SYMMETRY-BREAKING DE NOVO CRYSTAL GENERATION VIA MARKOVIAN JUMP DIFFUSION

Anonymous authors
Paper under double-blind review

ABSTRACT

Generating crystals has recently attracted significant interest due to their broad applications in materials science. However, existing generative models struggle to produce complete crystallographic specifications, limiting their ability to capture full crystal symmetries and structural dependencies. In particular, current state-of-the-art approaches generate crystals only up to site symmetries and rely on sampling space groups from empirical distributions during inference. Inspired by *spontaneous symmetry breaking* in physics, where crystals break symmetries of the underlying system, we propose a novel diffusion-based framework that generates full crystallographic specifications by reversing from the lowest-symmetry priors. Our method leverages a Markovian jump-diffusion process to model these symmetry-breaking dynamics, enabling it to traverse different space groups in a physically motivated manner. Our model, dubbed *Symmetry-breaking Crystal Diffusion* (SbCD), introduces a principled approach to explicitly incorporate inter-space-group transitions into the generative process. In de novo crystal generation experiments, SbCD substantially outperforms its symmetry-preserving counterpart and achieves state-of-the-art thermodynamic stability on MPTS-52, offering a promising perspective for generative modeling of crystalline materials.

1 INTRODUCTION

Table 1: Comparison of crystallographic structure specifications modeled by existing crystal generative frameworks. ✓denotes the modeling of a crystallographic structure specification, while ✗denotes its absence. Structure elements enclosed in braces $\{\}$ denote equivalent crystal representations.

Method	STRUCTURE SPECIFICATION $\{\mathbf{k}, \mathbf{L}\}$	$\{\mathbf{F}, \mathbf{X}\}$	$\mathbf{A}$	$\mathbf{S}$	ASYMMETRIC UNIT	SYMMETRY BREAKING G
CDVAE (Xie et al., 2021)	✓	✓	✓	✗	✗	✗
DiffCSP (Jiao et al., 2023), FlowMM (Miller et al., 2024)	✓	✓	✓	✗	✗	✗
DiffCSP++ (Jiao et al., 2024), SGFM (Puny et al., 2025)	✓	✓	✓	✗	✗	✗
SymmCD (Levy et al., 2025), SGEquiDiff (Chang et al., 2026)	✓	✓	✓	✓	✓	✗
SbCD (ours)	✓	✓	✓	✓	✓	✓

Crystals play a fundamental role in materials science, underpinning advances across diverse fields including electrochemical energy storage (Collins et al., 2016), semiconductor hardware design (Bakaul et al., 2016), and pharmaceuticals (Feig et al., 2025). Traditionally, discovering a useful crystal with targeted properties has required years or even decades of trial-and-error experimentation (Cheetham et al., 2022). Accelerating this discovery process (Szymanski et al., 2023; Merchant et al., 2023) thus holds immense potential for unlocking superior technologies. Generative models (Kingma & Welling, 2013; Ho et al., 2020; Song et al., 2020; Austin et al., 2021; Lipman et al., 2022) have emerged as a highly effective paradigm for the in silico screening of structured data with desired properties. Pioneering work by Xie et al. (2021) introduced latent diffusion models for raw crystal-structure representations. Subsequent approaches (Jiao et al., 2023; Miller et al., 2024; Okhotin et al., 2025) operate on more compact unit cells and model fractional coordinates on Riemannian manifolds, better respecting the underlying geometry. More recently, Jiao et al. (2024) further constrained the generative process to specific space groups, enabling template-conditioned generation of crystals with desirable properties. Notably, Levy et al. (2025) and Chang et al. (2026) proposed learning crystal site symmetries, yielding improved generalization to non-trivial symmetry point groups. Table 1 compares these approaches.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Despite this progress, many methods rely heavily on space groups and Wyckoff positions derived from empirical distributions during generation, limiting their ability to capture full crystal symmetries and structural dependencies. Specifically, Puny et al. (2025) showed that models not explicitly learning Wyckoff positions generalize poorly to unseen configurations. In addition, Chang et al. (2025) invoked Neumann's principle (Neumann & Meyer, 1885), which states that physical properties must be invariant under a crystal's point group symmetries, underscoring the importance of accurately modeling symmetry through Wyckoff positions and their associated space groups. In this work, we bridge this gap by learning to directly generate complete crystallographic specifications, including Wyckoff positions and space groups, rather than sampling from empirical distributions. To achieve this, we draw inspiration from *spontaneous symmetry breaking* (SSB) in physics, where ordered phases emerge through spontaneous breaking of a higher symmetry. For instance, vanadium dioxide ($\mathrm{VO_2}$) undergoes a reduction in crystal symmetry from the high-symmetry rutile phase ($\mathrm{P4_2/mnm}$) to lower-symmetry monoclinic phases such as M1 ($\mathrm{P2_1/c}$) or M2 ($\mathrm{C2/m}$) (Müller, 2011). We then leverage Markovian jump diffusion (Feller, 1949; Campbell et al., 2022) to learn space-group transitions, where forward processes correspond to simulating SSB processes that reduce crystal symmetry to lower space groups, while reverse (generative) processes sample crystals directly from the lowest-symmetry priors.

Here, we summarize our contributions: (i) We theoretically derive a variational bound objective that unifies the structural dependencies among crystal components, in which (ii) we leverage Markovian jump diffusion to model space-group distributions. We then (iii) derive novel symmetry-breaking diffusion processes on continuous and discrete state spaces, which adaptively enforce space-group constraints and admit analytical forms. To model site symmetries, (iv) we propose a simplified representation that facilitates capturing site-symmetry distributions and yields more stable generated crystal structures. Our model, dubbed *Symmetry-breaking Crystal Diffusion* (SbCD), represents the first generative framework capable of generating complete crystallographic specifications.

The remainder of this paper is organized as follows. Section 2 reviews the background on Markovian jump diffusion. Section 3 revisits existing symmetry-preserving crystal diffusion frameworks under space-group constraints. Building on these foundations, Section 4 develops a novel theoretical symmetry-breaking generative diffusion framework for crystalline materials. Finally, Section 5 presents promising results on de novo crystal generation tasks.

2 MARKOVIAN JUMP DIFFUSION

Many natural phenomena can be modeled via discontinuous Markov processes, such as Poisson processes. These contrast with diffusion processes, a class of continuous-time Markov processes where changes occur incrementally over infinitesimal time intervals (Anderson, 1982; Øksendal, 2003). Here, we consider processes that evolve via discrete jumps: the system remains in a given state for a random duration before transitioning abruptly to another state. Following Feller (1949), we introduce the instantaneous rate matrix $\mathbf{\Lambda}_t(\widetilde{x}, x)$ from state $\widetilde{x}$ to state x as follows:

$$\mathbf{\Lambda}_t(\widetilde{x}, x) \triangleq \lim_{\Delta t \to 0} \frac{q_{t|t-\Delta t}(x|\widetilde{x}) - \delta_{\widetilde{x}x}}{\Delta t} = \lambda_t(\widetilde{x})\mathbf{\Pi}_t(\widetilde{x}, x) - \delta_{\widetilde{x}x}\lambda_t(\widetilde{x}) \tag{1}$$

where $q_{t|t-\Delta t}(x|\widetilde{x})$ denotes the infinitesimal transition probability, $\sum_x q_{t|t-\Delta t}(x|\widetilde{x}) = 1$; $\lambda_t(\widetilde{x})$ is the total jump (exit) rate out of state $\widetilde{x}$ at the time t, how fast the process leaves the state; $\mathbf{\Pi}_t(\widetilde{x}, x)$ is the probability that, given a jump from state $\widetilde{x}$ occurs at the time t, the process jumps to state x, $\mathbf{\Pi}_t(\widetilde{x}, \widetilde{x}) = 0$, denoting where the process goes after jumping; and, $\mathbf{\Lambda}_t(\widetilde{x}, \widetilde{x}) < 0$. Then, the forward and backward Kolmogorov equations for a continuous-time Markov chain are given by Feller (1949):

$$\frac{\partial}{\partial t} q_{t|\widetilde{t}}(x|\widetilde{x}) = \sum_{x'} \mathbf{\Lambda}_t(x', x) q_{t|\widetilde{t}}(x'|\widetilde{x}), \quad \frac{\partial}{\partial \widetilde{t}} q_{t|\widetilde{t}}(x|\widetilde{x}) = -\sum_{x'} \mathbf{\Lambda}_{\widetilde{t}}(\widetilde{x}, x') q_{t|\widetilde{t}}(x|x') \tag{2}$$

We now define the time-reversal of the forward process with reverse transition rate $\hat{\mathbf{\Lambda}}_t(x, \widetilde{x})$. Campbell et al. (2022) further derive its relation to the forward transition rate by manipulating the forward and backward equations above, yielding:

$$\hat{\mathbf{\Lambda}}_t(x, \widetilde{x}) = \mathbf{\Lambda}_t(\widetilde{x}, x) \frac{q_t(\widetilde{x})}{q_t(x)} = \mathbf{\Lambda}_t(\widetilde{x}, x) \sum_{x_0} \frac{q_{t|0}(\widetilde{x}|x_0)}{q_{t|0}(x|x_0)} q_{0|t}(x_0|x) \tag{3}$$

However, computing this reverse transition rate involves marginalizing $q_{0|t}(x_0|x)$ over x_0, which is generally intractable. Campbell et al. (2022) proposes estimating this term via a parametric neural network $p^{\theta}_{0|t}(x_0|x)$, and learning the reverse transition rate $\hat{\mathbf{\Lambda}}^{\theta}_t(x, \widetilde{x})$ by optimizing the continuous-time evidence bound, which we formulate in the following lemma:

Lemma 2.1. *Following Campbell et al. (2022), the negative log-likelihood of a continuous-time jump diffusion process admits the variational upper bound.*

$$\mathbb{E}_{p_{\text{data}}(x_0)}\left[-\log p^{\theta}_0(x_0)\right] \leq \mathcal{L}_{\text{jump}}(\theta) + \text{Const.}$$
$$\mathcal{L}_{\text{jump}}(\theta) \triangleq T\mathbb{E}_{t\sim\mathcal{U}(0,T)\widetilde{x}\sim q_{t|0}(\widetilde{x}|x_0)x\sim\Pi_t(\widetilde{x},x)}\left[-\hat{\mathbf{\Lambda}}^{\theta}_t(\widetilde{x},\widetilde{x}) + \mathbf{\Lambda}_t(\widetilde{x},\widetilde{x})\log(\hat{\mathbf{\Lambda}}^{\theta}_t(x,\widetilde{x}))\right]$$

where the first term acts as a regularizer that penalizes the total rate of leaving state $\widetilde{x}$, thereby preventing the reverse process from excessive jumping, while the second term maximizes the log-rate of the correct reverse transition pair $(x, \widetilde{x})$, scaled by the total rate of leaving $\widetilde{x}$. Sampling can then be implemented efficiently using τ-leaping algorithms (Gillespie, 2001; Wilkinson, 2018).

Lemma 2.2. *Consider a finite-state continuous-time Markov chain with K ordinal states, $x_t \in \{1, \ldots, K\}$, and learned reverse transition-rate matrix $\hat{\mathbf{\Lambda}}^{\theta}_t$. Assume that for $i \neq j$, $\hat{\mathbf{\Lambda}}^{\theta}_t(i,j) \geq 0$, and $\hat{\mathbf{\Lambda}}^{\theta}_t(i,i) = -\sum_{j\neq i}\hat{\mathbf{\Lambda}}^{\theta}_t(i,j)$. Then the next state $x_{t-\Delta t}$ after a τ-leap of size Δt is updated via $x_{t-\Delta t} = x_t + \sum_{j=1}^{K}\mathbf{P}_j\,(j - x_t)$, where $\mathbf{P}$ is a vector drawn from independent Poisson random variables $\mathbf{P}_j \sim \text{Poisson}\big(\hat{\mathbf{\Lambda}}^{\theta}_t(x_t, j)\,\Delta t\big),\ \forall j \neq x_t$.*

3 Space-group-aware crystal generation

Crystal symmetry and representation. To characterize crystal symmetries, we define a *space group* $G \triangleq \{(\mathbf{O}, \mathbf{t}) \in \mathbf{E}(3) \mid \mathbf{x} \mapsto \mathbf{O}\mathbf{x} + \mathbf{t}\}$ as a discrete subgroup of the Euclidean group $\mathbf{E}(3)$ leaving a crystal invariant, and its *point group* $P \triangleq \{\mathbf{O} \mid (\mathbf{O}, \mathbf{t}) \in G\}$ by projecting out translations. In total, there are 230 space groups and 32 crystallographic point groups, the complete sets of which we denote by $\mathcal{G}$ and $\mathcal{P}$, respectively. Due to the periodic arrangement of atoms, crystal structures admit multiple equivalent representations. Prior works typically model either the full crystal (Xie et al., 2021) or its *unit cell* (Jiao et al., 2023; 2024). Here, we adopt the more compact *asymmetric unit* (Levy et al., 2025), the minimal non-redundant subset that generates the full crystal via symmetry operations. Moreover, an asymmetric unit can be mathematically described by the tuple $(\mathbf{F}, \mathbf{A}, \mathbf{L}, \mathbf{S})$, where $\mathbf{F} \in \mathbb{R}^{M\times 3}$ and $\mathbf{A} \in \mathbb{R}^{M\times Z}$ represent the fractional coordinates and atom types (over Z chemical species) of M representative atoms, respectively; $\mathbf{L} \in \mathbb{R}^{3\times 3}$ denotes the crystal lattice, from which the raw Cartesian atomic coordinates can be obtained as $\mathbf{X} = \mathbf{L}\mathbf{F}$; and $\mathbf{S} \in \mathcal{P}^M$ specifies the site-symmetry group of each atom. These components are used in prior space-group-aware, symmetry-preserving crystal generation methods.

Space group constraints. The crystal representation can be further constrained depending on its crystal family; the 230 space groups are classified into 6 crystal families. Jiao et al. (2024) derive conditions on the lattice using the polar decomposition (Hall, 2013). Concretely, any invertible matrix $\mathbf{L}$ admits a unique decomposition $\mathbf{L} = \mathbf{T}\exp(\mathbf{U})$, where $\mathbf{T} \in \mathbb{R}^{3\times 3}$ is orthogonal, and $\mathbf{U} \in \mathbb{R}^{3\times 3}$ is symmetric. This decomposition yields an *invariant* representation: applying an orthogonal transformation to $\mathbf{L}$ affects only $\mathbf{T}$, leaving $\mathbf{U}$ unchanged. Furthermore, $\mathbf{U}$ can be expanded over a fixed set of symmetric basis matrices $\{\mathbf{B}_i\}_{i=1}^{6}$ as $\mathbf{U} = \sum_{i=1}^{6} k_i\mathbf{B}_i$. Thus, the coefficient vector $\mathbf{k} = (k_i)_{i=1}^{6}$ provides a complete, invariant lattice representation. We then introduce a masking mechanism that respects lattice constraints during both forward and reverse diffusion processes:

Lemma 3.1. *Let $\mathbf{k} \in \mathbb{R}^6$ denote the invariant crystal-lattice representation, whose values are constrained by a given crystal family. A discrete-time, continuous-state diffusion process (Ho et al., 2020) satisfies the masking mechanism defined below to remain within the specified crystal family:*

$$\mathbf{k}_t = \mathfrak{m} \odot \left(\sqrt{\bar{\alpha}_t}\,\mathbf{k}_0 + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}\right) + \mathfrak{m}_{\mathfrak{b}}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(0, I), \tag{4}$$

where $\bar{\alpha}_t$ denotes the noise schedule from Ho et al. (2020), $\odot$ is the element-wise product, and $(\mathfrak{m}, \mathfrak{m}_{\mathfrak{b}}) \in \{0,1\}^6 \times \mathbb{R}^6$ is the space-group-informed mask–bias pair, e.g., $\mathfrak{m} = [0,0,0,0,1,1]$ and $\mathfrak{m}_{\mathfrak{b}} = [-\log\frac{3}{4}, 0, 0, 0, 0, 0]$ correspond to the mask–bias pair for the hexagonal family (Figure 1).

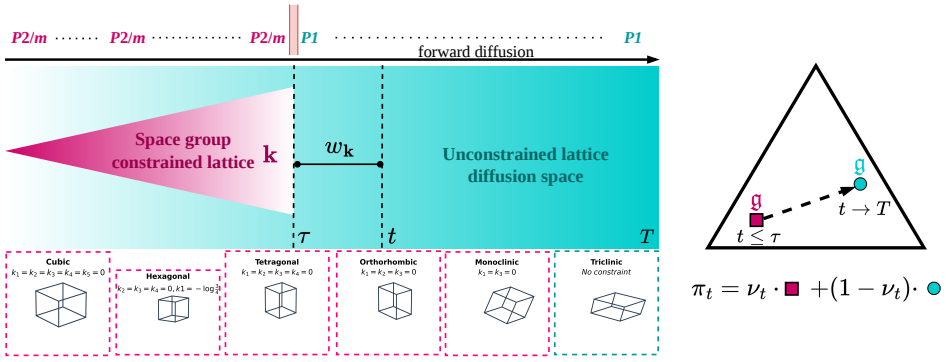


Figure 1: SbCD Conceptual Framework. (**Left**, *continuous state space*) A symmetry-breaking diffusion process that undergoes a reduction in symmetry from a higher space group P2/m to a lower space group P1 at time step τ. The colored regions correspond to specific space-group constraints applied to the lattice component $\mathbf{k}$. The upper panel depicts space-group transitions, while the lower panel visualizes lattice constraints for six crystal families. (**Right**, *discrete state space*) For $t > \tau$, the induced site-symmetry prior $\boldsymbol{\pi}_t$ evolves monotonically along the straight line segment connecting $\mathfrak{g}$, the marginal site-symmetry prior of P2/m, and $\mathfrak{g}$, the marginal site-symmetry prior of P1.

Symmetry preserving. We characterize the local symmetry at each atomic site via its *site symmetry*, i.e., the set of symmetry operations that leave the site invariant. Sites that are symmetry-equivalent under the action of the space group G form an orbit corresponding to a *Wyckoff position*. The site-symmetry group is a point group isomorphic to a subgroup of the point group P. Levy et al. (2025) encode the site symmetries $\mathbf{S}$ using a binary representation of oriented site-symmetry symbols (Hahn et al., 1983), which specify generators of the site-symmetry group along 15 possible crystallographic axes (including body and face diagonals), with up to 13 distinct symmetry operations per axis. To preserve symmetry within a given space group, they employ categorical diffusion (Austin et al., 2021) with space-group-specific marginal priors over site symmetries during both training and generation.

Lemma 3.2. *Let $s \in \{1, \cdots, 15\}$ index the crystallographic axes, and $\mathbf{s} \in \{0,1\}^{13} \subset \mathbf{S}$ denote the site-symmetry representation along the s-th axis. Let $\mathfrak{g} \in [0,1]^{13}$ be the marginal prior of site symmetry associated with the space group G over the s-th axis. A discrete-time, discrete-state diffusion process (Austin et al., 2021) that preserves crystal symmetry is defined via the forward transition:*

$$\mathbf{s_t} \sim \mathrm{Cat}(\mathbf{s_t}; \mathbf{s}_0 \bar{\mathbf{Q}}_{s,t}), \qquad \bar{\mathbf{Q}}_{s,t} = (1-\bar{\beta}_t)\mathbf{I} + \bar{\beta}_t \mathbf{1}\mathfrak{g}$$

where $\bar{\beta}_t$ denotes the categorical noise schedule from Austin et al. (2021). By construction, this space-group-conditioned prior ensures that generated site-symmetry states are consistent with the symmetry constraints imposed by G, which consequently restricts the candidate Wyckoff positions of representative atoms. If multiple candidates share the same site-symmetry state, we assign the Wyckoff position whose symmetry-equivalent sites are closest to generated fractional coordinates.

4 SbCD: Symmetry-Breaking Crystal Diffusion

4.1 Structure-Informed Variational Bound

Crystals are inherently heterogeneous structures comprising multiple data modalities. This challenge necessitates moving beyond a single generative modeling paradigm or data representation. Here, we unify diffusion-based training objectives across continuous and discrete time and state spaces into a single formulation. We recall the parametrization of an asymmetric unit as $\mathcal{M} = (\mathbf{F}, \mathbf{A}, \mathbf{k}, \mathbf{S}, G)$. As defined in Section 3, the site symmetries $\mathbf{S}$ and the lattice representation $\mathbf{k}$ are explicitly constrained by the space group G. Under these structural dependencies, we formulate the variational upper bound.

