Inpainting crystal structure generations with score-based denoising

Xinzhe Dai [1 2 *] **Peichen Zhong** [1 2 *] **Bowen Deng** [1 2] **Yifan Chen** [3] **Gerbrand Ceder** [1 2]

Abstract

Searching for the optimal atomic position of additive atoms in a given host structure is crucial in designing materials with intercalation chemistry for energy storage. In this study, we present an application of the SE(3)-equivariant diffusion model for such conditional crystal structure predictions using inpainting methods. The model, built upon the `e3nn` framework, was pre-trained on the Materials Project structure database via denoising score matching. By solving the reverse stochastic differential equation using the predictor-corrector method, the model is capable of *de novo* crystal generation as well as conditional generation – finding atomic sites of additive atoms within a host structure. We benchmarked the model performance on the WBM dataset and showcased examples of ion intercalation in different MnO_2 polymorphs. This efficient, probabilistic site-finding tool offers the potential for accelerating the materials discovery.

1. Introduction

Crystal structure generation is a fundamental problem in materials design. Recent advances have shown that generative models can significantly enhance the ability for *de novo* crystal generations, such as variational autoencoders (Ren et al., 2022; Zhu et al., 2023), language-based models (Gruver et al., 2024), and diffusion models (Xie et al., 2021; Jiao et al., 2024; Yang et al., 2023; Zeni et al., 2023), etc. In addition to *de novo* crystal generation, the conditional generation task of finding optimal sites of additive atoms within a given host crystal structure is another crucial application that can be potentially realized by generative models. This fundamental materials modeling step is closely related to intercalation chemistry in materials design. The intercalation manifests itself as a process in which ions (intercalants) can be inserted into or removed from a structure without significant changes to the host framework (i.e., topotactic reactions). For example, the charging and discharging of Li-ion batteries involves the intercalation of Li ions into the host structure of different electrodes. (e.g., the Li ion intercalates between Li_xCoO_2 as the cathode (Mizushima et al., 1980) and $Li_{1-x}C_6$ as the anode (Stevens & Dahn, 2001)).

Another important factor of ion insertions is that the additive ion can drastically change the property of host materials, therefore providing the chance to design functional materials with desired properties. For example, the superionic conductor $Li_{0.388}Ta_{0.238}La_{0.475}Cl_3$ is discovered as a lithiated $LaCl_3$-type structure with aliovalent substitution (Yin et al., 2023). Current methods for site finding often rely on topological analysis of structure (He et al., 2019) or require additional inputs such as charge density obtained from density functional theory (DFT) calculations (Shen et al., 2020). These approaches can be computationally expensive or may not be universally applicable to any given inorganic crystals.

We propose using SE(3)-equivariant diffusion model as a probabilistic-based method to search for optimal sites of the intercalants. The diffusion model is built upon the `e3nn` framework (Geiger & Smidt, 2022) and was pre-trained on the MP database (Jain et al., 2013). The crystal generation is accomplished by solving the reverse stochastic differential equation (SDE) using the predictor-corrector method (Song et al., 2022). We delivered preliminary benchmark tests based on the stable ionic crystals in the WBM dataset (Wang et al., 2021) and also demonstrated examples of Zn insertion to various MnO_2 polymorphs. By leveraging conditional generation with equivariant diffusion models, we propose this as an efficient and generalizable approach to intercalation-based materials generation, which can facilitate the design of advanced materials for energy storage applications.

2. Methods

2.1. Preliminaries of crystal structures

A 3D crystal structure can be represented as an infinite periodic arrangement using a structural unit known as a unit cell, which is defined by $\mathcal{M} = (\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{L})$. The

[*]Equal contribution [1]University of California Berkeley [2]Lawrence Berkeley National Laboratory [3]Courant Institute, New York University. Correspondence to: Peichen Zhong <zhongpc@berkeley.edu>.

Published at ICML 2024 AI for Science workshop.

atomic information is encapsulated in atomic numbers $\boldsymbol{A} = (a_1, a_2, \cdots, a_N) \in \mathbb{R}^N$, atomic positions in the unit cell $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N) \in \mathbb{R}^{3\times N}$, and the lattice matrix $\boldsymbol{L} = (\boldsymbol{l}_1, \boldsymbol{l}_2, \boldsymbol{l}_3) \in \mathbb{R}^{3\times 3}$ where the rows $\boldsymbol{l}_1, \boldsymbol{l}_2, \boldsymbol{l}_3$ are the translational vectors. The representation delineates the infinite periodic crystal structure as

$$\left\{(a_i', \boldsymbol{x}_i') | a_i' = a_i, \boldsymbol{x}_i' = \boldsymbol{x}_i + \boldsymbol{k}\boldsymbol{L}, \boldsymbol{k} \in \mathbb{Z}^{1\times 3}\right\}, \quad (1)$$

where arbitrary integer $\boldsymbol{k}$ denotes the 3D translation in units of $\boldsymbol{l}_i$. This representation serves as a complete basis for atomic arrangements in $\mathbb{R}^3$ with periodic boundary conditions.

2.2. Data preparation

Training: The training dataset in our work consists of 109,805 crystal structures from the Materials Project (MP) database with $E_{\text{hull}} < 0.1$ eV. The E_{hull} represents the energy above the convex hull in phase diagrams, which is a critical quantity to examine the thermodynamic stability of a material ($E_{\text{hull}} = 0$ for stable materials, and $E_{\text{hull}} > 0$ for meta-stable materials) (Bartel et al., 2020). The training set is divided into training and validation sets at an 8:2 ratio.

Test: The test dataset consists of 1,131 stable ionic crystal structures ($E_{\text{hull}} = 0$ eV, containing Li/Mg/Na/Ca ions) from the WBM dataset (Wang et al., 2021), which has no overlapping crystal structures with the MP database.

