

Cite this: *Mater. Horiz.*, 2025, 12, 9669Received 24th April 2025,
Accepted 31st July 2025

DOI: 10.1039/d5mh00774g

rsc.li/materials-horizons

Crystal structure prediction with host-guided inpainting generation and foundation potentials

Peichen Zhong,^{id}*^{ab} Xinzhe Dai,^{bc} Bowen Deng,^{id}^{bc} Gerbrand Ceder^{id}^{bc} and Kristin A. Persson^{id}*^{abc}

Unconditional crystal structure generation with diffusion models faces challenges in identifying symmetric crystals as the unit cell size increases. We present the crystal host-guided generation (CHGGen) framework to address this challenge through conditional generation using an inpainting method, which optimizes a fraction of atomic positions within a predefined and symmetrized host structure to improve the success rate for symmetric structure generation. By integrating inpainting structure generation with a foundation potential for structure optimization, we demonstrate the method on the ZnS–P₂S₅ and Li–Si chemical systems, where the inpainting method generates a higher fraction of symmetric structures than unconditional generation. The practical significance of CHGGen extends to enabling the structural modification of crystal structures, particularly for systems with partial occupancy or intercalation chemistry. The inpainting method also allows for seamless integration with other generative models, providing a versatile framework for accelerating materials discovery.

1 Introduction

Crystal structure prediction (CSP) is a foundational tool in computational materials discovery with wide-ranging applications in energy storage,¹ drug design,² and superconductors.³ The ability to predict stable atomic arrangements for a given chemical composition is critical for materials design, yet remains a challenge due to the high dimensionality of chemical and configurational space.⁴ Traditional computational approaches using density functional theory (DFT) calculations have achieved notable successes, such as random structure searches,⁵ genetic algorithms,⁶ particle-swarm optimization,⁷ substitution models,⁸ and exact lattice model approaches.⁹ However, DFT-based search algorithms can become

New concepts

Graph neural network-based diffusion models suffer from locality bias, generating reasonable local environments but failing to propagate long-range crystallographic order. We demonstrate that crystal-host guided inpainting generation (CHGGen) can mitigate this issue. The inpainting method is a conditional generation originally developed in computer vision for context-based image generation. In crystal structure prediction problems, it optimizes atomic positions within symmetrized host structures rather than generating complete structures unconditionally. Our approach achieves higher symmetry compared to unconditional methods, particularly for polyanion systems. Beyond structure prediction, CHGGen enables structural modification of materials with partial occupancy or intercalation chemistry. By integrating with foundation potentials for structure optimization, CHGGen provides a modular, practical framework for accelerating materials discovery across diverse chemical spaces.

computationally prohibitive, especially when applied to multi-component systems with complex compositional spaces.¹⁰

Recent advances in graph neural network (GNN)-based machine learning models have introduced promising alternatives to traditional CSP methods, with a key milestone being the development of foundation potentials, or universal machine learning interatomic potentials—offering accurate and transferable modeling across diverse material systems.^{11–14} These foundation potentials trained on millions of DFT calculations demonstrate remarkable generalizability in exploring vast chemical spaces for materials discovery.^{15–18} Another emerging direction is deep-generative models, particularly diffusion models, which learn the data manifold or probabilistic distribution and generate new configurations *via* stochastic or variational approaches.^{19–21} Xie *et al.*²² introduced CDVAE that uses a variational autoencoder to sample lattice parameters and compositions and a diffusion model to optimize atomic coordinates. Although promising, CDVAE-generated structures are predominantly thermodynamically unstable or lack symmetry.^{23,24} Kurz *et al.*²⁵ introduced a wrapped normal distribution to effectively couple lattice diffusion with fractional coordinates,

^a Bakar Institute of Digital Materials for the Planet, UC Berkeley, California 94720, USA. E-mail: zhongpc@berkeley.edu

^b Materials Sciences Division, Lawrence Berkeley National Laboratory, California 94720, USA. E-mail: kapersson@lbl.gov

^c Department of Materials Science and Engineering, UC Berkeley, California 94720, USA

a strategy that was successfully implemented in DiffCSP.^{25–27} Zeni *et al.*²⁸ further adapted the scheme with edge features for lattice scores in the MatterGen framework that enables diffusions on lattices, fractional coordinates, and chemical species. By learning from materials datasets such as Materials Project (MP)²⁹ and Alexandria,³⁰ MatterGen is capable of conducting scalable and universal exploration in high-dimensional design space and achieves excellent performance in structural stability, uniqueness, and novelty, albeit with an observed limitation to a smaller scale (*e.g.*, $N_{\text{atom}} \leq 20$). Beyond GNN-based models, other approaches such as U-net-based diffusion models,³¹ optimization of subcell structures from amorphous configurations,³² and large language models^{33,34} have shown promise in crystal structure generations without the need to limit structure sizes.

In this work, we extend GNN-based diffusion models to enable fractional crystal structure design *via* inpainting generation—that is, given a host or substrate structure, we optimize the placement of additional ‘guest’ atoms within the existing framework. This application is particularly valuable in several material domains, *e.g.*, defective materials, intercalation electrodes,³⁵ molecular absorption on catalyst surfaces,³⁶ and interfacial solid reactions where surfaces reconstruct while bulk structures remain unchanged.³⁷ We first summarize the fundamentals of diffusion models and inpainting generation, and discuss the locality bias of GNN-based diffusion models, particularly as a key challenge when generation is performed at large scales. To address these gaps, we introduce crystal host-guided generation (CHGGen), which integrates inpainting generation based on symmetrized frameworks and a foundation potential for structure optimization. We demonstrate the effectiveness of host-guided generation through a case study on CSP within the ZnS–P₂S₅ chemical space, and showcase the broader applicability of CHGGen across the continuous chemical space of the Li_x–Si alloy system. Finally, we discuss the limitations and potential opportunities of applying CHGGen with state-of-the-art generative models for CSP problems in future directions.

2 Theory

We first briefly recap the concepts of diffusion models with score-based denoising and then introduce inpainting as a conditional generation method for the structural modification of crystal structures.

2.1 Diffusion model

Generating samples from a probability density function $p(x)$ in high-dimensional space \mathbb{R}^d can be achieved by modeling the gradient of the log-probability density, known as the score function $\nabla_x \log p(x)$ in diffusion models. Song *et al.*³⁸ demonstrated that both the diffusion process and its reverse can be formulated as stochastic differential equations (SDE)

$$dx = f(x,t)dt + g(t)dw, \quad (1)$$

$$dx = [f(x,t) - g^2(t)\nabla_x \log p_t(x)]dt + g(t)d\bar{w}, \quad (2)$$

where w and \bar{w} represent the standard Brownian motion process and its time-reversed analogue, respectively. $f(x,t)$ is the drift coefficient and $g(t)$ is the diffusion coefficient of $x(t)$. $p_t(x)$ denotes the probability density of $x(t)$. Here t is the time variable $t \in [0, T]$ to describe the diffusion process $\{x(t)\}_{t=0}^T$.

Eqn (1) describes the forward process to corrupt the data distribution $x(0) \sim p_0(x)$ to obtain the prior distribution $x(T) \sim p_T(x)$, which follows a uniform distribution. Eqn (2) describes the reverse process to sample $x(0)$ by solving the reverse SDE with the score term $\nabla_x \log p(x)$. For crystal structure generation, we adopt the variance-exploding (VE) diffusion scheme for the atomic coordinates, where the process $\{x(t)\}_{t=0}^T$ is given by the SDE

$$dx = \sqrt{\frac{d[\sigma^2(t)]}{dt}}dw. \quad (3)$$

Here $\{\sigma(t)\}$ is a sequence of exponentially increasing standard deviations given $\sigma_{\min} = \sigma_1, \dots, \sigma_T = \sigma_{\max}$. The VE-SDE is particularly suitable for atomic coordinates in crystals, as VE-SDE does not induce disconnected graphs at the large noisy limit under periodic boundary conditions.

