SymADiT: Symmetry-Aware Crystal Generation with Latent Flow Matching

Anmar Karmush[*1], **Cedric Mathieu Brandenburg**[2,3], **Soheil Ershadrad**[2,3], **Johanna Rosen**[2,3], **Michael Felsberg**[1], **Filip Ekström Kelvinius**[4]

[1]Department of Electrical Engineering (ISY) & AI4x, Linköping University
[2]Department of Physics, Chemistry and Biology (IFM), Linköping University
[3]Wallenberg Initiative Materials Science for Sustainability (WISE), Linköping University
[4]Department of Computer and Information Science (IDA), Linköping University
[*]Corresponding author: anmar.karmush@liu.se

Abstract

Tackling the task of crystal generation, we aim to enhance the previously proposed, symmetry-agnostic, All-atom Diffusion Transformer (ADiT) by introducing SymADiT, a symmetry-aware variant. To do so, we use a representation of crystals based on Wyckoff positions. We follow ADiT and perform generative modelling in latent space, adapted to our symmetry-aware representation. By forcing the output of the generative model to adhere to the symmetry restrictions imposed by the generated crystal's space group and each atom's Wyckoff position, the generated crystals exhibit more realistic symmetry properties. We benchmark our method against both symmetry-aware and symmetry-agnostic models for crystal generation and show competitive performance, generating stable, symmetric, and physically plausible crystals with only an off-the-shelf Transformer architecture. Our code is available at https://github.com/Karmush21/SymADiT.

1 Introduction

The vast combinatorial space of yet unexplored materials holds considerable potential for advancing several fields, including the discovery of novel 2D materials, next-generation energy storage systems, and materials for carbon capture [1]. In exploration of this space, generative models have emerged as an attractive class of methods [2], largely due to their ability to rapidly generate large numbers of candidate materials. Within this broader materials space, crystalline materials are of particular interest, motivating substantial efforts toward their computational generation.

Early generative models for crystals [3–5], as well as more recent approaches [6, 7], typically do not explicitly model the intrinsic symmetries of crystalline materials, instead relying on learning such symmetries directly from data. These symmetries are fundamental to determining material properties; for example, crystal symmetry governs whether phenomena such as piezoelectricity can occur [8]. In addition, these *symmetry-agnostic* methods exhibit two fundamental shortcomings: first, their representations are inherently redundant, as they include more structural detail than necessary, given that much of it is determined by the material's symmetry; second, while these models are effective at generating novel crystal structures, the generated structures often exhibit only translational symmetry. This is highly implausible for real-world crystals and, from a generative modelling standpoint, inconsistent with crystal databases (e.g., [3, 9]), where usually only a small fraction of structures lack symmetry beyond pure translations.

To overcome these two issues, recent works have focused on *symmetry-aware* generative modelling of crystals [10–14]. These methods adopt a *Wyckoff*-based representation that explicitly encodes the

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

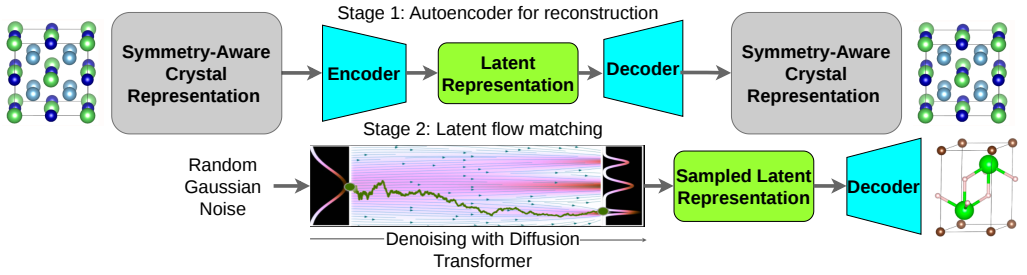


Figure 1: Illustration of our representation and framework, which is inspired by ADiT [6]: Training an autoencoder for reconstruction, followed by learning a generative model over its latent representations. However, in our approach, the decoded representation explicitly contains symmetry information, thus enabling the generation of crystals with realistic symmetry properties by design.

crystal's symmetry, retaining only the minimal information required to reconstruct the full structure via its symmetry operations. We contribute to this line of Wyckoff-based research by building on *All-atom Diffusion Transformer* (ADiT) [6], a recent generative framework for crystals and molecules. While ADiT employs a symmetry-agnostic representation of crystals, we propose *SymADiT*, a symmetry-aware alternative that can be directly substituted into such a framework. In this work, however, we focus exclusively on crystal generation. The central idea behind our representation is to restrict the model to predicting only the degrees of freedom (DOF) not constrained by the space group. By encoding symmetry directly through Wyckoff positions, the symmetry properties are enforced at the representation level rather than learned implicitly by the network. This allows the use of a standard off-the-shelf Transformer architecture [15] while reducing the number of tokens relative to ADiT's symmetry-agnostic representation.

Due to the nature of ADiT (which encodes both crystals and molecules), the generative modelling is performed in latent space. Following their design, we first train an autoencoder (AE) to encode crystals using our symmetry-aware representation, and subsequently perform generative modelling in the resulting latent space. This setting naturally enables us to investigate symmetry-aware generation in latent space, which, to the best of our knowledge, has not been previously explored. [1] In contrast, most existing methods for crystal generation operate directly in data space. As a result, they typically require multiple generative components (e.g., separate mechanisms for continuous and discrete variables) [3, 10, 17]. By operating in latent space, the generative process is restricted to a single mechanism. The proposed methodology is illustrated in Figure 1.

