

<https://doi.org/10.1038/s41524-025-01728-w>

Modeling crystal defects using defect informed neural networks

Ziduo Yang^{1,2}, Xiaoqing Liu², Xiuying Zhang², Pengru Huang³✉, Kostya S. Novoselov³ & Lei Shen^{2,4}✉

Most AI-for-Materials research to date has focused on ideal crystals, whereas real-world materials inevitably contain defects that play a critical role in modern functional technologies. The defects break geometric symmetry and increase interaction complexity, posing particular challenges for traditional ML models. Here, we introduce Defect-Informed Equivariant Graph Neural Network (DefiNet), a model specifically designed to accurately capture defect-related interactions and geometric configurations in point-defect structures. DefiNet achieves near-DFT-level structural predictions in milliseconds using a single GPU. To validate its accuracy, we perform DFT relaxations using DefiNet-predicted structures as initial configurations and measure the residual ionic steps. For most defect structures, regardless of defect complexity or system size, only 3 ionic steps are required to reach the DFT-level ground state. Finally, comparisons with scanning transmission electron microscopy (STEM) images confirm DefiNet's scalability and extrapolation beyond point defects, positioning it as a valuable tool for defect-focused materials research.

Studying crystalline materials and their devices necessarily requires investigating defects. On the one hand, defects are intrinsic and unavoidable in crystals, often significantly limiting device performance. On the other hand, defect engineering, the deliberate introduction of extrinsic defects into materials, is crucial for unlocking novel properties and functionalities in crystalline materials, enabling advancements in modern functional technologies^{1–4}.

The defect space is primarily defined by three variables: the host structure, the types of defects, and defect configurations². The types of defects are limited to a few categories, such as intrinsic vacancies and impurity substitutions. However, the space for defect configurations is immense, making thorough experimental or computational investigations very challenging⁵. These defects typically induce local lattice distortions. To optimize the defect structures, one typically performs conventional *ab initio* methods such as density functional theory (DFT), as depicted in Fig. 1a. DFT calculations involve iterative electronic and ionic steps that gradually converge the system to its lowest energy configuration. These steps are computationally expensive, with the time scaling approximately as N^3 where N is the number of atoms, making DFT calculations particularly challenging for large or complex systems.

The emerging technique of machine learning (ML) interatomic potentials^{6–11} has shown the potential in reducing computational demands associated with defect structure optimization. By training a

graph neural network (GNN) to iteratively approximate physical quantities such as energies, forces, and stresses, ML-potential relaxation bypasses the computationally intensive electronic step while retaining the ionic step, as shown in Fig. 1b. For example, Mosquera-Lois et al.¹² and Jiang et al.¹³ have demonstrated that ML interatomic potentials can provide both cost-effectiveness and accuracy in identifying the ground-state configurations of defect structures. Despite these advantages, two primary challenges remain in applying ML interatomic potentials to the study of defect structures. First, existing ML interatomic potentials do not explicitly consider the complicated defect-related interactions. Second, the development of ML interatomic potentials heavily relies on the availability of comprehensive databases with detailed labels for energy, forces, or stresses during structural relaxations, which may not always be available for complex defect systems.

To overcome these challenges, we develop the Defect-Informed Equivariant Graph Neural Network (DefiNet), a single-step ML model specifically designed for the rapid relaxation of defect crystal structures without requiring any iterative process, as shown in Fig. 1c. DefiNet offers four key advantages:

1) Defect-explicit representation—Conventional GNNs model defect structures using defect-implicit graphs, in which no explicit flags denote defect sites and the network must infer them implicitly from structures, as shown in Fig. 1d. DefiNet instead builds a single host-structure graph and

¹Department of Electronic Engineering, College of Information Science and Technology, Jinan University, Guangzhou, China. ²Department of Mechanical Engineering, National University of Singapore, Singapore, Singapore. ³Institute for Functional Intelligent Materials, National University of Singapore, Singapore, Singapore. ⁴National University of Singapore (Chongqing) Research Institute, Chongqing, China. ✉e-mail: pengru@nus.edu.sg; shenlei@nus.edu.sg

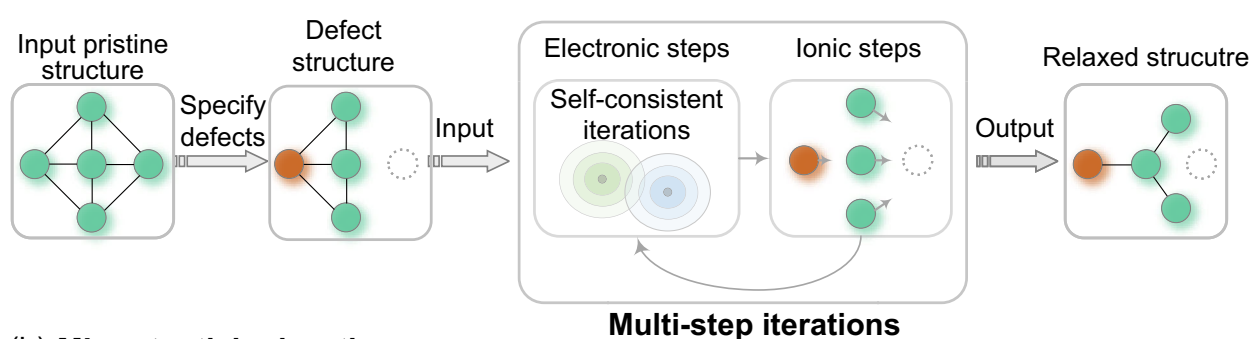
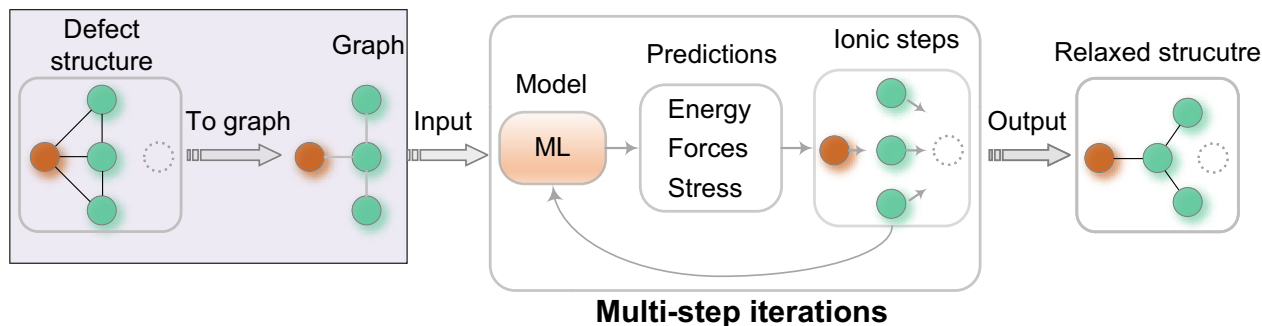
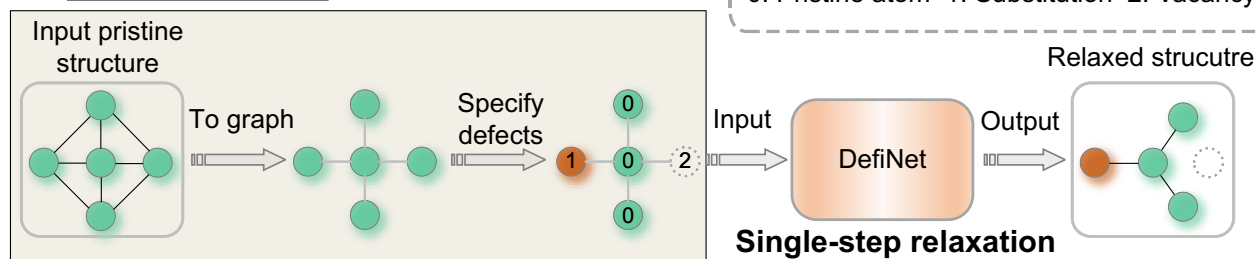
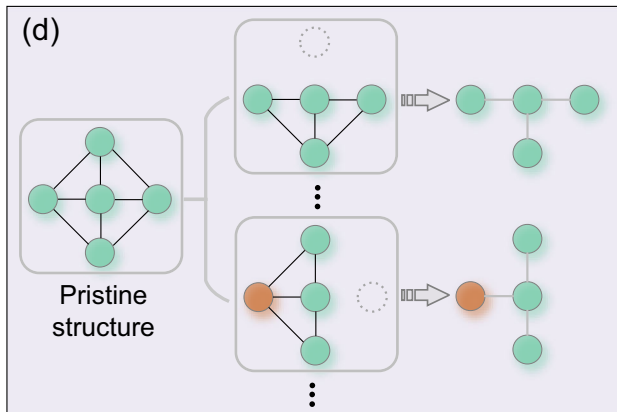
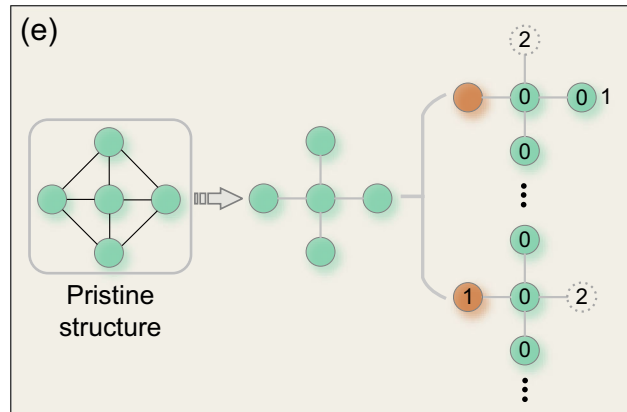
(a) DFT relaxation**(b) ML-potential relaxation****(c) DefiNet relaxation****(d)****Defect-implicit graph representation****(e)****Defect-explicit graph representation**