Proposition 4.1. *Let $\mathbb{Q}$, $\hat{\mathbb{Q}}$, and $\mathbb{P}^\theta$ denote the forward, exact reverse, and learned reverse path measures of the continuous-time Markov chain. The negative log-likelihood on crystal structures $\mathcal{M}_0$ is bounded above by the expected Kullback–Leibler divergence over the reverse trajectory:*

$$-\log p_\theta(\mathcal{M}_0) \le \mathbb{E}_{q_{T|0}}\Big[D_{\mathrm{KL}}\big(\hat{\mathbb{Q}}^{\mathcal{M}_T\to\mathcal{M}_0} \,\big\|\, \mathbb{P}^{\theta,\mathcal{M}_T}\big)\Big] + \mathrm{Const.}$$

The divergence decomposes additively into space group, lattice, site symmetry, atom type, and fractional coordinate components:

$$D_{\mathrm{KL}}\big(\hat{\mathbb{Q}}^{\mathcal{M}_T\to\mathcal{M}_0} \,\big\|\, \mathbb{P}^{\theta,\mathcal{M}_T}\big) = \underbrace{D_{\mathrm{KL}}\big(\hat{\mathbb{Q}}_{\mathbf{A}} \,\big\|\, \mathbb{P}^\theta_{\mathbf{A}}\big) + D_{\mathrm{KL}}\big(\hat{\mathbb{Q}}_{\mathbf{F}} \,\big\|\, \mathbb{P}^\theta_{\mathbf{F}}\big)}_{\text{I: Atom Types \& Fractional Coordinates}}$$
$$+ \underbrace{D_{\mathrm{KL}}\big(\hat{\mathbb{Q}}_G \,\big\|\, \mathbb{P}^\theta_G\big)}_{\text{II: Space Group}} + \underbrace{\mathbb{E}_{\hat{\mathbb{Q}}_G}\Big[D_{\mathrm{KL}}\big(\hat{\mathbb{Q}}_{\mathbf{k}|G} \,\big\|\, \mathbb{P}^\theta_{\mathbf{k}|G}\big) + D_{\mathrm{KL}}\big(\hat{\mathbb{Q}}_{\mathbf{S}|G} \,\big\|\, \mathbb{P}^\theta_{\mathbf{S}|G}\big)\Big]}_{\text{III: Lattice \& Site Symmetry}}$$

Here, the conditioning superscripts have been omitted on the right-hand side for brevity. The notation $\hat{\mathbb{Q}}_{\cdot}^{\mathcal{M}_T\to\mathcal{M}_0}$ denotes a bridge measure pinned at both endpoints $\mathcal{M}_T$ and $\mathcal{M}_0$, and $\mathbb{P}_{\cdot}^{\theta,\mathcal{M}_T}$ denotes a path measure conditioned on $\mathcal{M}_T$. While most existing de novo crystal generation frameworks neglect learning the space-group distribution (term II) and assume it is sampled from data during inference, our method explicitly models G as part of the generative process. Moreover, during the diffusion process, the space group transitions, resulting from term II, directly modify the constraints imposed on the lattice and site symmetry (term III), which we address in the following sections. Finally, the space-group-independent attributes (term I) can be modeled using standard techniques from prior work of Jiao et al. (2023); Austin et al. (2021); Levy et al. (2025).

4.2 Physics-inspired space group modeling via jump diffusion

Taking inspiration from *spontaneous symmetry breaking*, in which physical systems spontaneously break symmetry into lower-symmetry states, we learn the space-group distribution by mirroring this phenomenon via Markovian jump-diffusion. In the forward process, a crystal remains in a given space group $\widetilde{G}$ for a random duration before transitioning to a lower-symmetry space group G, denoted by $G \prec \widetilde{G}$. The transition dynamics are governed by a time-dependent rate matrix $\mathbf{\Lambda}_t(\widetilde{G}, G)$. The generative process approximates the reverse transitions by learning a parametric rate $\hat{\mathbf{\Lambda}}^\theta_t(G, \widetilde{G})$. For simplicity, we fix the terminal state of every jump to $G = \mathrm{P1}$, the lowest-symmetry space group (triclinic system) that contains only the identity operation. In the following proposition, we introduce the transition rate matrix such that (i) the forward process mixes rapidly toward a reference distribution and (ii) the Kolmogorov forward equation admits a closed-form solution.

Proposition 4.2. *Consider the discrete state space of 230 crystallographic space groups drawn from the six crystal families. Let $\mathbf{\Pi}, \mathbf{\Lambda}_t \in \mathbb{R}^{230\times 230}$ denote the time-invariant transition probability matrix and the instantaneous rate matrix, respectively, defined element-wise as follows:*

$$\mathbf{\Pi}(\widetilde{G}, G) = \begin{cases} 0 & \widetilde{G} = \mathrm{P1} \\ 0 & G = \widetilde{G} \\ 1 & G = \mathrm{P1} \neq \widetilde{G} \\ 0 & \textit{otherwise}. \end{cases} \qquad \mathbf{\Lambda}_t(\widetilde{G}, G) = \begin{cases} 0 & \widetilde{G} = \mathrm{P1} \\ -\lambda(t) & G = \widetilde{G} \\ \lambda(t) & G = \mathrm{P1} \neq \widetilde{G} \\ 0 & \textit{otherwise}. \end{cases}$$

Then, $\mathbf{\Lambda}_t$ and $\mathbf{\Lambda}_s$ commute for all t, s. Consequently, the forward transition probability of being in state G at time t, given state $\widetilde{G}$ at time $\widetilde{t} < t$, $q_{t|\widetilde{t}}(G|\widetilde{G})$, admits the matrix exponential form:

$$q_{t|\widetilde{t}}(G|\widetilde{G}) = \left(\mathbf{D}\exp\Big[\mathbf{\Sigma}\int_{\widetilde{t}}^{t}\lambda(s)ds\Big]\mathbf{D}^{-1}\right)_{\widetilde{G},G} \quad \textit{with} \quad \frac{\mathbf{\Lambda}_t}{\lambda(t)} = \mathbf{D}\mathbf{\Sigma}\mathbf{D}^{-1}$$

We defer the proof to the Appendix. Given the analytical forward jump kernel above, we learn the reverse rate $\hat{\mathbf{\Lambda}}^\theta_t(G, \widetilde{G})$ by minimizing the variational upper bound component corresponding to $D_{\mathrm{KL}}(\hat{\mathbb{Q}}_G^{\mathcal{M}_T\to\mathcal{M}_0} \| \mathbb{P}_G^{\theta,\mathcal{M}_T})$ in Proposition 4.1.

Posterior holding-time τ and symmetry-breaking time window w. Optimising the jump diffusion loss $\mathcal{L}_{\text{jump}}(\theta)$ via Lemma 2.1 first requires sampling a diffusion timestep $t \sim \mathcal{U}(0, T)$, followed by sampling the corresponding space group according to the analytical forward kernel $q_{t|0}$ from Proposition 4.2. For large diffusion times, $t \to T$, the forward process has already reached the space group P1 and remains there thereafter. In the following corollary, we define the posterior holding time τ as the duration for which the process remains in its initial space group before jumping, and the symmetry-breaking time window w as the elapsed time spent in the space group P1 up to the current diffusion timestep. Figure 1 depicts these induced time variables.

Corollary 4.3. *Let a crystal initially reside in a space group $\widetilde{G}$ and be observed in a lower-symmetry space group $G \prec \widetilde{G}$ at time t. Consider a constant transition rate, $\lambda(t) = \lambda\ \forall t$, then the posterior distribution of the holding time $\tau \in [0, t]$ is $\tau = -\frac{1}{\lambda}\log(1 - u(1 - e^{-\lambda t}))$, where $u \sim \mathcal{U}(0, 1)$. Consequently, the symmetry-breaking time window can be calculated as $w = t - \tau$.*

4.3 Symmetry-breaking diffusion in continuous-state space

We extend the space-group-constrained diffusion framework of Jiao et al. (2024) to learn the crystal-lattice distribution. In their framework, the space group is fixed during both forward and reverse diffusion, and an element-wise masking mechanism on the lattice representation $\mathbf{k}$ (see Lemma 3.1) ensures it remains confined to the space-group-constrained subspace. In contrast, our framework permits the space group to evolve throughout the diffusion process. We therefore introduce a symmetry-breaking diffusion process over the continuous-state lattice space, which adaptively enforces the lattice constraints of the prevailing crystal family whenever the process transitions to a new space group. Below, we provide the analytical forms of the forward and reverse sampling kernels, which correspond to the variational bound term $\mathbb{E}_{\hat{\mathbb{Q}}_G}\left[D_{\text{KL}}\left(\hat{\mathbb{Q}}_{\mathbf{k}|G} \,\middle\|\, \mathbb{P}^{\theta}_{\mathbf{k}|G}\right)\right]$ in Proposition 4.1.

Proposition 4.4. *Consider a forward diffusion process with a symmetry-breaking event at time τ, transitioning from space group $\widetilde{G}$ to G. From Lemma 3.1, let $(\mathfrak{m}, \mathfrak{m}_\mathfrak{b})$ and $(\widetilde{\mathfrak{m}}, \widetilde{\mathfrak{m}}_\mathfrak{b})$ denote the mask–bias pairs associated with G and $\widetilde{G}$, respectively. Then, the continuous-state forward diffusion kernel applied to the lattice parameters $\mathbf{k}$ is given analytically by:*

$$\begin{aligned}\mathbf{k}_t = \widetilde{\mathfrak{m}} \odot \left(\sqrt{\bar{\alpha}_t}\,\mathbf{k}_0 + \sqrt{1-\bar{\alpha}_t}\,\widetilde{\boldsymbol{\epsilon}}\right) + \widetilde{\mathfrak{m}}_\mathfrak{b} + \mathfrak{m} \odot \widetilde{\mathfrak{m}}_\mathfrak{b}\sqrt{\frac{\bar{\alpha}_t}{\bar{\alpha}_{t-w_\mathbf{k}}}} \\ - \left(\widetilde{\mathfrak{m}} \odot \sqrt{1 - \frac{\bar{\alpha}_t}{\bar{\alpha}_{t-w_\mathbf{k}}}}\,\widetilde{\boldsymbol{\epsilon}} + \widetilde{\mathfrak{m}}_\mathfrak{b}\right) + \left(\mathfrak{m} \odot \sqrt{1 - \frac{\bar{\alpha}_t}{\bar{\alpha}_{t-w_\mathbf{k}}}}\,\boldsymbol{\epsilon} + \mathfrak{m}_\mathfrak{b}\right),\end{aligned}$$

where $w_\mathbf{k} = t - \tau$. The corresponding reverse-time sampling step is simulated as:

$$\mathbf{k}_{t-1} = \widetilde{\mathfrak{m}} \odot \left(\frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_t}(1-\alpha_t)\,\hat{\mathbf{k}}^{\theta}_0(\mathcal{M}_t) + \mathfrak{m} \odot \frac{\sqrt{\alpha_t}}{1-\bar{\alpha}_t}(1-\bar{\alpha}_{t-1})\,\mathbf{k}_t + \mathfrak{m}_\mathfrak{b} + \sigma_t\boldsymbol{\epsilon}\right) + \widetilde{\mathfrak{m}}_\mathfrak{b}.$$

This novel masking mechanism respects the lattice constraints both before and after symmetry breaking occurs. Moreover, our formulation directly generalizes the prior work of Jiao et al. (2024), which corresponds to the special case of no symmetry breaking in Lemma 3.1, i.e., $\widetilde{G} = G$. Figure 1 (left) illustrates how lattice constraints adaptively change during a space-group transition.

4.4 Symmetry-breaking diffusion in discrete-state space

Novel site-symmetry representation learning. Levy et al. (2025) proposed learning site symmetries by using a binary matrix of size 15×13 (15 canonical symmetry axes and 13 possible symmetry operations per axis; see Lemma 3.2), based on the oriented site-symmetry symbols (Hahn et al., 1983). However, this encoding scheme does not distinguish Wyckoff positions sharing the same oriented site-symmetry symbol. Due to this ambiguity, the scheme may complicate the learning process and often requires projecting generated site symmetries onto the nearest valid point group (Levy et al., 2025). For instance, in space group Immm, the two Wyckoff positions $4\mathrm{i}$ and $4\mathrm{j}$ both possess the oriented site-symmetry symbol $\mathrm{mm2}$ and are therefore represented by the same 15×13 binary matrix. Here, we directly use a one-hot vector representation to model the oriented site-symmetry symbols. We identify 81 distinctive oriented site-symmetry symbols, inferred across all 230 space groups (see Appendix D.1 for the list).

In the *symmetry-preserving* framework of Levy et al. (2025), for a fixed space group, site-symmetry states are corrupted via a discrete-state diffusion process (Austin et al., 2021), with transition dynamics defined by the marginal site-symmetry prior induced by that space group (Lemma 3.2). In contrast, in our symmetry-breaking process, the ambient space group itself may transition from one group to another. This change alters the induced marginal prior over site symmetries, and therefore requires an interpolation between site-symmetry priors across space groups. In the following proposition, we introduce such an interpolant to enable learning of site symmetries in the symmetry-breaking diffusion process.

Proposition 4.5. *Let $\widetilde{\mathfrak{g}}, \mathfrak{g} \in \Delta^{81-1}$ be the marginal site-symmetry priors associated with the space groups $\widetilde{G}$ and G, respectively. Let $\bar{\alpha}_t$ be a strictly decreasing noise schedule with symmetry-breaking time τ satisfying $0 < \bar{\alpha}_t, \bar{\alpha}_\tau \leq 1$. For $t > \tau$, define*

$$\boldsymbol{\pi}_t = \nu_t \widetilde{\mathfrak{g}} + (1-\nu_t)\mathfrak{g} = \frac{\bar{\alpha}_t(1-\bar{\alpha}_{t-w_\mathbf{S}})\widetilde{\mathfrak{g}} + (\bar{\alpha}_{t-w_\mathbf{S}} - \bar{\alpha}_t)\mathfrak{g}}{\bar{\alpha}_{t-w_\mathbf{S}}(1-\bar{\alpha}_t)} \quad \text{where } w_\mathbf{S} = t-\tau,\ \nu_t \in [0,1]$$

Then, $\boldsymbol{\pi}_t$ lies in the convex hull of the two priors, $\boldsymbol{\pi}_\tau = \widetilde{\mathfrak{g}},\ \lim_{t\to T} \boldsymbol{\pi}_t = \mathfrak{g}$, and the trajectory $\{\boldsymbol{\pi}_t\}_{t>\tau}$ moves monotonically along the straight line segment connecting $\widetilde{\mathfrak{g}}$ to $\mathfrak{g}$ inside the simplex.

Figure 1 (right) visualizes the interpolated prior trajectory. Below, we derive corresponding forward and reverse discrete-state diffusion kernels and summarize the sampling of SbCD in Algorithm 1.

Corollary 4.6. *Let the discrete-state diffusion transition matrices of Austin et al. (2021) be redefined under the symmetry-breaking formulation as:*

$$\mathbf{Q}_t = \alpha_t I + (1-\alpha_t)\mathbf{1}(\mathfrak{g})^T, \qquad \bar{\mathbf{Q}}_{t-1} = \bar{\alpha}_{t-1} I + (1-\bar{\alpha}_{t-1})\mathbf{1}(\widetilde{\mathfrak{g}})^T$$

Then, the forward and reverse sampling processes on the site symmetry $\mathbf{S}$ can be derived as follows:

$$\mathbf{S}_t \sim \mathrm{Cat}\left(\mathbf{S}_t; \mathbf{S}_0 \bar{\mathbf{Q}}_t^{\widetilde{G}\to G}\right) \qquad \mathbf{S}_{t-1} \sim \mathrm{Cat}\left(\mathbf{S}_{t-1}; \frac{\mathbf{S}_t \mathbf{Q}_t^T \odot \hat{\mathbf{S}}_0^\theta(\mathcal{M}_t)\bar{\mathbf{Q}}_{t-1}}{\hat{\mathbf{S}}_0^\theta(\mathcal{M}_t)\bar{\mathbf{Q}}_t^{\widetilde{G}\to G}\mathbf{S}_t^T}\right)$$

5 Experiments

Task and metrics. Following Levy et al. (2025), we conduct de novo crystal generation experiments on the MP-20 dataset, a subset of the Materials Project (Jain et al., 2013), and the more challenging MPTS-52 dataset (Baird et al., 2024), which contains up to 52 atoms per primitive unit cell. We evaluate our results using standard metrics: (i) *Validity*, including structure ($\mathrm{Struct.}$), requiring a minimum pairwise atomic distance of 0.5Å, and composition ($\mathrm{Comp.}$), requiring overall charge neutrality; (ii) *Coverage*, measuring similarity between generated and test materials with respect to structure and composition cutoff distances; and (iii) *Property statistics*, measuring the Wasserstein-1 distance between property distributions, including density (d_ρ, unit $\mathrm{g/cm^3}$) and the number of unique elements ($\#\mathrm{elem.}$), and the Jensen Shannon distance between space group distributions ($\mathrm{d_{sg}}$). Finally, we assess the thermodynamic stability (Stable) of generated materials and compute their novelty and uniqueness ($\mathrm{S.U.N}$), following Miller et al. (2024). As in Levy et al. (2025), we employ the pretrained CHGNet model (Deng et al., 2023) to perform structural relaxations before calculating the energy above the hull (E^{hull}). We evaluate all methods on 10,000 sampled crystals, except that property statistics are computed on 1,000 valid structures, following prior works (Xie et al., 2021; Levy et al., 2025). We defer the experimental details to Appendix D.

Baselines. We benchmark SbCD against diffusion-based methods, including CDVAE (Xie et al., 2021), DiffCSP (Jiao et al., 2023), DiffCSP++ (Jiao et al., 2024), and SGEquiDiff (Chang et al., 2025) as well as flow-based methods, including FlowMM (Miller et al., 2024) and SGFM (Puny et al., 2025). In particular, we compare SbCD with its symmetry-preserving counterpart, SymmCD (Levy et al., 2025). We classify baselines by whether they use raw-crystal ($\mathrm{R.}$), unit-cell ($\mathrm{Unit.}$), or asymmetric-unit ($\mathrm{Asym.}$) representations. In addition, we ablate various SbCD variants: SbCD$^\sharp$, which learns site symmetries using the 15×13 binary matrix (as described in Lemma 3.2), SbCD$^\diamond$, which has a model architecture similar to that of DiffCSP/++ (Jiao et al., 2023; 2024), and a non symmetry-breaking SbCD$^\diamond$ ablation, SbCD$^\diamond$ (w/o SB), which explicitly conditions on empirical space groups in both training and inference.

Table 2: De novo crystal generation on MP-20. Baselines are from Levy et al. (2025) and Puny et al. (2025). Hyphens (−) indicate unreproducible results.

	Method	**Validity** (%)(↑)		**Coverage** (%)(↑)		**Property statistics** (↓)		**CHGNet** $E^{\text{hull}} < 0$ (%)(↑)	
		Struct.	Comp.	Recall	Precision	d_ρ	# elem.	Stable	S.U.N
R.	CDVAE	99.97	85.61	99.31	99.47	0.70	1.28	4.42	4.26
Unit.	DiffCSP	97.43	82.50	99.55	98.73	0.18	0.56	11.32	8.92
	FlowMM	96.67	83.25	99.49	99.58	0.23	0.08	9.05	6.49
	DiffCSP++ (empirical)	99.44	86.50	99.72	99.61	0.12	0.33	11.36	8.62
	DiffCSP++ (Wyformer)	99.66	80.34	–	–	0.67	0.10	–	–
	SGFM (empirical)	99.87	86.81	–	–	0.075	0.181	–	–
	SGFM (Wyformer)	99.87	84.76	–	–	0.24	0.23	–	–
Asym.	SymmCD	90.34	85.81	99.58	97.76	0.23	0.40	9.34	6.86
	SymmCD‡ (retrain)	87.37	86.30	99.46	97.21	0.41	0.28	7.40	6.79
	SbCD	89.80	90.32	99.44	97.91	0.18	0.03	11.28	7.15
	SbCD$^{\sharp}$	89.81	88.32	99.51	97.31	0.09	0.04	10.74	7.31
	SbCD$^{\diamond}$	88.25	87.47	99.44	97.71	0.16	0.07	10.23	7.67
	SbCD$^{\diamond}$(w/o SB)	78.22	84.93	99.76	96.33	0.15	0.21	8.90	6.19

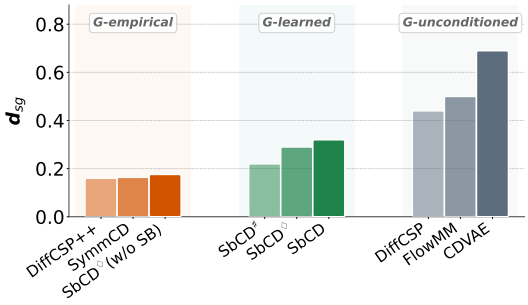


Figure 2: Results on the JS space-group distance between generated and test distributions on MP20. G-empirical and G-learned rely on empirical and learned distributions, while G-unconditioned denotes no space-group conditioning.

