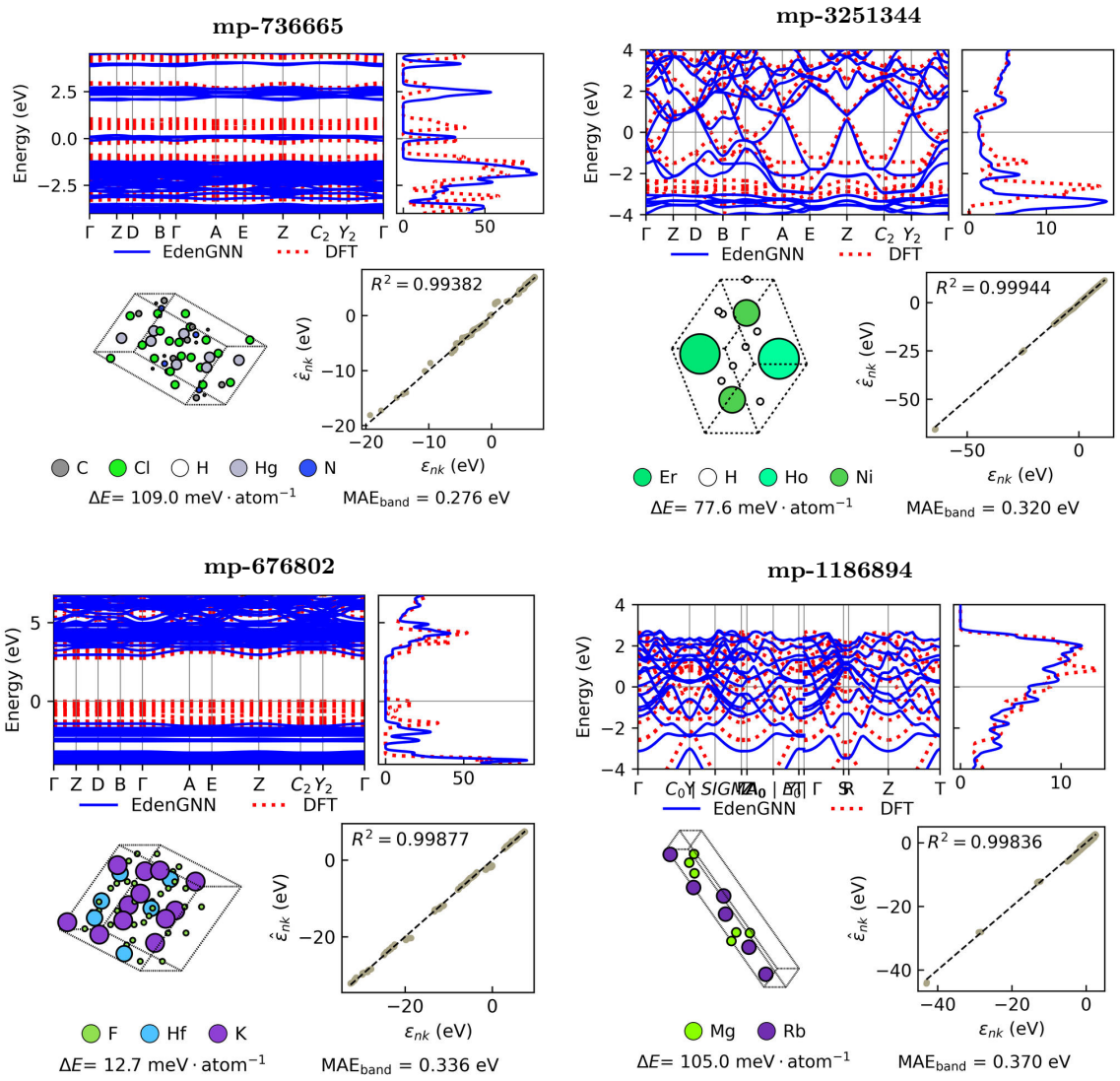

mp-1210046
Energy (eV)
Γ X S Y Γ Z U R T Z|Y|T|R|S
EdenGNN
DFT
Cl Pr S
$R^2 = 0.99992$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
ΔE= 1.0 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.061 eV
mp-1030
Γ X|U | K Γ L W X
Pt Sb
$R^2 = 0.99959$
ΔE= 3.0 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.062 eV
mp-3246000
Γ X M Γ Z|M$_0$|Z$_0$ X P N Γ
Cu Pa Zn
$R^2 = 0.99998$
ΔE= 17.5 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.063 eV
mp-1114135
Γ X|U | K Γ L W X
Ga I In Rb
$R^2 = 0.99023$
ΔE= 100.8 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.063 eV
mp-723093
Γ X | Y Γ Z | R_2 Γ T_2 | U_2 Γ V_2
H I Na O
$R^2 = 0.99987$
ΔE= 0.1 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.066 eV
mp-30205
C_0 | Y $SIGM$ A_0 Z_0 | E_0 Y | Γ S R Z T
Bi Cr O
$R^2 = 0.99993$
ΔE= 7.2 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.067 eV


mp-998297
Energy (eV)
Γ X | Y Γ Z | R Γ T | U Γ V
EdenGNN
DFT
$R^2 = 0.99985$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Br Cs Sr
ΔE= 1.7 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.073 eV
mp-541436
Γ Z D B Γ A E ZC_2 Y_2 Γ
$R^2 = 0.99977$
Nb O P Tl
ΔE= 4.8 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.074 eV
mp-3251650
Γ TH_2 | H_0 L Γ S_0 | S_2 Γ
$R^2 = 0.99994$
Ir La Ta
ΔE= 4.0 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.078 eV
mp-3300962
Γ Z D B Γ A E Z C_2Y_2 Γ
$R^2 = 0.99993$
F O Pr S
ΔE= 2.2 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.080 eV
mp-559727
Γ Z D B Γ A E Z C_2 Y_2 Γ
$R^2 = 0.99996$
As Cu O Tl
ΔE= 2.5 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.081 eV
mp-3194629
Γ X | Y Γ Z | R_2 Γ T_2 | U_2 Γ V_2
$R^2 = 0.99995$
Ni Re Ru Th
ΔE= 35.1 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.082 eV