2.3. SE(3)-equivariant graph neural network

We adopt a customized SE(3)-equivariant graph neural network (GNN) from the `graphite` library (Hsu et al., 2022) based on `NequIP` and `e3nn` frameworks (Batzner et al., 2022; Geiger & Smidt, 2022). The GNN model directly outputs a vector to represent the noising displacement. The initial embedding generates two attributes $\boldsymbol{h}_{i,x}$ and $\boldsymbol{h}_{i,z}$ transforming the type of the i-th atom through embedding layers. The interaction blocks update node attributes $\boldsymbol{h}_{i,x}$ by self-interaction and aggregating attributes of neighboring atoms with the weighted tensor product (wTP), whereas $\boldsymbol{h}_{i,z}$ does not change

$$\boldsymbol{h}_{i,x}^{(l+1)} = \text{wTP}(\boldsymbol{h}_{i,x}^{(l)}, \boldsymbol{h}_{i,z}^{(l)}), \quad (2)$$

$$\boldsymbol{h}_{i,x}^{(l+1)} = \frac{1}{Z}\sum_{j\in N(i)} \text{wTP}_{\|\boldsymbol{e}_{ij}\|}(\boldsymbol{h}_{j,x}^{(l)}, Y(\hat{\boldsymbol{e}}_{ij})). \quad (3)$$

Equations (2) and (3) represent self-interaction and convolution, respectively. $Y(\hat{\boldsymbol{e}}_{ij})$ is the spherical harmonic of a normalized vector $\hat{\boldsymbol{e}}_{ij}$ pointing from node i to j. $N(i)$ denotes neighbors of node i. The final self-interaction layer outputs a single vector representing the estimated score

$$\boldsymbol{h}_{i,x}^{(L)} = \text{wTP}(\boldsymbol{h}_{i,x}^{(L-1)}, \boldsymbol{h}_{i,z}^{(L-1)}) = (s_\theta)_i \quad (4)$$

2.4. Variance exploding diffusion model

Generating samples with probability density function $p(\boldsymbol{x})$ in high-dimensional space $\mathbb{R}^d$ can be achieved by modeling the gradient of the log-probability density $\nabla_{\boldsymbol{x}} \log p(\boldsymbol{x})$ (score function) using diffusion models. Both the diffusion process and its reverse can be formulated as stochastic differential equations (SDE) (Song et al., 2021).

$$\mathrm{d}\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, t)\mathrm{d}t + g(t)\mathrm{d}\boldsymbol{w}, \quad (5)$$

$$\mathrm{d}\boldsymbol{x} = \left[\boldsymbol{f}(\boldsymbol{x}, t) - g^2(t)\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})\right]\mathrm{d}t + g(t)\mathrm{d}\overline{\boldsymbol{w}}, \quad (6)$$

where $\boldsymbol{w}$ and $\overline{\boldsymbol{w}}$ represent the Brownian motions, $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{t})$ is drift coefficient and $g(t)$ is diffusion coefficient of $\boldsymbol{x}(t)$.

We chose the variance-exploding (VE) diffusion as the prior distribution $p(x_T)$ follows a simple uniform distribution. In addition, the coordinates of atoms in a crystal satisfy periodic boundary conditions, and the VE diffusion does not lead to extremely large displacements in the noisy limit. The VE diffusion can be formulated as follows (Song et al., 2021):

$$\mathrm{d}\boldsymbol{x} = \sqrt{\frac{\mathrm{d}[\sigma^2(t)]}{\mathrm{d}t}}\mathrm{d}\boldsymbol{w}, \quad (7)$$

where a sequence of exponentially increasing standard deviations is given $\sigma_{\min} = \sigma_1, \cdots, \sigma_T = \sigma_{\max}$. The samples can be drawn using the ancestral sampling

$$\boldsymbol{x}_{t-1} = \boldsymbol{x}_t + (\sigma_t^2 - \sigma_{t-1}^2)\boldsymbol{s}_{\boldsymbol{\theta}^*}(\boldsymbol{x}_t, t) + \boldsymbol{z}\sqrt{\frac{\sigma_{t-1}^2(\sigma_t^2 - \sigma_{t-1}^2)}{\sigma_t^2}} \quad (8)$$

where $x_T \sim \mathcal{N}(\boldsymbol{0}, \sigma_T^2\boldsymbol{I})$, and $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. In the continues 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 proposed by (Song et al., 2021) (see Algorithm 1).

2.5. Model training

To estimate $\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})$, we use score matching (SM) to learn the model θ^* that minimizes the loss function

$$\mathcal{L}_{\text{SM}} = \mathbb{E}_{p(\boldsymbol{x})}\left[\|\boldsymbol{s}_\theta(\boldsymbol{x}(t), t) - \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})\|^2\right]. \quad (9)$$

$p_t(\boldsymbol{x})$ can be approximated by a Gaussian transition probability $p(\boldsymbol{x_t}|\boldsymbol{x_0}) \propto e^{-(\boldsymbol{x}_t - \boldsymbol{x}_0)^2/2\sigma^2}$. Equation (9) can be formulated as denoising score matching (DSM) following Vincent (2011); Ho et al. (2020)

$$\mathcal{L}_{\text{DSM}} = \mathbb{E}_{p(\boldsymbol{x}_t|\boldsymbol{x_0})}\left[\left\|\boldsymbol{s}_\theta(\boldsymbol{x}(t), t) - \frac{\varepsilon}{\sigma}\right\|^2\right], \quad (10)$$

where $s_\theta(\boldsymbol{x}(t), t)$ is the score function given by the GNN model, $\boldsymbol{\varepsilon} = (\boldsymbol{x}_t - \boldsymbol{x}_0)/\sigma \sim \mathcal{N}(0, 1)$, and $\sigma \sim \mathcal{U}(\sigma_{\min}, \sigma_{\max})$. The training framework was built upon the architecture of `CDVAE` using the `pytorch-lighting` library (Xie et al., 2021).

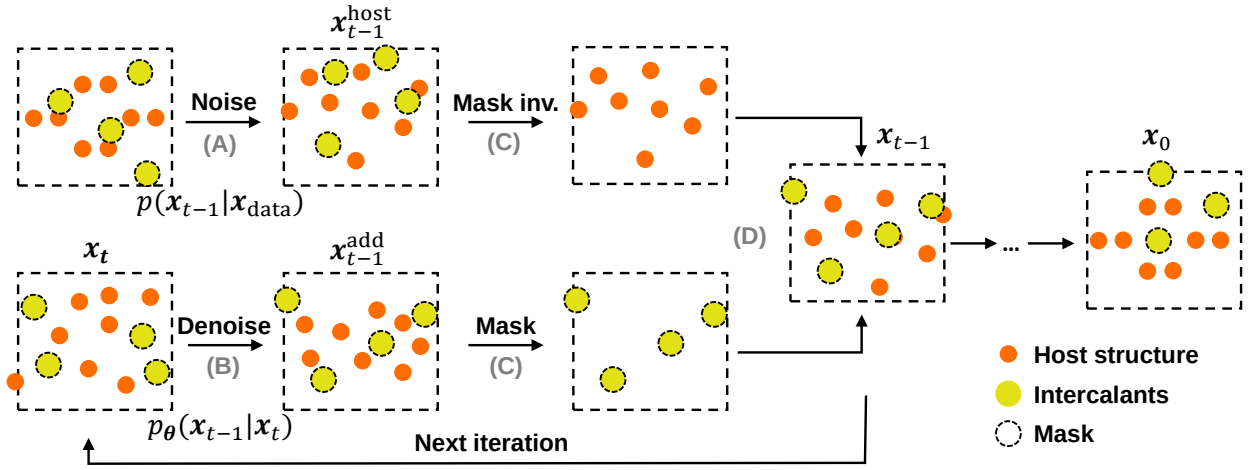

Figure 1. Iterative sampling strategy for crystal structure inpainting given a host structure. (A) Add noise to the host structure. (B) Compute the scores $\{\boldsymbol{s}_\theta(\boldsymbol{x}_t)\}$ of all atoms and perform denoising to update the atomic positions. (C) Apply masks on the host structure and the intercalants. For each iteration, the host structure perturbation starts from Process (A). Note that the symbols $\boldsymbol{x}^{\text{host}}$ and $\boldsymbol{x}^{\text{add}}$ represent the atomic positions of all atoms (including host and intercalants), where the superscripts distinguish the different masking operations on host and additive atoms, respectively.