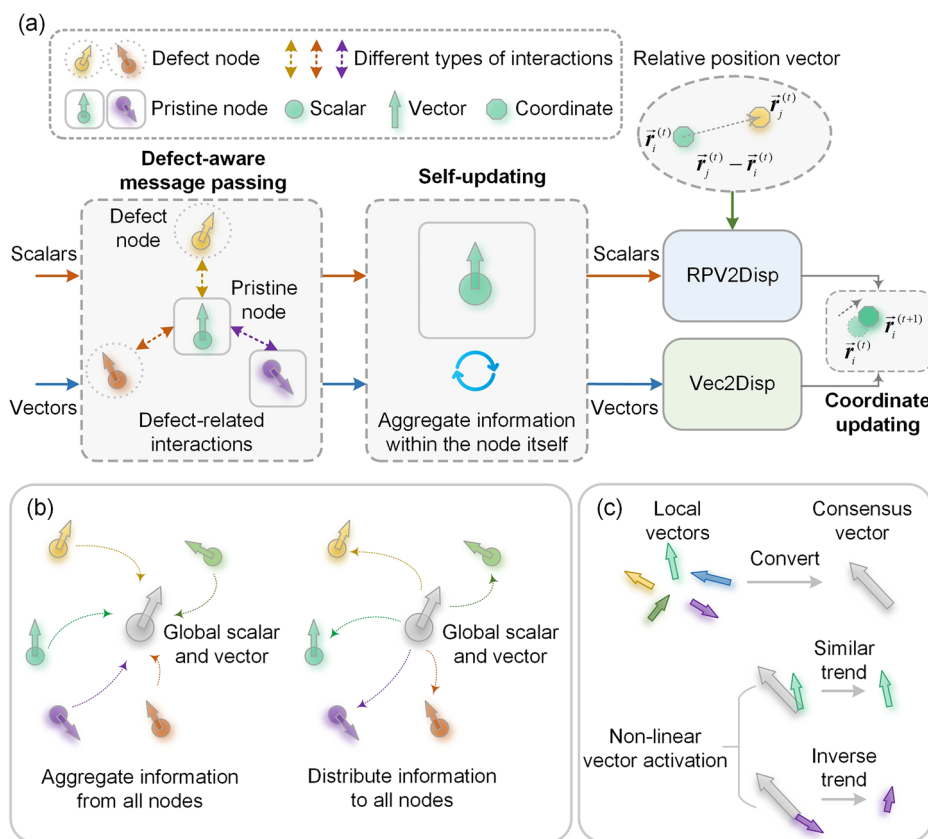
Fig. 1 | Overview of crystal defect structure relaxation methods. **a** Relaxation using DFT with multi-step iterations. **b** Relaxation using ML potentials with multi-step iterations. **c** Relaxation using our DefiNet with a single step. **d** Defect-implicit graph

used by standard GNN workflows, where defect sites are not explicitly labeled. **e** Defect-explicit graph introduced here, in which nodes carry explicit markers (0 = pristine atom, 1 = substitution, 2 = vacancy) to identify defects.

attaches markers to nodes to explicitly denote defects, yielding a defect-explicit graph (Fig. 1e). Combined with our defect-aware message passing scheme, this design captures complex defect-defect and defect-host interactions more accurately.

2) End-to-end trainability—DefiNet directly maps initial structures to relaxed configurations, enabling efficient end-to-end training and scalable parallel computing capabilities. This makes it highly suitable for large-scale calculations as it completely eliminates iterative relaxation steps.

Fig. 2 | Detailed designs of DefiNet. **a** Overview of the three-stage updating process, including defect-aware message passing, self-updating, and defect-aware coordinate updating. **b** Implementation of global node (including global scalar and global vector). **c** Non-linear vector activation technique.



3) **Equivariant representation**—The model leverages equivariant representation to ensure that rotational transformations of the input structure are consistently reflected throughout the network's layers and in the final output coordinates, leading to more precise geometric representations.

4) **Scalability**—It is well known that in conventional DFT or ML interatomic potential approaches, computational cost increases significantly with structural complexity and the total number of atoms due to their reliance on iterative algorithms. In contrast, DefiNet's single-step and end-to-end design enable it to accurately predict defect structures regardless of defect complexity or system size.

We evaluated DefiNet on 14,866 defect structures across six widely studied materials, including MoS_2 , WSe_2 , h-BN, GaSe, InSe, and black phosphorus (BP), each presenting a variety of defects. Our results show that with just a few hundred training samples per material, DefiNet achieves precise structural relaxation within tens of milliseconds using a single GPU, even without utilizing its parallel computing capabilities. To validate the accuracy and efficiency, we use the original unrelaxed structures and DefiNet-predicted structures as initial configurations for DFT calculations. DefiNet improves the computational efficiency by 87%, demonstrating DefiNet's efficiency in identifying energetically favorable configurations. Moreover, DefiNet efficiently scales from small to large systems while maintaining its ability to generalize between high- and low-defect-density scenarios. Comparisons with high-resolution scanning transmission electron microscopy (STEM) images of complex defects, such as line defects, further validate the model's scalability and extrapolation capabilities beyond point defects. Collectively, these advancements establish DefiNet as a powerful tool for defect-focused materials and device research.

Results

DefiNet architecture

Graph neural networks (GNNs) operate directly on graph-structured data, making them ideal for crystalline materials, where atoms map to nodes and

interatomic bonds to edges^{6–8,14–33}. DefiNet extends this paradigm with a defect-explicit representation: instead of relying on defect-implicit graphs, where defects must be inferred from structures, DefiNet augments a single host-structure graph with explicit markers (0 = pristine atom, 1 = substitution, 2 = vacancy) to indicate defect sites, thereby enabling the network to explicitly encode defect-related interactions during message passing.

The overall architecture of DefiNet is depicted in Fig. 2a. The model employs a vector-scalar-coordinate triplet representation for each node to encapsulate invariant, equivariant, and structural features, respectively. Scalar features encode information related to the material's properties that are invariant to geometric transformations. Vector features provide geometrical information that is equivariant to rotations. The initial coordinates are updated through successive layers to optimize the structure toward a more stable state.

DefiNet updates this triplet representation through a three-stage graph convolution process, as illustrated in Fig. 2a. The process begins with defect-aware message passing, in which neighboring nodes exchange information through marker-conditioned edges (i.e., defect-defect, defect-pristine, and pristine-pristine) so that the propagated messages explicitly encode both the presence and the category of each defect. The self-updating stage then updates the scalar and vector features using the node's internal information. The final stage, defect-aware coordinate updating, optimizes atom coordinates using two specific modules, namely the Relative Position Vector to Displacement (RPV2Disp) and Vector to Displacement (Vec2Disp). These modules predict the necessary displacements to move each atom toward an optimized structure.

DefiNet further incorporates two technologies to boost model performance. First, it adopts the global node (including global scalar and global vector) introduced by Yang et al.³⁴ to capture long-range interactions, as illustrated in Fig. 2b. These global components aggregate scalar and vector information from all nodes across the graph and subsequently redistribute it to each node, thereby enhancing the model's ability to identify long-range interactions effectively. Second, while non-linearity is crucial for the

expressive power of neural networks, introducing non-linearity into vector representations without compromising equivariance presents a challenge³⁵. To address this, we have introduced a novel nonlinear vector activation, as illustrated in Fig. 2c. This method computes a consensus vector by aggregating local vectors, capturing the overarching directional trend among them. Vectors that align with this consensus vector, as indicated by a dot product greater than zero, are deemed significant and retained without changes. In contrast, vectors that diverge from this consensus trend, shown by a dot product less than zero, are modified by adding the consensus vector, thus reorienting them closer to the dominant directional trend. The intuition behind this design is that if most directional features agree on a common trend, then outlier vectors that strongly deviate are likely to be noisy or weakly informative and should be softly regularized toward the consensus.