The samples can be generated using ancestral sampling, where successive states are sampled according to:

$$x_{t-1} = x_t + (\sigma_t^2 - \sigma_{t-1}^2)s_\theta(x_t, t) + z\sqrt{\frac{\sigma_{t-1}^2(\sigma_t^2 - \sigma_{t-1}^2)}{\sigma_t^2}} \quad (4)$$

where $x_T \sim \mathcal{N}(\mathbf{0}, \sigma_T^2 \mathbf{I})$, and $z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. In the continuous limit, $\sqrt{\sigma_{t-1}^2/\sigma_t^2} \approx 1$. The implementation is achieved using a predictor-corrector sampling strategy with the Langevin corrector. We refer the readers to ref. 38 for mathematical details of score-based SDE and sampling strategies.

Unconditional generation. Inputs: randomly initialized atomic positions x_T . Signal-to-noise ratio δ . Number of predictor steps T ; number of corrector steps M .

for $t = T, \dots, 1$ **do**

$$x_{t-1} \leftarrow x_t + (\sigma_t^2 - \sigma_{t-1}^2)s_\theta(x_t, t)$$

$$z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$$x_{t-1} \leftarrow x_{t-1} + \sqrt{\frac{\sigma_{t-1}^2(\sigma_t^2 - \sigma_{t-1}^2)}{\sigma_t^2}}z$$

for $j = 1, \dots, M$ **do**

$$z \sim \mathcal{N}(0, 1)$$

$$g \leftarrow s_\theta(x_{t-1}, t - 1)$$

$$\varepsilon \leftarrow 2(\sqrt{3}\delta/\|g\|_2)^2$$

$$x_{t-1} \leftarrow x_{t-1} + \varepsilon g + \sqrt{2\varepsilon}z$$

end for

end for

2.2 Denoising score matching

To estimate the score function $\nabla_x \log p_t(x)$, we use score matching (SM) to optimize the model parameters θ^* by minimizing

$$\mathcal{L}_{\text{SM}} = \mathbb{E}_{p_t(x)} [\|s_\theta(x(t), t) - \nabla_x \log p_t(x)\|^2]. \quad (5)$$

$\mathbb{E}_{p_t(x)}[\cdot]$ represents the expectation value with respect to the probability distribution $p_t(x)$, which can be approximated by a Gaussian transition probability $p(x(t)|x(0)) \propto e^{-[x(t) - x(0)]^2/2\sigma^2}$ such that eqn (5) is formulated as denoising score matching (DSM) with^{39,40}

$$\mathcal{L}_{\text{DSM}} = \mathbb{E}_{x(0)} \mathbb{E}_{x(t)|x(0)} \left[\left\| s_{\theta}(x(t), t) + \frac{e}{\sigma} \right\|^2 \right]. \quad (6)$$

Here $x(0) \sim p_0(x)$ and $x(t) \sim p(x(t)|x(0))$, $s_{\theta}(x(t), t)$ is the score function predicted by the GNN model, e represents the normalized noise $[x(t) - x(0)]/\sigma \sim \mathcal{N}(0, \mathbf{I})$. In the training process, σ is sampled uniformly from the interval $[\sigma_{\min}, \sigma_{\max}]$ to perturb the configuration $x(0)$ and obtain the noisy configuration $x(t)$ to construct the DSM loss in eqn (6).

2.3 Inpainting

Inpainting is a conditional generation process where a model completes missing elements within a given context. Inpainting has demonstrated significant applications in materials and chemistry, including the discovery of chemical reaction transition states⁴¹ and the generation of symmetry-constrained 2D materials.⁴² In CSP, inpainting enables the optimal placement of additional atoms (termed guest atoms) within a predefined host crystal structure, where a binary masking strategy applies different noise treatments to known regions (host structure) and unknown regions (areas to be inpainted – guest atoms).

Unlike training a certain distribution of the mask, Lugmayr *et al.*⁴³ introduced the repaint algorithm (inpainting + resampling) for high-quality 2D image inpainting using diffusion models. One can simply train the diffusion model with DSM to learn the joint distribution. During inference, the conditional distribution is approached using the resampling technique for inpainting generation. As shown in the Algorithm, in addition to the unconditional generation steps, (the resampling repeatedly “jumps back” in the diffusion process and resamples the unknown regions multiple times with r steps) at each timestep t , with a mask m to separate the host and guest atoms. This resampling procedure helps harmonize the generated content with existing regions by allowing multiple attempts at generating coherent inpainted content. For detailed implementation and theoretical foundations, we refer readers to ref. 43 and 44 for details of this approach.

Inpainting generation. Inputs: atomic positions of unperturbed host structure with randomly initialized guest atoms x_0^{host} ; atomic positions of all atoms sampled randomly in the unit cell x_r ; mask for guest atoms m ; signal-to-noise ratio δ ; number of predictor steps T ; number of corrector steps M ; number of resampling steps r .

for $t = T, \dots, 1$ **do**

for $n = 1, \dots, r$ **do**

$$x_{t-1} \leftarrow x_t + (\sigma_t^2 - \sigma_{t-1}^2) s_{\theta}(x_t, t)$$

$$z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$$x_{t-1} \leftarrow x_{t-1} + \sqrt{\frac{\sigma_{t-1}^2(\sigma_t^2 - \sigma_{t-1}^2)}{\sigma_t^2}} z$$

for $j = 1, \dots, M$ **do**

$$z \sim \mathcal{N}(0, 1)$$

$$g \leftarrow s_{\theta}(x_{t-1}, t - 1)$$

$$\varepsilon \leftarrow 2(\sqrt{3}\delta/\|g\|_2)^2$$

$$x_{t-1} \leftarrow x_{t-1} + \varepsilon g + \sqrt{2\varepsilon} z$$

end for

$$x_{t-1}^{\text{host}} \leftarrow x_0^{\text{host}} + \sigma_{t-1} z$$

$$x_{t-1} \leftarrow (1 - m) \odot x_{t-1}^{\text{host}} + m \odot x_{t-1}$$

if $n < r$ **and** $t > 1$ **then**

$$z \sim \mathcal{N}(0, 1)$$

$$x_t \leftarrow x_{t-1} + \sqrt{\sigma_{t-1}^2 - \sigma_{t-2}^2} z$$

end if

end for

end for

Fig. 1 illustrates the iterative sampling procedure for inpainting generation. During each reverse diffusion step, the process follows several distinct stages: first, the atoms in the host structure are perturbed using Gaussian noise determined by the noise scheduler of the subsequent step σ_{t-1} (process A):

$$x_{t-1}^{\text{host}} = x_0^{\text{host}} + \sigma_{t-1} z, \quad z \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (7)$$

This ensures that both the guest atoms and host structure in the current configuration x_{t-1} maintain comparable noise scales. The GNN model then computes the score using all atoms and executes the reverse diffusion *via* eqn (4) (process B). In our notation, x^{host} contains atomic positions of all atoms (including the framework and guest atoms) for indexing consistency. We apply masks $(1 - m)$ to x^{host} and m to x (process C) to construct x_{t-1} (process D):

$$x_{t-1} \leftarrow (1 - m) \odot x_{t-1}^{\text{host}} + m \odot x_{t-1} \quad (8)$$

Through this iterative reverse diffusion process, the noise scale $\{\sigma_t\}$ gradually decreases, resulting in a final crystal structure that is closely aligned with the original host structure with minimal deviation (*e.g.*, $\sigma_{\min} = 0.001 \text{ \AA}$). The positions of the guest atoms are therefore determined by the distribution conditioned on the host structure.