2.6. Crystal structure inpainting

Inpainting. Crystal structure inpainting is conceptualized as a *conditional generation* process, where the host structure is given and the intercalants are considered as missing parts. To approximate the conditional diffusion, we employed an unconditional diffusion model trained with the DSM strategy. The conditional probabilistic transitions were incorporated into the sampling process by adding noise to the host structure.

An illustration of the iterative sampling procedure is outlined in Figure 1. In each reverse diffusion step, the atoms in the host structure are first perturbed with a Gaussian noise given by the noise scheduler of the next step σ_{t-1} (Process A)

$$\boldsymbol{x}_{t-1}^{\text{host}} = \boldsymbol{x}_0^{\text{host}} + \sigma_{t-1}\boldsymbol{z}, \quad \boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}). \tag{11}$$

The additive atoms (intercalants) and host structure in the current configuration $\boldsymbol{x}_t$ therefore have a similar scale of noise. The denoising score can be computed using all atoms with the GNN model via Equation (4), and the diffusion is performed using Equation (8) (Process B). For the notation, x^{host} and x^{add} represent different operations on the noise/-denoise process and include all atomic positions of both the host structure and intercalants. We applied masks $(1-m)$ on x^{host} and m on x^{add} (Process C) to form $\boldsymbol{x}_{t-1}$ (Process D).

$$\boldsymbol{x}_{t-1} = (1-m) \odot \boldsymbol{x}_{t-1}^{\text{host}} + m \odot \boldsymbol{x}_{t-1}^{\text{add}} \tag{12}$$

The iterative reverse diffusion process reduces noise incrementally, thus ensuring the final crystal structure closely aligns with the original host structure.

Resampling. It is observed that direct application of the *replacement* method leads to locally harmonized structures but inconsistent global context. To address the issue, *resampling* is proposed, where atomic positions $\boldsymbol{x}_i$ are repeatedly diffused back and forth for r times in each diffusion step as detailed in Algorithm 1 (Lugmayr et al., 2022). The horizon of the backward and forward diffusion operation can be extended from one resampling step to longer resampling steps, denoted as jump length j, where $j = 1$ is the previous case. The technique allows the model to harmonize prediction for the newly generated additive atoms and the noisy atoms from the host structure. The number of resamples r and jump length j control the diffusion steps scheduler (shown in Code 1) thus the harmonizing effect with Algorithm 1.

The inpainting & resampling method was originally proposed in the imaging process (Lugmayr et al., 2022) and was further adopted to scientific fields such as structure-based drug molecule design (Schneuing et al., 2023) and objective-aware transition states generation for chemical reactions (Duan et al., 2023). In this study, we integrated the periodic boundary condition and SE(3)-equivariance with these techniques and applied them to crystalline materials in $\mathbb{R}^3$ space.

2.7. Machine learning interatomic potential

For the generated structures, we used the pre-trained CHGNet as an efficient and accurate calculator to evaluate the energy and interatomic forces without running DFT. The pre-trained CHGNet achieves mean-absolute errors of 30

Task	Reconst-Rate (%)	RMSE (Å)	F_{max}^{Median} (eV/Å)	$F_{max}^{75\%}$	ΔE^{Median} (meV/atom)	$\Delta E^{75\%}$
Li	82.54 ± 0.41	0.034 ± 0.005	0.108 ± 0.001	0.256 ± 0.012	1.0 ± 0.1	7 ± 1
Na	86.70 ± 1.02	0.044 ± 0.003	0.111 ± 0.001	0.213 ± 0.010	1.3 ± 0.2	5.2 ± 0.8
Mg	83.95 ± 2.34	0.039 ± 0.007	0.138 ± 0.003	0.244 ± 0.039	1.1 ± 0.0	4.2 ± 0.1
Ca	83.33 ± 2.12	0.061 ± 0.007	0.171 ± 0.004	0.382 ± 0.039	2.2 ± 0.2	15 ± 3

Table 1. Reconstruction rate and stability analysis of the unpainted structures containing Li/Na/Mg/Can in the WBM dataset. $F_{\max}$ is the maximum interatomic force in the generated structure. ΔE is the energy difference between the generated structure and its relaxed structure. The cutoff values for the median and 75% quantile from the distribution of the WBM-reconstructed compounds are presented. Both $F_{\max}$ and ΔE were calculated using CHGNet as the MLIP. The standard deviation arises from different realizations of diffusion using the same initial conditions for the median and 75% quantile metrics.

Algorithm 1 Inpainting Crystal Generation

Input: atomic positions of host structure and randomly initialized additive atoms $\boldsymbol{x}_0^{\text{host}}$; atomic positions sampled randomly in the unit cell $\boldsymbol{x}_T$; mask for additive atoms m; signal-to-noise ratio δ; number of resampling steps r
for $t = T, \cdots, 1$ **do**
  for $n = 1, \cdots, r$ **do**
    $\boldsymbol{x}_{t-1}^{\text{add}} \leftarrow \boldsymbol{x}_t + (\sigma_t^2 - \sigma_{t-1}^2)\boldsymbol{s}_\theta(\boldsymbol{x}_t, t)$
    $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
    $\boldsymbol{x}_{t-1}^{\text{add}} \leftarrow \boldsymbol{x}_{t-1}^{\text{add}} + \sqrt{\frac{\sigma_{t-1}^2(\sigma_t^2 - \sigma_{t-1}^2)}{\sigma_t^2}}\,\boldsymbol{z}$
    for $j = 1, \cdots, M$ **do**
      $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{1})$
      $\boldsymbol{g} \leftarrow \boldsymbol{s}_\theta(\boldsymbol{x}_{t-1}, t-1)$
      $\varepsilon \leftarrow 2(\sqrt{3}\delta / \|\boldsymbol{g}\|_2)^2$
      $\boldsymbol{x}_{t-1}^{\text{add}} \leftarrow \boldsymbol{x}_{t-1}^{\text{add}} + \varepsilon\,\boldsymbol{g} + \sqrt{2\varepsilon}\,\boldsymbol{z}$
    end for
    $\boldsymbol{x}_{t-1}^{\text{host}} \leftarrow \boldsymbol{x}_0^{\text{host}} + \sigma_{t-1}\boldsymbol{z}$
    $\boldsymbol{x}_{t-1} \leftarrow (1 - m) \odot \boldsymbol{x}_{t-1}^{\text{host}} + m \odot \boldsymbol{x}_{t-1}^{\text{add}}$
    if $n < r$ **and** $t > 1$ **then**
      $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{1})$
      $\boldsymbol{x}_t \leftarrow \boldsymbol{x}_{t-1} + \sqrt{\sigma_{t-1}^2 - \sigma_{t-2}^2}\,\boldsymbol{z}$
    end if
  end for
end for