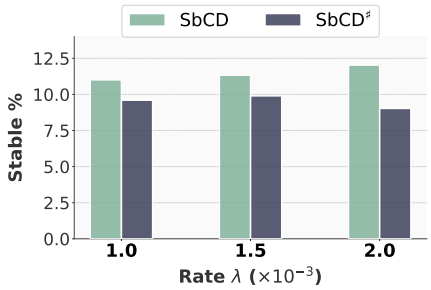


Figure 3: Thermodynamic stability results ($E^{\text{hull}} < 0$) across increasing transition rates λ: Comparison between the novel site-symmetry representation, SbCD, versus the canonical representation on symmetry axes and operations (Levy et al., 2025), SbCD$^{\sharp}$.

5.1 De novo crystal generation results

Table 2 demonstrates that SbCD generated crystals, which exhibit a significantly higher proportion of charge-neutral compositions and remain substantially more similar to the test materials in terms of the number of unique elements, indicating a better capture of the underlying compositional distribution. On the thermodynamic stability metric, SbCD performs competitively with DiffCSP and DiffCSP++, despite the latter additionally employing space-group constraints on unit-cell fractional coordinates. Notably, SbCD variants demonstrate clear improvements over their symmetry-preserving SymmCD counterparts across most evaluation criteria, underscoring the effectiveness of incorporating symmetry-breaking transitions into the diffusion process. Moreover, existing methods typically depend on external Wyckoff positions for generating crystalline materials. Table 2 demonstrates that most models suffer performance drops when replacing *empirical* Wyckoff positions with unseen ones generated by *Wyformer* (Kazeev et al., 2025). In contrast, SbCD represents the first framework capable of generating complete crystallographic structure specifications (see Table 1), while outperforming the generated Wyckoff position baselines across various metrics.

In Figure 2, SbCD variants narrow the performance gap with methods directly conditioned on space groups from empirical distributions, and outperform methods without space-group conditioning. In parallel, SbCD$^{\diamond}$ outperforms its variant without learning space group distributions through symmetry-breaking modeling, SbCD$^{\diamond}$ (w/o SB). These results highlight SbCD's ability to effectively capture a wide range of space-group distributions. Eventually, Table 3 shows SbCD's performance on more challenging, larger MPTS-52 crystalline materials. Despite generating lower structural-validity crystals compared to SGEquiDiff, SbCD effectively scales and achieves *the best performance on thermodynamic stability metrics*. This result emphasizes that high structural validity can not necessarily guarantee thermodynamic-stability improvement on generated materials.

Table 3: De novo crystal generation on MPTS-52. We scale SbCD at different sizes: -S, -L, -XL.

Method	**Validity** (%)(↑)		**Coverage** (%)(↑)		**Property statistics** (↓)			**CHGNet** $E^{\text{hull}} < 0$ (%)(↑)	
	Struct.	Comp.	Recall	Precision	d_ρ	# elem.	d_{sg}	Stable	S.U.N
SGEquiDiff (params=5.5M)	97.93	80.43	99.90	96.15	0.870	0.559	0.306	5.79	3.90
SymmCD (params=61M)	90.10	79.20	99.60	96.30	0.844	0.317	0.274	5.97	4.62
SymmCD‡ (params=61M)(retrain)	85.92	80.22	99.74	95.24	1.104	0.381	0.297	5.26	4.60
SbCD-S (params=14.8M)	90.42	78.62	99.57	99.57	0.818	0.585	0.358	**6.88**	**5.01**
SbCD-L (params=47.9M)	88.02	81.13	99.53	97.02	0.665	0.567	0.368	**8.23**	**5.29**
SbCD-XL (params=62.1M)	90.55	80.08	99.65	97.10	0.760	0.654	0.343	**8.36**	**5.68**

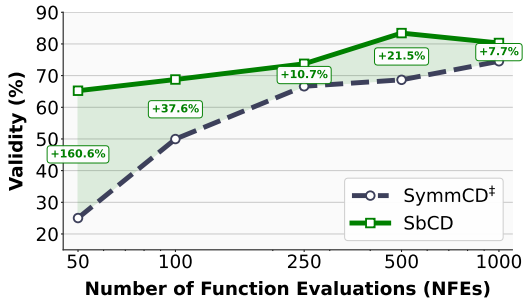


Figure 4: Comparison of the crystal validity, multiplying structural and compositional, between SbCD and SymmCD‡ for different NFEs on MP-20.

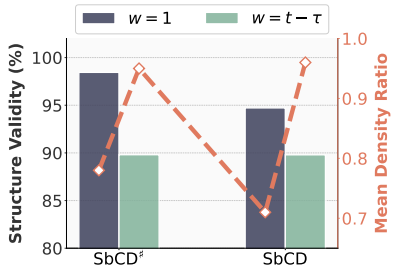


Figure 5: Results on structural validity and mean density ratio w.r.t. the test set for different symmetry-breaking time window sizes, w, on MP-20.

5.2 Ablation Studies

Symmetry-breaking window revealing the structural validity pitfalls. Figure 5 visualizes the structural validity results under a fixed time window ($w = 1$). We observe that SbCD variants significantly generate more structurally valid crystals ($\sim 98\%$). However, their mean density ratio relative to the training data decreases dramatically, indicating that low-density crystals tend to exhibit valid structures more readily. In Table 2, CDVAE (Xie et al., 2021), while achieving the best performance on structure validity, generates low-density crystals (d_ρ is high), whose structures are loosely packed. This can also be observed in generated structures with low-symmetry space groups, predominantly P1 from Levy et al. (2025). In practice, we would prioritize obtaining more compact crystal structures, which SbCD variants (with the theoretical window size from Corollary 4.3) can straightforwardly satisfy, as demonstrated by their mean density ratio results.

Robustness of site-symmetry learning under varying transition rates λ and NFEs. Rates determine how quickly crystals transition toward the least-constrained space group P1. Therefore, picking a suitable rate is essential so that the diffusion process has enough time to denoise in original space groups while still converging to the prior distributions. Figure 3 shows that SbCD achieves greater robustness than SbCD$^{\sharp}$, even at high transition rates, which underscores the advantage of our simplified site-symmetry representation in alleviating the challenges of modeling the complex site symmetries from Levy et al. (2025). In terms of inference speed, Figure 4 shows that SbCD significantly outperforms its symmetry-preserving counterpart SymmCD‡, particularly in low-NFE regimes, which highlights SbCD's potential for enabling few-step de novo crystal generation.

6 Conclusions

We propose SbCD, a novel symmetry-breaking crystal diffusion model that enables sampling full crystallographic structure specifications from minimal symmetry assumptions. As such, SbCD marks an important step for de novo crystal generation. SbCD takes inspiration from spontaneous symmetry breaking and introduces novel adaptive space group constraints on both continuous and discrete state spaces. Empirically, SbCD substantially outperforms its symmetry-preserving counterpart, offering a promising proof of concept. Future work could extend SbCD by (i) enforcing dynamically varying Wyckoff symmetry constraints on asymmetric-unit fractional coordinates and (ii) exploring more physically realistic transition pathways within its symmetry-breaking framework.

AI USE STATEMENT

In this work, we used generative AI tools for [assist in the writing of proofs]. We have not used generative AI tools for [help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, propose or refine hypotheses, design or provide feedback on research methodology or experiments, implement methods, interpret results], and [Generate synthetic data sets, assist with translation, clean and reformat dataset, support qualitative and thematic data analysis] are not applicable to this work. Additionally, we used generative AI tools for [edit a research paper to improve readability]. We have reviewed the only AI proof-writing assistance on stochastic path measure theories in Proposition 4.1. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

SbCD holds the potential to accelerate materials discovery by efficiently exploring large chemical and crystallographic design spaces, potentially enabling advances in energy storage, catalysis, semiconductors, and medicine. At the same time, it may also facilitate the discovery of harmful materials, highlighting the need for responsible use, safety screening, and rigorous experimental validation.

REPRODUCIBILITY STATEMENT

In the Appendix, we provide detailed descriptions of SbCD's training and sampling algorithms, complete proofs of the theoretical results, and citations for all datasets used in the experiments. We also include a comprehensive specification of the model architecture, hyperparameters, and computational resources. Upon acceptance, we commit to releasing both the source code and model weights publicly.

REFERENCES

Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.

Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow, and Rianne Van Den Berg. Structured denoising diffusion models in discrete state-spaces. *Advances in neural information processing systems*, 34:17981–17993, 2021.

Sterling G Baird, Hasan M Sayeed, Joseph Montoya, and Taylor D Sparks. matbench-genmetrics: A python library for benchmarking crystal structure generative models using time-based splits of materials project structures. *Journal of Open Source Software*, 9(97):5618, 2024.

Saidur Rahman Bakaul, Claudy Rayan Serrao, Michelle Lee, Chun Wing Yeung, Asis Sarker, Shang-Lin Hsu, Ajay Kumar Yadav, Liv Dedon, Long You, Asif Islam Khan, et al. Single crystal functional oxides on silicon. *Nature communications*, 7(1):10547, 2016.

Andrew Campbell, Joe Benton, Valentin De Bortoli, Tom Rainforth, George Deligiannidis, and Arnaud Doucet. A continuous time framework for discrete denoising models. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=DmT862YAieY.

Rees Chang, Angela Pak, Alex Guerra, Ni Zhan, Nick Richardson, Elif Ertekin, and Ryan P Adams. Space group equivariant crystal diffusion. *arXiv preprint arXiv:2505.10994*, 2025.

Rees Chang, Angela Pak, Alex Guerra, Ni Zhan, Nick Richardson, Elif Ertekin, and Ryan Adams. Space group equivariant crystal diffusion. *Advances in Neural Information Processing Systems*, 38:72772–72805, 2026.

Anthony K Cheetham, Ram Seshadri, and Fred Wudl. Chemical synthesis and materials discovery. *Nature Synthesis*, 1(7):514–520, 2022.

Gillian Collins, Eileen Armstrong, David McNulty, Sally O'Hanlon, Hugh Geaney, and Colm O'Dwyer. 2d and 3d photonic crystal materials for photocatalysis and electrochemical energy storage and conversion. *Science and Technology of advanced MaTerialS*, 17(1):563–582, 2016.

Bowen Deng, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Kevin Han, Christopher J Bartel, and Gerbrand Ceder. Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5(9):1031–1041, 2023.

Vivian R Feig, Sanghyun Park, Pier Giuseppe Rivano, Jinhee Kim, Benjamin Muller, Ashka Patel, Caroline Dial, Sofia Gonzalez, Hannah Carlisle, Flavia Codreanu, et al. Self-aggregating long-acting injectable microcrystals. *Nature Chemical Engineering*, 2(3):209–219, 2025.

William Feller. On the theory of stochastic processes, with particular reference to applications. 1949. URL https://api.semanticscholar.org/CorpusID:121027442.

Scott Fredericks, Kevin Parrish, Dean Sayre, and Qiang Zhu. Pyxtal: A python library for crystal structure generation and symmetry analysis. *Computer Physics Communications*, 261:107810, 2021.

Daniel T Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of chemical physics*, 115(4):1716–1733, 2001.

Theo Hahn, Uri Shmueli, and JC Wilson Arthur. *International tables for crystallography*, volume 1. Reidel Dordrecht, 1983.

Brian C Hall. Lie groups, lie algebras, and representations. In *Quantum Theory for Mathematicians*, pp. 333–366. Springer, 2013.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, 07 2013. ISSN 2166-532X. doi: 10.1063/1.4812323. URL https://doi.org/10.1063/1.4812323.

Rui Jiao, Wenbing Huang, Peijia Lin, Jiaqi Han, Pin Chen, Yutong Lu, and Yang Liu. Crystal structure prediction by joint equivariant diffusion. *Advances in Neural Information Processing Systems*, 36:17464–17497, 2023.

Rui Jiao, Wenbing Huang, Yu Liu, Deli Zhao, and Yang Liu. Space group constrained crystal generation. *arXiv preprint arXiv:2402.03992*, 2024.

Nikita Kazeev, Wei Nong, Ignat Romanov, Ruiming Zhu, Andrey E Ustyuzhanin, Shuya Yamazaki, and Kedar Hippalgaonkar. Wyckoff transformer: Generation of symmetric crystals. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=eFHfRQRjJo.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Daniel Levy, Siba Smarak Panigrahi, Sékou-Oumar Kaba, Qiang Zhu, Kin Long Kelvin Lee, Mikhail Galkin, Santiago Miret, and Siamak Ravanbakhsh. SymmCD: Symmetry-preserving crystal generation with diffusion models. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=xnssGv9rpW.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Amil Merchant, Simon Batzner, Samuel S Schoenholz, Muratahan Aykol, Gowoon Cheon, and Ekin Dogus Cubuk. Scaling deep learning for materials discovery. *Nature*, 624(7990):80–85, 2023.

Benjamin Kurt Miller, Ricky TQ Chen, Anuroop Sriram, and Brandon M Wood. Flowmm: Generating materials with riemannian flow matching. *arXiv preprint arXiv:2406.04713*, 2024.

Ulrich Müller. Symmetry relationships between related crystal structures. In *Symmetry relations between related crystal structures*, pp. 156–178. Springer, 2011.

Franz Neumann and Oskar Emil Meyer. *Vorlesungen über die Theorie der Elasticität der festen Körper und des Lichtäthers*, volume 5. BG Teubner, 1885.

Andrey Okhotin, Maksim Nakhodnov, Nikita Kazeev, Andrey E Ustyuzhanin, and Dmitry Vetrov. Miad: Mirage atom diffusion for de novo crystal generation. *arXiv preprint arXiv:2511.14426*, 2025.

Bernt Øksendal. Stochastic differential equations. In *Stochastic differential equations: an introduction with applications*, pp. 38–50. Springer, 2003.

Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L Chevrier, Kristin A Persson, and Gerbrand Ceder. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013.

Omri Puny, Yaron Lipman, and Benjamin Kurt Miller. Space group conditional flow matching. *arXiv preprint arXiv:2509.23822*, 2025.

Janosh Riebesell, Rhys EA Goodall, Philipp Benner, Yuan Chiang, Bowen Deng, Alpha A Lee, Anubhav Jain, and Kristin A Persson. Matbench discovery–a framework to evaluate machine learning crystal stability predictions. *arXiv preprint arXiv:2308.14920*, 2023.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pp. 2256–2265. pmlr, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

Nathan J Szymanski, Bernardus Rendy, Yuxing Fei, Rishi E Kumar, Tanjin He, David Milsted, Matthew J McDermott, Max Gallant, Ekin Dogus Cubuk, Amil Merchant, et al. An autonomous laboratory for the accelerated synthesis of inorganic materials. *Nature*, 624(7990):86, 2023.

Darren J Wilkinson. *Stochastic modelling for systems biology*. Chapman and Hall/CRC, 2018.

Tian Xie, Xiang Fu, Octavian-Eugen Ganea, Regina Barzilay, and Tommi Jaakkola. Crystal diffusion variational autoencoder for periodic material generation. *arXiv preprint arXiv:2110.06197*, 2021.

A VISUALIZATION OF DISCOVERED S.U.N MATERIALS

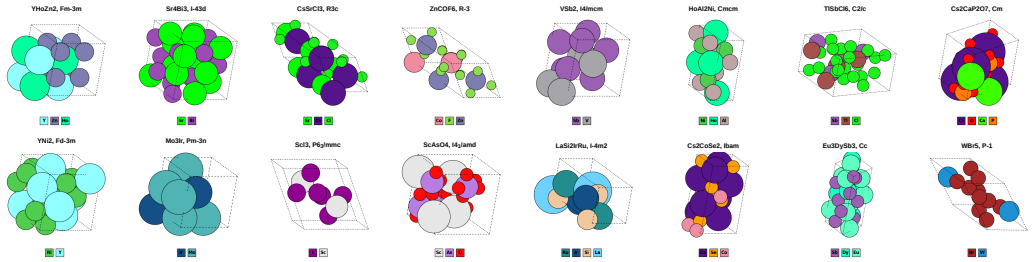


Figure 6: S.U.N. MP-20 materials discovered by SbCD.

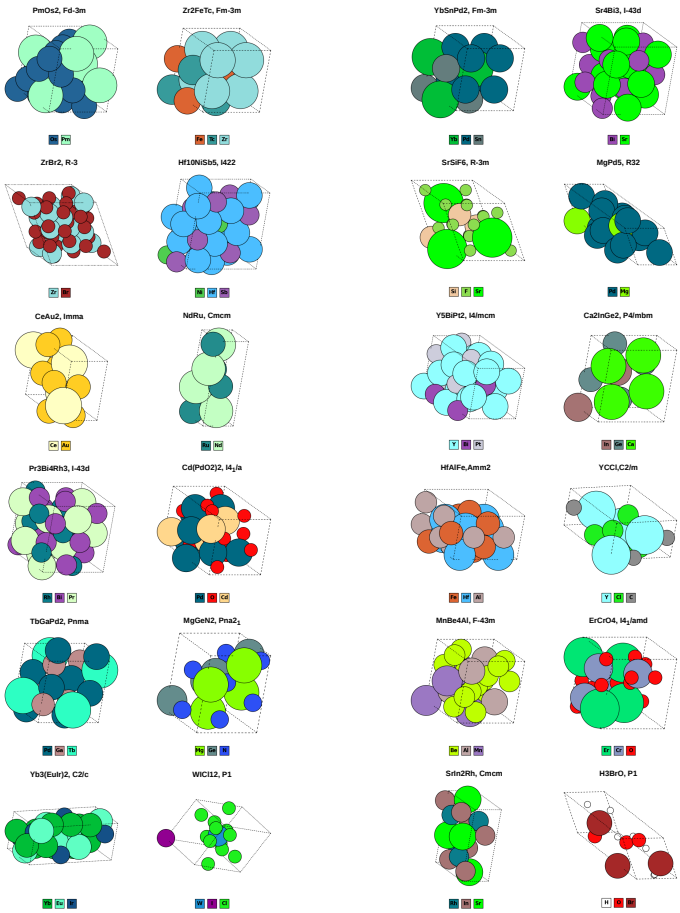


Figure 7: Visualization of S.U.N. crystalline materials on MP-20, including their chemical formulas and space group symmetries, discovered with SbCD variants: SbCD$^{\diamondsuit}$ (left) and SbCD$^{\sharp}$ (right).

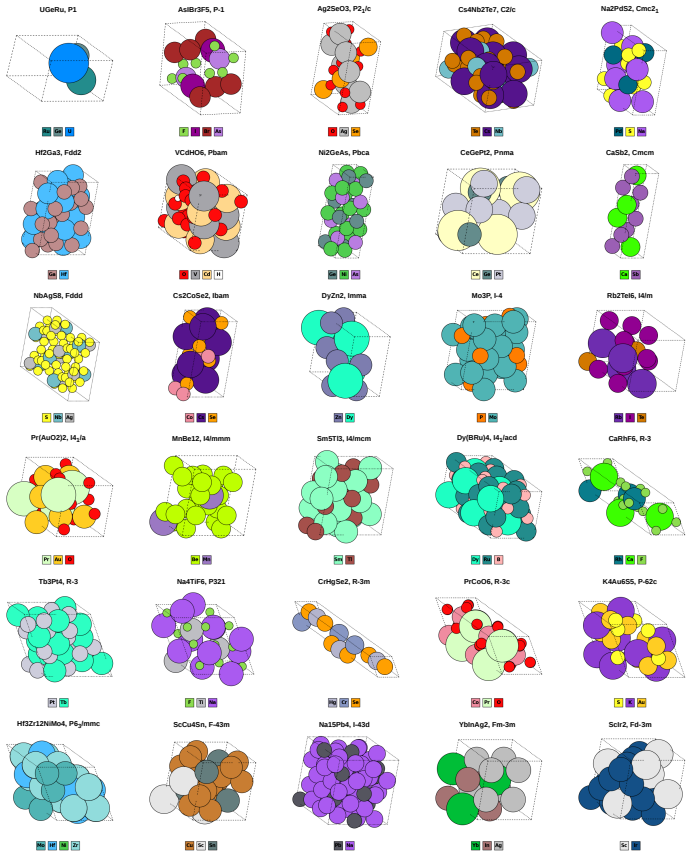


Figure 8: S.U.N. MPTS-52 materials discovered by SbCD-XL.