In summary, our first contribution is a representation based on Wyckoff positions, which we incorporate into the ADiT framework to form *SymADiT*, enabling symmetry-aware crystal generation with a simple Transformer backbone. We show how this enables more realistic generated crystals compared to ADiT, and we also benchmark our approach against a wide range of both symmetry-aware and symmetry-agnostic methods, showing competitive performance in crystal generation.

2 Preliminaries: Wyckoff Representation

Symmetry-Agnostic Representation of Crystals An ideal crystalline material is represented by its crystal structure as an infinite periodic arrangement of *unit cells*, allowing a single unit cell to serve as a finite representation. A widely used symmetry-agnostic approach [3–7] is to represent the material by explicitly representing one full unit cell. In this way, a crystal $\mathbf{C}$ with M atoms in its unit cell can be represented by a tuple $\mathbf{C} = (\mathbf{A}, \mathbf{X}, \mathbf{L})$, where $\mathbf{A} \in \mathbb{Z}^M$ denotes the chemical element associated with each atom, $\mathbf{X} \in \mathbb{R}^{M\times 3}$ are Cartesian atom coordinates, and $\mathbf{L} \in \mathbb{R}^{3\times 3}$ are the unit cell basis vectors. The atomic positions can equivalently be expressed in the lattice basis as fractional coordinates $\mathbf{F} \in [0, 1)^{M\times 3}$. In the remainder of this paper, we assume all atomic coordinates are given in fractional form. The lattice vector geometry can likewise be parametrised by six scalar

[1] Prior latent-space generative models have considered SE(3)-equivariance [16], in the context of molecular generation. Such models do not explicitly handle crystal-specific symmetries, such as space groups and Wyckoff positions. Here, *symmetry-aware* refers specifically to the explicit modelling of the material symmetry properties, rather than only enforcing equivariance to rigid-body transformations.

parameters $\ell = (a, b, c, \alpha, \beta, \gamma)$, which provide a sufficient description[2] of $\mathbf{L}$. Since the choice of unit cell is not unique, prior symmetry-agnostic methods [3, 5, 6] typically canonicalize it using a primitive cell obtained via Niggli reduction [18]. This symmetry-agnostic representation exhibits two main shortcomings (see Section 1), motivating the development of symmetry-aware approaches, which we discuss next and extend in our proposed representation.

Space group Fundamental for describing the symmetry of a crystal is its *space group*. A space group $G \in \mathbb{G}$, where $\mathbb{G}$ denotes the set of 230 space groups, is a group of isometries that tiles $\mathbb{R}^3$. G consists of two types of transformations: an infinite subgroup of discrete lattice translations, as well as a collection of symmetry operations $G(\cdot) = \{R_g(\cdot) + v_g | R_G \in O(3), v_g \in \mathbb{R}^3\} \in G$, where R_g is a point group operation (e.g., rotation, reflection), and v_g is a translation. Once the space group is specified, much of the crystal's structural information is largely constrained (see Paragraphs Wyckoff Positions, Unit Cell, and Appendix A). This insight forms the basis of our representation, whose implications we discuss in the following sections.

Wyckoff Positions[3] The Wyckoff positions of a space group G partition regions of the unit cell with specific symmetry constraints. When an atom occupies a Wyckoff position, the nature of that position constrains the coordinates to a point (0 DOF), a line (1 DOF), a plane (2 DOF), or a general position in the unit cell (3 DOF). More formally, the *orbit* of a point x is the set $\{gx | g \in G\}$, i.e., all positions obtained by applying every symmetry operation in G. All points in an orbit are symmetry-equivalent, and one element in that set is called a *representative*. The number of distinct elements in the orbit is called the *multiplicity*. Wyckoff positions are labelled by their multiplicity and a Wyckoff letter, with the letters assigned in lexicographic order corresponding (partially) to increasing multiplicity (e.g., `4a`, `8b`). These combinations are not unique across different space groups. A 2D illustration can be seen in Figure 2.

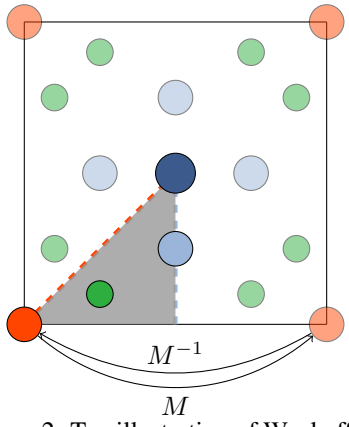


Figure 2: Toy illustration of Wyckoff positions in a 2D unit cell. Colours distinguish Wyckoff positions. Opaque atoms denote representatives, while faded atoms indicate symmetry-equivalent positions; together they form the corresponding orbit. M denotes a symmetry operation. The gray region highlights the asymmetric unit used in Wyckoff-based approaches, containing a single representative from each orbit. Image adapted from [19].