meV/atom for energy and 70 meV/Å for interatomic forces on the MPtrj dataset against the DFT (Deng et al., 2023). The structure relaxations were optimized by the `FIRE` optimizer (Bitzek et al., 2006) over the potential energy surface provided by CHGNet, where the atomic positions, cell shape, and cell volume were simultaneously optimized to reach converged interatomic forces of 0.1 eV/Å.

2.8. DFT calculations

We performed density functional theory (DFT) calculations with the `VASP` package using the projector-augmented wave method (Kresse & Furthmüller, 1996; Kresse & Joubert, 1999), a plane-wave basis set with an energy cutoff of 520 eV, and a reciprocal space discretization of 25 k-points per Å^{-1}. The calculations were converged to 10^{-6} eV in total energy for electronic loops and 0.02 eV/Å in interatomic forces for ionic loops. The computational setting was consistent with the `MPRelaxSet` (Jain et al., 2011).

3. Experiments

3.1. Crystal Reconstruction

Setup. We collected the stable materials ($E_{\text{hull}} = 0$ eV) containing Li, Na, Mg, and Ca as intercalants from the WBM dataset (Wang et al., 2021). The intercalants in the host structure were first removed and then randomly initialized back into the host structure with fractional coordinates $\{(f_x, f_y, f_z)\}$ drawn from a normal distribution $(f_x, f_y, f_z) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ to generate the corrupted structures. The inpainting tasks were performed using the diffusion model following Algorithm 1 to optimize the atomic positions of the intercalants. We evaluated the reconstruction rate by measuring the difference between the inpainted structure and the original structure using the `StructureMatcher` from `pymatgen` (Ong et al., 2013) with criteria `stol=0.3, ltol=0.2`. For successfully matched structures, we compared the root-mean-squared error (RMSE) in $\mathbb{R}^3$ space. The RMSE is normalized by the number of intercalants since the host structure remains fixed during reconstruction. We ran four different diffusions to obtain the average and standard deviation of the performance metrics. We also benchmarked the influence of the signal-to-noise ratio parameter on the crystal reconstruction task, with details provided in Appendix Figure 3.

Results. Table 1 presents the reconstruction rate and RMSE for the four types of intercalants, from which we observed over 80% of the Li-, Na-, and Mg-containing structures in the WBM dataset are recovered from their corresponding corrupted structures using the inpainting generative model. The recovered structures have an average RMSE below 0.05 Å with respect to the original structures. The energy and interatomic forces of the inpainted structures were sub-

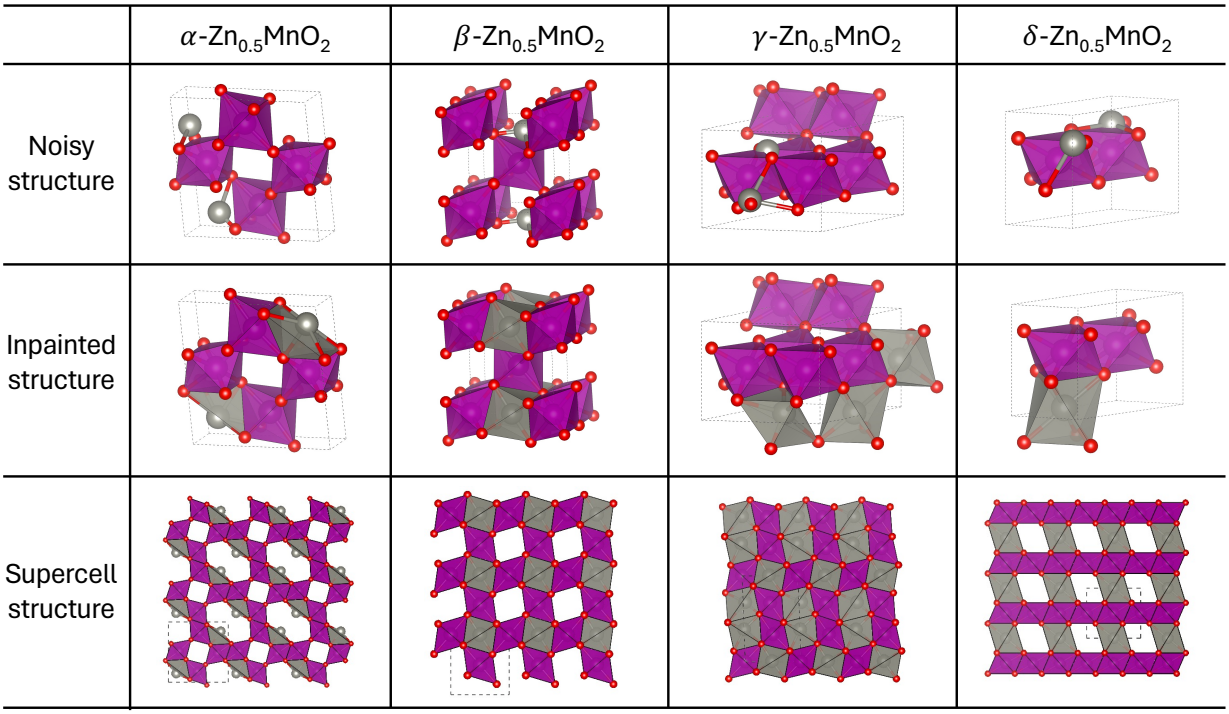

Figure 2. MnO_2 intercalated with Zn in α-, β-, γ-, and δ-polymorphs (red: oxygen, purple: Mn, gray: Zn). The noisy structure is initiated to perform reverse diffusion to obtain the inpainted structure. The dashed lines represent the lattice of unit cells. The supercell structures of inpainted structures illustrate the translation symmetry of crystals. The purple and gray polyhedra indicate the local atomic coordination of Mn and Zn. See (Zhang et al., 2024) for sites and diffusion channels identified in experiments for comparison.

sequently evaluated using CHGNet, where these residual forces were used as a metric to inspect whether the reconstructed structures were reasonably relaxed. We calculated the distribution of these metrics, Table 1 presents the median and 75% quantile of such distribution of maximum interatomic force ($F_{\max}$), with relatively small values (< 0.2 eV/Å for the median) observed for all four types of intercalants. A histogram plot of the $F_{\max}$ is presented in Appendix Figure 4.