B FUTURE WORK

Dynamic Wyckoff symmetry constraints on asymmetric-unit fractional coordinates. To our knowledge, existing approaches that enforce such constraints typically require *fixed Wyckoff positions and space groups and operate on unit cells*. For example, DiffCSP++ (Jiao et al., 2024) uses fixed empirical Wyckoff positions and projects noisy fractional coordinates onto the corresponding Wyckoff subspaces, ensuring that the generated coordinates satisfy the prescribed symmetry constraints. SGFM (Puny et al., 2025) introduces space-group-conditional flow matching, applying frame averaging to unit-cell fractional coordinates using the predefined symmetry operations of a given space group. More recently, SGEquiDiff (Chang et al., 2025) adopts a multi-stage training strategy in which earlier stages sample Wyckoff positions, which are then held fixed to constrain fractional-coordinate generation in subsequent stages. In contrast, SbCD jointly evolves space groups, Wyckoff positions, and fractional coordinates on asymmetric units. This setting poses an additional challenge because the associated Wyckoff symmetry constraints can change at every diffusion step. Enforcing these dynamically varying constraints on asymmetric-unit fractional coordinates remains, to our knowledge, an open problem that we leave for future work.

Physically realistic symmetry-breaking transition pathways. SbCD currently models direct transitions from higher-symmetry space groups to the lowest-symmetry space group, P1. Future work could explore more physically realistic pathways by redesigning the instantaneous rate matrix $\mathbf{\Lambda}_t$ (Proposition 4.2) alongside adaptive lattice and site-symmetry constraints. This would allow structures to evolve through intermediate space groups toward the P1 prior, potentially capturing symmetry-breaking dynamics more faithfully. Even in its current form, our physics-motivated framework enables generative models to learn complete crystal symmetry data distributions, offers flexibility and improves performance in generating crystals under minimal symmetry assumptions.

C THEORETICAL PROOFS

C.1 PROPOSITION 4.1

Proposition. *Let $\mathbb{Q}$, $\hat{\mathbb{Q}}$, and $\mathbb{P}^\theta$ denote the forward, exact reverse, and learned reverse path measures of the continuous-time Markov chain. The negative log-likelihood on crystal structures $\mathcal{M}_0$ is bounded above by the expected Kullback–Leibler divergence over the reverse trajectory:*

$$-\log p_\theta(\mathcal{M}_0) \le \mathbb{E}_{q_{T|0}}\Big[D_{\mathrm{KL}}\big(\hat{\mathbb{Q}}^{\mathcal{M}_T\to\mathcal{M}_0} \,\big\|\, \mathbb{P}^{\theta,\mathcal{M}_T}\big)\Big] + \mathrm{Const.}$$

The divergence decomposes additively into space group, lattice, site symmetry, atom type, and fractional coordinate components:

$$D_{\mathrm{KL}}\big(\hat{\mathbb{Q}}^{\mathcal{M}_T\to\mathcal{M}_0} \,\big\|\, \mathbb{P}^{\theta,\mathcal{M}_T}\big) = \underbrace{D_{\mathrm{KL}}\big(\hat{\mathbb{Q}}_{\mathbf{A}} \,\big\|\, \mathbb{P}^{\theta}_{\mathbf{A}}\big) + D_{\mathrm{KL}}\big(\hat{\mathbb{Q}}_{\mathbf{F}} \,\big\|\, \mathbb{P}^{\theta}_{\mathbf{F}}\big)}_{\text{I: Atom Types \& Fractional Coordinates}}$$
$$+ \underbrace{D_{\mathrm{KL}}\big(\hat{\mathbb{Q}}_{G} \,\big\|\, \mathbb{P}^{\theta}_{G}\big)}_{\text{II: Space Group}} + \underbrace{\mathbb{E}_{\hat{\mathbb{Q}}_G}\Big[D_{\mathrm{KL}}\big(\hat{\mathbb{Q}}_{\mathbf{k}|G} \,\big\|\, \mathbb{P}^{\theta}_{\mathbf{k}|G}\big) + D_{\mathrm{KL}}\big(\hat{\mathbb{Q}}_{\mathbf{S}|G} \,\big\|\, \mathbb{P}^{\theta}_{\mathbf{S}|G}\big)\Big]}_{\text{III: Lattice \& Site Symmetry}}$$

Proof: Let $\mathbb{Q}$ denote the forward CTMC path measure, $\hat{\mathbb{Q}}$ the exact reverse path measure, and $\mathbb{P}^\theta$ the learned reverse path measure. Let $\mathcal{P} = C([0,T],\mathbb{M})$ denote the path space on $\mathbb{M}$. For an assymmetric-unit crystal representation $\mathcal{M} = \{G, \mathbf{A}, \mathbf{k}, \mathbf{S}, \mathbf{F}\}$, we denote by $\mathcal{M}_t$ the noisy structure at diffusion timestep t, and by W_s the reverse diffusion path. Following Campbell et al. (2022), we derive the negative log-likelihood of the model.

$$\begin{aligned}
-\log p_\theta(\mathcal{M}_0) &= -\log \int_{\mathbb{M}} p_{\text{ref}}(d\mathcal{M}_T) \int_{\mathcal{P}} \mathbf{1}_{\{W_T=\mathcal{M}_0\}} \mathbb{P}^{\theta,\mathcal{M}_T}(dw) \\
&= -\log \int_{\mathbb{M}} q_{T|0}(d\mathcal{M}_T|\mathcal{M}_0) \int_{\mathcal{P}} \mathbf{1}_{\{W_T=\mathcal{M}_0\}} \hat{\mathbb{Q}}^{\mathcal{M}_T}(dw) \frac{dp_{\text{ref}}}{dq_{T|0}(\cdot|\mathcal{M}_0)}(\mathcal{M}_T) \frac{d\mathbb{P}^{\theta,\mathcal{M}_T}}{d\hat{\mathbb{Q}}^{\mathcal{M}_T}}(w) \\
&= -\log \int_{\mathbb{M}} q_{T|0}(d\mathcal{M}_T|\mathcal{M}_0) \int_{\mathcal{P}} \hat{\mathbb{Q}}^{\mathcal{M}_T}(dw|W_T=\mathcal{M}_0) \frac{dp_{\text{ref}}}{dq_{T|0}(\cdot|\mathcal{M}_0)}(\mathcal{M}_T) \frac{d\mathbb{P}^{\theta,\mathcal{M}_T}}{d\hat{\mathbb{Q}}^{\mathcal{M}_T}}(w) \hat{\mathbb{Q}}^{\mathcal{M}_T}(\{W_T=\mathcal{M}_0\})
\end{aligned} \tag{5}$$

Where $\hat{\mathbb{Q}}^{\mathcal{M}_T}, \mathbb{P}^{\theta,\mathcal{M}_T}$ denote the exact and learned reverse path measures starting from $\mathcal{M}_T$, respectively. In the second equation, we apply a change of measure, and in the third equation, we apply a bridge decomposition of $\mathbf{1}_{\{W_T=\mathcal{M}_0\}}\hat{\mathbb{Q}}^{\mathcal{M}_T}(dw)$. Specifically, this term is decomposed into the bridge path measure, $\hat{\mathbb{Q}}^{\mathcal{M}_T}(dw|W_T=\mathcal{M}_0)$ and the probability $\hat{\mathbb{Q}}^{\mathcal{M}_T}(\{W_T=\mathcal{M}_0\})$, which represents a probability that a reverse path starting from $\mathcal{M}_T$ terminates at $\mathcal{M}_0$. We then define the joint probability measure $\mu(d\mathcal{M}_T, dw)$ and the integrand $f(\mathcal{M}_T, w)$ as follows:

$$\begin{aligned}
\mu(d\mathcal{M}_T, dw) &= q_{T|0}(d\mathcal{M}_t|\mathcal{M}_0)\hat{\mathbb{Q}}^{\mathcal{M}_T}(dw|W_T=\mathcal{M}_0) \\
f(\mathcal{M}_T, w) &= \frac{dp_{\text{ref}}}{dq_{T|0}(\cdot|\mathcal{M}_0)}(\mathcal{M}_T) \frac{d\mathbb{P}^{\theta,\mathcal{M}_T}}{d\hat{\mathbb{Q}}^{\mathcal{M}_T}}(w) \hat{\mathbb{Q}}^{\mathcal{M}_T}(\{W_T=\mathcal{M}_0\})
\end{aligned} \tag{6}$$

We can rewrite the negative log-likelihood:

$$-\log p_\theta(\mathcal{M}_0) = -\log \int f(\mathcal{M}_T, w)\, \mu(d\mathcal{M}_T, dw) \tag{7}$$

By Jensen's inequality ($-\log$ is convex, so $-\log \mathbb{E}[X] \le \mathbb{E}[-\log X]$)

$$-\log p_\theta(\mathcal{M}_0) \le \int -\log f(\mathcal{M}_T, w) \mu(d\mathcal{M}_T, dw) \tag{8}$$

where $-\log f(\mathcal{M}_T, w) = -\log \frac{dp_{\text{ref}}}{dq_{T|0}(\cdot|\mathcal{M}_0)}(\mathcal{M}_T) - \log \frac{d\mathbb{P}^{\theta,\mathcal{M}_T}}{d\hat{\mathbb{Q}}^{\mathcal{M}_T}}(w) - \log \hat{\mathbb{Q}}^{\mathcal{M}_T}(\{W_T=\mathcal{M}_0\})$

The last term is independent of the model parameter θ, and the first term does not depend on the path realization w. Therefore, taking the expectation with respect to μ yields

$$
\begin{aligned}
-\log p_\theta(\mathcal{M}_0) &\le \mathbb{E}_\mu\left[-\log \frac{d\mathbb{P}^{\theta,\mathcal{M}_T}}{d\widehat{\mathbb{Q}}^{\mathcal{M}_T}}(w)\right] + C_0 \\
&= \mathbb{E}_{q_{T|0}(\cdot|\mathcal{M}_0)}\left[D_{\mathrm{KL}}\Big(\widehat{\mathbb{Q}}^{\mathcal{M}_T\to\mathcal{M}_0} \,\|\, \mathbb{P}^{\theta,\mathcal{M}_T}\Big)\right] + C,
\end{aligned} \tag{9}
$$

where

$$
C_0 = \mathbb{E}_{q_{T|0}(\cdot|\mathcal{M}_0)}\left[-\log \widehat{\mathbb{Q}}^{\mathcal{M}_T}(\{W_T = \mathcal{M}_0\}) - \log \frac{dp_{\text{ref}}}{dq_{T|0}(\cdot|\mathcal{M}_0)}(\mathcal{M}_T)\right],
$$

After rewriting the bridge expectation as a KL divergence, the bridge-normalization term cancels with the corresponding term in C_0.

$$
C = \mathbb{E}_{q_{T|0}(\cdot|\mathcal{M}_0)}\left[-\log \frac{dp_{\text{ref}}}{dq_{T|0}(\cdot|\mathcal{M}_0)}(\mathcal{M}_T)\right].
$$

Finally, $\widehat{\mathbb{Q}}^{\mathcal{M}_T\to\mathcal{M}_0} := \widehat{\mathbb{Q}}^{\mathcal{M}_T}(\cdot|W_T = \mathcal{M}_0)$ denotes the bridge measure connecting $\mathcal{M}_T$ and $\mathcal{M}_0$.

We assume that the bridge measure factorizes according to the dependency structure of the crystal representation, with G governing $\mathbf{k}$ and $\mathbf{S}$, while $\mathbf{A}$ and $\mathbf{F}$ are modeled independently. Under the same factorization for the learned reverse measure, the KL chain rule yields the following decomposition.

$$
\begin{aligned}
\hat{\mathbb{Q}}^{\mathcal{M}_T\to\mathcal{M}_0}(dG, d\mathbf{A}, d\mathbf{k}, d\mathbf{S}, d\mathbf{F}) &= \hat{\mathbb{Q}}_G^{\mathcal{M}_T\to\mathcal{M}_0}(dG)\hat{\mathbb{Q}}_{\mathbf{k}|G}^{\mathcal{M}_T\to\mathcal{M}_0}(d\mathbf{k}|G)\cdot\hat{\mathbb{Q}}_{\mathbf{S}|G}^{\mathcal{M}_T\to\mathcal{M}_0}(d\mathbf{S}|G)\hat{\mathbb{Q}}_{\mathbf{A}}^{\mathcal{M}_T\to\mathcal{M}_0}(d\mathbf{A})\hat{\mathbb{Q}}_{\mathbf{F}}^{\mathcal{M}_T\to\mathcal{M}_0}(d\mathbf{F}) \\
\mathbb{P}^{\theta,\mathcal{M}_T}(dG, d\mathbf{A}, d\mathbf{k}, d\mathbf{S}, d\mathbf{F}) &= \mathbb{P}_G^{\theta,\mathcal{M}_T}(dG)\cdot\mathbb{P}_{\mathbf{k}|G}^{\theta,\mathcal{M}_T}(d\mathbf{k}|G)\mathbb{P}_{\mathbf{S}|G}^{\theta,\mathcal{M}_T}(d\mathbf{S}|G)\mathbb{P}_{\mathbf{A}}^{\theta,\mathcal{M}_T}(d\mathbf{A})\mathbb{P}_{\mathbf{F}}^{\theta,\mathcal{M}_T}(d\mathbf{F})
\end{aligned} \tag{10}
$$

We apply the KL chain rule that yields:

$$
\begin{aligned}
D_{\mathrm{KL}}\left(\hat{\mathbb{Q}}^{\mathcal{M}_T\to\mathcal{M}_0}\|\mathbb{P}^{\theta,\mathcal{M}_T}\right) &= D_{\mathrm{KL}}\left(\hat{\mathbb{Q}}_G^{\mathcal{M}_T\to\mathcal{M}_0}\|\mathbb{P}_G^{\theta,\mathcal{M}_T}\right) \\
&\quad + \mathbb{E}_{\hat{\mathbb{Q}}_G^{\mathcal{M}_T\to\mathcal{M}_0}}\left[D_{\mathrm{KL}}\left(\hat{\mathbb{Q}}_{\mathbf{k}|G}^{\mathcal{M}_T\to\mathcal{M}_0}\|\mathbb{P}_{\mathbf{k}|G}^{\theta,\mathcal{M}_T}\right) + D_{\mathrm{KL}}\left(\hat{\mathbb{Q}}_{\mathbf{S}|G}^{\mathcal{M}_T\to\mathcal{M}_0}\|\mathbb{P}_{\mathbf{S}|G}^{\theta,\mathcal{M}_T}\right)\right] \\
&\quad + D_{\mathrm{KL}}\left(\hat{\mathbb{Q}}_{\mathbf{A}}^{\mathcal{M}_T\to\mathcal{M}_0}\|\mathbb{P}_{\mathbf{A}}^{\theta,\mathcal{M}_T}\right) + D_{\mathrm{KL}}\left(\hat{\mathbb{Q}}_{\mathbf{F}}^{\mathcal{M}_T\to\mathcal{M}_0}\|\mathbb{P}_{\mathbf{F}}^{\theta,\mathcal{M}_T}\right)
\end{aligned} \tag{11}
$$

Now, we consider a discretized version of the continuous-time exact reverse bridge path measure on $\mathbf{A}, \mathbf{F}$:

$$
\begin{aligned}
\hat{\mathbb{Q}}_{\mathbf{A}}^{\mathcal{M}_T\to\mathcal{M}_0}(dw) &\Rightarrow q(\mathbf{A}_{t_1}, \mathbf{A}_{t_2}, \ldots, \mathbf{A}_{t_{N-1}}|\mathbf{A}_0, \mathbf{A}_T) \\
\hat{\mathbb{Q}}_{\mathbf{F}}^{\mathcal{M}_T\to\mathcal{M}_0}(dw) &\Rightarrow q(\mathbf{F}_{t_1}, \mathbf{F}_{t_2}, \ldots, \mathbf{F}_{t_{N-1}}|\mathbf{F}_0, \mathbf{F}_T)
\end{aligned} \tag{12}
$$

And, the learned reverse process continuous-time path measure can be discretized as:

$$
\begin{aligned}
\mathbb{P}_{\mathbf{A}}^{\theta,\mathcal{M}_T}(dw) &\Rightarrow p_\theta(\mathbf{A}_{t_{N-1}}, \ldots, \mathbf{A}_{t_0}|\mathcal{M}_T) = \prod_{j=1}^{N} p_\theta(\mathbf{A}_{t_{j-1}}|\mathcal{M}_{t_j}), \text{ where } \mathcal{M}_{\mathrm{t_j}} \text{ contains } \mathbf{A}_{\mathrm{t_j}} \\
\mathbb{P}_{\mathbf{F}}^{\theta,\mathcal{M}_T}(dw) &\Rightarrow p_\theta(\mathbf{F}_{t_{N-1}}, \ldots, \mathbf{F}_{t_0}|\mathcal{M}_T) = \prod_{j=1}^{N} p_\theta(\mathbf{F}_{t_{j-1}}|\mathcal{M}_{t_j}), \text{ where } \mathcal{M}_{\mathrm{t_j}} \text{ contains } \mathbf{F}_{\mathrm{t_j}}
\end{aligned} \tag{13}
$$

We obtain the discretized Radon–Nikodym derivative

$$
\begin{aligned}
\log \frac{d\hat{\mathbb{Q}}_{\mathbf{A}}^{\mathcal{M}_T\to\mathcal{M}_0}}{d\mathbb{P}_{\mathbf{A}}^{\theta,\mathcal{M}_T}}(w) &\Rightarrow \sum_{j=1}^{N} \log \frac{q(\mathbf{A}_{t_{j-1}}|\mathbf{A}_0, \mathbf{A}_{t_j})}{p_\theta(\mathbf{A}_{t_{j-1}}|\mathcal{M}_{t_j})} \\
\log \frac{d\hat{\mathbb{Q}}_{\mathbf{F}}^{\mathcal{M}_T\to\mathcal{M}_0}}{d\mathbb{P}_{\mathbf{F}}^{\theta,\mathcal{M}_T}}(w) &\Rightarrow \sum_{j=1}^{N} \log \frac{q(\mathbf{F}_{t_{j-1}}|\mathbf{F}_0, \mathbf{F}_{t_j})}{p_\theta(\mathbf{F}_{t_{j-1}}|\mathcal{M}_{t_j})}
\end{aligned} \tag{14}
$$

These discretized Radon–Nikodym derivatives recover the KL terms appearing in the discrete-time evidence upper bound of Sohl-Dickstein et al. (2015), which underlies diffusion models in continuous-state spaces (Ho et al., 2020) and discrete-state spaces (Austin et al., 2021). Similarly, we obtain the discretized Radon–Nikodym derivatives for $\mathbf{k}$ and $\mathbf{S}$:

$$\log \frac{d\hat{\mathbb{Q}}^{\mathcal{M}_T\to\mathcal{M}_0}_{\mathbf{k}|G}}{d\mathbb{P}^{\theta,\mathcal{M}_T}_{\mathbf{k}|G}}(w) \Rightarrow \sum_{j=1}^{N} \log \frac{q(\mathbf{k}_{t_{j-1}}|\mathbf{k}_0, \mathbf{k}_{t_j}, G_{t_j}, G_{t_{j-1}})}{p_\theta(\mathbf{k}_{t_{j-1}}|\mathcal{M}_{t_j}, G_{t_{j-1}})}, \text{ where } \mathcal{M}_{\mathrm{t_j}} \text{ contains } \mathrm{G_{t_j}} \text{ and } \mathbf{k}_{\mathrm{t_j}}$$

$$\log \frac{d\hat{\mathbb{Q}}^{\mathcal{M}_T\to\mathcal{M}_0}_{\mathbf{S}|G}}{d\mathbb{P}^{\theta,\mathcal{M}_T}_{\mathbf{S}|G}}(w) \Rightarrow \sum_{j=1}^{N} \log \frac{q(\mathbf{S}_{t_{j-1}}|\mathbf{S}_0, \mathbf{S}_{t_j}, G_{t_j}, G_{t_{j-1}})}{p_\theta(\mathbf{S}_{t_{j-1}}|\mathcal{M}_{t_j}, G_{t_{j-1}})}, \text{ where } \mathcal{M}_{\mathrm{t_j}} \text{ contains } \mathrm{G_{t_j}} \text{ and } \mathbf{S}_{\mathrm{t_j}}$$

Finally, the remaining term $D_{\mathrm{KL}}\left(\hat{\mathbb{Q}}^{\mathcal{M}_T\to\mathcal{M}_0}_G \| \mathbb{P}^{\theta,\mathcal{M}_T}_G\right)$ is optimized using CTMC Markov jump diffusion from Feller (1949); Campbell et al. (2022).

We conclude the proof.