Unit Cell Similar to Wyckoff positions, the space group also constrains the unit cell geometry, meaning that certain parameters in ℓ are no longer independent. For example, in cubic space groups (Nos. 195–230), the unit cell is constrained to be cubic, enforcing $a = b = c$ and $\alpha = \beta = \gamma = 90°$, so that only a single lattice parameter remains independent (a in ℓ). In a Wyckoff-based representation, it is necessary to work with a larger symmetric unit cell, referred to as *conventional*.

Asymmetric Unit An alternative to the symmetry-agnostic representation (where the full unit cell is described) is to encode only a minimal subset of the unit cell, known as the *asymmetric unit* (ASU), from which the full crystal can be reconstructed given the space group and its symmetry operations. This is the standard representation used in Wyckoff-based approaches. The ASU consists of one representative of each symmetry orbit (see Figure 2). This shift to only considering symmetrically inequivalent atoms reduces the dimension of the generative modelling substantially. In Section 3, we present our representation of the ASU, while prior approaches are reviewed in Section 4.

Generative Models Generative frameworks [22–27] aim to learn the underlying (unknown) data distribution $p_{\text{data}}(\cdot)$ and generate new samples from it. In essence, the core idea is based on starting from a pure "noise" sample $x_0 \sim p_0(x_0)$, and turning this into a "clean" sample $x_1 \sim p_\theta(x_1) \approx$

[2]We lose the orientation of the cell with this parametrisation, but this can be arbitrarily chosen. This parametrisation of $\mathbf{L}$ is typically used in practice.

[3]We cover only the essential aspects of Wyckoff positions needed to motivate our method; see, e.g., [20, 21] for a more comprehensive treatment.

Algorithm 1 Pseudocode for AE encoder $\mathcal{E}$

Input: ASU of orbit i: $(\{a_i\}, \{w_i\}, \{f_i\}, \ell, G)$
Output: Latent representations $\{z_i\}$

Encode local information
$h_l = \text{Embed}(a_i)$ $\quad h_l \in \mathbb{R}^{d_m}$
$h_l = h_l + \text{Embed}(w_i)$
$h_l = h_l + \text{MLP}(f_i)$
Encode global information
$h_g = \text{Embed}(G)$ $\quad h_g \in \mathbb{R}^{d_m}$
$h_g = h_g + \text{MLP}(\ell)$
Combine local and global information
$h_i = h_l + h_g$
Apply encoder network
$\{h_i\} = \text{Enc}(\{h_i\})$
Down-project to latent representation and saturate
$z'_i = \text{Linear}(\{h_i\})$ $\quad z'_i \in \mathbb{R}^d$
$z_i = z'_i \Big/ \sqrt{1 + \left(\frac{z'_i}{5}\right)^2}$

Algorithm 2 Pseudocode for AE decoder $\mathcal{D}$

Input: Latent representations $\{z_i\}$
Output: Reconstructed ASU $(\{a_i\}, \{w_i\}, \{f_i\}, \ell)$

Up-project latents to d_{model}
$h_i = \text{Linear}(z_i)$ $\quad h_i \in \mathbb{R}^{d_m}$
Apply decoder network
$\{h_i\} = \text{Dec}(\{h_i\})$
Predict outputs
$\mathcal{S}_G[\,\cdot\,]$ and $\mathcal{S}_w[\,\cdot\,]$ set some information deterministically based on space group or Wyckoff position
$w'_i \sim p_{\text{Wyckoff}}(\cdot \mid \mathcal{S}_G[\text{Linear}_w(h_i)])$ $\quad w'_i \in \mathbb{W}$
$a'_i \sim p_{\text{atom}}(\cdot \mid \text{Linear}_a(h_i))$ $\quad a'_i \in \mathbb{Z}$
$f'_i = \mathcal{S}_{w'_i}[\text{Linear}_f(h_i)]$ $\quad f'_i \in \mathbb{R}^3$
$\ell' = \mathcal{S}_G[\text{Linear}_\ell(\sum_i h_i)]$ $\quad \ell' \in \mathbb{R}^6$

$p_{\text{data}}(x_1)$, via a learned (deterministic or stochastic) map $x_1 = f_\theta(x_0)$.[4] The distribution $p_0(x_0)$ should be something we can easily sample, typically a standard Gaussian distribution. We use flow matching [28] as our generative framework, adjusted for operating in latent space, and describe this in more detail in Section 3.2.

3 Proposed Method: SymADiT

In this section, we introduce our proposed method: Symmetry-aware ADiT (SymADiT). The method is based on a representation of crystals whose key idea is to exploit the fact that once the space group is specified, many structural degrees of freedom for the Wyckoff positions and unit cell become constrained, thereby reducing the dimensionality of the generative problem. Rather than predicting the full ASU structure, we instead model and predict only the symmetry-independent (*free*) parameters, while all remaining quantities are deterministically inferred from the space group. Furthermore, by operating on symmetry-orbits rather than individual atoms, SymADiT substantially reduces the number of tokens processed by the Transformer compared with ADiT, resulting in a more compact representation and faster generation.