Database

We have developed a database for 2D material defects (2DMD)²⁴, to facilitate the training and evaluation of ML models for defect structure analysis. This database includes structures with point defects for commonly used 2D materials including MoS₂, WSe₂, h-BN, GaSe, InSe, and black phosphorous (BP). Details of these point defects with supercell specifications are presented in Table 1. All defects in our dataset are in the neutral charge state.

The database is divided into two sections: one with a low-density of structured defect configurations, and another with a high-density of randomly configured defects, according to the defect concentration. The low-density section includes 5933 structures each for MoS₂ and WSe₂, with defect concentration lower than 1.6% (1 to 3 defects) per structure, covering all potential configurations within an 8 × 8 supercell. The high-density section comprises randomly generated substitution and vacancy defects across all six materials. For each defect concentration—2.5%, 5%, 7.5%, 10%, and 12.5%—100 structures were created, resulting in a total of 500 configurations per material and 3000 in total. In total, the dataset contains 14,866 structures, each comprising 120–192 atoms after applying supercell expansion.

The database is stratified by material and defect density (low vs. high) and then randomly split into training, validation, and test sets in an 8:1:1 ratio. Each subset maintains the same overall data distribution but contains non-overlapping defect configurations.

Evaluation metric

We use the coordinate MAE between the ML-relaxed and DFT-relaxed structures to evaluate the model's performance. Since structural variations between unrelaxed and relaxed defect structures are primarily localized near the defect sites, we further introduce localized MAE statistics for a more precise assessment of model's performance. Specifically, we denote atoms within an $x\text{Å}$ radius of the defect sites as A_x , where x is set to 3, 4, 5, and 6 in our experiments. For example, the coordinate MAE for A_5 considers only atoms within a 5 Å radius of the defect site when calculating the MAE.

Model performance on structures with low-density defects

We first benchmark DefiNet on structures with low-density defects (defect concentration below 1.6%), comparing it against the state-of-the-art (SOTA) single-step ML model, DeepRelax³⁶. A concise comparison of the

key differences between DefiNet and DeepRelax is provided in Supplementary Note 2. As a baseline, we introduce a Dummy model that simply returns the input initial structure as its output, serving as a control reference for evaluation. All models are trained, validated, and tested on identical datasets.

Figure 3a, b presents the performance of the models, showing that both DeepRelax and DefiNet significantly outperform the Dummy model. DefiNet surpasses DeepRelax notably, achieving improvements of 78.38%, 61.86%, 64.77%, 66.67%, and 70.37% in coordinate MAE for all atoms, A_3 , A_4 , A_5 , and A_6 , respectively, across all defect structures in both materials. Additionally, DefiNet is approximately 26.2 times more computationally efficient than DeepRelax in terms of inference speed, as shown in Supplementary Table 1.

We also assess DefiNet's performance using different percentages of the training data, as shown in Supplementary Fig. 9, to investigate the relationship between dataset size and model accuracy. The results show that performance improves rapidly when increasing the training size in the low-data regime (e.g., from 10% to 30%), but the gains become increasingly marginal beyond that point. This trend suggests that DefiNet can learn effectively from limited data, while additional data primarily serves to fine-tune predictions rather than drive major improvements.

Model performance on structures with high-density defects

While low-density defects are more commonly studied, they represent only a small portion of the entire defect space. High-density defects can reveal important and unique physical phenomena that low-density studies may not capture. In particular, interactions between multiple defects can significantly influence material properties in ways that isolated defects cannot. These complex defect-related interactions pose a significant challenge for ML models.

Here, we demonstrate that DefiNet also achieves strong performance on structures with high-density defects (defect concentrations between 2.5 and 12.5%), as shown in Fig. 3c–h. We make three key observations: First, DefiNet proves to be robust across multiple materials. Second, compared to the results in Fig. 3a, b, both DeepRelax and DefiNet show less significant improvements. This is likely due to two factors: (1) the high-density defect datasets contain significantly fewer samples (only 500 per material), limiting learning capacity; and (2) the space of possible defect configurations increases substantially with defect density, making the task more complex. Third, DefiNet still significantly outperforms DeepRelax, with improvements of 32.82%, 35.88%, 34.08%, 33.88%, and 33.33% in coordinate MAE for all atoms, A_3 , A_4 , A_5 , and A_6 , respectively, across all defect structures in the six materials.

Figure 4 provides a visual comparison of the unrelaxed, DFT-relaxed, and DefiNet-predicted structures. As can be seen, the DefiNet-predicted structure closely matches the DFT-relaxed structure, demonstrating the model's effectiveness in handling complex defect configurations.

DFT validation

Validating the energetic favorability of ML-predicted structures is essential to ensuring their physical relevance, accuracy, and efficiency. While coordinate errors provide insight into geometrical accuracy, further analysis is needed to confirm that the predicted structures correspond to local minima on the potential energy surface. We conduct DFT validations to assess whether the structures relaxed by DefiNet are the same as or very similar to DFT ones. For this validation, we randomly selected 25 WSe₂ and 25 MoS₂ structures from the low-density defect test set for DFT calculations. Detailed settings for the DFT calculations are provided in Section "DFT calculations". These two materials were chosen because they appear in both low- and high-density defect categories, making them well-suited for evaluating DFT validation across different defect densities.

We compared the number of ionic steps required for convergence in two cases: starting from unrelaxed structures and starting from DefiNet-predicted structures. The results, shown in Fig. 5a, indicate that using DefiNet-predicted structures as starting points significantly reduces the computational effort required for DFT relaxation, with the number of ionic

Table 1 | Overview of point defect types and DFT calculation parameters for the 2DMD datasets

Materials	Substitutions	Vacancies	Supercell	Cell size (Å)
MoS ₂	S → Se; Mo → W	Mo; S	8 × 8	(25.52, 25.52, 20)
WSe ₂	Se → S; W → Mo	W; Se	8 × 8	(26.62, 26.62, 20)
h-BN	B → C; N → C	B; N	8 × 8	(20.10, 20.10, 20)
GaSe	Ga → In; Se → S	Ga; Se	6 × 6	(22.91, 22.91, 20)
InSe	In → Ga; Se → S	In; Se	6 × 6	(24.58, 24.58, 20)
BP	P → N	P	6 × 6	(19.80, 27.61, 20)