3 Results

We developed an SE(3)-equivariant graph neural network (GNN) based on the NequIP architecture to predict the score function for both unconditional and inpainting diffusion processes. The model training followed the DSM scheme described in eqn (6), where the dataset was prepared using crystal structures from the MP database with energy above hull $E_{\text{hull}} < 0.1 \text{ eV}$ (see Methods in SI).

In the following sections, we first examine the locality bias encountered when generating structures with large unit cells through unconditional generation. These insights led to the development of the CHGGen framework. We demonstrate the

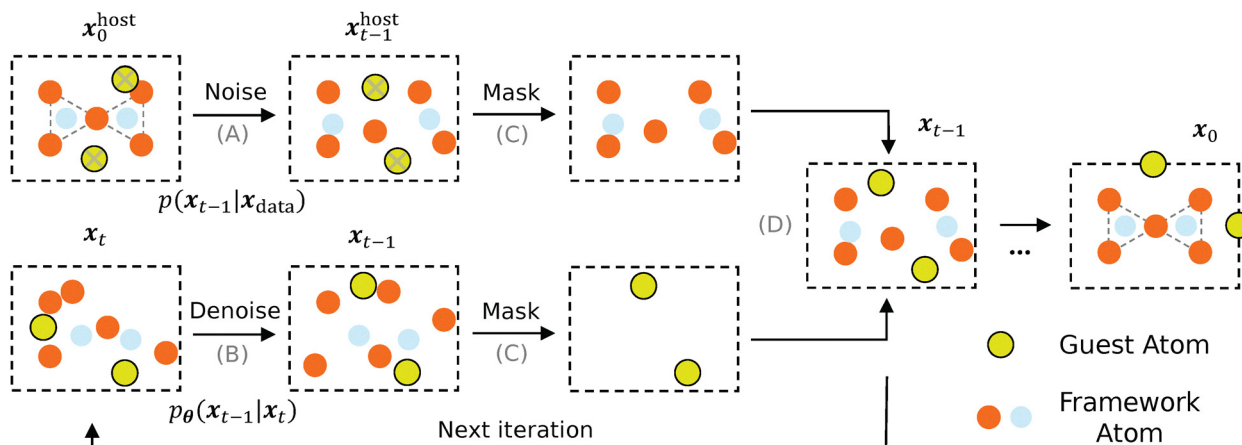


Fig. 1 Illustration of the iterative sampling strategy for structure inpainting with a host framework. Process (A): add noise to the host framework. Process (B): denoise atomic configuration from x_t to x_{t-1} using scores $\{s_\theta(x_t)\}$ predicted by the GNN. Process (C): apply masks to the host structure and the guest atoms. Process (D): combine the host structure and guest atoms to form the configuration x_t for the next iteration. Note that the symbol x^{host} contains atomic positions of all atoms (including both host and guest atoms). The gray crosses on the guest atoms indicate that their information is not used in the processes but is retained for indexing consistency.

effectiveness of CHGGen on two example chemical systems: Zn-P-S and Li-Si, which are complemented by CHGNet as a foundation potential for iterative structural optimizations and thermodynamic stability screening. Additionally, we also demonstrate example studies of CSP of 16 compositions used in DiffCSP²⁷ and solid-solid interface in SI.

3.1 Locality bias of GNN-based generative models

To investigate the limitations of unconditional generation at large scales, we generated 10 supercell configurations of

$(\text{Li}_4\text{S}_2\text{-P}_2\text{S}_5)_8$ with the parametrized diffusion models. Fig. 2(a) shows a snapshot of the generated structures, where the long-range periodicity is absent and an amorphous configuration is exhibited. Fig. 2(c) and (d) present the radial distribution functions (RDF) of the generated structures and the structures from the MP database, showing similar major peaks in both RDF plots. The RDF suggests that the generated structures exhibit physically reasonable Li-S and P-S bonding environments, which represent the learned local distribution from the dataset.

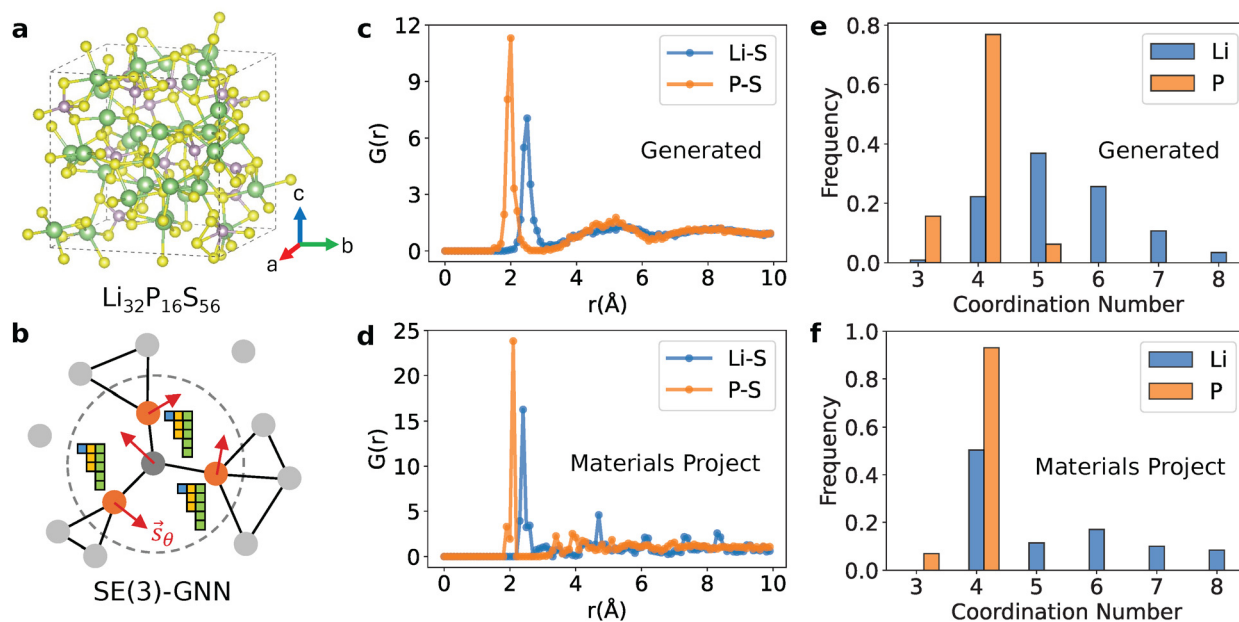


Fig. 2 Analysis of the locality bias in GNN-based diffusion models. (a) A generated supercell structure of $\text{Li}_{32}\text{P}_{16}\text{S}_{56}$ exhibiting an amorphous configuration. (b) Illustration of SE(3)-equivariant graph neural networks, where the score function $\{s_\theta(x_t)\}$ is predicted as a vector from each graph node (red arrows). (c) and (d) Radial distribution functions (RDF) of Li-S and P-S in generated structures compared with database structures from the MP. (e) and (f) Comparison of local chemical environments grouped by coordination number between generated structures and MP structures.

We evaluated local coordination environments using the LocalGeometryFinder toolkit⁴⁵ and classified the local environments based on coordination numbers. In Fig. 2(e), P atoms predominantly occupy tetrahedral sites (4-coordinated), consistent with known Li–P–S crystal structures.⁴⁶ In contrast, Li exhibits a broad distribution of coordination numbers, with peaks at 5-fold (~40%), 4-fold (~25%), and 6-fold (~30%) geometries. Notably, these coordination statistics qualitatively align with patterns observed in the MP training dataset (Fig. 2(f)), where P atoms maintain rigid tetrahedral coordination while Li atoms display more variable environments.