3.1 Stage 1: Autoencoder

A crystal structure[5], $\mathbf{C} = (G, \mathbf{W}, \mathbf{A}, \mathbf{F}, \ell)$ consisting of N symmetry orbits, can be represented via its ASU as:

$$\text{Space group } G \in \{1, \ldots, 230\}$$
$$\text{Atom types } \mathbf{A} = \{a_i\}_{i=1}^N \in \mathbb{Z}^N, \quad \text{Wyckoff pos. } \mathbf{W} = \{w_i\}_{i=1}^N \in \mathbb{W}^N,$$
$$\text{Fractional coords. } \mathbf{F} = \{f_i\}_{i=1}^N \in [0, 1)^{N\times 3},$$
$$\text{Lattice params. } \ell = (a, b, c, \alpha, \beta, \gamma) \in \mathbb{R}^6,$$

where $\mathbb{W}$ denotes the set of all 1731 Wyckoff positions across all space groups, and $\mathbb{Z}$ denotes the set of 100 atom types considered in this work. Starting from the crystal ASU, an encoder $\mathcal{E}$ maps the attributes of each orbit to a latent representation z_i:

$$\mathbf{Z} = \mathcal{E}(G, \mathbf{A}, \mathbf{W}, \mathbf{F}, \ell), \qquad \mathbf{Z} = \{z_i\}^N \in \mathbb{R}^{N\times d}, \tag{1}$$

which is described in Algorithm 1. Each latent vector of an orbit encodes both local information about the orbit $(\mathbf{A}, \mathbf{W}, \mathbf{F})$, as well as global information about the crystal (G, ℓ), with cross-orbit dependencies captured through self-attention [15]. While ADiT uses KL divergence [26] as a regularizer to structure the latent space, we follow [30] and replace this with a saturating function (adopting their setting, we fix the saturation parameter to 5, see last line in Algorithm 1), which we found performed better. See Appendix B for an ablation study on different regularisation techniques.

[4] Denoting noise as x_0 and the clean sample as x_1 is consistent with the framework we use, but the general description also covers other deep generative models.

[5] An example of how the representation is aligned with a crystal library (`PyXtal` [29]) can be seen in Appendix A.

The decoder $\mathcal{D}$, described in Algorithm 2, maps the latent embedding back to components of $\mathbf{C}$ as

$$\mathbf{A}', \mathbf{F}', \mathbf{W}', \ell' = \mathcal{D}(\mathbf{Z}, G). \tag{2}$$

Similar to ADiT [6], we use a standard off-the-shelf Transformer [15] for both the encoder $\mathcal{E}$ and decoder $\mathcal{D}$, with the only modification being the removal of unnecessary positional embeddings. The decoder predicts a full parameterization of the ASU, after which symmetry-constrained quantities are masked out or determined deterministically from the space group and predicted Wyckoff positions to ensure adherence to symmetry. These operations are represented by the symmetrizing functions $\mathcal{S}_G\big[\cdot\big]$ and $\mathcal{S}_w\big[\,\cdot\,\big]$ in Algorithm 2. For example, although the network predicts over the full enumeration of space group specific Wyckoff positions (1731 in total), only those compatible with the current space group are retained. Similarly, if a Wyckoff position permits only a single free coordinate x, the predicted y and z coordinates are excluded from decoding, with their values instead determined by the symmetry constraints. Likewise, for a cubic space group, only the lattice constant a in ℓ constitutes an independent degree of freedom; the remaining lengths (b, c) and angles (α, β, γ) are determined by the space group constraints. Details on the network architectures are provided in Appendix C, while Appendix D visualizes the encoder's latent space via dimensionality reduction.

Training The autoencoder is trained using a weighted objective comprising a reconstruction loss and additional regularization terms. The reconstruction loss itself consists of four terms, each corresponding to different components of the crystal's ASU:

$$\mathcal{L} = \lambda_{\mathbf{A}}\mathcal{L}_{\mathbf{A}} + \lambda_{\mathbf{W}}\mathcal{L}_{\mathbf{W}} + \lambda_{\mathbf{F}}\mathcal{L}_{\mathbf{F}} + \lambda_{\ell}\ell_{\ell}, \quad (\lambda_A, \lambda_W, \lambda_F, \lambda_\ell) = (1, 1, 5, 1). \tag{3}$$

As atoms and Wyckoff positions are discrete variables, their corresponding reconstruction terms $\mathcal{L}_{\mathbf{A}}$ and $\mathcal{L}_{\mathbf{W}}$ are standard cross-entropy losses. We use a periodic cosine loss for the fractional coordinates $\mathbf{F}$ and standard mean squared error (MSE) for the cell parameters ℓ. The loss terms are weighted to emphasise accurate reconstruction of the fractional coordinates, with weights chosen empirically. Importantly, the loss is applied only to the free parameters of $\mathbf{W}$, $\mathbf{F}$, and ℓ. Symmetry-constrained parameters are excluded from the loss, since their values are determined by the symmetry constraints and are not used during decoding.

Regularisation Similar to ADiT [6], we use a dimension of the latent representation d that is significantly smaller than the encoder/decoder hidden dimension d_m. Futhermore, we apply augmentations to both the free lattice parameters and atomic positions to improve robustness. The perturbations are restricted to the unconstrained degrees of freedom imposed by symmetry. For example, in a cubic unit cell only the side-length (a in ℓ) is perturbed, while an atom constrained to move along a line (1 DOF) is perturbed only along that direction.