mp-557599
Energy (eV)
EdenGNN
DFT
$R^2 = 0.99983$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Bi Cl O Rb
ΔE= 0.6 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.096 eV
mp-7152
$R^2 = 0.99993$
Cs Cu Se Zr
ΔE= 5.0 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.100 eV
mp-3202136
$R^2 = 0.99995$
Hf N Ni Zr
ΔE= 2.2 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.100 eV
mp-1218930
$R^2 = 0.99993$
Bi Ca Cu O Sr
ΔE= 2.1 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.102 eV
mp-3204923
$R^2 = 0.99994$
Al Ni Sc Zr
ΔE= 10.1 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.104 eV
mp-3293052
$R^2 = 0.99913$
Cs O Si Zn
ΔE= 12.8 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.108 eV


mp-3201372
Energy (eV)
EdenGNN
DFT
$R^2 = 0.99985$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
Cd Ga O Rh
ΔE= 3.2 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.112 eV
mp-1173490
$R^2 = 0.99984$
Al N O Si Sm
ΔE= 22.0 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.113 eV
mp-755621
$R^2 = 0.99971$
Ba N Nb O
ΔE= 7.0 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.115 eV
mp-3209842
$R^2 = 0.99907$
Au Cs Si
ΔE= 5.4 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.122 eV
mp-14374
$R^2 = 0.99949$
N Na O Rh
ΔE= 19.9 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.126 eV
mp-556697
$R^2 = 0.99962$
Bi Ca O Ta
ΔE= 7.0 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.129 eV


mp-6760
Energy (eV)
Γ M K Γ A L H A | L M | H K
EdenGNN
DFT
$R^2 = 0.99978$
$\hat{\varepsilon}_{nk}$ (eV)
ε_{nk} (eV)
B La Na O
ΔE= 3.4 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.137 eV
mp-867110
Γ X U | K Γ L W X
$R^2 = 0.99993$
Co Os Sc
ΔE= 104.4 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.144 eV
mp-1028791
Γ M K A | L H A | L H
$R^2 = 0.99971$
Mo S Se Te W
ΔE= 17.6 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.176 eV
mp-622613
Γ X S Y Γ Z U R T Z | U X | Y R
$R^2 = 0.99917$
C Fe N O Rb
ΔE= 40.6 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.215 eV
mp-30926
Γ Z D B Γ A E Z C_2 Y_2 Γ
$R^2 = 0.99891$
B La O
ΔE= 8.0 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.234 eV
mp-555792
Γ X U | K Γ L W X
$R^2 = 0.99965$
Na O Os
ΔE= 8.9 meV · atom^{-1}
MAE$_{\text{band}}$ = 0.274 eV



Supplementary Fig. 7: Band structure and density of state comparisons for 100 randomly sampled structures from the Materials Project database. Each panel is labeled by its corresponding Materials Project ID. The plots are arranged in ascending order in terms of the MAE_{band} metric to illustrate the model's performance distribution.

References

[1] Mario Geiger and Tess Smidt. e3nn: Euclidean Neural Networks, 2022.

[2] Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, May 2022.

[3] Andrea Grisafi, Alberto Fabrizio, Benjamin Meyer, David M. Wilkins, Clemence Corminboeuf, and Michele Ceriotti. Transferable Machine-Learning Model of the Electron Density. *ACS Central Science*, 5(1):57–64, January 2019.

[4] Sheng Gong, Tian Xie, Taishan Zhu, Shuo Wang, Eric R. Fadel, Yawei Li, and Jeffrey C. Grossman. Predicting charge density distribution of materials using a local-environment-based graph convolutional network. *Physical Review B*, 100(18):184103, November 2019.

[5] Peter Bjørn Jørgensen and Arghya Bhowmik. Equivariant graph neural networks for fast electron density estimation of molecules, liquids, and solids. *npj Computational Materials*, 8(1):183, August 2022.

[6] Teddy Koker, Keegan Quigley, Eric Taw, Kevin Tibbetts, and Lin Li. Higher-order equivariant neural networks for charge density prediction in materials. *npj Computational Materials*, 10(1):161, July 2024.

[7] Aleksandrovich Varshalovich, Dmitriĭ, Anatolï Nikolaevitch Moskalev, and Valerij Kel'manovĭc Khersonskii. *Quantum theory of angular momentum*. World Scientific, 1988.

[8] Weiqing Zhou, Daye Zheng, Qianrui Liu, Denghui Lu, Yu Liu, Peize Lin, Yike Huang, Xingliang Peng, Jie J Bao, Chun Cai, et al. Abacus: An electronic structure analysis package for the ai era. *The Journal of Chemical Physics*, 163(19), 2025.

[9] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL materials*, 1(1), 2013.

[10] Rhys EA Goodall, Abhijith S Parackal, Felix A Faber, Rickard Armiento, and Alpha A Lee. Rapid discovery of stable materials by coordinate-free coarse graining. *Science advances*, 8(30):eabn4117, 2022.

[11] Amil Merchant, Simon Batzner, Samuel S. Schoenholz, Muratahan Aykol, Gowoon Cheon, and Ekin Dogus Cubuk. Scaling deep learning for materials discovery. *Nature*, 2023.