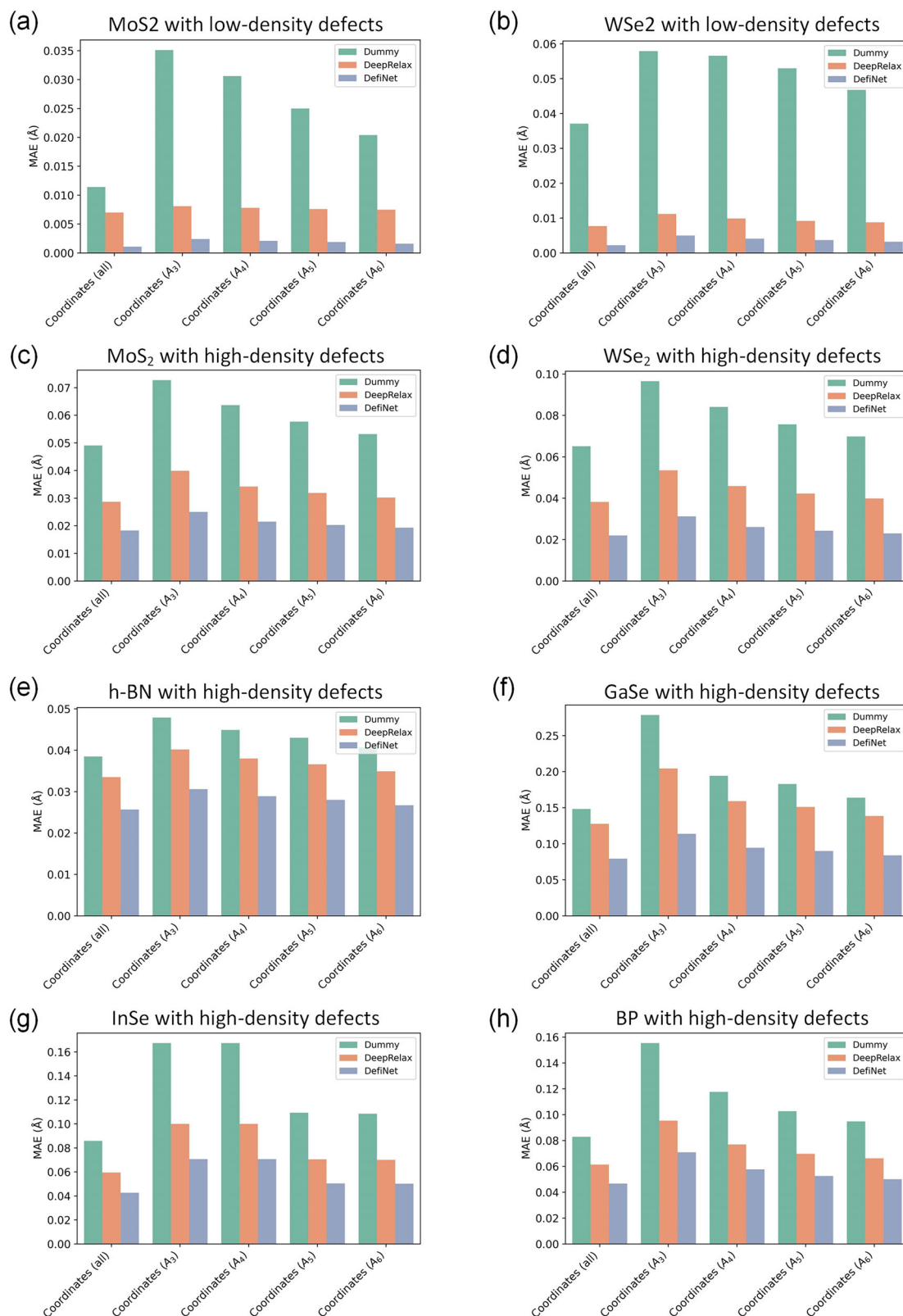
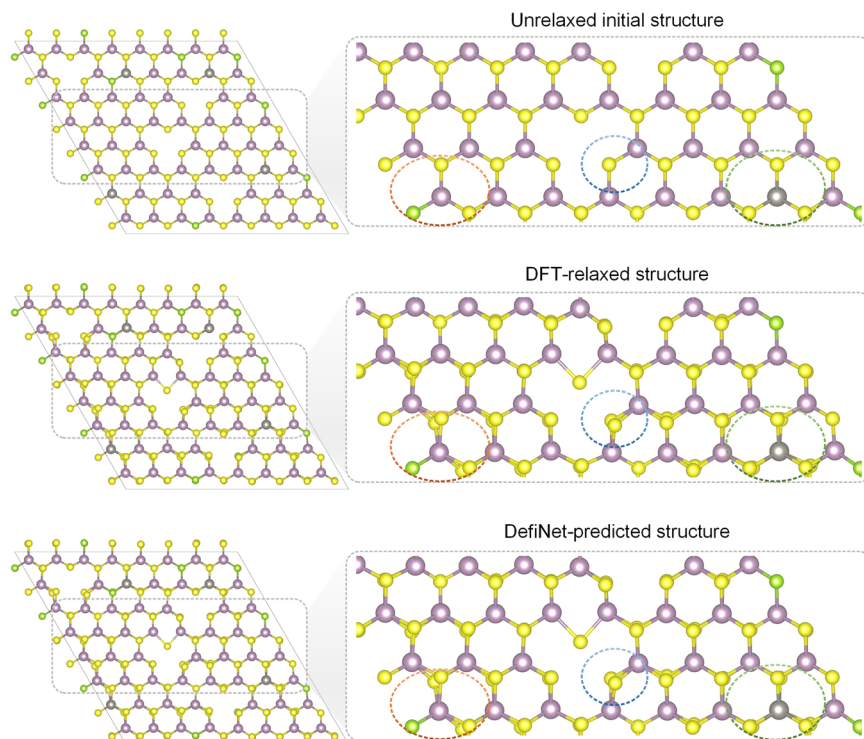


Fig. 3 | Model performance for structures with low- and high-density defects. **a** MoS₂ and **b** WSe₂ with low-density defects, and **c** MoS₂, **d** WSe₂, **e** h-BN, **f** GaSe, **g** InSe, and **h** BP with high-density defects. A₃, A₄, A₅, and A₆ represent MAE calculations using only atoms within 3 Å, 4 Å, 5 Å, and 6 Å radii around defect sites, respectively.

steps decreasing by approximately 87%. Notably, these residual ionic steps also remain nearly constant, regardless of defect complexity. The very low residual ionic steps demonstrate the high accuracy of DefiNet. The steady residual ionic steps, even for highly complex defects, highlight the exceptional efficiency of DefiNet. Importantly, both initialization strategies

(starting from unrelaxed structures and from DefiNet-predicted configurations) converge to the same final DFT-relaxed configurations, with a coordinate MAE of zero across all samples. Additional DFT validation results for high-density defect scenarios are available in Supplementary Fig. 1, which also demonstrates DefiNet's promising performance.

Fig. 4 | Visual comparison of defect structures relaxed by DFT and DefiNet. Example of an MoS_2 crystal structure containing both substitutional and vacancy defects, alongside the corresponding DFT-relaxed and DefiNet-predicted structures.



To evaluate the scalability of DefiNet, we tested its performance across different supercell sizes. Specifically, we randomly selected five defect structures from the test set containing different types of defects. We then created supercells with sizes of 8×8 , 12×12 , and 16×16 , resulting in structures with around 190 atoms, 430 atoms, and 770 atoms, respectively. DefiNet was used to predict the relaxed structures for these unrelaxed configurations. We assessed both the residual ionic steps and the CPU core hours required for the DefiNet-predicted structures, comparing these results to those of the unrelaxed structures.

As illustrated in Fig. 5b, DefiNet consistently achieves constant ionic steps of 3, irrespective of the system size, demonstrating its ability to scale effectively with increasing system size. We further compare the CPU core hours required for the relaxation of both unrelaxed and DefiNet-predicted structures. As shown in Fig. 5c, the computational cost for the large-scale unrelaxed structure is extremely high. In contrast, the relaxation time for the DefiNet-predicted structures is significantly reduced, highlighting DefiNet's capability for large systems by dramatically decreasing the computational cost. Further scalability evaluations are detailed in Supplementary Note 5.

Experimental validation

To further validate the accuracy and extrapolation of DefiNet using experimental results, we conducted comparisons with STEM images, assessing the alignment between DefiNet-relaxed structures and actual experimental observations. Figure 6a–c shows STEM images (overlaid with the DefiNet-relaxed structure) of MoS_2 ³⁷ and WSe_2 ³⁸ with different types of complex defects, including in a line defect (sequential S vacancies), mixed single Se (SV_{Se}) with double Se vacancies (DV_{Se}), and a threefold symmetric trefoil defect. The strong alignment between the DefiNet-predicted and experimentally observed structures highlights DefiNet's accuracy and extrapolation in capturing such complex defects beyond the training point defects. We provide a comparison among the unrelaxed structures, DefiNet-predicted structures, and the STEM image, as shown in Supplementary Fig. 4.

Comparison to ML-potential relaxation

ML-potential relaxation is a popular alternative to DFT-based relaxation methods. To demonstrate the superiority of DefiNet, we

compare it against two well-known ML-potential models: M3GNet and CHGNet. These methods typically require large datasets to train GNN surrogate models that iteratively approximate physical quantities such as energies, forces, and stresses. For this comparison, we used the MoS_2 low-density defect dataset, which contains a sufficient number of samples (5933) with detailed information obtained during DFT-based relaxation. All methods were trained, validated, and tested on the same data splits. Detailed experimental settings are provided in Supplementary Note 7. As shown in Supplementary Fig. 5, DefiNet significantly outperforms M3GNet, CHGNet, and DeepRelax in terms of coordinate MAE and robustness. This result is further validated by DFT calculations, with detailed comparisons available in Supplementary Fig. 6.

Ablation study

To elucidate the contributions of DefiNet's key architectural components, we performed an ablation study focusing on its two main innovations:

- Defect-Aware Message Passing (DAMP): This component allows the model to capture complex interactions involving defects.
- Defect-Aware Coordinate Updating (DACU): The RPV2Disp and Vec2Disp modules are designed to update atomic coordinates effectively, taking into account the unique influences of defects on the surrounding lattice.

We created two ablated versions of DefiNet to assess the impact of these components:

- Vanilla Model: This version removes both main components, DAMP and DACU.
- Vanilla + DAMP: This version includes the DAMP but removes the Defect-Aware Coordinate Updating modules.
- Vanilla + DAMP + DACU (DefiNet): This is the full DefiNet model incorporating both components.