We further relaxed these structures using CHGNet and recorded the energy difference that is dissipated in the ionic relaxation process ($\Delta E = E_{\text{unrelax}} - E_{\text{relax}}$). The median ΔE is approximately 1 meV/atom for Li, Na, and Mg, and the 75% quantile shows a higher value of less than 10 meV/atom. The Ca-containing structures demonstrate slightly higher values in $F_{\max}$ and ΔE (2.2 meV/atom for the median and 15 meV/atom for the 75% quantile). The standard deviation of energy and forces were calculated from the different trials of diffusion using the same initial conditions. The small residual energy and interatomic forces indicate the inpainted structures generated by the diffusion model are close to their local energy minima, evidencing the reasonable intercalation sites predicted by Algorithm 1.

To assess the quality of the generated structures compared to ground truth with *ab initio* accuracy, we performed DFT calculations to relax the Li-containing crystal structures produced by our inpainting method. The analysis of interatomic forces is presented in Appendix Figure 5. These converged DFT calculations confirm that the generated structures closely approximate local energy minima. Notably, the maximum force ($F_{\max}$) observed in our generated structures is lower than that predicted by the pretrained CHGNet. This outcome is consistent with expectations, given that CHGNet has a known force prediction error of 70 meV/Å, while our generated structures exhibit only a small RMSE deviation (< 0.05 Å) from the DFT ground truth in the WBM dataset (Table. 1).

3.2. Example: inpainting MnO_2 frameworks

Setup. MnO_2 exhibits various polymorphs (α, β, γ, and δ) and are promising candidates for intercalation-based battery cathode design. These polymorphs serve as host structures for various intercalants (e.g., Li/Na/Mg, etc) (Kitchaev et al., 2017). To demonstrate an example study using the site searching function with the diffusion model, we ran inpainting generations using Zn as intercalant for $Zn_{0.5}MnO_2$. The host structures of the four polymorphs were obtained from the MP database (α: `mp-1080238`, β: `mp-29159`, γ: `mp-626068`, and δ: `mp-1002573`). The signal-to-

noise ratio was set to 0.2, as it achieves the lowest RMSE according to Appendix Figure 3 on the test of the WBM dataset.

Results. Figure 2 presents the inpainted structures of $Zn_{0.5}MnO_2$ for the four polymorphs. The inserted Zn atoms are located in the interstitial vacancy sites of the different MnO_2 polymorphs, consistent with the diffusion channels demonstrated in an experimentally reported study (Zhang et al., 2024). Other intercalants, including Li, H, and Mg, are shown in Appendix Figure 6 for illustration.

4. Discussion

Crystal inpainting generation is of great importance for materials discovery, as the optimal intercalation site determines the stability and property of a material. In this work, we propose the application of an SE(3)-equivariant diffusion model for crystal inpainting generation. The model was pre-trained on the Materials Project structure database using an unconditional generation strategy, i.e., the *de novo* diffusion generation for crystals. By employing resampling techniques, the model can conditionally generate and optimize unknown atomic positions within a given host structure while preserving local stability.

Identifying insertion positions has traditionally been addressed by various physics-based models. Structural topology or charge density analysis (He et al., 2019; Shen et al., 2020), for instance, can search sites in continuous vacant spaces but often at high computational cost. The cluster expansion method, while mathematically rigorous for searching the lattice ground-state configuration (Huang et al., 2016), requires a well-defined crystalline structure (e.g., with Wyckoff positions) including vacancies (Barroso-Luque et al., 2022). In contrast, our probability-based inpainting approach for crystal generation is well-pretrained and can be efficiently implemented without specifying additional crystalline information beyond the host structure. This approach combines computational efficiency with flexibility, making it a useful alternative to traditional physics-based methods for atom insertion tasks.

One limitation of the current model is its inability to account for volume changes during diffusion, which is especially significant for elements with large ionic radii (e.g., Na/Ca). Although lattice relaxation using MLIP or DFT can partially address this issue post-generation, it would be promising to develop future models that can accommodate diffusion for both host structures with lattice changes and the atoms of interest. We anticipate the adoption of inpainting crystal generation along with further improvements can accelerate the identification of promising candidates for energy storage in computational material design.

References

Barroso-Luque, L., Zhong, P., Yang, J. H., Xie, F., Chen, T., Ouyang, B., and Ceder, G. Cluster expansions of multicomponent ionic materials: Formalism and methodology. *Physical Review B*, 106(14):144202, oct 2022.

Bartel, C. J., Trewartha, A., Wang, Q., Dunn, A., Jain, A., and Ceder, G. A critical examination of compound stability predictions from machine-learned formation energies. *npj Computational Materials*, 6:97, 7 2020.

Batzner, S., Musaelian, A., Sun, L., Geiger, M., Mailoa, J. P., Kornbluth, M., Molinari, N., Smidt, T. E., and Kozinsky, B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, May 2022.

Bitzek, E., Koskinen, P., Gähler, F., Moseler, M., and Gumbsch, P. Structural relaxation made simple. *Physical Review Letters*, 97(17):1–4, 2006.

Deng, B., Zhong, P., Jun, K., Riebesell, J., Han, K., Bartel, C. J., and Ceder, G. Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling, 9 2023.

Duan, C., Du, Y., Jia, H., and Kulik, H. J. Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model. *Nature Computational Science*, 3(12):1045–1055, December 2023.

Geiger, M. and Smidt, T. e3nn: Euclidean neural networks. *arXiv preprint arXiv:2207.09453*, 2022.

Gruver, N., Sriram, A., Madotto, A., Wilson, A. G., Zitnick, C. L., and Ulissi, Z. Fine-tuned language models generate stable inorganic materials as text. *arXiv preprint arXiv:2402.04379*, 2024.

He, X., Bai, Q., Liu, Y., Nolan, A. M., Ling, C., and Mo, Y. Crystal structural framework of lithium super-ionic conductors. *Advanced Energy Materials*, 9(43):1902078, 2019.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Hsu, T., Sadigh, B., Bertin, N., Park, C. W., Chapman, J., Bulatov, V., and Zhou, F. Score-based denoising for atomic structure identification. *arXiv preprint arXiv:2212.02421*, 2022.