Based on the successful learning and reconstruction of local distribution from the generative model, we hypothesize that the failure to propagate long-range order in generated structures stems from two interrelated factors: (1) locality bias in GNNs: while the model effectively captures short- to medium-range atomic correlations, its finite receptive field constrains the learning of global crystallographic patterns. (2) Stochasticity in reverse diffusion: the stochastic differential equation for reverse diffusion processes inherently samples from a learned distribution of the entire dataset. Without coupling the atomic arrangements and supercell parameters, the diffusion process tends to sample from the entire distribution of the dataset in the generated structures, rather than from a narrowed distribution in specific crystal systems. Consequently, structures with large unit cells manifest as “mosaics” of local structure motifs rather than coherent crystalline structures. These limitations may be universal even with lattice diffusion, as the GNN architecture lacks explicit mechanisms to maintain long-range crystallographic order when featuring the atomic configurational space.⁴⁷

3.2 Crystal host-guided generation

Motivated by our observation that GNN-based diffusion models struggle to generate crystalline structures with long-range periodic structures, we developed the CHGGen framework as a targeted approach to mitigate this limitation. CHGGen integrates three key components: (1) unconditional structure generation, (2) inpainting generation based on symmetry-refined host structure, and (3) structural optimization using the CHGNet.

Fig. 3 illustrates the CHGGen computational workflow. The process begins with sampling various Bravais lattices at a fixed volume through a random search over lattice constant ratios and angles (see Methods in SI). The unit cell volume is determined as $N \times V_0$, where N represents the number of atoms and V_0 denotes the atomic volume. The V_0 can be initialized either from related crystalline phases or predicted by composition-based regression models.⁴⁸ This atomic volume serves as prior information subject to optimization in subsequent steps. Following lattice determination, fractional coordinates for all atoms are initialized with random numbers drawn from $\mathcal{N}(0, 1)$. The diffusion process then proceeds by solving the reverse SDE using scores predicted by the SE(3)-GNN (unconditional generation). Given that the volumes and lattices of the generated structures are drawn from simple priors and random search, the CHGNet is employed for structure relaxation to optimize both unit cells and atomic coordinates. This process represents a well-defined local energy minima search task and does not suffer from the locality bias encountered during the diffusion process.

The next phase is initiated by removing atoms that exhibit broad local environment distributions (*e.g.*, Li). The remaining

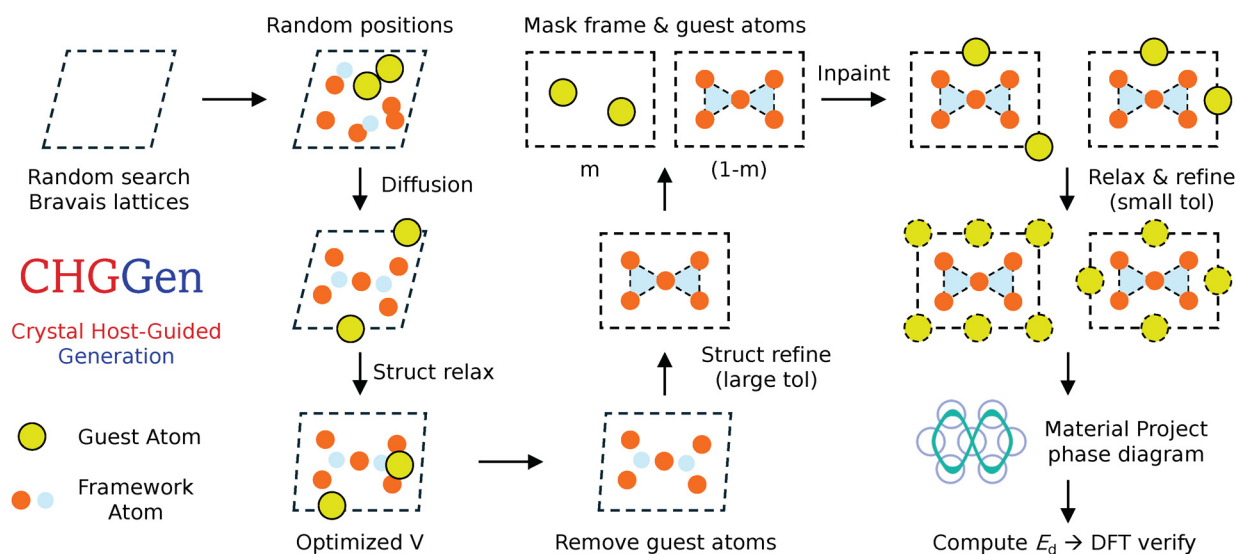


Fig. 3 Computational workflow of CHGGen. The process begins with a random search for Bravais lattices containing a specified number of atoms, followed by an unconditional generation with reverse diffusion and structure relaxation using CHGNet. Structure refinement is applied after removing guest atoms to obtain a symmetrized framework. Inpainting generation is then performed based on this refined framework to guide the creation of complete crystal structures. Finally, the generated structures undergo relaxation to determine decomposition energy, with promising candidates (those exhibiting low decomposition energy) selected for DFT verification. The dashed circles represent crystallographically equivalent atomic positions in a crystal structure.

structure (framework) undergoes symmetry refinement using spglib through incremental structural matching tolerance to obtain a space group with higher symmetry (*i.e.*, until the space group is not P1). Since the guest atoms exhibit diverse local environment distributions, refinement without them is more feasible for obtaining the symmetric structure. The fractional coordinates of the removed guest atoms are then reinitialized from $\mathcal{N}(\mathbf{0}, \mathbf{I})$ within the symmetrized framework, and inpainting generation is performed using masks m and $(1 - m)$ for the guest and framework atoms, respectively.

The inpainting-generated structures are further relaxed using CHGNet and structure refinement is performed with a small tolerance to obtain the space group. The CHGNet-calculated energy for the relaxed structure is used to determine the decomposition energy E_d relative to the MP phase diagram at the GGA/GGA+ U level of accuracy. Finally, structures with E_d within a specified threshold (*e.g.*, $E_d < 0.1$ eV per atom) are submitted for DFT calculations to obtain more accurate thermodynamic stability assessments. In our studies, we used the r^2 SCAN functional to evaluate the DFT decomposition energy against the MP r^2 SCAN phase diagram.

3.3 Example: Zn–P–S

The first example predicts the crystal structure in the ZnS–P₂S₅ chemical space, which represents a logical extension of the related Li–P–S system that exhibits various stable and metastable polymorphs along the Li₂S–P₂S₅ compositional line.⁴⁹

Understanding phase stability in analogous Zn-based systems is important for advancing Zn-based solid-state batteries.

We focused on the CSP of ZnP₂S₆ and Zn₂P₂S₇ using CHGNet. To assess the local stability of the generated structures, we evaluated the structural and energetic differences between the initially generated structures and their CHGNet-relaxed counterparts. In Fig. 4(a) and (b), we present the energy and geometrical differences between the relaxed structures and generated structures. Most of the generated structures exhibit energy changes of $\Delta E < 0.1$ eV per atom, with a median value of 0.08 eV per atom. The geometric differences, quantified by maximum pair-wise root-mean-squared distance (RMSD), show a median value of 0.10 Å between relaxed and generated structures. As illustrated in Fig. 4(a) and (b), outliers with large energy changes correspond to RMSD values exceeding 0.3 Å, indicating that most of the generated structures are close to the local minima and can be reasonably searched using foundation potential structure relaxation.

Fig. 4(c) presents the CHGNet-predicted decomposition energies (E_d) to quantify the thermodynamic stability. The distribution shows a median E_d of 0.07 eV per atom, indicating that the majority of generated structures are metastable in this chemical space, while 6.5% of the generated structures are identified as stable with $E_d < 0$ at the CHGNet level of accuracy.