The results on high-density datasets, as shown in Supplementary Fig. 7, indicate that both ablated models exhibit decreased performance compared

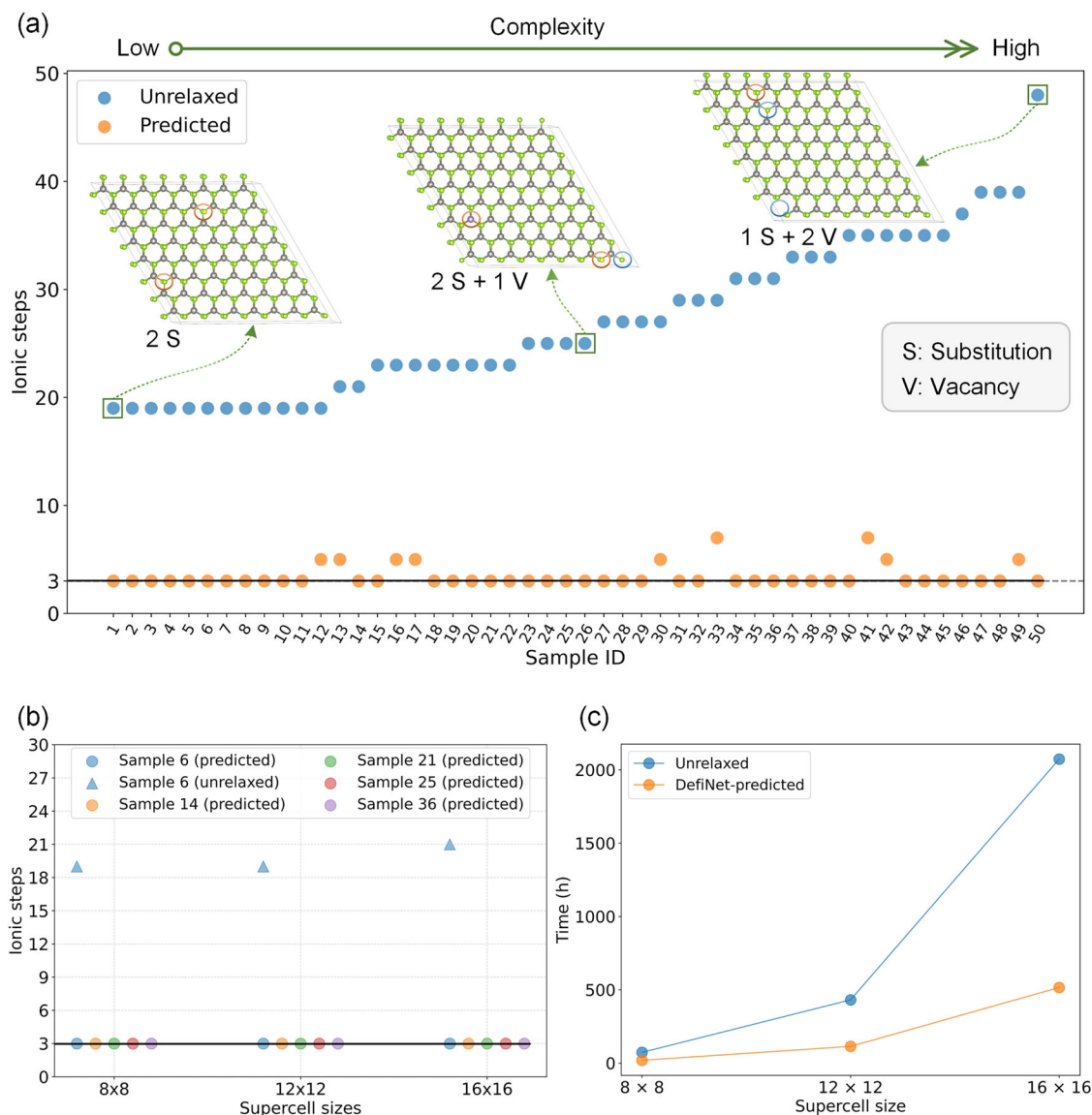


Fig. 5 | DFT validation on DefiNet's accuracy, efficiency, and scalability.

a Comparison of the number of DFT ionic steps required to relax structures starting from the initial unrelaxed configurations and from the DefiNet-predicted structures for low-density defects. The steady residual ionic steps against the defect complexity are indicated by a horizontal black solid line. The sample ID is sorted based on the number of ionic steps required by the unrelaxed structures for better observation. **b** Residual ionic steps for five randomly selected defect structures from the 50 samples across different supercell sizes, starting from DefiNet-predicted

configurations. Only a single reference run is shown for unrelaxed structures due to the high computational cost of initiating DFT relaxation from unrelaxed configurations. The steady residual ionic steps against the structural size are indicated by a horizontal black solid line. **c** Comparison of DFT CPU core hours on large supercells using unrelaxed and DefiNet-predicted configurations. Due to the extremely high computational cost associated with the unrelaxed structure of the 16×16 supercell size with 770 atoms, only one sample was selected as an example for this experiment.

to the full DefiNet. These findings confirm that both components are critical for DefiNet's superior performance.

We also conduct an additional ablation study to evaluate two auxiliary components: global nodes and nonlinear vector activation. As shown in Supplementary Fig. 8, both mechanisms improve model performance, supporting their inclusion in the final architecture.

Discussion

Recently, GNNs have been used for defect property and structure analysis^{12,13,39–44}, showing great potential to reduce the high computational cost of DFT calculations. Two recent works^{12,13} have demonstrated that employing machine learning (ML) interatomic potentials can achieve both cost-effectiveness and accuracy in searching ground-state configurations of defect structures. Those approaches, however, require large databases

annotated with energies, forces, and stresses, and they treat defect sites only implicitly, leaving the network to infer defect-defect interactions on its own. DefiNet avoids these limitations. First, it is trained solely on pairs of initial and relaxed structures, which makes it easier to implement in real applications. Second, it explicitly considers complex defect-related interactions, leading to more accurate relaxation of defect crystal structures. We also benchmark DefiNet against the previous single-step model, DeepRelax. DefiNet not only achieves a significantly lower coordinate MAE but also runs nearly 26× faster than DeepRelax.

Our scalability tests demonstrate that DefiNet maintains high accuracy when applied to larger systems beyond the sizes used during training. Moreover, we perform two transferability evaluations: (1) Train DefiNet on high-density defect structures and test on low defect-density structures, and vice versa. (2) Train DefiNet on structures with short average defect-defect

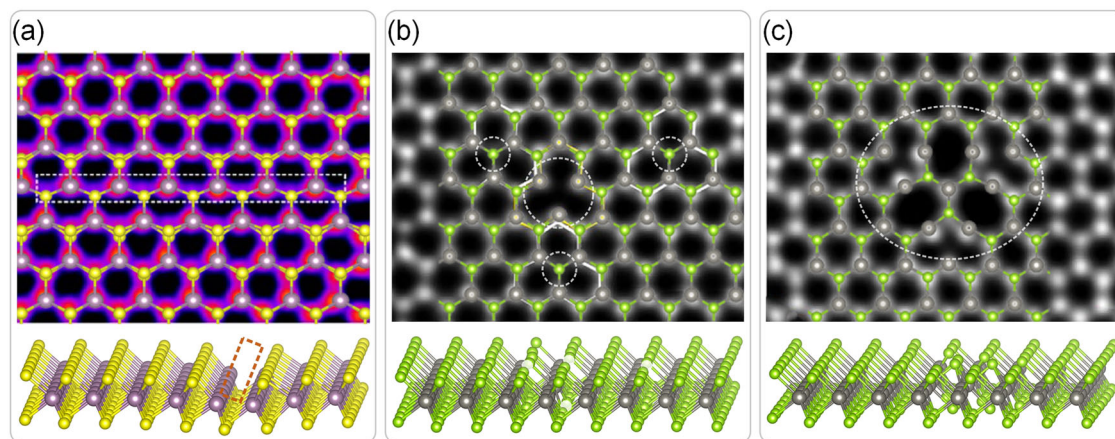


Fig. 6 | Comparison between STEM images and DefiNet-relaxed structures. Ball-and-stick models of the corresponding DefiNet-relaxed structures are shown below each STEM image, with a line defect marked by orange rectangles. **a** STEM image of MoS₂ featuring a line defect (sequential S vacancies), overlaid with the DefiNet-relaxed structure. Reprinted with permission from³⁷. Copyright 2016 American

Chemical Society. **b** STEM image of WSe₂ with mixed SV_{Se} and DV_{Se} defects, overlaid with the DefiNet-relaxed structure. **c** STEM image of WSe₂ with a three-fold symmetrical trefoil defect, overlaid with the DefiNet-relaxed structure. Defect sites are highlighted with white dotted lines for clarity.

distances and test on those with long distances, and vice versa. These experiments demonstrate DefiNet's good transferability (see Supplementary Note 9).

The DFT validations confirm that the structures predicted by DefiNet are energetically favorable. Importantly, initiating DFT calculations from DefiNet-predicted structures significantly reduces the number of required ionic steps by approximately 87%, irrespective of defect complexity or system size. This hybrid approach leverages the speed of DefiNet and the precision of DFT, offering an efficient pathway for exploring defect structures in materials. While DefiNet demonstrates remarkable performance, certain limitations warrant discussion.

First, this study focuses exclusively on 2D materials with point defects, and only six materials comprising a limited subset of elements from the periodic table are considered. As a result, the trained DefiNet model cannot be directly generalized to materials containing previously unseen elements. Expanding DefiNet to support a broader range of materials, including both 2D and 3D systems, as well as more complex defect types, would significantly enhance its applicability and generalization capability.