3.2 Stage 2: Latent Flow Matching

After the first training stage, the AE is frozen and its encoder $\mathcal{E}$ is used to produce latent representations of the crystals for training the latent flow matching model. The model learns to generate N latent representations $\mathbf{Z} = \{z_i \in \mathbb{R}^d\}_{i=1}^N$, one per symmetry orbit in the ASU, where each z_i encodes local and global structural information. Together, these can be decoded using the decoder $\mathcal{D}$ to produce a crystal's ASU.

Training We closely follow the notation of ADiT [6]. During training, we begin by encoding a crystal's ASU into a latent representation $\mathbf{Z}$. We define $\mathbf{Z}^{(1)}$ as being a clean sample at time $t = 1$. We then sample a noisy latent sample $\mathbf{Z}^{(0)} \sim \mathcal{N}(0, I_d)$ and construct an intermediate latent by linearly interpolating the two samples as

$$\mathbf{Z}^{(t)} = (1 - t)\mathbf{Z}^{(0)} + t\mathbf{Z}^{(1)}, \;\; t \in [0, 1]. \tag{4}$$

This defines a continuous path from noise to data. Flow matching aims to learn a vector field $u_t(Z^{(t)})$ that describes how a sample should move along the path over time. The ground truth conditional vector field is given by

$$u_t(\mathbf{Z}^{(t)}|\mathbf{Z}^{(1)}) = \frac{\mathbf{Z}^{(1)} - \mathbf{Z}^{(t)}}{1 - t},$$

which corresponds to moving directly toward the clean latent sample. The goal of flow matching is to train a denoiser neural network $\mathcal{F}$ to match the conditional vector field. The network takes as input

Algorithm 3 Pseudocode for DiT sampling

Input: G, integration steps T, cfg. scale γ
Output: Generated sample $(\mathbf{A}, \mathbf{W}, \mathbf{F}, \ell)$
Sample priors and initial noisy latents $Z^{(0)}$ at $t = 0$
Sample $G \sim \hat{p}_{\text{data}}(G)$
Sample $O \sim \hat{p}_{\text{data}}(O \mid G)$ # orbit (index set)
$\mathbf{Z}_O^{(0)} = \{z_i^{(0)} \sim \mathcal{N}(0,1)^d : i \in O\}$
$\Delta t = 1/T$
Denoising loop for orbit O
for $t \in \text{linspace}(0.0, 1.0, T)$:
  $\mathbf{Z}'_{O,\text{cond}} = \mathcal{F}(\mathbf{Z}_O^{(t)}, t, G)$ *# Conditional prediction*
  $\mathbf{Z}'_{O,\text{uncond}} = \mathcal{F}(\mathbf{Z}_O^{(t)}, t, \phi)$ *# Unconditional prediction*
  $\mathbf{Z}'_O = (1-\gamma)\,\mathbf{Z}'_{O,\text{uncond}} + \gamma\,\mathbf{Z}'_{O,\text{cond}}$ *# Classifier-free guidance*
  $\mathbf{Z}_O^{(t+\Delta t)} = \mathbf{Z}_O^{(t)} + \Delta t \cdot \frac{\mathbf{Z}'_O - \mathbf{Z}_O^{(t)}}{1-t}$ *# Euler step*
$(\mathbf{A}, \mathbf{W}, \mathbf{F}, \ell) = \mathcal{D}(\mathbf{Z}^{(1)}, G)$ # Decode latents to ASU (Algorithm 2)
$C = \text{PyXtal}(G, \mathbf{W}, \mathbf{A}, \mathbf{F}, \ell)$ # Construct crystal

the intermediate sample, the time step, and additional information to predict a clean sample $\mathbf{Z}'^{(1)}$. Training of the denoiser network is done by minimizing MSE between $\mathbf{Z}^{(1)}$ and $\mathbf{Z}'^{(1)}$.

Generation During generation, we aim to sample from the (unknown) distribution $p_{\text{data}}(G, \mathbf{A}, \mathbf{W}, \mathbf{F}, \ell)$. Given that the space group constrains many degrees of freedom, we start by sampling a space group G. In practice, during crystal generation, each orbit will correspond to a token in the Transformer, and we therefore introduce the number of orbits O as an auxiliary variable, which is sampled conditioned on the space group. We then sample the latents $\mathbf{Z}$ (which are also auxiliary variables) from our flow matching model, and the remaining variables are sampled by the (probabilistic) decoder $\mathcal{D}$ given $\mathbf{Z}$, G, and (implicitly, as it determines the number of orbits) O. For both G and O, we can use the empirical training data distribution $\hat{p}_{\text{data}}(G)$ and $\hat{p}_{\text{data}}(O|G)$. By defining $\mathbf{X}{:=}(\mathbf{A}, \mathbf{W}, \mathbf{F}, \ell)$, our model can be written as

$$p_\theta(G, O, \mathbf{Z}, \mathbf{X}) = \hat{p}_{data}(G)\; \hat{p}_{data}(O|G)\; p_{\theta_{\text{FM}}}(\mathbf{Z}|G, O)\; p_{\theta_{\mathcal{D}}}(\mathbf{X}|\mathbf{Z}, G, O), \tag{5}$$