Huang, W., Kitchaev, D. A., Dacek, S. T., Rong, Z., Urban, A., Cao, S., Luo, C., and Ceder, G. Finding and proving the exact ground state of a generalized Ising model by convex optimization and MAX-SAT. *Physical Review B*, 94(13):134424, oct 2016.

Jain, A., Hautier, G., Ong, S. P., Moore, C. J., Fischer, C. C., Persson, K. A., and Ceder, G. Formation enthalpies by mixing GGA and GGA+U calculations. *Physical Review B*, 84(4):045115, jul 2011.

Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., and Persson, K. A. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1:011002, 7 2013.

Jiao, R., Huang, W., Lin, P., Han, J., Chen, P., Lu, Y., and Liu, Y. Crystal structure prediction by joint equivariant diffusion. *Advances in Neural Information Processing Systems*, 36, 2024.

Kitchaev, D. A., Dacek, S. T., Sun, W., and Ceder, G. Thermodynamics of Phase Selection in MnO $_2$ Framework Structures through Alkali Intercalation and Hydration. *Journal of the American Chemical Society*, 139(7):2672–2681, February 2017.

Kresse, G. and Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science*, 6(1):15–50, jul 1996.

Kresse, G. and Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*, 59(3):1758–1775, jan 1999.

Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., and Van Gool, L. RePaint: Inpainting using Denoising Diffusion Probabilistic Models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11451–11461, New Orleans, LA, USA, June 2022. IEEE. ISBN 978-1-66546-946-3.

Mizushima, K., Jones, P., Wiseman, P., and Goodenough, J. Li_xCoO_2 ($0 < x < 1$): A new cathode material for batteries of high energy density. *Materials Research Bulletin*, 15(6):783–789, jun 1980.

Ong, S. P., Richards, W. D., Jain, A., Hautier, G., Kocher, M., Cholia, S., Gunter, D., Chevrier, V. L., Persson, K. A., and Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, February 2013.

Ren, Z., Tian, S. I. P., Noh, J., Oviedo, F., Xing, G., Li, J., Liang, Q., Zhu, R., Aberle, A. G., Sun, S., Wang, X., Liu, Y., Li, Q., Jayavelu, S., Hippalgaonkar, K., Jung, Y., and Buonassisi, T. An invertible crystallographic representation for general inverse design of inorganic crystals with targeted properties. *Matter*, 5:314–335, 1 2022.

Schneuing, A., Du, Y., Harris, C., Jamasb, A., Igashov, I., Du, W., Blundell, T., Lió, P., Gomes, C., Welling, M., Bronstein, M., and Correia, B. Structure-based drug design with equivariant diffusion models, June 2023. arXiv:2210.13695 [cs, q-bio].

Shen, J.-X., Horton, M., and Persson, K. A. A charge-density-based general cation insertion algorithm for generating new li-ion cathode materials. *npj Computational Materials*, 6:161, 10 2020.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-Based Generative Modeling through Stochastic Differential Equations, February 2021. arXiv:2011.13456 [cs, stat].

Song, Y., Shen, L., Xing, L., and Ermon, S. Solving Inverse Problems in Medical Imaging with Score-Based Generative Models, June 2022. arXiv:2111.08005 [cs, eess, stat].

Stevens, D. and Dahn, J. The mechanisms of lithium and sodium insertion in carbon materials. *Journal of The Electrochemical Society*, 148(8):A803, 2001.

Vincent, P. A Connection Between Score Matching and Denoising Autoencoders. *Neural Computation*, 23(7): 1661–1674, July 2011.

Wang, H.-C., Botti, S., and Marques, M. A. L. Predicting stable crystalline compounds using chemical similarity. *npj Computational Materials*, 7(1):12, 2021.

Xie, T., Fu, X., Ganea, O.-E., Barzilay, R., and Jaakkola, T. Crystal diffusion variational autoencoder for periodic material generation. *arXiv preprint arXiv:2110.06197*, 2021.

Yang, M., Cho, K., Merchant, A., Abbeel, P., Schuurmans, D., Mordatch, I., and Cubuk, E. D. Scalable diffusion for materials generation. *arXiv preprint arXiv:2311.09235*, 2023.

Yin, Y.-C. et al. A LaCl3-based lithium superionic conductor compatible with lithium metal. *Nature*, 616:77–83, 4 2023.

Zeni, C., Pinsler, R., Zügner, D., Fowler, A., Horton, M., Fu, X., Shysheya, S., Crabbé, J., Sun, L., Smith, J., et al. Mattergen: a generative model for inorganic materials design. *arXiv preprint arXiv:2312.03687*, 2023.

Zhang, Q., Zhao, J., Chen, X., Yang, R., Ying, T., Cheng, C., Liu, B., Fan, J., Li, S., and Zeng, Z. Unveiling the Energy Storage Mechanism of MnO $_2$ Polymorphs for Zinc-Manganese Dioxide Batteries. *Advanced Functional Materials*, pp. 2306652, March 2024.

Zhu, R., Nong, W., Yamazaki, S., and Hippalgaonkar, K. Wycryst: Wyckoff inorganic crystal generator framework. *arXiv preprint arXiv:2311.17916*, 2023.

A. Code implementation for inpainting crystal generation

```
t_T = 250
jump_len = 10
jump_n_sample = 10

jumps = {}
for j in range(0, t_T - jump_len, jump_len):
    jumps[j] = jump_n_sample - 1

t = t_T
ts = []

while t >= 1:
    t = t-1
    ts.append(t)
    if jumps.get(t, 0) > 0:
        jumps[t] = jumps[t] - 1
        for _ in range(jump_len):
        t=t+1
        ts.append(t)
ts.append(-1)
```

Listing 1. Scheduler algorithm

B. Architecture of the model

The architecture of the SE(3)-equivariant model is summarized in Table 2.

Name	Value
Number of convolution layers	4
Radius cutoff of crystal graph	6.0 Å
Irreps for initial node attributes $\boldsymbol{h}_{i,x}^{(0)}$	32x0e
Irreps for auxiliary node attributes $\boldsymbol{h}_{i,z}^{(0)}$	32x0e
Irreps for hidden node attributes $\boldsymbol{h}_{i,x}^{(l)}$	32x0e + 32x1e
Irreps for edge attributes	32x0e + 32x1e + 8x2e
Number of basis functions for expanding edge distance $\|\hat{\boldsymbol{e}}_{ij}\|$	16
Number of hidden neurons for the MLP	64
Number of trainable parameters	2,655,808

Table 2. Architecture of the SE(3)-equivariant diffusion model.

C. Hyperparameters for training and sampling

The pre-trained model was trained for 0.98 million steps with a batch size of 4 on one NVIDIA A100 GPU using the Adam optimizer. The learning rate is initialized and kept at 0.001. Gradient clipping was applied by a value of 0.1.