Second, in this work, we only focus on defects in the neutral charge state. It is worth noting that point defects in semiconductors frequently adopt multiple charge states, each with distinct geometric relaxations. Because existing ML approaches struggle to encode charge directly, most studies to date also limit themselves to neutral or fixed ionic states^{13,45}. There are two possible directions for extending DefiNet to charged defects: (1) transfer learning from a neutral-trained model to charged configurations, or (2) introducing a global charge-state embedding as an additional input feature. Unfortunately, the lack of sufficiently large, labeled datasets of charged-defect geometries prevents us from exploring these strategies here, and we therefore leave this as an important direction for future work.

Third, point defects in low-symmetry semiconductors can occupy several energetically competitive local minima (i.e., metastable configurations) with distinct geometries and functional behaviors^{46,47}. Since DefiNet outputs only a single relaxed structure per defect, the current version of DefiNet is unable to capture these alternative metastable states.

Methods

Input representation

In this work, the defect structure is represented as a defect-explicit graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{M})$, where \mathcal{V} and \mathcal{E} are sets of nodes and edges corresponding to atoms and bonds within the pristine structure, and \mathcal{M} is a set of markers representing defect types. Each marker $m_i \in \mathcal{M}$ is a categorical variable that takes a value from the set $\{0, 1, 2\}$, where 0 denotes a pristine atom, 1

indicates a substitution, and 2 represents a vacancy. By contrast, a conventional defect-implicit graph $\tilde{\mathcal{G}} = (\mathcal{V}, \mathcal{E})$ omits defect information \mathcal{M} . In principle, a sufficiently expressive GNN could infer defect sites from structures alone, but doing so is often inefficient, as representation learning is empirically data-hungry⁴⁸. Providing explicit markers imposes a strong inductive bias: the network no longer has to learn a feature extractor that separates pristine atoms from defect sites, enabling the model to reach the same generalization error with fewer training examples. Importantly, for vacancies, a placeholder node is retained at the position of the missing atom in the pristine lattice and marked with $m_i = 2$. This node is treated as an active part of the graph and participates in message passing. By explicitly incorporating vacancy sites into the graph structure, the model can directly learn spatial relationships between vacancies and neighboring atoms, rather than relying on implicit inference from the structure.

Each node $v_i \in \mathcal{V}$ contains three feature types: scalar $\mathbf{x}_i \in \mathbb{R}^F$, vector $\tilde{\mathbf{x}}_i \in \mathbb{R}^{F \times 3}$, and coordinates $\vec{\mathbf{r}}_i \in \mathbb{R}^3$, which encapsulate invariant, equivariant, and structural features, respectively. The number of features F is kept constant throughout the network. The scalar feature is initialized as an embedding dependent solely on the atomic number, given by $\mathbf{x}_i^{(0)} = E(z_i) \in \mathbb{R}^F$, where z_i is the atomic number and E is an embedding layer that takes z_i as input and returns an F -dimensional feature. The vector feature is initially set to $\tilde{\mathbf{x}}_i^{(0)} = \vec{\mathbf{0}} \in \mathbb{R}^{F \times 3}$. To capture long-range interactions, we introduce a global node v_G , which includes a global scalar $\mathbf{x}_G \in \mathbb{R}^F$ and a global vector $\tilde{\mathbf{x}}_G \in \mathbb{R}^{F \times 3}$. These are initialized as a trainable F -dimensional feature and $\vec{\mathbf{0}}$, respectively. We also define the relative position vector as $\vec{\mathbf{r}}_{ij} = \vec{\mathbf{r}}_j - \vec{\mathbf{r}}_i$ to introduce directional information into the edges. Each node is connected to its closest neighbors within a cutoff distance D , with a maximum number of neighbors N , where D and N are predefined constants.

DefiNet workflow

The proposed DefiNet consists of four layers, each of which updates the node representation through a three-stage graph convolution process that includes defect-aware message passing, self-updating, and defect-aware coordinate updating. This process incorporates message distribution and aggregation to capture long-range interactions, as illustrated in Fig. 7.

Defect-aware message passing

At layer t each node v_i aggregates information from its neighbors v_j in a defect-aware manner. This process results in intermediate scalar and vector

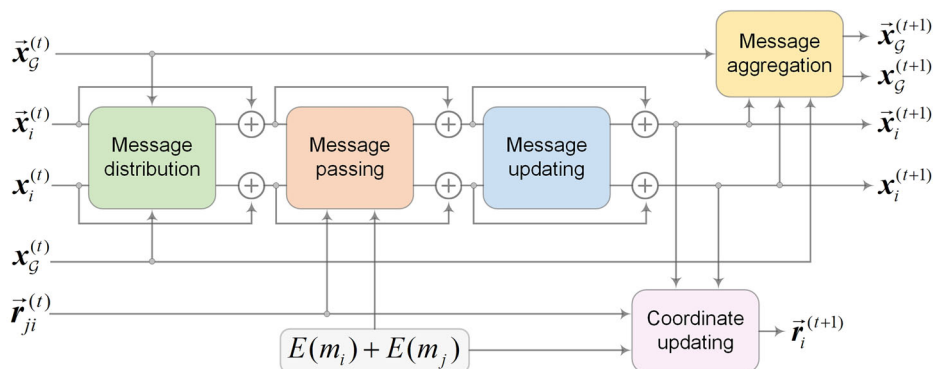


Fig. 7 | Workflow of the t -th graph convolution layer in DefiNet. The process begins with message distribution, where the global scalar $\mathbf{x}_G^{(t)}$ and global vector $\bar{\mathbf{x}}_G^{(t)}$ are globally distributed to each scalar $\mathbf{x}_i^{(t)}$ and vector $\bar{\mathbf{x}}_i^{(t)}$. This is followed by defect-aware message passing, which locally collects messages from neighboring nodes v_j , weighting messages according to interatomic distances and the defect markers m_i

and m_j . Next, message updating refines the node representation using the information within the node itself, resulting in $\mathbf{x}_i^{(t+1)}$ and $\bar{\mathbf{x}}_i^{(t+1)}$. Coordinate updating then further refines the atomic coordinates, resulting in the updated coordinates $\bar{\mathbf{r}}_i^{(t+1)}$. Finally, message aggregation is performed to update the global scalar and vector, resulting in $\mathbf{x}_G^{(t+1)}$ and $\bar{\mathbf{x}}_G^{(t+1)}$.

variables \mathbf{q}_i and $\bar{\mathbf{q}}_i$, defined as follows:

$$\mathbf{q}_i = \sum_{v_j \in \mathcal{N}(v_i)} \phi_h(\mathbf{x}_j^{(t)}) \circ \lambda_h(\|\bar{\mathbf{r}}_{ji}^{(t)}\|) \circ \gamma_h(E(m_i) + E(m_j)) \quad (1)$$

$$\begin{aligned} \bar{\mathbf{q}}_i &= \sum_{v_j \in \mathcal{N}(v_i)} \phi_u(\mathbf{x}_j^{(t)}) \circ \lambda_u(\|\bar{\mathbf{r}}_{ji}^{(t)}\|) \circ \gamma_u(E(m_i) + E(m_j)) \circ \bar{\mathbf{x}}_j^{(t)} \\ &\quad + \phi_v(\mathbf{x}_j^{(t)}) \circ \lambda_v(\|\bar{\mathbf{r}}_{ji}^{(t)}\|) \circ \gamma_v(E(m_i) + E(m_j)) \circ \frac{\bar{\mathbf{r}}_{ji}^{(t)}}{\|\bar{\mathbf{r}}_{ji}^{(t)}\|} \end{aligned} \quad (2)$$

Here, \circ denotes the element-wise product, E is an embedding layer that maps the marker m_i to an F -dimensional feature, and $\phi_h, \phi_u, \phi_v, \gamma_h, \gamma_u$ and γ_v are multilayer perceptrons (MLPs). The functions λ_h, λ_u and λ_v are linear combinations of Gaussian radial basis functions²¹. The pair-wise gate $\gamma(\cdot)$ re-weights each message according to the marker pair (m_i, m_j) , thereby distinguishing pristine-pristine, defect-pristine, and defect-defect interactions.

Self-updating

We employ the self-updating mechanism proposed by Yang et al.³⁴. During this phase, the F scalars and F vectors within \mathbf{q}_i and $\bar{\mathbf{q}}_i$, respectively, are aggregated to generate the updated scalar $\mathbf{x}_i^{(t+1)}$ and vector $\bar{\mathbf{x}}_i^{(t+1)}$. Specifically, the scalar representation $\mathbf{x}_i^{(t+1)}$ and vector representation $\bar{\mathbf{x}}_i^{(t+1)}$ are updated according to the following equations:

$$\mathbf{x}_i^{(t+1)} = \phi_s(\mathbf{q}_i \oplus \|\mathbf{V}\bar{\mathbf{q}}_i\|) + \tanh(\phi_g(\mathbf{q}_i \oplus \|\mathbf{V}\bar{\mathbf{q}}_i\|))\mathbf{q}_i \quad (3)$$

$$\bar{\mathbf{x}}_i^{(t+1)} = \phi_h(\mathbf{q}_i \oplus \|\mathbf{V}\bar{\mathbf{q}}_i\|) \circ \mathbf{U}\bar{\mathbf{q}}_i \quad (4)$$

where \oplus denotes concatenation, $\phi_s, \phi_g, \phi_h: \mathbb{R}^{2F} \rightarrow \mathbb{R}^F$ are MLPs, and $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{F \times F}$ are trainable matrices.