where $\theta = (\theta_{\text{FM}}, \theta_{\mathcal{D}})$ are the parameters of the flow matching and decoder neural networks. Similar to ADiT, we use a Diffusion Transformer (DiT) [31] as our denoiser network $\mathcal{F}$, where conditional information is incorporated through adaptive layer norm. Moreover, we use self-conditioning [32], in which the denoiser's prediction from the previous timestep is concatenated with the current input, with a 50% dropout probability during training. We condition on the space group via classifier-free guidance (CFG) [33] to incorporate the conditioning on space group into the latent generative process (see Appendix E for an ablation study on this conditioning).[6] At each denoising step, the model produces two predictions: one conditioned on the space group and one without. These two predictions are combined using guidance scale γ as a weighted combination, controlling the influence of the space group conditioning. Ablation studies on self-conditioning and CFG scale are provided in Appendix B. Consistent with prior work [10, 11], we jointly sample Wyckoff positions and atomic species, ensuring that Wyckoff positions with zero degrees of freedom are not sampled more than once for different orbits, as this would result in overlapping atomic positions. Pseudocode for generation is presented in Algorithm 3.

4 Related Work

4.1 Symmetry-agnostic approaches

Symmetry-agnostic methods do not explicitly enforce crystallographic symmetries, but instead rely on learning them from data. CDVAE [3] is an early example that applies diffusion to generate atom coordinates, while a variational autoencoder models the unit cell shape. DiffCSP [4] extends on CDVAE, defining a joint diffusion process over atoms and unit cell. MatterGen [17] is similar to

[6] This is also implemented in ADiT, although it is not explicitly described in the paper.

DiffCSP, but also allows conditioning on properties. The authors investigated conditioning on space groups, but still, in their experiments only 58 % of the generated structures are classified as the space group they were conditioned on. FlowMM [5] is a model based on flow matching that extends flow matching to respect symmetries such as translation, rotation, permutation, and periodic boundaries, but does not consider the types of symmetry we do in our work. The model we build on, ADiT [6], forms a unified generative framework that generates both periodic crystals and non-periodic molecular systems. We focus exclusively on crystal generation, as the crystallographic symmetries considered in this work are specific to periodic structures. CrystalDiT [7], similarly to ADiT, uses a Diffusion Transformer [31] approach, but performs its diffusion in data space. OMatG [34] unifies diffusion and flow matching through stochastic interpolants, jointly generating atomic positions and lattice vectors alongside discrete flow matching for atomic species.

4.2 Symmetry-aware approaches

DiffCSP++ [14] is an early symmetry-aware method: an extension of DiffCSP where generation starts with choosing a *template*, i.e., space group and occupied Wyckoff positions, from the training set. Generation then proceeds by generating elements, atomic positions, and unit cell similar to DiffCSP, while respecting the symmetries determined by the template. SymmCD [12] is a method that, similar to SymADiT, focuses on modelling of the ASU. However, its generative procedure still operates on and processes the full expanded material. WyCryst [35] and WyckoffDiff [11] take a similar approach in that they generate the symmetry properties of crystals (space group, Wyckoff positions, elements), but they do not model the remaining degrees of freedom, while we generate the complete geometry. SGEquiDiff [10] uses a multi-stage approach, where the space group and unit cell are first generated, and an autoregressive model then generates the occupied Wyckoff positions and elements. Conditioned on this information, a space group-equivariant diffusion process generates the atomic positions, respecting the geometric constraints enforced by the Wyckoff positions.

5 Experiments and Results

In this section, we present SymADiT model details, datasets, evaluation pipeline, baselines used for comparison, and experimental results.

Training and Hyperparameters Similar to ADiT, the encoder $\mathcal{E}$ and decoder $\mathcal{D}$ use a standard Transformer architecture. Unlike ADiT, however, we do not use positional embeddings. We use an embedding dimension of $d_m = 512$, a latent space dimension of $d = 32$, 8 attention heads, and 8 layers, for a total of 52M parameters. The second-stage DiT denoiser, with the same latent dimension and 12 heads and layers, contains 130M parameters. Due to the nature of the Transformer being easily scalable, we also explore increasing the DiT to 450M parameters (DiT_L), similar to ADiT. For the generation, we use cfg scale $\gamma = 2$ (focusing more on the conditional path), with the number of time steps $T = 500$. Models are trained for a maximum of 7,000 epochs, with batch size 512, on either A100 (80GB) or H200 (141GB) GPUs.

Datasets We evaluate our model on three datasets. The first is the commonly used MP20 dataset [3], which contains $45{,}229$ crystals with a maximum of 20 atoms in the primitive unit cell. For a fair comparison, we use the same split used in the original paper, meaning $27{,}138$ crystals are used for training, and $9{,}046$ for validation, with the remaining reserved for testing (which are not used). In addition, we evaluate our own models on the MPTS52 [9] dataset, containing $40{,}476$ experimentally verified crystals with a maximum of 52 atoms in the primitive unit cell. The training split contains $27{,}380$ crystals, with approximately 32% overlap with the MP20 training set. Furthermore, we explore merging the two datasets (MP20+MPTS52), creating a dataset with $65{,}886$ crystals. We divide the data into training and validation sets using an $80/20$ split, ensuring similar space group distributions across both sets.