C.2 Proposition 4.2

Proposition. *Consider the discrete state space of* 230 *crystallographic space groups drawn from the six crystal families. Let* $\mathbf{\Pi}, \mathbf{\Lambda}_t \in \mathbb{R}^{230\times 230}$ *denote the time-invariant transition probability matrix and the instantaneous rate matrix, respectively, defined element-wise as follows:*

$$\mathbf{\Pi}(\widetilde{G}, G) = \begin{cases} 0 & \widetilde{G} = \mathrm{P1} \\ 0 & G = \widetilde{G} \\ 1 & G = \mathrm{P1} \neq \widetilde{G} \\ 0 & \textit{otherwise}. \end{cases} \qquad \mathbf{\Lambda}_t(\widetilde{G}, G) = \begin{cases} 0 & \widetilde{G} = \mathrm{P1} \\ -\lambda(t) & G = \widetilde{G} \\ \lambda(t) & G = \mathrm{P1} \neq \widetilde{G} \\ 0 & \textit{otherwise}. \end{cases}$$

Then, $\mathbf{\Lambda}_t$ *and* $\mathbf{\Lambda}_s$ *commute for all* t, s*. Consequently, the forward transition probability of being in state* G *at time* t*, given state* $\widetilde{G}$ *at time* $\widetilde{t} < t$*,* $q_{t|\widetilde{t}}(G|\widetilde{G})$*, admits the matrix exponential form:*

$$q_{t|\widetilde{t}}(G|\widetilde{G}) = \left(\mathbf{D}\exp\left[\mathbf{\Sigma}\int_{\widetilde{t}}^{t}\lambda(s)ds\right]\mathbf{D}^{-1}\right)_{\widetilde{G},G} \quad \textit{with} \quad \frac{\mathbf{\Lambda}_t}{\lambda(t)} = \mathbf{D}\mathbf{\Sigma}\mathbf{D}^{-1}$$

Proof: We consider a continuous-time Markov process with time-dependent transition rates. Let $q_{t|\widetilde{t}}(G|\widetilde{G})$ denote the transition probability to be in the space group G at time t, if the process was in the space group $\widetilde{G}$ at the earlier time $\widetilde{t} < t$. We start from the Kolmogorov forward equation (Feller, 1949) in the jump-rate form:

$$\frac{\partial}{\partial t} q_{t|\widetilde{t}}(G|\widetilde{G}) = -\lambda_t(G) q_{t|\widetilde{t}}(G|\widetilde{G}) + \sum_{G'} \lambda_t(G')\mathbf{\Pi}_t(G', G) q_{t|\widetilde{t}}(G'|\widetilde{G}) \tag{15}$$

Here,

- $\lambda_t(G')$ is the total jump rate out of state G' at time t.
- $\mathbf{\Pi}_t(G', G)$ is the probability that, given a jump occurs from the space group G' at time t, the process jumps to the space group G (and $\Pi_t(G', G') = 0$).

We recall the *instantaneous* transition rate (generator) matrix $\mathbf{\Lambda}_t(G', G)$ with entries:

$$\mathbf{\Lambda}_t(G', G) = \lambda_t(G')\mathbf{\Pi}_t(G', G) - \delta_{G'G}\lambda_t(G') = \begin{cases} \lambda_t(G')\mathbf{\Pi}_t(G', G) & G \neq G' \\ -\lambda_t(G') & G = G' \end{cases} \tag{16}$$

Apply the generator matrix to the forward equation, we have:

$$\begin{aligned}\frac{\partial}{\partial t}q_{t|\widetilde{t}}(G|\widetilde{G}) &= -\lambda_t(G)q_{t|\widetilde{t}}(G|\widetilde{G}) + \sum_{G'}\lambda_t(G')\mathbf{\Pi}_t(G',G)q_{t|\widetilde{t}}(G'|\widetilde{G}) \\ &= \mathbf{\Lambda}_t(G,G)q_{t|\widetilde{t}}(G|\widetilde{G}) + \sum_{G'}\lambda_t(G')\mathbf{\Pi}_t(G',G)q_{t|\widetilde{t}}(G'|\widetilde{G}) \\ &= \mathbf{\Lambda}_t(G,G)q_{t|\widetilde{t}}(G|\widetilde{G}) + \sum_{G'\neq G}\mathbf{\Lambda}_t(G',G)q_{t|\widetilde{t}}(G'|\widetilde{G}) \\ &= \sum_{G'}\mathbf{\Lambda}_t(G',G)q_{t|\widetilde{t}}(G'|\widetilde{G}) = \left[q_{t|\widetilde{t}}\mathbf{\Lambda}_t\right]_{\widetilde{G},G} \end{aligned} \tag{17}$$

In the matrix form, the forward equation is equivalent to the matrix ODE:

$$\frac{\partial}{\partial t}q_{t|\widetilde{t}} = q_{t|\widetilde{t}}\mathbf{\Lambda}_t \tag{18}$$

Integrating both sides gives:

$$q_{t|\widetilde{t}} = I + \int_{\widetilde{t}}^{t} q_{t|\widetilde{t}}\mathbf{\Lambda}_s ds \tag{19}$$

Iterating the integral gives the Dyson series:

$$\begin{aligned} q_{t|\widetilde{t}} &= I + \int_{\widetilde{t}}^{t}\mathbf{\Lambda}_{s_1}ds_1 + \int_{\widetilde{t}}^{t}\int_{\widetilde{t}}^{s_1}\mathbf{\Lambda}_{s_1}\mathbf{\Lambda}_{s_2}ds_1ds_2 + \dots \\ &= I + \sum_{n=1}^{\infty}\int_{\widetilde{t}}^{t}\int_{\widetilde{t}}^{s_1}\cdots\int_{\widetilde{t}}^{s_{n-1}}\mathbf{\Lambda}_{s_n}\cdots\mathbf{\Lambda}_{s_1}ds_n\cdots ds_1 \end{aligned} \tag{20}$$

with the ordering $\widetilde{t} \le s_n \le \cdots \le s_1 \le t$. Now, assuming that the instantaneous transition rate matrix $\mathbf{\Lambda}_{s_i}$ commutes with $\mathbf{\Lambda}_{s_j}$ for all s_i, s_j, i.e., $\mathbf{\Lambda}_{s_i}\mathbf{\Lambda}_{s_j} = \mathbf{\Lambda}_{s_j}\mathbf{\Lambda}_{s_i}$. With this assumption, the integrand is symmetric in $(s_1, s_2, \cdots, s_n)$. The integration can be further simplified to the matrix exponential form:

$$\begin{aligned} q_{t|\widetilde{t}} &= I + \sum_{n=1}^{\infty}\frac{1}{n!}\left(\int_{\widetilde{t}}^{t}\mathbf{\Lambda}_s ds\right)^n \\ &= \exp\left(\int_{\widetilde{t}}^{t}\mathbf{\Lambda}_s ds\right) \end{aligned} \tag{21}$$

We now consider the space-group *relative* transition probability $\mathbf{\Pi}_t$ that is constant over time:

$$\mathbf{\Pi}_t(\widetilde{G},G) = \mathbf{\Pi}(\widetilde{G},G) = \begin{cases} 0 & \widetilde{G} = \mathrm{P1} \\ 0 & G = \widetilde{G} \\ 1 & G = \mathrm{P1} \neq \widetilde{G} \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

This design forces every jump to the lowest symmetry space group P1. The instantaneous rate matrix thus can be obtained by applying its definition with $\mathbf{\Pi}_t$ above:

$$\mathbf{\Lambda}_t(\widetilde{G},G) = \lambda(t)\mathbf{\Pi}^+(\widetilde{G},G) = \begin{cases} 0 & \widetilde{G} = \mathrm{P1} \\ -\lambda(t) & G = \widetilde{G} \\ \lambda(t) & G = \mathrm{P1} \neq \widetilde{G} \\ 0 & \text{otherwise.} \end{cases} \text{ with } \quad \mathbf{\Pi}^+(\widetilde{G},G) = \begin{cases} 0 & \widetilde{G} = \mathrm{P1} \\ -1 & G = \widetilde{G} \\ 1 & G = \mathrm{P1} \neq \widetilde{G} \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

We can easily verify the commutativity of the instantaneous rate matrix for all t, s:

$$\mathbf{\Lambda}_t\mathbf{\Lambda}_s = \lambda(t)\mathbf{\Pi}^+\lambda(s)\mathbf{\Pi}^+ = \lambda(t)\lambda(s)\mathbf{\Pi}^+\mathbf{\Pi}^+ = \mathbf{\Lambda}_s\mathbf{\Lambda}_t \tag{24}$$

Thus, we can apply Equation 21 to get the analytical form of space-group transition probabilities.

$$\begin{aligned} q_{t|\widetilde{t}} &= \exp\left(\int_{\widetilde{t}}^{t} \Lambda_s ds\right) \\ &= \exp\left(\mathbf{\Pi}^{+} \int_{\widetilde{t}}^{t} \lambda(s) ds\right) \\ &= \mathbf{D} \exp\left(\Sigma \int_{\widetilde{t}}^{t} \lambda(s) ds\right) \mathbf{D}^{-1} \end{aligned} \tag{25}$$

with the eigen-decomposition of $\mathbf{\Pi}^{+} = \mathbf{D}\Sigma\mathbf{D}^{-1}$

We conclude the proof.

C.3 COROLLARY 4.3

Corollary. *Let a crystal initially reside in a space group $\widetilde{G}$ and be observed in a lower-symmetry space group $G \prec \widetilde{G}$ at time t. Consider a constant transition rate, $\lambda(t) = \lambda\ \forall t$, then the posterior distribution of the holding time $\tau \in [0, t]$ is $\tau = -\frac{1}{\lambda}\log(1 - u(1 - e^{-\lambda t}))$, where $u \sim \mathcal{U}(0, 1)$. Consequently, the symmetry-breaking time window can be calculated as $w = t - \tau$.*

Proof: Consider a stochastic system with two states, $\widetilde{G}$ and G. The system is initialized in state $\widetilde{G}$ and is observed to be in state G at time t. The posterior distribution of the holding time τ (the transition event $\widetilde{G} \to G$) is derived via Bayes' theorem as follows:

$$p(\tau \mid G, \widetilde{G}) = \frac{P(G \mid \tau, \widetilde{G}) \cdot P(\tau \mid \widetilde{G})}{P(G \mid \widetilde{G})}. \tag{26}$$

The likelihood term equals one if the transition occurred before the observation time t:

$$P(G \mid \tau, \widetilde{G}) = \mathbf{1}[\tau < t]. \tag{27}$$

The standard survival distribution gives the prior density of the absorption time:

$$p(\tau) = P(\tau \mid \widetilde{G}) = \lambda(\tau) \exp\left(-\int_0^{\tau} \lambda(u)\, du\right). \tag{28}$$

The evidence (marginal likelihood) evaluates to:

$$\begin{aligned} P(G \mid \widetilde{G}) &= \int_0^{\infty} P(G \mid \tau)\, p(\tau \mid \widetilde{G})\, d\tau \\ &= \int_0^{t} 1 \cdot p(\tau)\, d\tau \\ &= \int_0^{t} \lambda(\tau) \exp\left(-\int_0^{\tau} \lambda(u)\, du\right) d\tau \\ &= 1 - \exp\left(-\int_0^{t} \lambda(u)\, du\right). \end{aligned} \tag{29}$$

Substituting these components, the posterior density is:

$$p(\tau \mid G, \widetilde{G}) = \frac{\mathbf{1}[\tau < t] \cdot \lambda(\tau) \exp\left(-\int_0^{\tau} \lambda(u)\, du\right)}{1 - \exp\left(-\int_0^{t} \lambda(u)\, du\right)}. \tag{30}$$

In the specific case of a constant transition rate $\lambda(t) = \lambda$, this simplifies to a truncated exponential distribution:

$$p(\tau \mid G, \widetilde{G}) = \frac{\lambda e^{-\lambda\tau}}{1 - e^{-\lambda t}}, \quad \tau \in [0, t]. \tag{31}$$

The corresponding cumulative distribution function (CDF) is:

$$F(\tau) = \frac{1 - e^{-\lambda\tau}}{1 - e^{-\lambda t}}. \tag{32}$$

By applying the inverse transform sampling method, samples of τ can be generated using:

$$\tau = -\frac{1}{\lambda} \log\left(1 - u(1 - e^{-\lambda t})\right), \quad u \sim \mathcal{U}(0, 1). \tag{33}$$

The symmetry-breaking time window is thus defined as $w = t - \tau$.

We conclude the proof.

C.4 Proposition 4.4

Proposition. *Consider a forward diffusion process with a symmetry-breaking event at time τ, transitioning from space group $\widetilde{G}$ to G. From Lemma 3.1, let $(\mathfrak{m}, \mathfrak{m}_\mathfrak{b})$ and $(\widetilde{\mathfrak{m}}, \widetilde{\mathfrak{m}}_\mathfrak{b})$ denote the mask–bias pairs associated with G and $\widetilde{G}$, respectively. Then, the continuous-state forward diffusion kernel applied to the lattice parameters $\mathbf{k}$ is given analytically by:*

$$\begin{aligned}\mathbf{k}_t = \widetilde{\mathfrak{m}} \odot \left(\sqrt{\bar{\alpha}_t}\,\mathbf{k}_0 + \sqrt{1 - \bar{\alpha}_t}\,\widetilde{\boldsymbol{\epsilon}}\right) + \widetilde{\mathfrak{m}}_\mathfrak{b} + \mathfrak{m} \odot \widetilde{\mathfrak{m}}_\mathfrak{b} \sqrt{\frac{\bar{\alpha}_t}{\bar{\alpha}_{t-w_\mathbf{k}}}} \\ - \left(\widetilde{\mathfrak{m}} \odot \sqrt{1 - \frac{\bar{\alpha}_t}{\bar{\alpha}_{t-w_\mathbf{k}}}}\,\widetilde{\boldsymbol{\epsilon}} + \widetilde{\mathfrak{m}}_\mathfrak{b}\right) + \left(\mathfrak{m} \odot \sqrt{1 - \frac{\bar{\alpha}_t}{\bar{\alpha}_{t-w_\mathbf{k}}}}\,\boldsymbol{\epsilon} + \mathfrak{m}_\mathfrak{b}\right),\end{aligned}$$

where $w_\mathbf{k} = t - \tau$. The corresponding reverse-time sampling step is simulated as:

$$\mathbf{k}_{t-1} = \widetilde{\mathfrak{m}} \odot \left(\frac{\sqrt{\bar{\alpha}_{t-1}}}{1 - \bar{\alpha}_t}(1 - \alpha_t)\,\hat{\mathbf{k}}_0^\theta(\mathcal{M}_t) + \mathfrak{m} \odot \frac{\sqrt{\alpha_t}}{1 - \bar{\alpha}_t}(1 - \bar{\alpha}_{t-1})\,\mathbf{k}_t + \mathfrak{m}_\mathfrak{b} + \sigma_t \boldsymbol{\epsilon}\right) + \widetilde{\mathfrak{m}}_\mathfrak{b}.$$

Proof: The set of symmetric basis matrices $\{\mathbf{B}_i\}_{i=1}^6$, derived by Jiao et al. (2024), is given as follows.

$$\mathbf{B}_1 = \begin{pmatrix}0 & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}, \quad \mathbf{B}_2 = \begin{pmatrix}0 & 0 & 1\\ 0 & 0 & 0\\ 1 & 0 & 0\end{pmatrix}, \quad \mathbf{B}_3 = \begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 1\\ 0 & 1 & 0\end{pmatrix},$$

$$\mathbf{B}_4 = \begin{pmatrix}1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 0\end{pmatrix}, \quad \mathbf{B}_5 = \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & -2\end{pmatrix}, \quad \mathbf{B}_6 = \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}.$$

We can cast the space-group-constrained diffusion process over lattice parameters proposed by Jiao et al. (2024) as a masking problem. Below, we describe the lattice masking mechanism applied to the six crystal families, ordered by increasing symmetry.

$$\begin{aligned}\text{i)}\ \text{Triclinic}: &\quad \mathfrak{m} = [1,1,1,1,1,1] && \mathfrak{m}_\mathfrak{b} = [0,0,0,0,0,0]\\ \text{ii)}\ \text{Monoclinic}: &\quad \mathfrak{m} = [0,1,0,1,1,1] && \mathfrak{m}_\mathfrak{b} = [0,0,0,0,0,0]\\ \text{iii)}\ \text{Orthorhombic}: &\quad \mathfrak{m} = [0,0,0,1,1,1] && \mathfrak{m}_\mathfrak{b} = [0,0,0,0,0,0]\\ \text{iv)}\ \text{Tetragonal}: &\quad \mathfrak{m} = [0,0,0,0,1,1] && \mathfrak{m}_\mathfrak{b} = [0,0,0,0,0,0]\\ \text{v)}\ \text{Hexagonal}: &\quad \mathfrak{m} = [0,0,0,0,1,1] && \mathfrak{m}_\mathfrak{b} = [-\log\tfrac{3}{4},0,0,0,0,0]\\ \text{vi)}\ \text{Cubic}: &\quad \mathfrak{m} = [0,0,0,0,0,1] && \mathfrak{m}_\mathfrak{b} = [0,0,0,0,0,0]\end{aligned} \tag{34}$$

Masking operation axioms Let $(\mathfrak{m}, \mathfrak{m}_b)$ and $(\widetilde{\mathfrak{m}}, \widetilde{\mathfrak{m}}_b)$ denote the mask-bias pairs associated with G and $\widetilde{G}$, respectively. Suppose that $\widetilde{G}$ belongs to a crystal family with higher symmetry than that of G, denoted by $G \prec \widetilde{G}$. Then the following properties hold:

$$\text{i)} \quad \widetilde{\mathfrak{m}} \odot \mathfrak{m} = \mathfrak{m} \odot \widetilde{\mathfrak{m}} = \widetilde{\mathfrak{m}} \qquad \text{ii)} \quad \mathfrak{m} \odot \mathfrak{m}_\mathfrak{b} = \widetilde{\mathfrak{m}} \odot \widetilde{\mathfrak{m}}_\mathfrak{b} = 0$$
$$\text{iii)} \quad C\mathfrak{m}_\mathfrak{b} = \mathfrak{m}_\mathfrak{b} \quad \text{for any scalar value } C \qquad \text{iv)} \quad \widetilde{\mathfrak{m}} \odot \mathfrak{m}_\mathfrak{b} = \mathfrak{m}_\mathfrak{b} \odot \widetilde{\mathfrak{m}} = 0 \tag{35}$$

Inductive proof At $t = \tau$, the diffusion process is in the space group $\widetilde{G}$. Applying Lemma 3.1, we have:

$$\mathbf{k}_\tau = \widetilde{\mathfrak{m}} \odot \left(\sqrt{\bar{\alpha}_\tau}\mathbf{k}_0 + \sqrt{1-\bar{\alpha}_\tau}\boldsymbol{\epsilon}_\tau\right) + \widetilde{\mathfrak{m}}_\mathfrak{b} \tag{36}$$