We discretized the reverse diffusion process (backward passing) into $T = 200$ steps. The initial unknown coordinates are sampled from prior Gaussian distribution and mixed with the corrupted host structure by several forward and backward passing with the resamples parameter r as 3 and the jump length parameter j as 10. For each backward step, we use ancestor sampling to sample x_{t-1} given x_t using the pre-trained model. After each predictor step, five corrector steps were applied via the Langevin corrector with signal-to-noise ratio parameter as 0.2 if not otherwise specified in the paper.

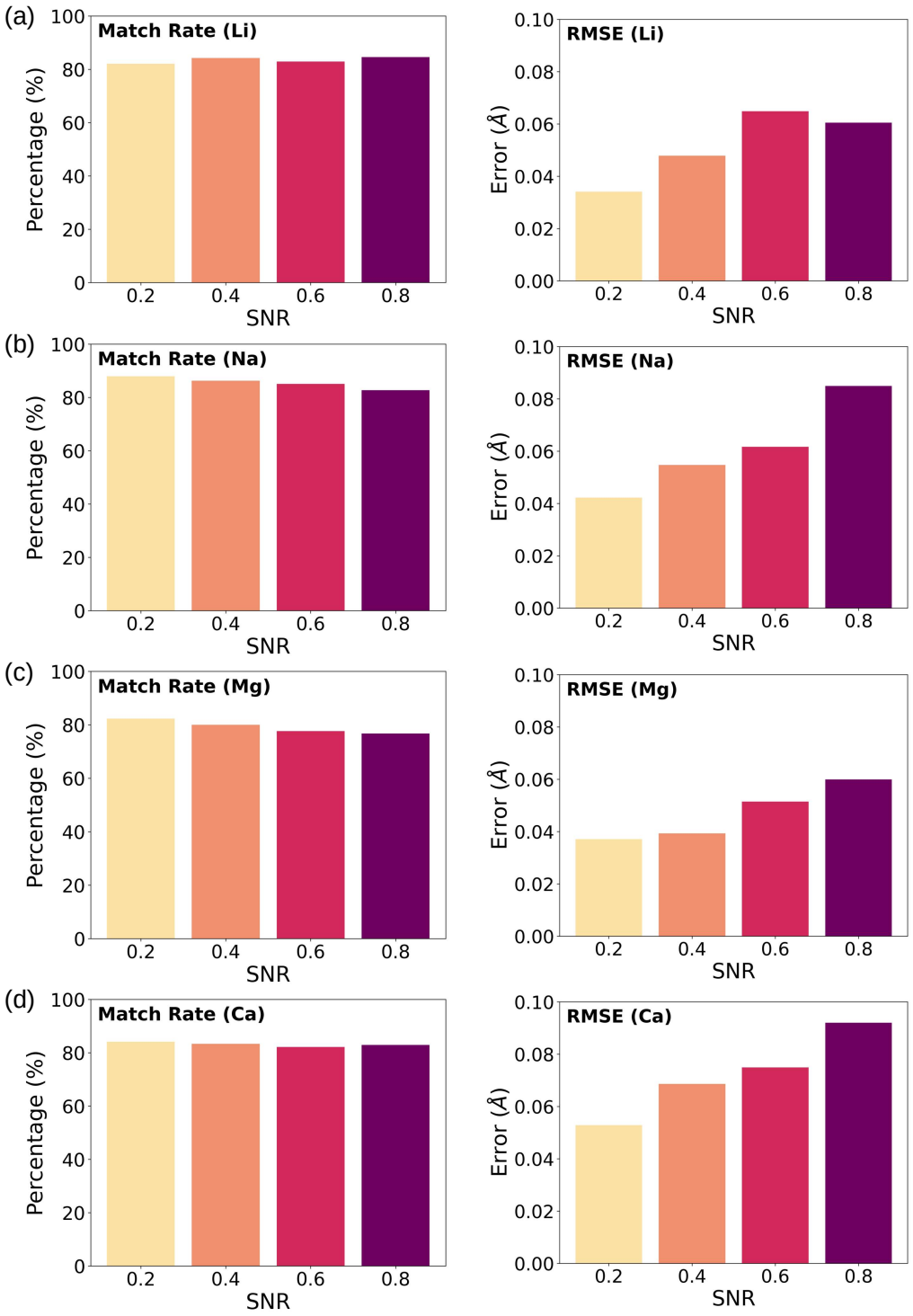

Figure 3. Reconstruction and error analysis of the inpainting generation based on the selected structures from the WBM dataset. (a-d) Reconstruction match rate and RMSE of Li-, Na-, Mg-, and Ca-containing materials. The signal-to-noise ratio (SNR) is set to 0.2, 0.4, 0.6, and 0.8, from left to right in each panel.