After r^2 SCAN-DFT calculations and reevaluation of E_d with respect to the MP- r^2 SCAN phase diagram, the structure with negative E_d was confirmed to be metastable. This discrepancy can be attributed to CHGNet's approximate 30 meV per atom

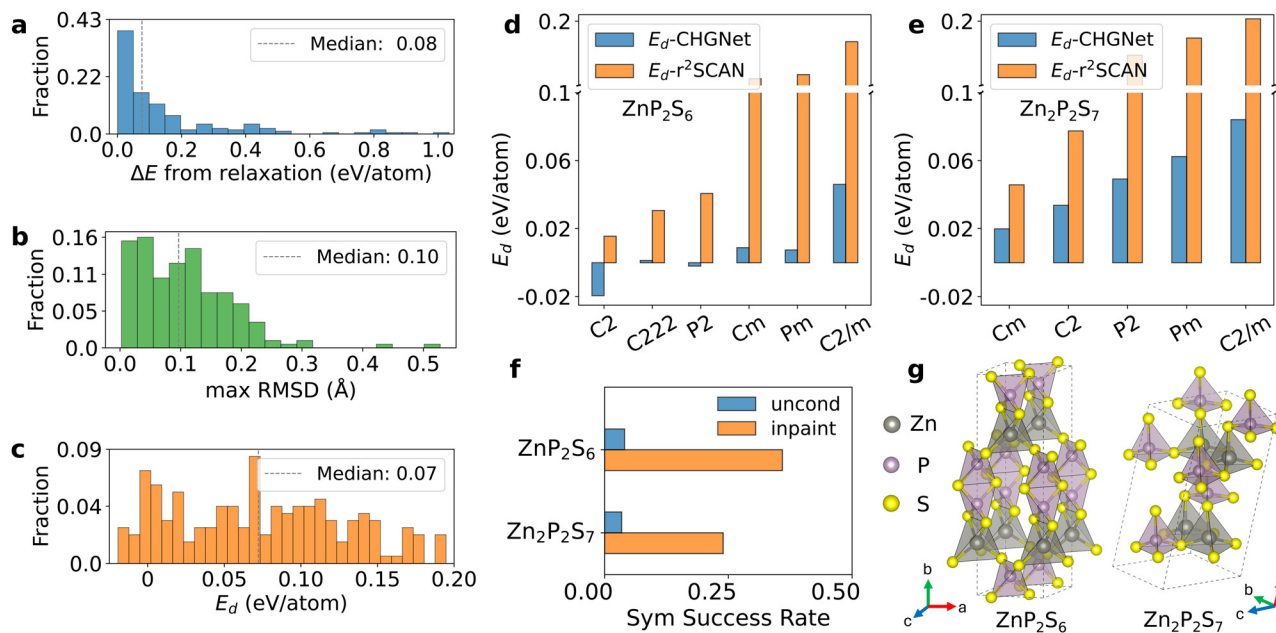


Fig. 4 Generation results in the Zn–P–S system demonstrating local stability, global stability, and capability for identifying symmetric structures. (a) Energy change (ΔE) following structure relaxation. (b) Maximum pair-wise root-mean-squared displacement (RMSD) representing differences between generated and relaxed structures. (c) Distribution of decomposition energies (E_d) of generated structures relative to the MP phase diagram as predicted by CHGNet. (d) and (e) Comparison of decomposition energies predicted by CHGNet (blue) and r^2 SCAN-DFT (orange) relative to the MP phase diagram at GGA/GGA+ U (blue) and r^2 SCAN (orange) levels of accuracy. (f) Success rate of identifying symmetric crystal structures using unconditional vs. inpainting generation. (g) Examples of generated structures with the lowest DFT decomposition energy in ZnP₂S₆ (C2, $E_d = 0.015$ eV per atom) and Zn₂P₂S₇ (Cm, $E_d = 0.046$ eV per atom).

prediction error, which may alter the predicted phase stability when E_d values are small. As illustrated in Fig. 4(d) and (e), while CHGNet demonstrates good overall agreement with DFT, it systematically underestimates the decomposition energy for structures with high values ($E_d > 0.1$ eV per atom), consistent with the known softening effect of foundation potentials.⁵⁰ This finding suggests that practitioners could consider using a lower threshold (e.g., $E_d < 30$ meV per atom) when screening generated structures for DFT validation.

To evaluate how diffusion models perform in generating symmetric crystal structures, we define a simple metric symmetry success rate as the fraction of generated structures possessing space groups with higher symmetry than $P1$ or $\bar{P}1$ after relaxation and refinement (see Methods in SI). Fig. 4(f) compares the symmetry success rates from unconditional (blue) and inpainting (orange) generations. The inpainting approach uses the refined P-S frameworks to achieve significantly higher success rates compared to the unconditional ones (<5%), which highlights the effectiveness and practical advantages of conditional generation with structural priors. The structures with the lowest- E_d for ZnP_2S_6 and $\text{Zn}_2\text{P}_2\text{S}_7$ are illustrated in Fig. 4(g). While the generated structures are predicted to be metastable by DFT calculations, this example demonstrates CHGGen's potential for exploring crystal structures in the chemical space that is currently absent from existing databases.

3.4 Example: Li-Si alloys

The second example extends to a binary system – Li-Si alloys. The Li-Si system has particular significance as a high-capacity anode in Li-ion batteries.⁵¹ The MP database contains 13

DFT-calculated structures, of which 4 are thermodynamically stable ($E_{\text{hull}} = 0$) at zero K.

To demonstrate the effectiveness of identifying low-energy polymorphs, we performed structure generation for Li-rich phases (Li_xSi_y , where $x > y$, a composition range known to contain many stable phases⁵²) following the workflow outlined in Fig. 3 (see Methods in SI). After CHGGen generation and $r^2\text{SCAN-DFT}$ calculations, we constructed the phase diagram using DFT formation energies from both MP structures and the generated structures. In Fig. 5(a), green circles represent new stable polymorphs identified by CHGGen, while blue squares indicate the reported stable MP structures. We calculated decomposition energies using the MP phase diagram (without the on-the-hull structure of Li_5Si_2). Negative values of E_d therefore indicate compounds that break the existing MP convex hull. Notably, CHGGen successfully predicted Li_5Si_2 ($R\bar{3}m$, $E_d = -7$ meV per atom) as a thermodynamically stable polymorph beyond the known MP stable structures (Fig. 5(b)). Interestingly, this structure had also been identified through previous studies, including work by Tipton *et al.*⁵³ using genetic algorithms and Morris *et al.*⁵⁴ using random structure search coupled with DFT calculations.

Using generative models, one can also identify metastable polymorphs with low decomposition energies in related compositions. For example, structures for compositions corresponding to Li_5Si_2 ($E_d = -0.006$ eV per atom), Li_2Si ($E_d = 0.003$ eV per atom), and Li_4Si ($E_d = 0.010$ eV per atom) with $C2/m$ space group are shown in Fig. 5(c). These low-energy polymorphs with different local structure motifs provide a more detailed mapping of the potential energy landscape, which benefits the training of MLIPs to understand the Si network