For $t > \tau$, the diffusion process jumps to the space group G. Considering $t = \tau + 1$, we have:

$$\begin{aligned}
\mathbf{k}_{\tau+1} &= \mathfrak{m} \odot \left(\sqrt{\alpha_{\tau+1}}\mathbf{k}_\tau + \sqrt{1-\alpha_{\tau+1}}\boldsymbol{\epsilon}_{\tau+1}\right) + \mathfrak{m}_b \\
&= \mathfrak{m} \odot \left[\sqrt{\alpha_{\tau+1}}\left(\widetilde{\mathfrak{m}} \odot \left(\sqrt{\bar{\alpha}_\tau}\mathbf{k}_0 + \sqrt{1-\bar{\alpha}_\tau}\boldsymbol{\epsilon}_\tau\right) + \widetilde{\mathfrak{m}}_\mathfrak{b}\right) + \sqrt{1-\alpha_{\tau+1}}\boldsymbol{\epsilon}_{\tau+1}\right] + \mathfrak{m}_b \\
&= \mathfrak{m} \odot \left[\widetilde{\mathfrak{m}} \odot \left(\sqrt{\bar{\alpha}_{\tau+1}}\mathbf{k}_0 + \sqrt{\alpha_{\tau+1}(1-\bar{\alpha}_\tau)}\boldsymbol{\epsilon}_\tau\right) + \widetilde{\mathfrak{m}}_\mathfrak{b}\sqrt{\alpha_{\tau+1}} + \sqrt{1-\alpha_{\tau+1}}\boldsymbol{\epsilon}_{\tau+1}\right] + \mathfrak{m}_b \\
&=^{(\star)} \widetilde{\mathfrak{m}} \odot \left(\sqrt{\bar{\alpha}_{\tau+1}}\mathbf{k}_0 + \sqrt{\alpha_{\tau+1}(1-\bar{\alpha}_\tau)}\boldsymbol{\epsilon}_\tau + \sqrt{1-\alpha_{\tau+1}}\boldsymbol{\epsilon}_\tau - \sqrt{1-\alpha_{\tau+1}}\boldsymbol{\epsilon}_\tau\right) \\
&\qquad + \mathfrak{m} \odot \widetilde{\mathfrak{m}}_\mathfrak{b}\sqrt{\alpha_{\tau+1}} + \mathfrak{m} \odot \sqrt{1-\alpha_{\tau+1}}\boldsymbol{\epsilon}_{\tau+1} + \mathfrak{m}_b \\
&=^{(\star\star)} \widetilde{\mathfrak{m}} \odot \left(\sqrt{\bar{\alpha}_{\tau+1}}\mathbf{k}_0 + \sqrt{1-\bar{\alpha}_{\tau+1}}\boldsymbol{\epsilon}_\tau\right) + \widetilde{\mathfrak{m}}_\mathfrak{b} - \widetilde{\mathfrak{m}}_\mathfrak{b} - \widetilde{\mathfrak{m}} \odot \sqrt{1-\alpha_{\tau+1}}\boldsymbol{\epsilon}_\tau \\
&\qquad + \mathfrak{m} \odot \widetilde{\mathfrak{m}}_\mathfrak{b}\sqrt{\alpha_{\tau+1}} + \mathfrak{m} \odot \sqrt{1-\alpha_{\tau+1}}\boldsymbol{\epsilon}_{\tau+1} + \mathfrak{m}_b \\
&= \widetilde{\mathfrak{m}} \odot \left(\sqrt{\bar{\alpha}_{\tau+1}}\mathbf{k}_0 + \sqrt{1-\bar{\alpha}_{\tau+1}}\boldsymbol{\epsilon}_\tau\right) + \widetilde{\mathfrak{m}}_\mathfrak{b} - \left(\widetilde{\mathfrak{m}} \odot \sqrt{1-\alpha_{\tau+1}}\boldsymbol{\epsilon}_\tau + \widetilde{\mathfrak{m}}_\mathfrak{b}\right) \\
&\qquad + \left(\mathfrak{m} \odot \sqrt{1-\alpha_{\tau+1}}\boldsymbol{\epsilon}_{\tau+1} + \mathfrak{m}_b\right) + \mathfrak{m} \odot \widetilde{\mathfrak{m}}_\mathfrak{b}\sqrt{\alpha_{\tau+1}} \\
&= \widetilde{\mathfrak{m}} \odot \left(\sqrt{\bar{\alpha}_{\tau+1}}\mathbf{k}_0 + \sqrt{1-\bar{\alpha}_{\tau+1}}\boldsymbol{\epsilon}_\tau\right) + \widetilde{\mathfrak{m}}_\mathfrak{b} - \left(\widetilde{\mathfrak{m}} \odot \sqrt{1-\frac{\bar{\alpha}_{\tau+1}}{\bar{\alpha}_\tau}}\boldsymbol{\epsilon}_\tau + \widetilde{\mathfrak{m}}_\mathfrak{b}\right) \\
&\qquad + \left(\mathfrak{m} \odot \sqrt{1-\frac{\bar{\alpha}_{\tau+1}}{\bar{\alpha}_\tau}}\boldsymbol{\epsilon}_{\tau+1} + \mathfrak{m}_b\right) + \mathfrak{m} \odot \widetilde{\mathfrak{m}}_\mathfrak{b}\sqrt{\frac{\bar{\alpha}_{\tau+1}}{\bar{\alpha}_\tau}}
\end{aligned} \tag{37}$$

In $(\star)$, we use $\mathfrak{m} \odot \widetilde{\mathfrak{m}} = \widetilde{\mathfrak{m}}$. In $(\star\star)$, we use the fact that the sum of two independent Gaussian random variables with variances $\alpha_{\tau+1}(1-\bar{\alpha}_\tau)$ and $1-\alpha_{\tau+1}$ is Gaussian with variance $1-\bar{\alpha}_{\tau+1}$.

Similarly, for $t = \tau + 2$, we can estimate $\mathbf{k}_{\tau+2}$ as follows:

$$
\begin{aligned}
\mathbf{k}_{\tau+2} &= \mathfrak{m} \odot \left(\sqrt{\alpha_{\tau+2}}\mathbf{k}_{\tau+1} + \sqrt{1-\alpha_{\tau+2}}\boldsymbol{\epsilon}_{\tau+2}\right) + \mathfrak{m}_b \\
&=^{(\dagger)} \mathfrak{m} \odot \Big[\sqrt{\alpha_{\tau+2}}\widetilde{\mathfrak{m}} \odot \left(\sqrt{\bar{\alpha}_{\tau+1}}\mathbf{k}_0 + \sqrt{1-\bar{\alpha}_{\tau+1}}\boldsymbol{\epsilon}_\tau\right) \\
&\quad - \sqrt{\alpha_{\tau+2}}\widetilde{\mathfrak{m}} \odot \sqrt{1-\alpha_{\tau+1}}\boldsymbol{\epsilon}_\tau + \sqrt{\alpha_{\tau+2}}\mathfrak{m} \odot \widetilde{\mathfrak{m}}_{\mathfrak{b}}\sqrt{\alpha_{\tau+1}} + \sqrt{\alpha_{\tau+2}}\mathfrak{m} \odot \sqrt{1-\alpha_{\tau+1}}\boldsymbol{\epsilon}_{\tau+1} \\
&\quad + \sqrt{\alpha_{\tau+2}}\mathfrak{m}_b + \sqrt{1-\alpha_{\tau+2}}\boldsymbol{\epsilon}_{\tau+2}\Big] + \mathfrak{m}_b \\
&=^{(\dagger\dagger)} \widetilde{\mathfrak{m}} \odot \left(\sqrt{\bar{\alpha}_{\tau+2}}\mathbf{k}_0 + \sqrt{\alpha_{\tau+2}(1-\bar{\alpha}_{\tau+1})}\boldsymbol{\epsilon}_\tau\right) - \widetilde{\mathfrak{m}} \odot \sqrt{\alpha_{\tau+2}(1-\alpha_{\tau+1})}\boldsymbol{\epsilon}_\tau \\
&\quad + \mathfrak{m} \odot \sqrt{\alpha_{\tau+2}(1-\alpha_{\tau+1})}\boldsymbol{\epsilon}_{\tau+1} + \mathfrak{m} \odot \widetilde{\mathfrak{m}}_{\mathfrak{b}}\sqrt{\alpha_{\tau+1}\alpha_{\tau+2}} + \mathfrak{m} \odot \sqrt{1-\alpha_{\tau+2}}\boldsymbol{\epsilon}_{\tau+2} + \mathfrak{m}_b \\
&= \widetilde{\mathfrak{m}} \odot \left(\sqrt{\bar{\alpha}_{\tau+2}}\mathbf{k}_0 + \sqrt{\alpha_{\tau+2}(1-\bar{\alpha}_{\tau+1})}\boldsymbol{\epsilon}_\tau\right) + \widetilde{\mathfrak{m}} \odot \sqrt{1-\alpha_{t+2}}\boldsymbol{\epsilon}_\tau \\
&\quad - \widetilde{\mathfrak{m}} \odot \sqrt{1-\alpha_{t+2}}\boldsymbol{\epsilon}_\tau - \widetilde{\mathfrak{m}} \odot \sqrt{\alpha_{\tau+2}(1-\alpha_{\tau+1})}\boldsymbol{\epsilon}_\tau \\
&\quad + \mathfrak{m} \odot \sqrt{\alpha_{\tau+2}(1-\alpha_{\tau+1})}\boldsymbol{\epsilon}_{\tau+1} + \mathfrak{m} \odot \sqrt{1-\alpha_{\tau+2}}\boldsymbol{\epsilon}_{\tau+2} \\
&\quad + \mathfrak{m} \odot \widetilde{\mathfrak{m}}_{\mathfrak{b}}\sqrt{\alpha_{\tau+1}\alpha_{\tau+2}} + \mathfrak{m}_b \\
&=^{(\dagger\dagger\dagger)} \widetilde{\mathfrak{m}} \odot \left(\sqrt{\bar{\alpha}_{\tau+2}}\mathbf{k}_0 + \sqrt{(1-\bar{\alpha}_{\tau+2})}\boldsymbol{\epsilon}_\tau\right) \\
&\quad - \widetilde{\mathfrak{m}} \odot \sqrt{(1-\alpha_{\tau+1}\alpha_{\tau+2})}\boldsymbol{\epsilon}_\tau \\
&\quad + \mathfrak{m} \odot \sqrt{1-\alpha_{\tau+1}\alpha_{\tau+2}}\boldsymbol{\epsilon}_{\tau+2} \\
&\quad + \mathfrak{m} \odot \widetilde{\mathfrak{m}}_{\mathfrak{b}}\sqrt{\alpha_{\tau+1}\alpha_{\tau+2}} + \mathfrak{m}_b \\
&= \widetilde{\mathfrak{m}} \odot \left(\sqrt{\bar{\alpha}_{\tau+2}}\mathbf{k}_0 + \sqrt{1-\bar{\alpha}_{\tau+2}}\boldsymbol{\epsilon}_\tau\right) + \widetilde{\mathfrak{m}}_{\mathfrak{b}} - \left(\widetilde{\mathfrak{m}} \odot \sqrt{1-\alpha_{\tau+1}\alpha_{\tau+2}}\boldsymbol{\epsilon}_\tau + \widetilde{\mathfrak{m}}_{\mathfrak{b}}\right) \\
&\quad + \left(\mathfrak{m} \odot \sqrt{1-\alpha_{\tau+1}\alpha_{\tau+2}}\boldsymbol{\epsilon}_{\tau+2} + \mathfrak{m}_b\right) + \mathfrak{m} \odot \widetilde{\mathfrak{m}}_{\mathfrak{b}}\sqrt{\alpha_{\tau+1}\alpha_{\tau+2}} \\
&= \widetilde{\mathfrak{m}} \odot \left(\sqrt{\bar{\alpha}_{\tau+2}}\mathbf{k}_0 + \sqrt{1-\bar{\alpha}_{\tau+2}}\boldsymbol{\epsilon}_\tau\right) + \widetilde{\mathfrak{m}}_{\mathfrak{b}} - \left(\widetilde{\mathfrak{m}} \odot \sqrt{1-\frac{\bar{\alpha}_{\tau+2}}{\bar{\alpha}_\tau}}\boldsymbol{\epsilon}_\tau + \widetilde{\mathfrak{m}}_{\mathfrak{b}}\right) \\
&\quad + \left(\mathfrak{m} \odot \sqrt{1-\frac{\bar{\alpha}_{\tau+2}}{\bar{\alpha}_\tau}}\boldsymbol{\epsilon}_{\tau+2} + \mathfrak{m}_b\right) + \mathfrak{m} \odot \widetilde{\mathfrak{m}}_{\mathfrak{b}}\sqrt{\frac{\bar{\alpha}_{\tau+2}}{\bar{\alpha}_\tau}}
\end{aligned} \tag{38}
$$

In $(\dagger)$, we substitute the expression for $\mathbf{k}_{\tau+1}$ obtained in step $(\star)$. In $(\dagger\dagger)$, we use $\mathfrak{m} \odot \widetilde{\mathfrak{m}} = \widetilde{\mathfrak{m}}$ and $\mathfrak{m} \odot \mathfrak{m}_{\mathfrak{b}} = 0$. In $(\dagger\dagger\dagger)$, we use the fact that the sum of two independent Gaussian random variables with variances $\alpha_{\tau+2}(1-\bar{\alpha}_{\tau+1})$ and $1-\alpha_{\tau+2}$ is Gaussian with variance $1-\bar{\alpha}_{\tau+2}$, and that the sum of two independent Gaussian random variables with variances $\alpha_{\tau+2}(1-\alpha_{\tau+1})$ and $1-\alpha_{\tau+2}$ is Gaussian with variance $1-\alpha_{\tau+1}\alpha_{\tau+2}$. For a general case $t > \tau$, and $w = t - \tau$, we have:

$$
\begin{aligned}
\mathbf{k}_t = \widetilde{\mathfrak{m}} \odot \left(\sqrt{\bar{\alpha}_t}\mathbf{k}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_\tau\right) + \widetilde{\mathfrak{m}}_{\mathfrak{b}} - \left(\widetilde{\mathfrak{m}} \odot \sqrt{1-\frac{\bar{\alpha}_t}{\bar{\alpha}_{t-w}}}\boldsymbol{\epsilon}_\tau + \widetilde{\mathfrak{m}}_{\mathfrak{b}}\right) \\
+ \left(\mathfrak{m} \odot \sqrt{1-\frac{\bar{\alpha}_t}{\bar{\alpha}_{t-w}}}\boldsymbol{\epsilon}_t + \mathfrak{m}_b\right) + \mathfrak{m} \odot \widetilde{\mathfrak{m}}_{\mathfrak{b}}\sqrt{\frac{\bar{\alpha}_t}{\bar{\alpha}_{t-w}}}
\end{aligned} \tag{39}
$$

Sampling is performed using DDIM samplers from Song et al. (2020). The reverse process proceeds toward increasing symmetry levels, from G to $\widetilde{G}$.

$$
\begin{aligned}
\mathbf{k}_{t-1} &= \widetilde{\mathfrak{m}} \odot \left(\frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_t}(1-\alpha_t)\,\hat{\mathbf{k}}_0^\theta(\mathcal{M}_t) + (\mathfrak{m} \odot \mathbf{k}_t + \mathfrak{m}_{\mathfrak{b}})\frac{\sqrt{\alpha_t}}{1-\bar{\alpha}_t}(1-\bar{\alpha}_{t-1}) + \sigma_t\boldsymbol{\epsilon}\right) + \widetilde{\mathfrak{m}}_{\mathfrak{b}} \\
&= \widetilde{\mathfrak{m}} \odot \left(\frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_t}(1-\alpha_t)\,\hat{\mathbf{k}}_0^\theta(\mathcal{M}_t) + \mathfrak{m} \odot \frac{\sqrt{\alpha_t}}{1-\bar{\alpha}_t}(1-\bar{\alpha}_{t-1})\,\mathbf{k}_t + \mathfrak{m}_{\mathfrak{b}} + \sigma_t\boldsymbol{\epsilon}\right) + \widetilde{\mathfrak{m}}_{\mathfrak{b}}
\end{aligned} \tag{40}
$$

We apply $C\mathfrak{m}_{\mathfrak{b}} = \mathfrak{m}_{\mathfrak{b}}$ for any scalar value C,i.e., the biases remain constant across crystal families, and $\hat{\mathbf{k}}_0^\theta(\mathcal{M}_t)$ denotes the data-prediction parameterization of a neural network.

We conclude the proof.

C.5 PROPOSITION 4.5

Proposition. *Let $\widetilde{\mathfrak{g}}, \mathfrak{g} \in \Delta^{81-1}$ be the marginal site-symmetry priors associated with the space groups $\widetilde{G}$ and G, respectively. Let $\bar{\alpha}_t$ be a strictly decreasing noise schedule with symmetry-breaking time τ satisfying $0 < \bar{\alpha}_t, \bar{\alpha}_\tau \leq 1$. For $t > \tau$, define*

$$\boldsymbol{\pi}_t = \nu_t \widetilde{\mathfrak{g}} + (1-\nu_t)\mathfrak{g} = \frac{\bar{\alpha}_t(1-\bar{\alpha}_{t-w_\mathbf{S}})\widetilde{\mathfrak{g}} + (\bar{\alpha}_{t-w_\mathbf{S}} - \bar{\alpha}_t)\mathfrak{g}}{\bar{\alpha}_{t-w_\mathbf{S}}(1-\bar{\alpha}_t)} \quad \text{where } w_\mathbf{S} = t-\tau,\ \nu_t \in [0,1]$$

Then, $\boldsymbol{\pi}_t$ lies in the convex hull of the two priors, $\boldsymbol{\pi}_\tau = \widetilde{\mathfrak{g}},\ \lim_{t\to T}\boldsymbol{\pi}_t = \mathfrak{g}$, and the trajectory $\{\boldsymbol{\pi}_t\}_{t>\tau}$ moves monotonically along the straight line segment connecting $\widetilde{\mathfrak{g}}$ to $\mathfrak{g}$ inside the simplex.

Proof: We introduce a *non-stationary prior mixture process* that defines *a forward process* which changes its target distribution midway through corruption. Let denote $\widetilde{\mathfrak{g}}$ and $\mathfrak{g}$ the marginal site symmetry prior of space group $\widetilde{G}$ and G, respectively. We still assume that $\widetilde{G}$ is a higher-symmetry space group than G, i.e., $G \prec \widetilde{G}$, and recall that τ is the symmetry-breaking event; we rewrite the forward categorical diffusion from Austin et al. (2021) as follows.

$$q\left(\mathbf{S}_t|\mathbf{S}_0, \tau, G, \widetilde{G}\right) = \mathrm{Cat}\left(\mathbf{S}_t; \mathbf{S}_0 \bar{\mathbf{Q}}_{t,\tau}^{\widetilde{G}\to G}\right) \tag{41}$$

We have the following transition matrices within a space group.

$$\bar{Q}_{\tau\leftarrow 0}^{\widetilde{G}} = \bar{\alpha}_\tau I + (1-\bar{\alpha}_\tau)\mathbf{1}(\widetilde{\mathfrak{g}})^T \tag{42}$$

$$\bar{Q}_{t\leftarrow \tau}^{G} = \frac{\bar{\alpha}_t}{\bar{\alpha}_\tau} I + (1-\frac{\bar{\alpha}_t}{\bar{\alpha}_\tau})\mathbf{1}(\mathfrak{g})^T \tag{43}$$

Then,

$$\begin{aligned}
\bar{\mathbf{Q}}_t^{\widetilde{G}\to G} = \bar{Q}_{\tau\leftarrow 0}^{\widetilde{G}} \bar{Q}_{t\leftarrow\tau}^{G} &= \left(\bar{\alpha}_\tau I + (1-\bar{\alpha}_\tau)\mathbf{1}(\widetilde{\mathfrak{g}})^T\right)\left(\frac{\bar{\alpha}_t}{\bar{\alpha}_\tau} I + (1-\frac{\bar{\alpha}_t}{\bar{\alpha}_\tau})\mathbf{1}(\mathfrak{g})^T\right) \\
&= \bar{\alpha}_t I + (\bar{\alpha}_\tau - \bar{\alpha}_t)\mathbf{1}(\mathfrak{g})^T + (1-\bar{\alpha}_\tau)\frac{\bar{\alpha}_t}{\bar{\alpha}_\tau}\mathbf{1}(\widetilde{\mathfrak{g}})^T \\
&\quad + (1-\bar{\alpha}_\tau)(1-\frac{\bar{\alpha}_t}{\bar{\alpha}_\tau})\mathbf{1}(\mathfrak{g})^T, \qquad \text{as} \quad (\widetilde{\mathfrak{g}})^T\mathbf{1} = 1 \\
&= \bar{\alpha}_t I + \frac{1}{\bar{\alpha}_\tau}\left[\bar{\alpha}_t(1-\bar{\alpha}_\tau)\mathbf{1}(\widetilde{\mathfrak{g}})^T + (\bar{\alpha}_\tau - \bar{\alpha}_t)\mathbf{1}(\mathfrak{g})^T\right] \\
&= \bar{\alpha}_t I + (1-\bar{\alpha}_t)\mathbf{1}(\boldsymbol{\pi}_t)^T \\
\text{where} \qquad \boldsymbol{\pi}_t &= \frac{\bar{\alpha}_t(1-\bar{\alpha}_\tau)\widetilde{\mathfrak{g}} + (\bar{\alpha}_\tau - \bar{\alpha}_t)\mathfrak{g}}{\bar{\alpha}_\tau(1-\bar{\alpha}_t)}
\end{aligned} \tag{44}$$

$\boldsymbol{\pi}_t$ is a symmetry-breaking (SB) prior that smoothly interpolates between two given site-symmetry priors.

Convexity

We rewrite the SB prior as a convex hull of the two priors

$$\boldsymbol{\pi}_t = \nu_t\widetilde{\mathfrak{g}} + (1-\nu_t)\mathfrak{g}, \qquad \text{with} \quad 0 \leq \nu_t = \frac{\bar{\alpha}_t(1-\bar{\alpha}_\tau)}{\bar{\alpha}_\tau(1-\bar{\alpha}_t)} \leq 1 \tag{45}$$

The convexity property ensures that $\boldsymbol{\pi}_t$ is a valid categorical distribution that lies within the probability simplex. The forward kernel thus remains row-stochastic and requires no renormalization trick.

Boundary consistency

For $t \leq \tau$, there is no transition between priors, i.e. $\mathfrak{g} = \widetilde{\mathfrak{g}}$, thus $\boldsymbol{\pi}_{t\leq\tau} = \widetilde{\mathfrak{g}}$. And for $t \to T, \bar{\alpha}_t \to 0, \nu_t \to 0$, which leads to $\lim_{t\to T}\boldsymbol{\pi}_t = \mathfrak{g}$. And, the path connects two priors continuously.

Monotone geometric path

By differentiating $\boldsymbol{\pi}_t$ w.r.t time t, we have

$$\frac{d\boldsymbol{\pi}_t}{dt} = \frac{d\nu_t}{dt}(\widetilde{\mathfrak{g}} - \mathfrak{g}) \tag{46}$$

Since $\bar{\alpha}_t$ is decreasing, ν_t is decreasing, thus $\frac{d\nu_t}{dt} < 0$ and the path moves monotonically toward $\mathfrak{g}$.