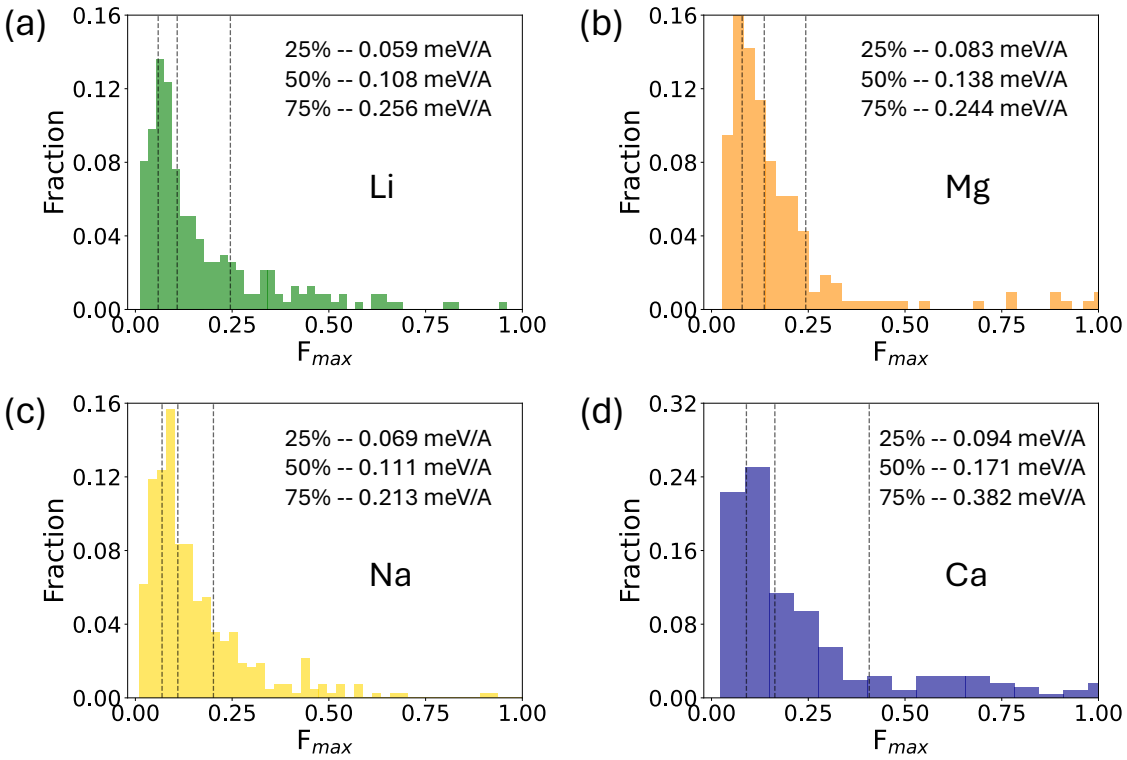

Figure 4. The maximum interatomic forces ($F_{\max}$) in the structures generated using the inpainting algorithm for different intercalants: Li (green), Mg (orange), Na (yellow), and Ca (dark blue). The three dashed lines in each panel represent the 25% quantile, median, and 75% quantile, from left to right.

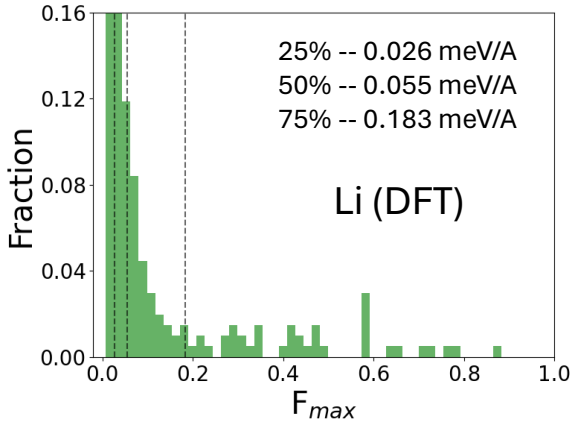

Figure 5. The maximum interatomic forces ($F_{\max}$) calculated by DFT for the inpainting generated Li-containing structures. The three dashed lines represent the 25% quantile, median, and 75% quantile, from left to right.

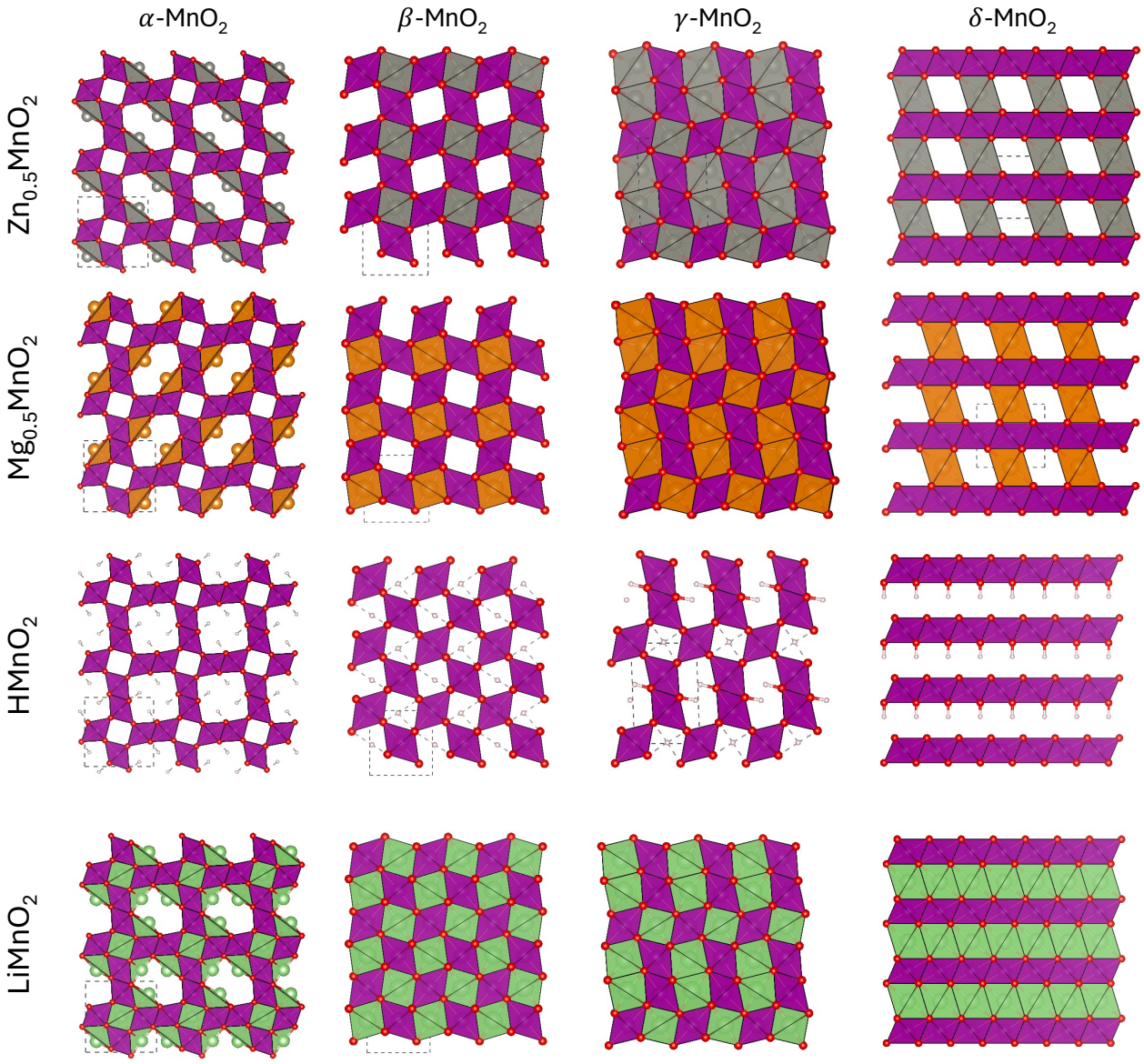

Figure 6. Inpainted structures using α-, β-, γ-, and δ-MnO_2 polymorphs with Zn (gray, $Zn_{0.5}MnO_2$), Mg (orange, $Mg_{0.5}MnO_2$), H (white, $HMnO_2$), and Li (green, $LiMnO_2$) as intercalants. The dashed lines represent the lattices of a unit cell and a supercell view is shown to illustrate the translation symmetry of crystals. The polyhedra indicate the local atomic coordination of Mn and intercalant.