Evaluation Following the literature, we evaluate all methods on 10,000 generated crystals and report structural and compositional validity. To provide a more comprehensive assessment of the generated crystals, we additionally use the following evaluation pipeline: Of the 10,000 sampled

crystals, we filter out crystals that do not pass the *structural validity* check.[7] We then compute *uniqueness* (U) on those crystals, which compares the generated crystals to each other to check if they are duplicates. Following that, we sample 1,000 crystals from those unique and structurally valid crystals, to calculate *novelty* (N), which measures whether the unique crystals are present in the training data. Both uniqueness and novelty are checked through `StructureMatcher` [37] with `stol`=0.3, `angle_tol`=10.0 and `ltol`=0.2, following ADiT [6]. These are reported as conditional probabilities.[8] Finally, we randomly select 500 structurally valid, unique, and novel (U | N) structures and perform expensive Density Functional Theory (DFT) [38, 39] calculations to assess thermodynamic stability (S) and metastability (M.S). This forms the S.U.N. and M.S.U.N. metrics, for which we report how many of the 500 structures satisfy the corresponding DFT-based stability criteria. We follow a similar DFT evaluation as others [5, 6, 10], with details presented in Appendix F for reproducibility.

To further assess how well the methods capture the underlying symmetry in the data, we investigate the fraction of generated crystals belonging to space group P1. These structures exhibit only translational symmetry. Moreover, we evaluate distributional properties by reporting Jensen–Shannon divergences (JSD) for both the space group (JSD_G) and Wyckoff positions (JSD_Wy) distributions. For the Wyckoff analysis, divergences are averaged across space groups, weighted by the number of generated samples in each group. Moreover, we report the Wasserstein distance for the distribution of the number of atoms in the unit cell W_A.

Metric applicability. Some reported metrics are not applicable to all methods. In particular, P1 and JSD_G are not applicable to symmetry-aware methods, since these methods all begin their generation pipeline by sampling the space group directly from the empirical training distribution.[9] Similarly, W_A is not applicable to the symmetry-agnostic methods or to DiffCSP++, since these methods sample the number of atoms from the empirical training distribution. Appendix G contains all details on the evaluation metrics.

Baselines On MP20, we compare our model to both symmetry-aware and symmetry-agnostic approaches. The symmetry-aware methods include SGEquiDiff [10], SymmCD [12], and DiffCSP++ [14]. For symmetry-agnostic methods, we include ADiT [6], CrystalDiT [7], MatterGen [17], FlowMM [5], DiffCSP [4], and OMatG [34]. For ADiT, SymmCD, SGEquiDiff, and OMatG, we trained the models from scratch and ran their generation pipelines, with no modifications made to any of the codebases. The remaining crystals were obtained through CrystalDiT [7].[10] See Appendix H for Licenses.

MP20 Results Table 1 compares SymADiT to both symmetry-aware and symmetry-agnostic approaches, showing competitive performance. First off, it is evident that the symmetry-agnostic methods struggle with capturing the symmetry in the data: The fraction of P1 crystals (P1 column) in MP20 is 2.23%, but the symmetry-agnostic model closest to this is MatterGen with 67%, and for ADiT[11] and CrystalDiT, which are the symmetry-agnostic methods most related to SymADiT, this number is $> 98\%$. Appendix I showcases further the proportion of space group symmetries for each method and highlights more clearly how symmetry-agnostic methods in general fail to capture the space group distribution in relation to MP20 training data. Furthermore, SymADiT most closely follows the training distribution among the symmetry-aware methods. This suggests that, for some competing symmetry-aware methods, the final generated structures may not always preserve the space group initially sampled during generation, thereby altering the resulting space group distribution.

The distribution metrics in Table 1 further highlight that symmetry-agnostic methods are considerably less effective at capturing the symmetry characteristics of the data in terms of Wyckoff positions (JSD_Wy). An interesting observation is that symmetry-aware methods only sample implicitly from the atom distribution (W_A) when sampling the occupied Wyckoff positions, but still manage to capture

[7] As others have noted [10, 13], only $\sim 90\%$ of the MP20 training set passes the compositional validity check based on SMACT [36], and we do therefore not filter based on this criterion.

[8] E.g., structural validity = $9500/10000$, uniqueness = $8500/9500$, novelty = $500/1000$.

[9] MP20 contains 2.23% P1 structures.

[10] CIF files can be found on HuggingFace: `https://huggingface.co/xiaohan-yi/CrystalDiT`.

[11] While ADiT is listed as symmetry-agnostic, it tries to incorporate symmetry via conditioning on space group during generation, but this does not seem to be effective.

Table 1: **MP20 dataset.** Metrics are reported over 10,000 generated crystals. Validity is evaluated on all samples. Uniqueness is measured as the fraction of structurally valid samples that are distinct from each other. Novelty is computed over 1,000 unique structures to assess duplication with the training data. (Meta)stability is evaluated by DFT on 500 samples that are structurally valid, unique, and novel. The M/S.U.N columns report the number of (meta)stable structures among the 500 samples evaluated with DFT. Dash indicates that the method uses the empirical training distribution and the metric is therefore not meaningful.