We conclude the proof.

C.6 COROLLARY 4.6

Corollary. *Let the discrete-state diffusion transition matrices of Austin et al. (2021) be redefined under the symmetry-breaking formulation as:*

$$\mathbf{Q}_t = \alpha_t I + (1-\alpha_t)\mathbf{1}(\mathfrak{g})^T, \qquad \bar{\mathbf{Q}}_{t-1} = \bar{\alpha}_{t-1} I + (1-\bar{\alpha}_{t-1})\mathbf{1}(\widetilde{\mathfrak{g}})^T$$

Then, the forward and reverse sampling processes on the site symmetry $\mathbf{S}$ *can be derived as follows:*

$$\mathbf{S}_t \sim \mathrm{Cat}\left(\mathbf{S}_t; \mathbf{S}_0 \bar{\mathbf{Q}}_t^{\widetilde{G}\to G}\right) \qquad \mathbf{S}_{t-1} \sim \mathrm{Cat}\left(\mathbf{S}_{t-1}; \frac{\mathbf{S}_t \mathbf{Q}_t^T \odot \hat{\mathbf{S}}_0^\theta(\mathcal{M}_t)\bar{\mathbf{Q}}_{t-1}}{\hat{\mathbf{S}}_0^\theta(\mathcal{M}_t)\bar{\mathbf{Q}}_t^{\widetilde{G}\to G}\mathbf{S}_t^T}\right)$$

Proof: The forward noise kernel follows from the proof in Appendix C.5, which we recall below:

$$\begin{aligned}
&q\left(\mathbf{S}_t|\mathbf{S}_0, \tau, G, \widetilde{G}\right) = \mathrm{Cat}\left(\mathbf{S}_t; \mathbf{S}_0 \bar{\mathbf{Q}}_{t,\tau}^{\widetilde{G}\to G}\right) \\
&\text{where } \bar{\mathbf{Q}}_t^{\widetilde{G}\to G} = \bar{\alpha}_t I + (1-\bar{\alpha}_t)\mathbf{1}(\boldsymbol{\pi}_t)^T \\
&\text{and } \boldsymbol{\pi}_t = \frac{\bar{\alpha}_t(1-\bar{\alpha}_\tau)\widetilde{\mathfrak{g}} + (\bar{\alpha}_\tau - \bar{\alpha}_t)\mathfrak{g}}{\bar{\alpha}_\tau(1-\bar{\alpha}_t)} = \nu_t\widetilde{\mathfrak{g}} + (1-\nu_t)\mathfrak{g}, \text{ with } \quad 0 \le \nu_t = \frac{\bar{\alpha}_t(1-\bar{\alpha}_\tau)}{\bar{\alpha}_\tau(1-\bar{\alpha}_t)} \le 1
\end{aligned} \tag{47}$$

We rewrite the reverse sampling kernel of Austin et al. (2021) under the symmetry-breaking formulation, in which the process proceeds toward higher symmetry orders in crystals. Specifically, we assume that the process jumps from space group $G_t = G$ to space group $G_{t-1} = \widetilde{G}$, where $G \prec \widetilde{G}$. The true posterior can be written as

$$q(\mathbf{S}_{t-1} \mid \mathbf{S}_t, \mathbf{S}_0, G_{t-1}, G_t) = \frac{q(\mathbf{S}_t \mid \mathbf{S}_{t-1}, G_t) q(\mathbf{S}_{t-1} \mid \mathbf{S}_0, G_{t-1})}{q(\mathbf{S}_t \mid \mathbf{S}_0, G_t)}. \tag{48}$$

We then parameterize the model posterior as follows:

$$p_\theta(\mathbf{S}_{t-1} \mid \mathcal{M}_t, G_{t-1}) = \frac{q(\mathbf{S}_t \mid \mathbf{S}_{t-1}, G_t)\tilde{p}_\theta(\mathbf{S}_{t-1} \mid \mathcal{M}_t, G_{t-1})}{\sum_{\mathbf{S}_{t-1}} q(\mathbf{S}_t \mid \mathbf{S}_{t-1}, G_t)\tilde{p}_\theta(\mathbf{S}_{t-1} \mid \mathcal{M}_t, G_{t-1})}. \tag{49}$$

This defines a valid distribution for any non-negative, unnormalized prior $\tilde{p}_\theta$ over $\mathbf{S}_{t-1}$. The true posterior is recovered by setting

$$\tilde{p}_\theta(\cdot \mid \mathcal{M}_t, G_{t-1}) = q(\cdot \mid \mathbf{S}_0, G_{t-1}). \tag{50}$$

Since $q(\mathbf{S}_{t-1} \mid \mathbf{S}_0, G_{t-1})$ is a closed-form forward kernel, we parameterize the model to predict $\mathbf{S}_0$, denoted by $\hat{\mathbf{S}}_0^\theta(\mathcal{M}_t)$. Then,

$$\tilde{p}_\theta(\cdot \mid \mathcal{M}_t, G_{t-1}) = q(\cdot \mid \hat{\mathbf{S}}_0^\theta(\mathcal{M}_t), G_{t-1}). \tag{51}$$

Thus, we can parameterize the model posterior as:

$$p_\theta(\mathbf{S}_{t-1} \mid \mathcal{M}_t, G_{t-1}) = \frac{q(\mathbf{S}_t \mid \mathbf{S}_{t-1}, G_t) q(\mathbf{S}_{t-1} \mid \hat{\mathbf{S}}_0^\theta(\mathcal{M}_t), G_{t-1})}{\sum_{\mathbf{S}_{t-1}} q(\mathbf{S}_t \mid \mathbf{S}_{t-1}, G_t) q(\mathbf{S}_{t-1} \mid \hat{\mathbf{S}}_0^\theta(\mathcal{M}_t), G_{t-1})}. \tag{52}$$

Alternatively, we can rewrite the model posterior using the corresponding space groups defined above.

$$p_\theta(\mathbf{S}_{t-1} \mid \mathcal{M}_t, \widetilde{G}) = \frac{q(\mathbf{S}_t \mid \mathbf{S}_{t-1}, G) q(\mathbf{S}_{t-1} \mid \hat{\mathbf{S}}_0^\theta(\mathcal{M}_t), \widetilde{G})}{\sum_{\mathbf{S}_{t-1}} q(\mathbf{S}_t \mid \mathbf{S}_{t-1}, G) q(\mathbf{S}_{t-1} \mid \hat{\mathbf{S}}_0^\theta(\mathcal{M}_t), \widetilde{G})}. \tag{53}$$

We now analyze each component:

(i)

$$\begin{aligned} &q(\mathbf{S}_t \mid \mathbf{S}_{t-1}, G), \text{ equivalent to. } \mathbf{S}_t \sim \mathrm{Cat}(\mathbf{S}_t; \mathbf{S}_{t-1}\mathbf{Q}_t) \Rightarrow \mathbf{S}_{t-1} \sim \mathrm{Cat}(\mathbf{S}_{t-1}; \mathbf{S}_t \mathbf{Q}_t^T) \\ &\text{where} \quad \mathbf{Q}_t = \alpha_t I + (1-\alpha_t)\mathbf{1}(\mathfrak{g})^T \end{aligned} \tag{54}$$

(ii)

$$\begin{aligned} &q(\mathbf{S}_{t-1} \mid \hat{\mathbf{S}}_0^\theta(\mathcal{M}_t), \widetilde{G}) \text{ equivalent to. } \mathbf{S}_{t-1} \sim \mathrm{Cat}(\mathbf{S}_{t-1}; \hat{\mathbf{S}}_0^\theta(\mathcal{M}_t)\bar{\mathbf{Q}}_{t-1}) \\ &\text{where} \quad \bar{\mathbf{Q}}_{t-1} = \bar{\alpha}_t I + (1-\bar{\alpha}_t)\mathbf{1}(\widetilde{\mathfrak{g}})^T \end{aligned} \tag{55}$$

(iii) The scalar normalization term.

$$\begin{aligned} &\sum_{\mathbf{S}_{t-1}} q(\mathbf{S}_t \mid \mathbf{S}_{t-1}, G) q(\mathbf{S}_{t-1} \mid \hat{\mathbf{S}}_0^\theta(\mathcal{M}_t), \widetilde{G}) \\ &\text{with} \quad q(\mathbf{S}_t \mid \mathbf{S}_{t-1}, G) \Rightarrow \mathbf{S}_{t-1} \sim \mathrm{Cat}(\mathbf{S}_{t-1}; \mathbf{S}_t \mathbf{Q}_t^T) \\ &\text{and} \quad q(\mathbf{S}_{t-1} \mid \hat{\mathbf{S}}_0^\theta(\mathcal{M}_t), \widetilde{G}) \Rightarrow \mathbf{S}_{t-1} \sim \mathrm{Cat}(\mathbf{S}_{t-1}; \hat{\mathbf{S}}_0^\theta(\mathcal{M}_t)\bar{\mathbf{Q}}_{t-1}) \\ &\text{the outer sum over } \mathbf{S}_{t-1} \text{ can be written as an inner product as follows:} \\ &\quad \hat{\mathbf{S}}_0^\theta(\mathcal{M}_t)\bar{\mathbf{Q}}_{t-1}(\mathbf{S}_t\mathbf{Q}_t^T)^T = \hat{\mathbf{S}}_0^\theta(\mathcal{M}_t)\bar{\mathbf{Q}}_{t-1}\mathbf{Q}_t\mathbf{S}_t^T = \hat{\mathbf{S}}_0^\theta(\mathcal{M}_t)\bar{\mathbf{Q}}_t^{\widetilde{G}\to G}\mathbf{S}_t^T \end{aligned} \tag{56}$$

From (i), (ii), (iii), we obtain the reverse sampling process on the site symmetry $\mathbf{S}$ as:

$$\mathbf{S}_{t-1} \sim \mathrm{Cat}\left(\mathbf{S}_{t-1}; \frac{\mathbf{S}_t\mathbf{Q}_t^T \odot \hat{\mathbf{S}}_0^\theta(\mathcal{M}_t)\bar{\mathbf{Q}}_{t-1}}{\hat{\mathbf{S}}_0^\theta(\mathcal{M}_t)\bar{\mathbf{Q}}_t^{\widetilde{G}\to G}\mathbf{S}_t^T}\right) \tag{57}$$

We conclude the proof.

D Experimental setup

D.1 Oriented site-symmetry symbols

We provide the list of oriented site-symmetry symbols used for one-hot encoding of site symmetries.

```
{-3.:  0, ..m:  1, 6/mm2/m:  2, m.mm:  3, ..2:  4, 4/mm.m:  5,
.m.:  6, .-3m:  7, 32.:  8, mm2..:  9,  -6mm2m:  10, -3..:  11,
2.22:  12, mmm..:  13, 3..:  14, m.m2:  15, mmm.:  16, -43m:  17,
6/m..:  18, 432:  19, .2.:  20, m..:  21, 4..:  22, 2mm.:  23,
4/mmm:  24, m2.:  25, 23.:  26, 222:  27, m.2m:  28, m-3m:  29,
2m.:  30, 3mm:  31, -3m2/m:  32, .3m:  33, 222.:  34, -32/m:  35,
3m.:  36, 222..:  37, 622:  38, -4m2:  39, 22.:  40, -4..:  41, 2:
42, -4m.2:  43, mmm:  44, 3m:  45, .m:  46, -622m2:  47, 2/m2/m.:
48, 32:  49, ..2/m:  50, 2mm:  51, -6..:  52, 4m.m:  53, 2.mm:
54, .32:  55, 6mm:  56, m:  57, 422:  58, mm2:  59, 4mm:  60,
1:  61, -42m:  62, m-3.:  63, 2/m..:  64, 4/m..:  65, -42.m:  66,
42.2:  67, m2m.:  68, mm.:  69, 6..:  70, 2/m:  71, -32/m.:  72,
.2/m.:  73, .-3.:  74, 322:  75, .3.:  76, 3.:  77, 2..:  78, m2m:
79, -1:  80}
```

D.2 STABLE, UNIQUENESS, AND NOVELTY (S.U.N) METRICS

Generated crystalline materials should ideally be thermodynamically stable. To assess stability, we calculate their energies and compare them against competing phases on a convex hull constructed from the lowest-energy materials at each composition. Following prior work (Levy et al., 2025; Miller et al., 2024), we utilize a pretrained CHGNet model (Deng et al., 2023) to estimate crystal energies and use the convex hull derived from the Materials Project (Riebesell et al., 2023). Among the stable generated structures, we compute the number of unique stable structures and assess their novelty with respect to the training dataset. For uniqueness and novelty evaluation, we utilize $\mathrm{StructureMatcher}$ (Ong et al., 2013) to determine structural similarity.

D.3 WYCKOFF POSITION RETRIEVAL

Since the proposed site-symmetry representation directly maps to a valid point group, we do not need to project the generated site symmetries onto the closest point group, unlike in Levy et al. (2025). However, multiple Wyckoff positions may share the same underlying oriented site-symmetry symbol, even though they correspond to different sets of symmetry-equivalent sites. As a result, the oriented site-symmetry symbol alone is not sufficient to uniquely determine the appropriate Wyckoff position for a generated fractional coordinate. We therefore compare the generated coordinate against all candidate Wyckoff positions of a given space group with the matching oriented site symmetry and select the one whose symmetry-equivalent sites lie closest to the generated coordinate. In practice, this nearest-site search is performed using the $\mathrm{search_closest_wp}$ function from PyXtal (Fredericks et al., 2021), and distances are evaluated under periodic boundary conditions in fractional-coordinate space.

D.4 ARCHITECTURE

We adopt the equivariant graph neural network-based denoiser from DiffCSP/++ (Jiao et al., 2023; 2024). For site-symmetry prediction, SbCD$^\sharp$ follows the approach from Levy et al. (2025), which predicts the probabilities of 13 symmetry elements over each of 15 possible axes, since both employ the same site-symmetry representation. In contrast, SbCD and SbCD$^\diamondsuit$ modify the existing architecture to predict probabilities over 81 possible oriented site-symmetry symbols only. For space groups, we also use a one-hot encoding over 230 discrete space groups instead of the complex space-group encoding scheme proposed by Levy et al. (2025).

Table 4: Hyperparameters used in the main experiments of Table 2 on MP-20.

Hyperparameter	SbCD	SbCD$^\sharp$	SbCD$^\diamondsuit$
numLayer	6	8	6
hiddenDim	1024	1024	512
embedDim	512	512	256
learnRate	1×10^{-3}	1×10^{-3}	1×10^{-3}
optim	Adam	Adam	Adam
scheduler	ReduceLROnPlateau	ReduceLROnPlateau	ReduceLROnPlateau
batchSize	256	256	256
numEpoch	3000	3000	3000
λ	2×10^{-3}	1.5×10^{-3}	1.5×10^{-3}
$l_{\mathbf{F}}, l_{\mathbf{k}}, l_{\mathbf{A}}, l_{\mathbf{S}}, l_G, l_{\text{jump}}$	$\{3, 10, 1, 1, 4, 1\}$	$\{3, 10, 1, 12, 4, 1\}$	$\{3, 10, 1, 1, 4, 1\}$

D.5 HYPERPARAMETER SEARCH

We conduct a hyperparameter search over architectural parameters, including the number of layers, $\mathrm{numLayer} \in \{6, 8\}$, the hidden dimension, $\mathrm{hiddenDim} \in \{512, 1024\}$, and the symmetry embedding dimension (for site symmetry and space groups), $\mathrm{embedDim} \in \{256, 512\}$. Moreover, we ablate the learning rate, $\mathrm{learnRate} \in \{5 \times 10^{-4}, 1 \times 10^{-3}\}$, and the transition rate, $\lambda \in \{0.8 \times 10^{-3}, 1 \times 10^{-3}, 1.5 \times 10^{-3}, 2 \times 10^{-3}\}$. For a fair comparison, we set the diffusion schedules (Austin et al., 2021; Ho et al., 2020) for $\mathbf{k}, \mathbf{S}, \mathbf{A}$, and $\mathbf{F}$ similarly to Levy et al. (2025). Tables 4 and 5 summarize the configuration used in the main experiments.

Table 5: Hyperparameters used in the main experiments of Table 3 on MPTS-52.

Hyperparameter	SbCD-S	SbCD-L	SbCD-XL
numLayer	6	6	8
hiddenDim	512	1024	1024
embedDim	256	512	512
learnRate	1×10^{-3}	1×10^{-3}	1×10^{-3}
optim	Adam	Adam	Adam
scheduler	ReduceLROnPlateau	ReduceLROnPlateau	ReduceLROnPlateau
batchSize	256	256	256
numEpoch	2000	2000	2000
λ	2×10^{-3}	1.5×10^{-3}	1.5×10^{-3}
$l_{\mathbf{F}}, l_{\mathbf{k}}, l_{\mathbf{A}}, l_{\mathbf{S}}, l_G, l_{\text{jump}}$	$\{3, 10, 1, 1, 4, 1\}$	$\{3, 10, 1, 10, 4, 1\}$	$\{3, 10, 1, 1, 4, 1\}$

Table 6: Comparison of memory usage, training time per epoch, and architecture size between SbCD variants at a batch size of 256 on an NVIDIA GeForce RTX 3060 with 20 CPU cores.

Method	Memory(MB)(↓)	Training time per epoch (s)(↓)	Number of parameters (M)(↓)
SbCD$^\sharp$	7731	81	61.5
SbCD$^\diamondsuit$	4323	34	14.8
SbCD	6775	62	47.9

For the SymmCD‡ baselines reported in Tables 2 and 3, we retrained the models using the hyperparameters and experimental settings recommended by Levy et al. (2025), as no official pretrained model weights are publicly available.

D.6 Computational analysis

SbCD variants utilize the asymmetric-unit structure as Levy et al. (2025), which is notably memory efficient for crystal generation. Moreover, SbCD and SbCD$^\diamondsuit$ employ a simplified site-symmetry representation, further reducing memory overhead and improving training efficiency. Table 6 compares the computational analysis of these models at a batch size of 256 on an NVIDIA GeForce RTX 3060.

D.7 Training and sampling algorithms

We provide the training and sampling procedures of SbCD in Algorithms 2 and 1, respectively. Notably, SbCD initiates the sampling process from the minimal symmetry prior, namely the space group P1 with its trivial site symmetry 1.

Algorithm 1 Sampling of SbCD

Require: Trained model θ.

$G_T \leftarrow$ P1 ▷ Lowest symmetry space group

$\mathbf{S}_T \leftarrow \mathbf{1}$ ▷ Trivial site symmetry of P1

$\mathbf{k}_T \sim \mathcal{N}(0, I), \quad \mathbf{F}_T \sim \mathcal{U}(0, I), \quad \mathbf{A}_T \sim p_{\text{marginal}}(\text{A})$

for t = T to 1 **do**

$G_{t-1} = G_t + \sum_{i=1}^{230} \mathbf{P}_i \left(G_i - G_t\right)$ with $\mathbf{P}_i = \text{Poisson}\left(\hat{\Lambda}_t^\theta(G_t, G_i)\right)$ ▷ Lemma 2.2

$/ *$ Proposition 4.4 $* /$

$\mathbf{k}_{t-1} \leftarrow \mathfrak{m} \odot \left(\frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_t}(1-\alpha_t)\hat{\mathbf{k}}_0^\theta(\mathcal{M}_t) + \mathfrak{m} \odot \frac{\sqrt{\alpha_t}}{1-\bar{\alpha}_t}(1-\bar{\alpha}_{t-1})\mathbf{k}_t + \mathfrak{m}_\mathfrak{b} + \sigma_t \boldsymbol{\epsilon}\right) + \mathfrak{m}_\mathfrak{b}$

$\mathbf{S}_{t-1} \sim \text{Cat}\left(\mathbf{S}_{t-1}; \frac{\mathbf{S}_t \mathbf{Q}_t^T \odot \hat{\mathbf{S}}_0^\theta(\mathcal{M}_t)\bar{\mathbf{Q}}_{t-1}}{\hat{\mathbf{S}}_0^\theta(\mathcal{M}_t)\bar{\mathbf{Q}}_t^{G_{t-1}\to G_t}\mathbf{S}_t^T}\right)$ with $\mathbf{Q}_t, \bar{\mathbf{Q}}_{t-1}, \bar{\mathbf{Q}}_t^{G_{t-1}\to G_t}$ in Corollary 4.6

$\mathbf{F}_{t-1} \leftarrow$ follow Jiao et al. (2023), $\mathbf{A}_{t-1} \leftarrow$ follow Austin et al. (2021); Levy et al. (2025)

end for

return Generated crystal unit cell: $\mathbf{k}_0$, G_0, and replicating $\mathbf{F}_0$, $\mathbf{A}_0$ via $\mathbf{S}_0$.