Defect-aware coordinate updating

The defect-aware coordinate updating step aims to refine the atomic coordinates using two modules, RVP2Disp and Vec2Disp, which represent two distinct contributions to the coordinate update. Specifically, RVP2Disp converts the relative position vector $\bar{\mathbf{r}}_{ji}^{(t)}$ into a displacement, while Vec2Disp translates the vector representation $\bar{\mathbf{x}}_i^{(t+1)}$ into a displacement. Together, these determine the displacement of each atom at the current

stage, as described by the following equations:

$$\bar{\mathbf{d}}_i^{(\text{RPV})} = \sum_{v_j \in \mathcal{N}(v_i)} \phi_q(\phi_v(\mathbf{x}_j^{(t+1)}) \circ \lambda_v(\|\bar{\mathbf{r}}_{ji}^{(t)}\|) \circ \gamma_v(E(m_i) + E(m_j))) \circ \frac{\bar{\mathbf{r}}_{ji}^{(t)}}{\|\bar{\mathbf{r}}_{ji}^{(t)}\|} \quad (5)$$

$$\bar{\mathbf{d}}_i^{(\text{Vec})} = \mathbf{W}_{\text{Vec}} \bar{\mathbf{x}}_i^{(t+1)} \quad (6)$$

Here, $\phi_v: \mathbb{R}^F \rightarrow \mathbb{R}^F$ and $\phi_q: \mathbb{R}^F \rightarrow \mathbb{R}$ are MLPs; γ_v is the pair-wise defect gate that re-weights messages according to the marker pair (m_i, m_j) ; and $\mathbf{W}_{\text{Vec}} \in \mathbb{R}^{1 \times F}$ integrates all the vectors within $\bar{\mathbf{x}}_i^{(t+1)}$. Finally, the coordinates are updated as follows:

$$\bar{\mathbf{r}}_i^{(t+1)} = \bar{\mathbf{r}}_i^{(t)} + \bar{\mathbf{d}}_i^{(\text{RPV})} + \bar{\mathbf{d}}_i^{(\text{Vec})} \quad (7)$$

The initial coordinate $\bar{\mathbf{r}}_i^{(0)}$ is set to the atom coordinate of the unrelaxed structure. The updated coordinates $\bar{\mathbf{r}}_i^{(t+1)}$ are equivariant to both rotation and translation, with a formal proof provided in Supplementary Note 1.

Message distribution and aggregation

To establish a more effective global communication channel across the entire graph, we implement a message distribution and aggregation scheme using global node technology³⁴. The message distribution process propagates the global scalar and vector at the current step to each node using the following equations:

$$\mathbf{x}_i^{(t)} = \phi(\mathbf{x}_i^{(t-1)} \oplus \mathbf{x}_G^{(t-1)}) + \mathbf{x}_i^{(t-1)} \quad (8)$$

$$\bar{\mathbf{x}}_i^{(t)} = \mathbf{W}(\bar{\mathbf{x}}_i^{(t-1)} + \bar{\mathbf{x}}_G^{(t-1)}) + \bar{\mathbf{x}}_i^{(t-1)} \quad (9)$$

where $\phi: \mathbb{R}^{2F} \rightarrow \mathbb{R}^F$ is an MLP, and $\mathbf{W} \in \mathbb{R}^{F \times F}$ is a trainable matrix.

The message aggregation step updates the global scalar and vector based on the node representations at the current step, as described by the following equations:

$$\mathbf{x}_G^{(t+1)} = \phi\left(\left(\frac{1}{|\mathcal{G}|} \sum_{v_i \in \mathcal{G}} \mathbf{x}_i^{(t)}\right) \oplus \mathbf{x}_G^{(t)}\right) + \mathbf{x}_G^{(t)} \quad (10)$$

$$\bar{\mathbf{x}}_{\mathcal{G}}^{(t+1)} = \mathbf{W} \left(\left(\frac{1}{|\mathcal{G}|} \sum_{v_i \in \mathcal{G}} \bar{\mathbf{x}}_i^{(t)} \right) + \bar{\mathbf{x}}_{\mathcal{G}}^{(t)} \right) + \bar{\mathbf{x}}_{\mathcal{G}}^{(t)} \quad (11)$$

It is important to note that this global communication pathway does not incorporate interatomic distances and thus does not model short-range interactions directly. Instead, such interactions—including those modified by defects—are explicitly captured by the localized, distance-aware, defect-sensitive message passing mechanism (defect-aware message passing) described in the previous section.

Non-linear vector activation

Non-linearity is essential for enhancing the expressive power of neural networks. Here, we introduce non-linearity into vector representations while preserving equivariance. Specifically, we first aggregate the F vectors within a node to obtain a consensus vector for each node:

$$\bar{\mathbf{x}}_i^{\mathcal{G}} = \mathbf{W}_p \bar{\mathbf{x}}_i \quad (12)$$

where $\mathbf{W}_p \in \mathbb{R}^{1 \times F}$ integrates all vectors within $\bar{\mathbf{x}}_i$ to produce the consensus vector $\bar{\mathbf{x}}_i^{\mathcal{G}} \in \mathbb{R}^3$, capturing the overarching trend across all vectors in the node. Next, each vector $\bar{\mathbf{x}}_i^j$ within $\bar{\mathbf{x}}_i$ is updated as follows:

$$\bar{\mathbf{x}}_i^j = \begin{cases} \mathbf{W}^j \bar{\mathbf{x}}_i^j, & \text{if } \langle \bar{\mathbf{x}}_i^{\mathcal{G}}, \mathbf{W}^j \bar{\mathbf{x}}_i^j \rangle > 0 \\ \mathbf{W}^j \bar{\mathbf{x}}_i^j + \bar{\mathbf{x}}_i^{\mathcal{G}}, & \text{otherwise} \end{cases} \quad (13)$$

Here, $\mathbf{W}^j \in \mathbb{R}^{1 \times F}$, and $\langle \cdot, \cdot \rangle$ denotes the dot product. The idea is that if the vectors align with the consensus trend, as indicated by a dot product greater than zero, they are considered significant and retained without modification. Conversely, vectors that diverge from the consensus trend (dot product less than or equal to zero) are considered potentially noisy or weakly informative and are softly regularized by adding the consensus vector. This adjustment encourages alignment with the dominant directional trend. Every time the vectors have been updated, we apply a non-linear vector activation to them.

Implementation details

The DefiNet model is implemented using PyTorch, and experiments are conducted on an NVIDIA RTX A6000 with 48 GB of memory. The training objective is to minimize the mean absolute error (MAE) loss between the ML-relaxed and DFT-relaxed structures, defined as follows:

$$\mathcal{L} = \frac{1}{N} \frac{1}{M} \sum_{i=1}^M |\vec{\mathbf{r}}_i^{(T)} - \tilde{\mathbf{r}}_i| \quad (14)$$

where N and M denote the sample size and the number of atoms in each sample, respectively. Here, T represents the total number of layers in the model, and $\tilde{\mathbf{r}}_i$ is the DFT-relaxed atomic coordinate. We use the AdamW optimizer with a learning rate of 0.0001 to update the model's parameters. Additionally, a learning rate decay strategy is implemented, reducing the learning rate if there is no improvement in coordinate MAE for 5 consecutive epochs.

DFT calculations

Our calculations are performed using DFT with the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional, as implemented in the Vienna Ab Initio Simulation Package (VASP)⁴⁹. The interaction between valence electrons and ionic cores is treated using the projector augmented wave (PAW) method⁵⁰, with a plane-wave energy cutoff of 500 eV. Initial crystal structures were taken from the Materials Project database. Given the large supercells required for defect calculations, structural relaxations were carried out using a Γ -point only Monkhorst-Pack grid. To prevent interactions between neighboring layers, a vacuum space of at least 15 Å was

introduced. During structural relaxation, atomic positions were optimized until the forces on all atoms were below 0.01 eV/Å, with an energy tolerance of 10^{-6} eV. For defect structures with unpaired electrons, we used standard collinear spin-polarized calculations, initializing magnetic ions in a high-spin ferromagnetic state, with the possibility of relaxation to a low-spin state during the ionic and electronic relaxation processes.

Data availability

The data that support the findings of this study are available in <https://zenodo.org/records/14027373>.

Code availability

The source code for DefiNet is available at <https://github.com/Shen-Group/DefiNet>.