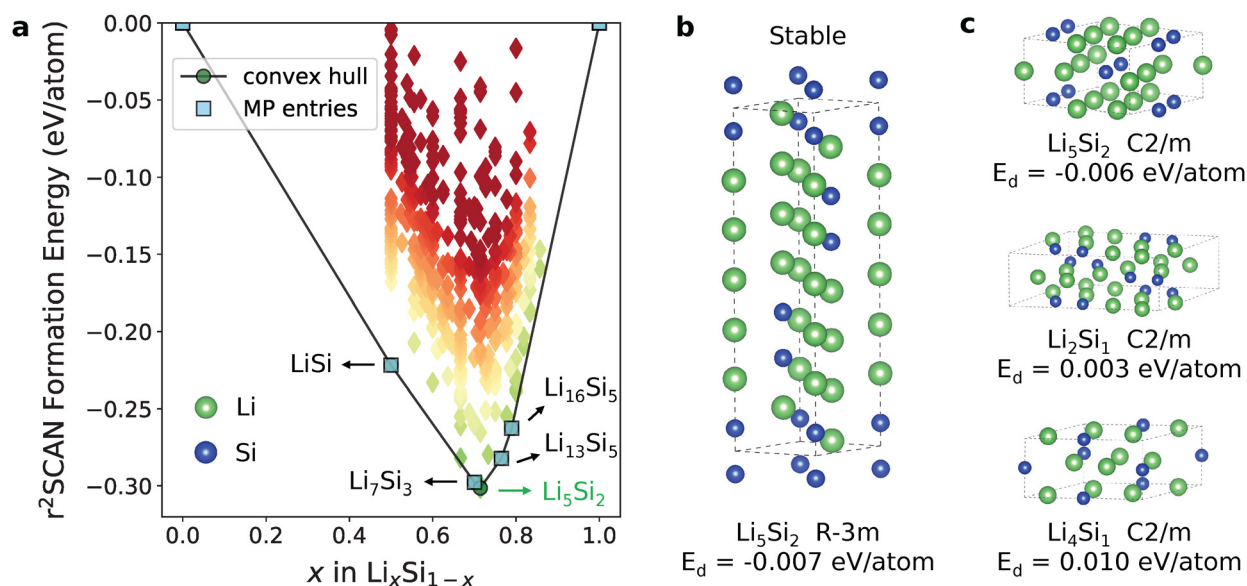


Fig. 5 Formation energies and crystal structures in the Li-Si chemical system. (a) Formation energy phase diagram calculated with generated structures in the Li-Si chemical system using $r^2\text{SCAN-DFT}$. Blue squares represent stable structures from the MP database, green dots indicate stable compounds on the formation energy convex hull, and the diamonds are the metastable compounds above the convex hull. (b) Stable generated structure (Li_5Si_2 , $R\bar{3}m$) confirmed by DFT calculations ($E_d = -0.007$ eV per atom). (c) Generated metastable structures with $C2/m$ space group in Li_5Si_2 , Li_2Si , and Li_4Si compositions, respectively.

aggregation and its effect on Li transport kinetics.⁵⁵ The case study illustrates how practitioners can benefit from combining diffusion models and foundation potentials to explore continuous unknown chemical spaces, such as those encountered in alloy design.

4 Discussion

Recent advances in diffusion models have shown their promise for crystal structure prediction (CSP), leveraging their ability to learn geometric features beyond elemental substitution heuristics.⁵⁶ Graph neural networks have emerged as the preferred architectural choice for diffusion-based approaches, primarily due to their inherent capability to incorporate rotational equivariance.^{57,58} A particularly significant theoretical insight is that the score function derived from the GNN-based diffusion models is mathematically equivalent to interatomic forces under harmonic potential approximation.²² This equivalence reveals that denoising pretraining provides substantial benefits for interatomic potential modeling,⁵⁹ which enhances the efficiency of local energy minima exploration as evidenced by the leading performance in MatBench Discovery benchmarks.^{60,61}

Based on this concept, diffusion models with GNNs represent a logical framework for generating reasonable local structure motifs (*e.g.*, atomic bonding patterns), which proves valuable when optimizing local atomic arrangements in CSP problems. Nonetheless, their inherent locality bias limits their ability to capture long-range periodic orders. Gong *et al.*⁴⁷ revealed that state-of-the-art GNNs fall short of accurately capturing the periodicity of crystal structures, *i.e.*, lattice parameters. This fundamental limitation explains the diminished performance of GNN-based diffusion models when generating structures with large unit cells, where long-period crystallinity is significant but hard to capture. This locality bias is exacerbated for species that can adapt to diverse local environments (*e.g.*, Li), which have a broad distribution of stable coordination geometries. The unconditional generation defaults to producing a disordered “mosaic” of these competing local motifs rather than a coherent crystal structure.

The practical advantage of host-guided generation with inpainting is to augment the symmetric structure generation. To evaluate this, we compared the success rates of obtaining symmetric crystals across different approaches in Zn–P–S and Li–Si chemical spaces: MatterGen (green), CHGGen with unconditional generation (blue), and CHGGen with inpainting generation (orange). MatterGen demonstrates superior performance compared to our baseline model for Li–Si alloys (Fig. 6(b)), highlighting the importance of explicitly modeling lattice diffusion in conjunction with atomic arrangements. However, for more complex systems such as Zn–P–S that contain polyanions, MatterGen performs less effectively when the number of atoms exceeds 10, with success rates lower than CHGGen with inpainting generation. This highlights the importance of inpainting generation when dealing with chemical systems involving polyanions. Notably, MatterGen exhibits declining success rates as system size increases in both cases, suggesting that it is likely all GNN-based diffusion models face scalability challenges.

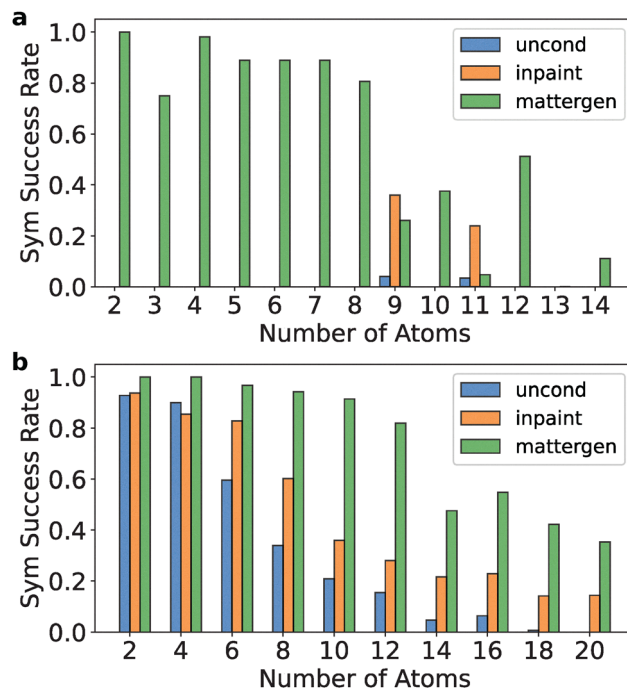


Fig. 6 Comparison of success rates for identifying symmetric crystal structures in (a) the Zn–P–S system and (b) the Li–Si system (blue: unconditional generation with CHGGen; orange: inpainting generation with GHGGen; green: MatterGen with chemical-system-guidance generation).

This scalability constraint potentially limits their ability to predict complex structural prototypes (*e.g.*, NASICON-type frameworks) across various solid-state materials.

Although CHGGen has not yet achieved industry-level performance as MatterGen in stability or general symmetry success rate, it shows clear improvements in generating symmetric crystal structures compared to unconditional generation methods. Assessing crystallographic symmetry in generated structures is critical for CSP, as discovering new structural prototypes is a key step toward materials discovery.²³ These prototypes can subsequently guide elemental substitution strategies¹⁵ or evolutionary searches⁶² for compositional optimization. The inpainting-based generation approach samples from the conditional distribution of unknown structural components within a given framework, yielding more well-defined local atomic arrangements than *de novo* sampling from a fully unconstrained distribution. Importantly, this inpainting strategy is simple to implement, as both conditional and unconditional generation operate within the same unified model—distinguished only by the masking strategy used during the reverse diffusion process. This modular design enables seamless integration with existing foundational generative models and provides a flexible mechanism for enforcing symmetry constraints during structure generation.