E ADDITIONAL RESULTS

E.1 SYMMETRY-BREAKING TIME WINDOW

Table 7 presents the full results of the ablation study on the symmetry-breaking time window, partially illustrated in Figure 5. As noted, SbCD variants with a fixed time window achieve high structural validity, indicating that such a schedule effectively preserves physically plausible atomic configurations. However, their property statistics and S.U.N. scores are significantly worse than the results in Table 2, obtained using the theoretically derived posterior symmetry-breaking time window. This suggests that while a fixed schedule improves structural validity, the adaptive posterior schedule provides a better trade-off between structural plausibility and overall generation quality.

Table 7: Ablation results with symmetry-breaking time window $w = 1$.

Method	**Validity** (%)(↑)		**Coverage** (%)(↑)		**Property statistics** (↓)		**CHGNet** $E^{\text{hull}} < 0$ (%)(↑)	
	Struct.	Comp.	Recall	Precision	d_ρ	# elem.	Stable	S.U.N
SbCD	94.27	81.48	99.26	99.01	1.66	0.18	7.38	5.25
SbCD$^\sharp$	98.13	84.00	93.38	98.89	1.3	0.21	7.44	4.74

E.2 STATISTICAL SIGNIFICANCE

Due to the high computational cost of validation, particularly for evaluating the stability of sampled crystals, we note that several baselines, including FlowMM, DiffCSP, DiffCSP++, CDVAE, and SGFM, report results from a single evaluation over 10000 samples. Under this sufficiently large sample size, we evaluate SbCD using three independent sampling runs with 10000 samples per run. Across these evaluations in Table 8, we observe that the performance of SbCD remains consistent, and the reported results are robust to sampling variance.

Table 8: Statistical robustness over three independent runs on the MP-20 dataset.

Method	**Validity** (%)(↑)		**Coverage** (%)(↑)		**Property statistics** (↓)			**CHGNet** $E^{\text{hull}} < 0$ (%)(↑)	
	Struct.	Comp.	Recall	Precision	d_ρ	# elem.	d_{sg}	Stable	S.U.N
SbCD	89.81 (±0.02)	90.44 (±0.21)	99.48 (±0.07)	98.18 (±0.46)	0.1910 (±0.0123)	0.0395 (±0.0086)	0.3168 (±0.0114)	11.273 (±0.012)	7.207 (±0.098)

Figure 3 clarification. Due to the high computational cost of structure relaxation and energy-above-hull evaluation, we conduct the ablation study on 3,000 generated crystalline materials.

E.3 NUMBER OF FUNCTION EVALUATIONS (NFES)

We compare SbCD and SymmCD under different numbers of function evaluations (NFEs) during sampling on the MP-20 benchmark. Since pretrained SymmCD weights are not publicly available, we retrain SymmCD using the training and evaluation protocol of Levy et al. (2025) to ensure a fair comparison. We evaluate structural validity (Struct.), compositional validity (Comp.), and their product (Validity) over 3,000 generated crystals. The results are summarized in Figure 4, with the corresponding numerical values reported in Tables 9, 10, and 11. We observe that, in the low-NFE regime, SbCD generates significantly more valid crystal structures than its symmetry-preserving counterpart, SymmCD. This demonstrates the benefit of learning space groups as part of crystal generative modeling and suggests the potential of SbCD for efficient few-step crystal generation.

Table 9: Structure Validity

Method	**NFEs**				
	50	100	250	500	1000
SymmCD‡	31.83	63.13	80.53	80.93	87.27
SbCD	85.00	87.17	89.73	91.67	89.63

Table 10: Composition Validity

Method	NFEs				
	50	100	250	500	1000
SymmCD‡	44.93	70.30	82.83	84.70	85.40
SbCD	76.74	78.33	81.43	90.60	89.73

Table 11: Validity

Method	NFEs				
	50	100	250	500	1000
SymmCD‡	25.03	49.97	66.63	68.67	74.53
SbCD	65.23	68.77	73.73	83.43	80.30

E.4 Analysis of space-group distributions

In Figure 2, we observe that DiffCSP++ and SymmCD directly condition on space groups sampled from the training data distribution, which naturally encourages generated samples to have space-group distributions close to the data distribution. SbCD variants achieve comparable values d_{sg} to DiffCSP++ and SymmCD, while outperforming DiffCSP, FlowMM, and CDVAE, which do not explicitly model or utilize crystal symmetries during inference. Furthermore, Figures 10 and 9 show that SbCD variants generate structures spanning a broad range of space groups observed in the data distribution. We also provide some examples of SbCD-generated crystal structures across different space groups in Figures 6, 7, and 8.

We additionally evaluate SbCD's performance on the top 10 most common space groups among generated structures in Table 12. Interestingly, SbCD$^\sharp$, which adopts the site-symmetry representation from Levy et al. (2025), achieves higher structural validity on the top 10 most common space groups compared with its average validity across all space groups. In contrast, SbCD$^\diamond$ and SbCD, which use our proposed site-symmetry representation, generate more valid structures across less frequent space groups, which compensates for their slightly lower structural validity on the most common space groups. Most property metrics worsen for the top 10 most common space groups, due to the limited coverage of these space groups in the data distribution. Nevertheless, we observe that structures generated from the most common space groups tend to exhibit higher stability.

Table 12: De novo generation results averaged across the 10 most representative space groups in crystalline materials generated on the MP-20 dataset.

Method	**Validity** (%)(↑)		**Coverage** (%)(↑)		**Property statistics** (↓)			**CHGNet** $E^{\text{hull}} < 0$ (%)(↑)	
	Struct.	Comp.	Recall	Precision	d_ρ	# elem.	d_{sg}	Stable	S.U.N
SbCD$^\sharp$	89.81	88.32	99.51	97.31	0.09	0.04	0.22	10.74	7.31
SbCD$^\sharp$ (10 SGs)	97.98	89.46	96.97	99.45	0.13	0.045	0.51	12.84	9.60
SbCD$^\diamond$	88.25	87.47	99.44	97.71	0.16	0.07	0.28	10.23	7.67
SbCD$^\diamond$ (10 SGs)	83.78	88.46	94.66	96.63	0.37	0.12	0.55	10.33	8.17
SbCD	89.80	90.32	99.44	97.91	0.18	0.03	0.32	11.28	7.15
SbCD (10 SGs)	86.60	91.02	97.71	98.51	0.23	0.042	0.54	11.76	7.79

E.5 Histogram of the energy above the convex hull E_{hull}

Figures 11 and 12 visualize the distributions of E_{hull} computed after relaxation with CHGNET (Deng et al., 2023) for different methods. As shown, SbCD performs competitively, generating a significantly higher proportion of structures that are thermodynamically stable ($E_{\text{hull}} < 0$) or metastable ($E_{\text{hull}} < 0.1$).

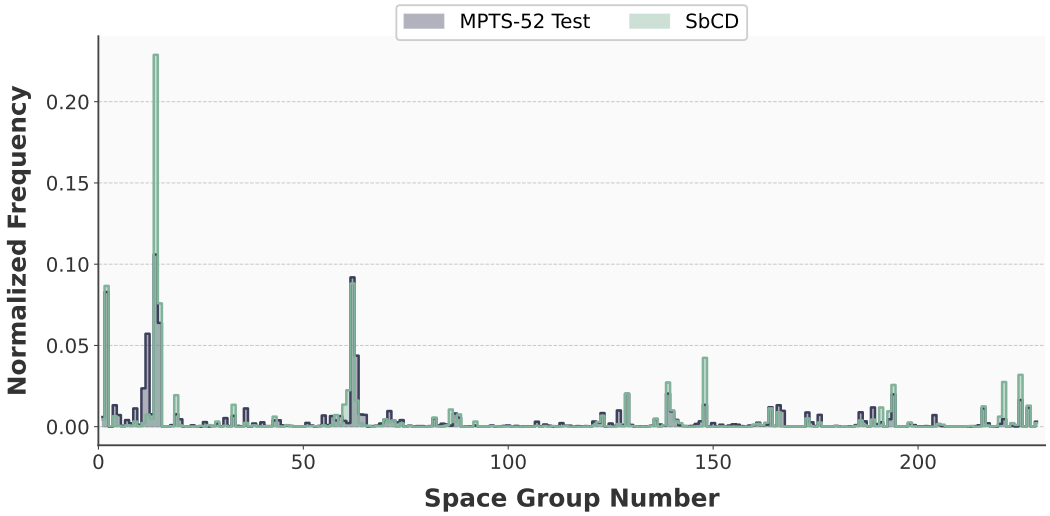


Figure 9: Space group distributions of the MPTS-52 test set and crystals sampled by SbCD.

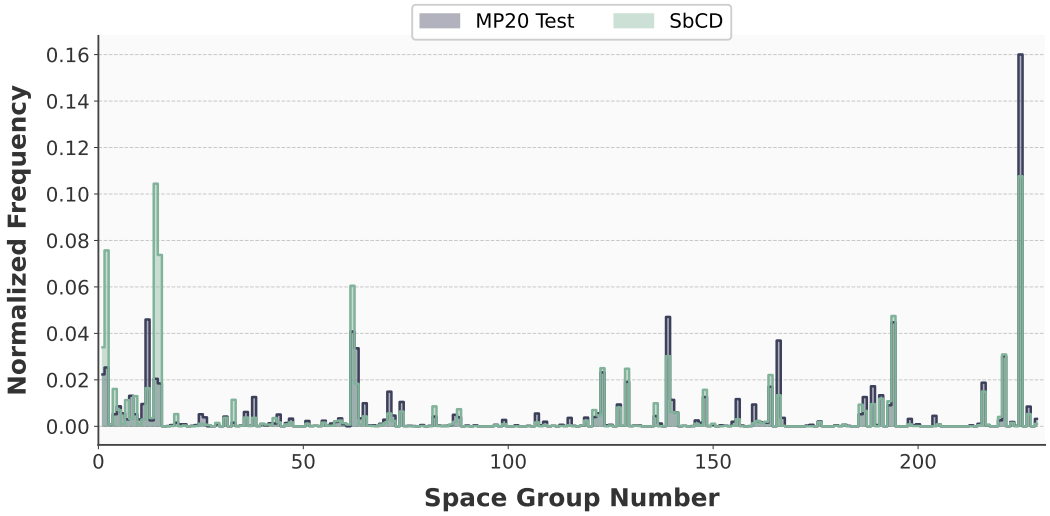


Figure 10: Space group distributions of the MP-20 test set and crystals sampled by SbCD.

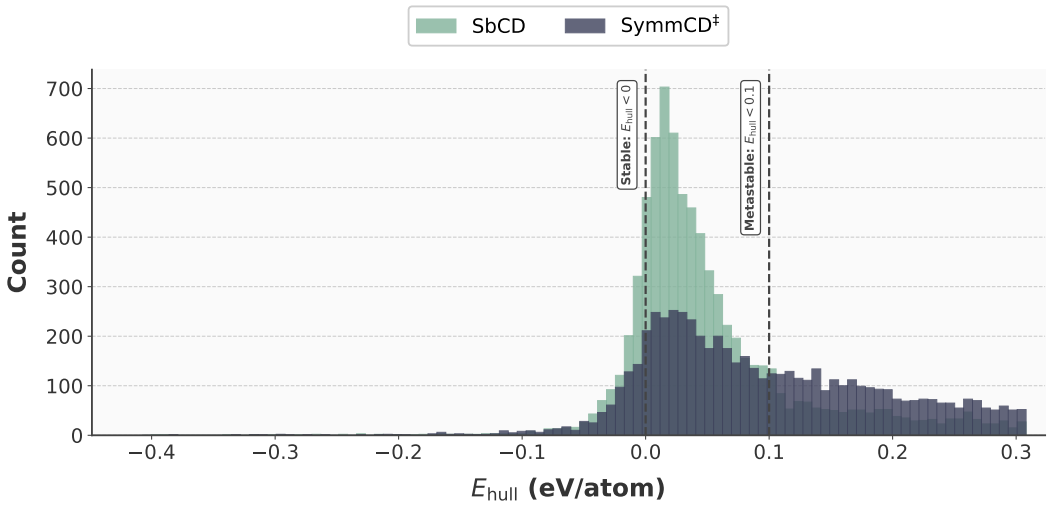


Figure 11: Comparison of the E_{hull} distributions on the MP-20 dataset.

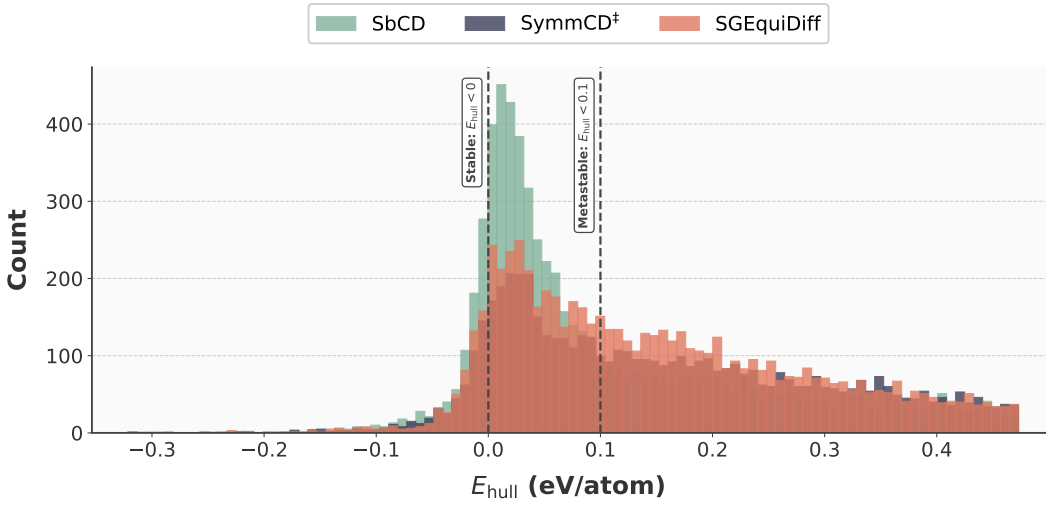


Figure 12: Comparison of the E_{hull} distributions on the MPTS-52 dataset.

Algorithm 2 Training of SbCD

Require: Model θ, transition-rate function λ, dataset of crystal asymmetric units $\mathcal{D}$

1: **while** not converged **do**

2: Sample a clean asymmetric unit and diffusion timestep

$$\mathcal{M}_0 = (\mathbf{F}_0, \mathbf{A}_0, \mathbf{k}_0, \mathbf{S}_0, G_0) \sim \mathcal{D}, \qquad t \sim \mathcal{U}(0, T)$$

3: Sample τ and w via Corollary 4.3

$$\tau = -\frac{1}{\lambda} \log\left(1 - u\left(1 - e^{-\lambda t}\right)\right), \quad u \sim \mathcal{U}(0,1)$$
$$w_{\mathbf{S}} = w_{\mathbf{k}} = t - \tau$$

4: Sample the space-group state at time t via Proposition 4.2

$$G_t \sim q_{t|0}(G_t|G_0) \to G_{t+1} \sim \Pi_t(G_t, G_{t+1})$$

5: Corrupt $\mathbf{k}_0$ using the symmetry-breaking continuous kernel via Proposition 4.4

$$\mathbf{k}_t = \mathfrak{m} \odot \left(\sqrt{\bar{\alpha}_t}\mathbf{k}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}\right) + \mathfrak{m}_\mathfrak{b} + \mathfrak{m} \odot \mathfrak{m}_\mathfrak{b}\sqrt{\frac{\bar{\alpha}_t}{\bar{\alpha}_{t-w_{\mathbf{k}}}}}$$
$$- \left(\mathfrak{m} \odot \sqrt{1 - \frac{\bar{\alpha}_t}{\bar{\alpha}_{t-w_{\mathbf{k}}}}}\boldsymbol{\epsilon} + \mathfrak{m}_\mathfrak{b}\right) + \left(\mathfrak{m} \odot \sqrt{1 - \frac{\bar{\alpha}_t}{\bar{\alpha}_{t-w_{\mathbf{k}}}}}\boldsymbol{\epsilon} + \mathfrak{m}_\mathfrak{b}\right)$$
with $\epsilon, \boldsymbol{\epsilon} \sim \mathcal{N}(0, I)$

6: Corrupt $\mathbf{S}_0$ using the symmetry-breaking categorical kernel in Proposition 4.5.

$$\mathbf{S}_t = \mathrm{Cat}(\mathbf{S}_t; \mathbf{S}_0 \bar{\mathbf{Q}}_t^{G_0 \to G_t}), \quad \bar{\mathbf{Q}}_t^{G_0 \to G_t} = \bar{\alpha}_t I + (1 - \bar{\alpha}_t)\mathbf{1}(\boldsymbol{\pi}_t)^T$$
$$\text{with } \boldsymbol{\pi}_t = \frac{\bar{\alpha}_t(1 - \bar{\alpha}_{t-w_{\mathbf{S}}})\mathfrak{g} + (\bar{\alpha}_{t-w_{\mathbf{S}}} - \bar{\alpha}_t)\mathfrak{g}}{\bar{\alpha}_{t-w_{\mathbf{S}}}(1 - \bar{\alpha}_t)}$$

7: Corrupt $\mathbf{F}_0$ with wrapped Gaussian noises (Jiao et al., 2023), yielding $\mathbf{F}_t$

8: Corrupt $\mathbf{A}_0$ with categorical noises (Austin et al., 2021; Levy et al., 2025), yielding $\mathbf{A}_t$.

9: Evaluate the training losses on the corrupted samples $\mathcal{M}_t, \mathcal{M}_{t+1}$, defined as:

$$\mathcal{M}_t = (\mathbf{F}_t, \mathbf{A}_t, \mathbf{k}_t, \mathbf{S}_t, G_t)$$
$$\mathcal{M}_{t+1} = (\mathbf{F}_{t+1}, \mathbf{A}_{t+1}, \mathbf{k}_{t+1}, \mathbf{S}_{t+1}, G_{t+1}) \text{ by repeating steps } 5, 6, 7, 8$$

$$\mathcal{L}_{\mathbf{k}} \leftarrow \mathrm{MSE}(\hat{\mathbf{k}}_0^\theta(\mathcal{M}_t), \mathbf{k}_0), \ \mathcal{L}_{\mathbf{S}} \leftarrow \mathrm{CE}(\hat{\mathbf{S}}_0^\theta(\mathcal{M}_t), \mathbf{S}_0), \ \mathcal{L}_{\mathbf{A}} \leftarrow \mathrm{CE}(\hat{\mathbf{A}}_0^\theta(\mathcal{M}_t), \mathbf{A}_0)$$
$$\mathcal{L}_G \leftarrow \mathrm{CE}(\hat{G}_0^\theta(\mathcal{M}_t), G_0)$$
$$\mathcal{L}_{\text{jump}} \leftarrow -\hat{\boldsymbol{\Lambda}}_t^\theta(G_t, G_t) + \boldsymbol{\Lambda}_t(G_t, G_t)\log(\hat{\boldsymbol{\Lambda}}_t^\theta(G_{t+1}, G_t)), \quad \text{Lemma 2.1}$$
$$\text{where } \hat{\boldsymbol{\Lambda}}_t^\theta(G_t, G_t) = -\sum_{G \neq G_t} \hat{\boldsymbol{\Lambda}}_t^\theta(G_t, G)$$
$$\text{with } \hat{\boldsymbol{\Lambda}}_t^\theta(G_t, G) = \boldsymbol{\Lambda}_t(G, G_t)\sum_{G_0} \frac{q_{t|0}(G|G_0)}{q_{t|0}(G_t|G_0)} p_{0|t}^\theta(G_0|\mathcal{M}_t)$$
$$\text{and } \hat{\boldsymbol{\Lambda}}_t^\theta(G_{t+1}, G_t) = \boldsymbol{\Lambda}_t(G_t, G_{t+1})\sum_{G_0} \frac{q_{t|0}(G_t|G_0)}{q_{t|0}(G_{t+1}|G_0)} p_{0|t}^\theta(G_0|\mathcal{M}_{t+1})$$
$$\mathcal{L}_{\mathbf{F}} \leftarrow \text{follow (Jiaoet al., 2023)}$$

10: Update θ by descending the full objective weighted by $\{l_{\mathbf{k}}, l_{\mathbf{S}}, l_{\mathbf{A}}, l_G, l_{\mathbf{F}}, l_{\text{jump}}\}$

$$\mathcal{L} = l_{\mathbf{k}}\mathcal{L}_{\mathbf{k}} + l_{\mathbf{S}}\mathcal{L}_{\mathbf{S}} + l_{\mathbf{A}}\mathcal{L}_{\mathbf{A}} + l_G\mathcal{L}_G + l_{\text{jump}}\mathcal{L}_{\text{jump}} + l_{\mathbf{F}}\mathcal{L}_{\mathbf{F}}.$$

11: **end while**