Received: 12 March 2025; Accepted: 3 July 2025;

Published online: 15 July 2025

References

- Davidsson, J., Bertoldo, F., Thygesen, K. S. & Armiento, R. Absorption versus adsorption: high-throughput computation of impurities in 2d materials. *npj 2D Mater. Appl.* **7**, 26 (2023).
- Huang, P. et al. Unveiling the complex structure-property correlation of defects in 2d materials based on high throughput datasets. *npj 2D Mater. Appl.* **7**, 6 (2023).
- Mosquera-Lois, I., Kavanagh, S. R., Walsh, A. & Scanlon, D. O. Identifying the ground state structures of point defects in solids. *npj Comput. Mater.* **9**, 25 (2023).
- Kazeev, N. et al. Sparse representation for machine learning the properties of defects in 2d materials. *npj Comput. Mater.* **9**, 113 (2023).
- Thomas, J. C. et al. A substitutional quantum defect in ws2 discovered by high-throughput computational screening and fabricated by site-selective stm manipulation. *Nat. Commun.* **15**, 3556 (2024).
- Chen, C. & Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* **2**, 718–728 (2022).
- Deng, B. et al. Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat. Mach. Intell.* **5**, 1031–1041 (2023).
- Batzner, S. et al. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).
- Batatia, I., Kovacs, D. P., Simm, G., Ortner, C. & Csányi, G. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. *Adv. Neural Inf. Process. Syst.* **35**, 11423–11436 (2022).
- Park, Y., Kim, J., Hwang, S. & Han, S. Scalable parallel algorithm for graph neural network interatomic potentials in molecular dynamics simulations. *J. Chem. Theory Comput.* **20**, 4857–4868 (2024).
- Ko, T. W. & Ong, S. P. Recent advances and outstanding challenges for machine learning interatomic potentials. *Nat. Comput. Sci.* **3**, 998–1000 (2023).
- Mosquera-Lois, I., Kavanagh, S. R., Ganose, A. M. & Walsh, A. Machine-learning structural reconstructions for accelerated point defect calculations. *npj Comput. Mater.* **10**, 121 (2024).
- Jiang, C., Marianetti, C. A., Khafizov, M. & Hurley, D. H. Machine learning potential assisted exploration of complex defect potential energy surfaces. *npj Comput. Mater.* **10**, 21 (2024).
- Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).
- Musaelian, A. et al. Learning local equivariant representations for large-scale atomistic dynamics. *Nat. Commun.* **14**, 579 (2023).
- Li, H. et al. Deep-learning density functional theory hamiltonian for efficient ab initio electronic-structure calculation. *Nat. Comput. Sci.* **2**, 367–377 (2022).

17. Gong, X. et al. General framework for e (3)-equivariant neural network representation of density functional theory hamiltonian. *Nat. Commun.* **14**, 2848 (2023).
18. Zhong, Y., Yu, H., Su, M., Gong, X. & Xiang, H. Transferable equivariant graph neural networks for the hamiltonians of molecules and solids. *npj Comput. Mater.* **9**, 182 (2023).
19. Zhong, Y. et al. Universal machine learning kohn-sham hamiltonian for materials. *Chin. Phys. Lett.* **41**, 077103 (2024).
20. Zhong, Y. et al. Accelerating the calculation of electron-phonon coupling strength with machine learning. *Nat. Comput. Sci.* **4**, 615–625 (2024).
21. Schütt, K. et al. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in neural information processing systems* **30** (2017).
22. Chen, C., Ye, W., Zuo, Y., Zheng, C. & Ong, S. P. Graph networks as a universal machine learning framework for molecules and crystals. *Chem. Mater.* **31**, 3564–3572 (2019).
23. Gastegger, J., Groß, J. & Günnemann, S. Directional message passing for molecular graphs. In *International Conference on Learning Representations (ICLR)* (2020).
24. Choudhary, K. & DeCost, B. Atomistic line graph neural network for improved materials property predictions. *npj Comput. Mater.* **7**, 185 (2021).
25. Unke, O. T. et al. Spookynet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nat. Commun.* **12**, 7273 (2021).
26. Zhang, X., Zhou, J., Lu, J. & Shen, L. Interpretable learning of voltage for electrode design of multivalent metal-ion batteries. *npj Comput. Mater.* **8**, 175 (2022).
27. Satorras, V. G., Hoogeboom, E. & Welling, M. E (n) equivariant graph neural networks. In *International conference on machine learning*, 9323–9332 (PMLR, 2021).
28. Omee, S. S. et al. Scalable deeper graph neural networks for high-performance materials property prediction. *Patterns* **3**, 100491 (2022).
29. Schütt, K., Unke, O. & Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, 9377–9388 (PMLR, 2021).
30. Han, J. et al. A survey of geometric graph neural networks: Data structures, models and applications. *Front. Comput. Sci.* **19**, 1–38 (2025).
31. Dong, L., Zhang, X., Yang, Z., Shen, L. & Lu, Y. Accurate piezoelectric tensor prediction with equivariant attention tensor graph neural network. *npj Comput. Mater.* **11**, 63 (2025).
32. Wu, X. et al. Graph neural networks for molecular and materials representation. *J. Mater. Inform.* **3**, N–A (2023).
33. Li, Y. et al. Local environment interaction-based machine learning framework for predicting molecular adsorption energy. *J. Mater. Inform.* **4** (2024).
34. Yang, Z. et al. Efficient equivariant model for machine learning interatomic potentials. *npj Comput. Mater.* **11**, 49 (2025).
35. Deng, C. et al. Vector neurons: A general framework for so (3)-equivariant networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12200–12209 (2021).
36. Yang, Z. et al. Scalable crystal structure relaxation using an iteration-free deep generative model with uncertainty quantification. *Nat. Commun.* **15**, 8148 (2024).
37. Wang, S., Lee, G.-D., Lee, S., Yoon, E. & Warner, J. H. Detailed atomic reconstruction of extended line defects in monolayer mos₂. *ACS nano* **10**, 5419–5430 (2016).
38. Lin, Y.-C. et al. Three-fold rotational defects in two-dimensional transition metal dichalcogenides. *Nat. Commun.* **6**, 6736 (2015).
39. Witman, M. D., Goyal, A., Ogitsu, T., McDaniel, A. H. & Lany, S. Defect graph neural networks for materials discovery in high-temperature clean-energy applications. *Nat. Computational Sci.* **3**, 675–686 (2023).
40. Way, L. et al. Defect diffusion graph neural networks for materials discovery in high-temperature, clean energy applications (2024).
41. Mannodi-Kanakkithodi, A. et al. Universal machine learning framework for defect predictions in zinc blende semiconductors. *Patterns* **3** (2022).
42. Frey, N. C., Akinwande, D., Jariwala, D. & Shenoy, V. B. Machine learning-enabled design of point defects in 2d materials for quantum and neuromorphic information processing. *ACS nano* **14**, 13406–13417 (2020).
43. Rahman, M. H. et al. Accelerating defect predictions in semiconductors using graph neural networks. *APL Mach. Learn.* **2** (2024).
44. Fang, Z. & Yan, Q. Leveraging persistent homology features for accurate defect formation energy predictions via graph neural networks. *Chem. Mater.* **37**, 1531–1540 (2025).
45. Kavanagh, S. R. Identifying split vacancies with foundation models and electrostatics. *arXiv preprint arXiv:2412.19330* (2024).
46. Huang, M., Wang, S. & Chen, S. Metastability and anharmonicity enhance defect-assisted nonradiative recombination in low-symmetry semiconductors. *arXiv preprint arXiv:2312.01733* (2023).
47. Kavanagh, S. R., Scanlon, D. O., Walsh, A. & Freysoldt, C. Impact of metastable defect structures on carrier recombination in solar cells. *Faraday Discuss.* **239**, 339–356 (2022).
48. Shalev-Shwartz, S. & Ben-David, S. *Understanding machine learning: From theory to algorithms* (Cambridge University Press, 2014).
49. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169 (1996).
50. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953 (1994).

Acknowledgements

This research was supported by the Natural Science Foundation of Guangdong Province (Grant No. 2025A1515011487), Ministry of Education, Singapore, Tier 1 (Grant No. A-8001194-00-00), Tier 2 (Grant No. A-8001872-00-00), and under its Research Center of Excellence award to the Institute for Functional Intelligent Materials (I-FIM, project No. EDUNC-33-18-279-V12). K.S.N. is grateful to the Royal Society (UK, grant number RSRP\R\190000) for support.

Author contributions

L.S., K.S.N., P.H., and Z.Y. designed the research. Z.Y. conduct the experiment. Z.Y., X.L., X.Z., P.H., and L.S. analyzed the data and results. Z.Y. and L.S. wrote the manuscript together. All authors reviewed and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41524-025-01728-w>.

Correspondence and requests for materials should be addressed to Pengru Huang or Lei Shen.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025