In addition, we highlight the practical significance of CHGGen for the structural modification of materials, which often relies on existing database structures. A notable example is the superionic conductor $\text{Li}_{0.388}\text{Ta}_{0.238}\text{La}_{0.475}\text{Cl}_3$, discovered through lithiation of the LaCl_3 -type host structure with

additional aliovalent substitution.⁶³ CHGGen framework offers a probabilistic approach to such design task that circumvents the need for topological analysis⁶⁴ or additional inputs such as DFT-derived charge densities,⁶⁵ which are often computationally prohibitive or not universally applicable.

Finally, while our framework demonstrates useful augmentation to diffusion-based generative models, the current symmetry refinement approach remains preliminary as it relies on spglib by simply increasing the tolerance threshold. As a proof-of-concept, this method predominantly yields structures with moderate symmetry (e.g., *C2*, *Cm* in monoclinic systems), which may limit the discovery of novel structure prototypes (see Fig. S1 and S2). Looking forward, several promising approaches have emerged for novel framework generation, including symmetry-constrained diffusion²⁴ and prototype-based generation using Wyckoff position-based representations.^{66–68} The integration of these advanced symmetry handling strategies with CHGGen could enhance the discovery of crystal structures particularly with intercalation chemistry.

In summary, we present CHGGen as an integrated framework that combines unconditional and inpainting generation with foundation potential optimization for CSP. While challenges in scaling complexity persist, the inpainting method provides a useful approach for generating symmetric crystals and incorporating intercalants into existing database structures. We anticipate broad adoption of this framework as the modular design of the inpainting methodology enables seamless integration with emerging diffusion models, which will ultimately accelerate materials discovery across diverse chemical spaces.

Conflicts of interest

The authors declare no conflicts of interest.

Data availability

The codebase and pretrained model for CHGGen are available at <https://github.com/zhongpc/chggen>.

The SI includes methods and supporting results for CHGGen-based structure generation. See DOI: <https://doi.org/10.1039/d5mh00774g>

Acknowledgements

This work was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division under Contract No. DE-AC0205CH11231 (Materials Project program KC23MP). The computations were supported by the National Energy Research Scientific Computing Center (NERSC) under the GenAI Project and the National Renewable Energy Laboratory (NREL) clusters under silimorphous allocation. P. Z. acknowledges funding support from the BIDMaP Postdoctoral Fellowship. The authors thank Bingqing Cheng, Yifan Chen, and Aaron Kaplan for valuable discussions.

Notes and references

- Z. Lu, B. Zhu, B. W. B. Shires, D. O. Scanlon and C. J. Pickard, *J. Chem. Phys.*, 2021, **154**, 174111.
- D. Zhou, I. Bier, B. Santra, L. D. Jacobson, C. Wu, A. Garaizar Suarez, B. R. Almaguer, H. Yu, R. Abel, R. A. Friesner and L. Wang, *Nat. Commun.*, 2025, **16**, 2210.
- Z. Zhang, T. Cui, M. J. Hutcheon, A. M. Shipley, H. Song, M. Du, V. Z. Kresin, D. Duan, C. J. Pickard and Y. Yao, *Phys. Rev. Lett.*, 2022, **128**, 047001.
- V. V. Gusev, D. Adamson, A. Deligkas, D. Antypov, C. M. Collins, P. Krysta, I. Potapov, G. R. Darling, M. S. Dyer, P. Spirakis and M. J. Rosseinsky, *Nature*, 2023, **619**, 68–72.
- C. J. Pickard and R. Needs, *J. Phys.: Condens. Matter*, 2011, **23**, 053201.
- A. R. Oganov and C. W. Glass, *J. Chem. Phys.*, 2006, **124**, 244704.
- Y. Wang, J. Lv, L. Zhu and Y. Ma, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2010, **82**, 094116.
- G. Hautier, C. Fischer, V. Ehrlicher, A. Jain and G. Ceder, *Inorg. Chem.*, 2011, **50**, 656–663.
- W. Huang, D. A. Kitchaev, S. T. Dacek, Z. Rong, A. Urban, S. Cao, C. Luo and G. Ceder, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2016, **94**, 134424.
- A. Ferrari, F. Körmann, M. Asta and J. Neugebauer, *Nat. Comput. Sci.*, 2023, **3**, 221–229.
- C. Chen and S. P. Ong, *Nat. Comput. Sci.*, 2022, **2**, 718–728.
- B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel and G. Ceder, *Nat. Mach. Intell.*, 2023, **5**, 1031–1041.
- I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, W. J. Baldwin and N. Bernstein, *et al.*, *arXiv*, 2023, preprint, arXiv:2401.00096, DOI: [10.48550/arXiv.2401.00096](https://doi.org/10.48550/arXiv.2401.00096).
- J. Kim, J. Kim, J. Kim, J. Lee, Y. Park, Y. Kang and S. Han, *J. Am. Chem. Soc.*, 2025, **147**, 1042–1054.
- A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon and E. D. Cubuk, *Nature*, 2023, **624**, 80–85.
- D. Zhang, X. Liu, X. Zhang, C. Zhang, C. Cai, H. Bi, Y. Du, X. Qin, A. Peng and J. Huang, *et al.*, *npj Comput. Mater.*, 2024, **10**, 293.
- H. Yang, C. Hu, Y. Zhou, X. Liu, Y. Shi, J. Li, G. Li, Z. Chen, S. Chen and C. Zeni, *et al.*, *arXiv*, 2024, preprint, arXiv:2405.04967, DOI: [10.48550/arXiv.2405.04967](https://doi.org/10.48550/arXiv.2405.04967).
- X. Fu, B. M. Wood, L. Barroso-Luque, D. S. Levine, M. Gao, M. Dzamba and C. L. Zitnick, *arXiv*, 2025, preprint, arXiv:2502.12147, DOI: [10.48550/arXiv.2502.12147](https://doi.org/10.48550/arXiv.2502.12147).
- Z. Ren, S. I. P. Tian, J. Noh, F. Oviedo, G. Xing, J. Li, Q. Liang, R. Zhu, A. G. Aberle, S. Sun, X. Wang, Y. Liu, Q. Li, S. Jayavelu, K. Hippalgaonkar, Y. Jung and T. Buonassisi, *Matter*, 2022, **5**, 314–335.
- H. Park, A. Onwuli and A. Walsh, *Nat. Commun.*, 2025, **16**, 4379.
- B. Cheng, *J. Chem. Theory Comput.*, 2024, **20**, 9259–9266.
- T. Xie, X. Fu, O.-E. Ganea, R. Barzilay and T. Jaakkola, International Conference on Learning Representations (ICLR), 2021.
- N. J. Szymanski and C. J. Bartel, *Mater. Horiz.*, 2025, DOI: [10.1039/D5MH00010F](https://doi.org/10.1039/D5MH00010F).

- 24 D. Levy, S. S. Panigrahi, S.-O. Kaba, Q. Zhu, K. L. K. Lee, M. Galkin, S. Miret and S. Ravanbakhsh, International Conference on Learning Representations (ICLR), 2025.
- 25 G. Kurz, I. Gilitschenski and U. D. Hanebeck, 2014 Sensor Data Fusion: Trends, Solutions, Applications (SDF), 2014, pp. 1-5.
- 26 V. D. Bortoli, E. Mathieu, M. Hutchinson, J. Thornton, Y. W. Teh and A. Doucet, *Adv. Neural Inf. Process. Syst.*, 2022, 2406–2422.
- 27 R. Jiao, W. Huang, P. Lin, J. Han, P. Chen, Y. Lu and Y. Liu, *Adv. Neural Inf. Process. Syst.*, 2023, 17464–17497.
- 28 C. Zeni, R. Pinsler, D. Zügner, A. Fowler, M. Horton, X. Fu, Z. Wang, A. Shysheya, J. Crabbé, S. Ueda, R. Sordillo, L. Sun, J. Smith, B. Nguyen, H. Schulz, S. Lewis, C.-W. Huang, Z. Lu, Y. Zhou, H. Yang, H. Hao, J. Li, C. Yang, W. Li, R. Tomioka and T. Xie, *Nature*, 2025, 1–56.
- 29 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Mater.*, 2013, 1, 011002.
- 30 J. Schmidt, N. Hoffmann, H. Wang, P. Borlido, P. J. M. A. Carriço, T. F. T. Cerqueira, S. Botti and M. A. L. Marques, *Adv. Mater.*, 2023, 35, 2210788.
- 31 S. Yang, K. Cho, A. Merchant, P. Abbeel, D. Schuurmans, I. Mordatch and E. D. Cubuk, *arXiv*, 2023, preprint, arXiv:2311.09235, DOI: [10.48550/arXiv.2311.09235](https://doi.org/10.48550/arXiv.2311.09235).
- 32 M. Aykol, A. Merchant, S. Batzner, J. N. Wei and E. D. Cubuk, *Nat. Comput. Sci.*, 2024, 5, 105–111.
- 33 N. Gruver, A. Sriram, A. Madotto, A. G. Wilson, C. L. Zitnick and Z. Ulissi, International Conference on Learning Representations (ICLR), 2024.
- 34 L. M. Antunes, K. T. Butler and R. Grau-Crespo, *Nat. Commun.*, 2024, 15, 10570.
- 35 H. H. Li, J.-X. Shen and K. A. Persson, *Energy Adv.*, 2024, 3, 255–262.
- 36 Y.-L. Liao, B. Wood, A. Das and T. Smidt, International Conference on Learning Representations (ICLR), 2024.
- 37 N. Rønne, A. Aspuru-Guzik and B. Hammer, *Phys. Rev. B*, 2024, 110, 235427.
- 38 Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon and B. Poole, International Conference on Learning Representations (ICLR), 2020.
- 39 P. Vincent, *Neural Comput.*, 2011, 23, 1661–1674.
- 40 J. Ho, A. Jain and P. Abbeel, *Adv. Neural Inf. Process. Syst.*, 2020, 33, 6840–6851.
- 41 C. Duan, Y. Du, H. Jia and H. J. Kulik, *arXiv*, 2023, preprint, arXiv:2304.06174, DOI: [10.48550/arXiv.2304.06174](https://doi.org/10.48550/arXiv.2304.06174).
- 42 M. Li, R. Okabe, M. Cheng, A. Chottratanapituk, N. T. Hung, X. Fu, B. Han, Y. Wang, W. Xie and R. Cava, *et al.*, 2024.
- 43 A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte and L. Van Gool, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
- 44 X. Dai, P. Zhong, B. Deng, Y. Chen and G. Ceder, ICML 2024 AI for Science Workshop, 2024.
- 45 D. Waroquiers, X. Gonze, G. M. Rignanese, C. Welker-Nieuwoudt, F. Rosowski, M. Göbel, S. Schenk, P. Degelmann, R. André, R. Glaum and G. Hautier, *Chem. Mater.*, 2017, 29, 8346–8360.
- 46 B. Lee, K. Jun, B. Ouyang and G. Ceder, *Chem. Mater.*, 2023, 35, 891–899.
- 47 S. Gong, K. Yan, T. Xie, Y. Shao-Horn, R. Gomez-Bombarelli, S. Ji and J. C. Grossman, *Sci. Adv.*, 2023, 9, eadi3245.
- 48 R. E. A. Goodall and A. A. Lee, *Nat. Commun.*, 2020, 11, 6280.
- 49 C. Szczuka, B. Karasulu, M. F. Groh, F. N. Sayed, T. J. Sherman, J. D. Bocarsly, S. Vema, S. Menkin, S. P. Emge, A. J. Morris and C. P. Grey, *J. Am. Chem. Soc.*, 2022, 144, 16350–16365.
- 50 B. Deng, Y. Choi, P. Zhong, J. Riebesell, S. Anand, Z. Li, K. Jun, K. A. Persson and G. Ceder, *npj Comput. Mater.*, 2025, 11, 9.
- 51 D. H. S. Tan, Y.-T. Chen, H. Yang, W. Bao, B. Sreenarayanan, J.-M. Doux, W. Li, B. Lu, S.-Y. Ham, B. Sayahpour, J. Scharf, E. A. Wu, G. Deysher, H. E. Han, H. J. Hah, H. Jeong, J. B. Lee, Z. Chen and Y. S. Meng, *Science*, 2021, 373, 1494–1499.
- 52 N. Artrith, A. Urban and G. Ceder, *J. Chem. Phys.*, 2018, 148, 241711.
- 53 W. W. Tipton, C. R. Bealing, K. Mathew and R. G. Hennig, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, 87, 184114.
- 54 A. J. Morris, C. P. Grey and C. J. Pickard, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2014, 90, 054111.
- 55 E. Sivonxay, M. Aykol and K. A. Persson, *Electrochim. Acta*, 2020, 331, 135344.
- 56 H. Park, Z. Li and A. Walsh, *Matter*, 2024, 7, 2355–2367.
- 57 J. Gasteiger, F. Becker and S. Günnemann, *Adv. Neural Inf. Process. Syst.*, 2021, 6790–6802.
- 58 S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt and B. Kozinsky, *Nat. Commun.*, 2022, 13, 2453.
- 59 S. Zaidi, M. Schaarschmidt, J. Martens, H. Kim, Y. W. Teh, A. Sanchez-Gonzalez, P. Battaglia, R. Pascanu and J. Godwin, International Conference on Learning Representations (ICLR), 2023.
- 60 J. Riebesell, R. E. A. Goodall, P. Benner, Y. Chiang, B. Deng, G. Ceder, M. Asta, A. A. Lee, A. Jain and K. A. Persson, *Nat. Mach. Intell.*, 2025, 7, 836–847.
- 61 Y.-L. Liao, T. Smidt, M. Shuaiyi and A. Das, Transactions on Machine Learning Research (TMLR), 2024.
- 62 J. Gan, P. Zhong, Y. Du, Y. Zhu, C. Duan, H. Wang, C. P. Gomes, K. A. Persson, D. Schwalbe-Koda and W. Wang, *arXiv*, 2025, preprint, arXiv:2502.20933, DOI: [10.48550/arXiv.2502.20933](https://doi.org/10.48550/arXiv.2502.20933).
- 63 Y.-C. Yin, *et al.*, *Nature*, 2023, 616, 77–83.
- 64 X. He, Q. Bai, Y. Liu, A. M. Nolan, C. Ling and Y. Mo, *Adv. Energy Mater.*, 2019, 9, 1902078.
- 65 J.-X. Shen, M. Horton and K. A. Persson, *npj Comput. Mater.*, 2020, 6, 161.
- 66 R. Zhu, W. Nong, S. Yamazaki and K. Hippalgaonkar, *arXiv*, 2023, preprint, arXiv:2311.17916, DOI: [10.48550/arXiv.2311.17916](https://doi.org/10.48550/arXiv.2311.17916).
- 67 Z. Cao, X. Luo, J. Lv and L. Wang, *arXiv*, 2024, preprint, arXiv:2403.15734, DOI: [10.48550/arXiv.2403.15734](https://doi.org/10.48550/arXiv.2403.15734).
- 68 N. Kazeev, W. Nong, I. Romanov, R. Zhu, A. Ustyuzhanin, S. Yamazaki and K. Hippalgaonkar, *arXiv*, 2025, preprint, arXiv:2503.02407, DOI: [10.48550/arXiv.2503.02407](https://doi.org/10.48550/arXiv.2503.02407).