Method	Sym.	Validity (%) ↑		Distr. distance ↓			% P1	% U | N ↑	# S.U.N. ↑	# M.S.U.N. ↑
		Struct.	Comp.	JSD_G	JSD_{Wy}	W_{A}				
SGEquiDiff [10]	Yes	99.65	84.56	–	0.05	0.48	–	98.35 | 89.50	51	89
SymmCD [12]	Yes	92.74	81.53	–	0.13	1.17	–	98.63 | 92.70	49	78
DiffCSP++ [14]	Yes	99.96	84.48	–	0.04	–	–	98.27 | 89.80	64	105
ADiT [6]	No	99.56	90.44	0.95	0.70	–	99.01	88.23 | 48.60	103	170
CrystalDiT [7]	No	97.82	84.34	0.94	0.67	–	98.25	92.10 | 65.10	47	152
MatterGen [17]	No	100.00	83.58	0.81	0.32	–	67.12	98.37 | 91.80	41	126
FlowMM [5]	No	99.23	82.19	0.91	0.35	–	90.93	98.52 | 90.90	36	106
DiffCSP [4]	No	99.93	81.94	0.85	0.40	–	70.60	97.35 | 91.10	53	107
OMatG [34]	No	99.13	94.52	0.95	0.55	–	99.20	97.62 | 80.10	87	134
SymADiT	Yes	96.50	88.23	–	0.04	0.22	–	87.42 | 54.60	85	136

this distribution well. In Appendix J, we show additional statistical properties of novel crystals generated by each method.

Taken together, these results highlight the need for models like SymADiT that incorporate symmetry as an inductive bias, rather than learning it purely from data. Furthermore, the results suggest that the inductive bias incorporated into the representation is sufficient to generate symmetric crystals, without requiring additional equivariance constraints in the generative model: *a simple Transformer combined with latent flow matching is sufficient*.

Investigating the M/S.U.N. columns in Table 1, we note that *SymADiT is able to achieve the highest number of M/S.U.N. crystals among all symmetry-aware methods*, showcasing its strong performance. SymADiT does not, by comparison, achieve higher M/S.U.N. numbers than ADiT. However, as stated and shown above, symmetry-agnostic methods struggle to capture the underlying symmetry in the data, neglecting a key characteristic of realistic and synthesizable crystals: *most experimentally observed crystals exhibit substantial symmetry and do not belong to the P1 space group.*

Thus, while the M/S.U.N. metric is an important measure of performance at this stage of the crystal generation pipeline, it is not the ultimate measure of success. The final goal is the experimental synthesis of the proposed crystals, where SymADiT (and other symmetry-aware methods) may provide a more suitable foundation by explicitly capturing the symmetry characteristics commonly observed in experimentally realised crystals. A statistical comparison of SymADiT's M/S.U.N. performance against other methods is presented in Appendix K.

The uniqueness and novelty fractions of crystals generated by SymADiT and ADiT are both comparably low. This suggests that the latent-space design could be the contributing factor. One possible interpretation is that the lower-dimensional latent space used by SymADiT and ADiT makes memorization or concentration around the training manifold more likely. Such a bias toward the training distribution may also help explain the strong M/S.U.N. performance of ADiT and SymADiT. Samples generated closer to the training manifold may be more likely to remain physically realistic and stable, suggesting that the latent-space bottleneck may trade some novelty for improved structural realism.

From a practical perspective, a lower uniqueness and novelty simply mean that one would have to generate more crystals to reach a desired number of crystals to be evaluated. Since generative models are computationally cheap compared to DFT, this extra sampling cost is not a major concern in practice, and it follows that what matters most is the quality of the crystals that make it to the final evaluation stage (DFT), with the previous steps, like novelty, uniqueness, and structural validity, acting as filters. In this regard, SymADiT is the best among the symmetry-aware methods. This does not, of course, mean that these metrics are unnecessary, but rather that they require a more nuanced interpretation. We further discuss the novelty metric in more detail in Appendix L.

Table 2: **MPTS52 and MP20+MPTS52.** Metrics are reported over 10,000 generated crystals, with M/S.U.N. computed using 100 DFT calculations. $\mathrm{SymADiT}_L$ refers to the larger 450M DiT model. See caption of Table 1 for additional details.

Method	Data	Validity (%) ↑		Distr. distance ↓			% P1	% U | N ↑	# S.U.N. ↑	# M.S.U.N. ↑
		Struct.	Comp.	JSD_G	$\mathrm{JSD}_{\mathrm{Wy}}$	W_{A}				
SymADiT	MPTS52	94.34	84.23	–	0.04	0.40	–	85.04 | 48.50	8	20
SymADiT	MP20 + MPTS52	95.31	86.60	–	0.04	0.32	–	93.94 | 54.30	9	30
$\mathrm{SymADiT}_L$	MP20 + MPTS52	94.96	87.27	–	0.05	0.49	–	92.57 | 49.50	7	29

Comparison of Inference Speed between SymADiT and ADiT By operating on symmetry orbits rather than individual atoms, SymADiT substantially reduces the number of tokens processed by the Transformer compared with ADiT, resulting in a more compact representation and faster generation. Specifically, on the MP20 dataset, SymADiT processes an average of 4.74 tokens per sample, compared with 10.34 for ADiT, resulting in a $2.40\times$ inference speed-up over ADiT. Further details about how the experiment was conducted are provided in Appendix M.

Additional Results Table 2 shows additional results for SymADiT on larger datasets. On both datasets, due to computational constraints, we ran only 100 DFT calculations on these models. SymADiT still performs reasonably in the distribution metrics, showing that the model can robustly scale up to larger structures. We found that the larger DiT did not improve the generation of novel structures, indicating that the 130M model is sufficient. Visual results of M/S.U.N. structures generated by SymADiT are presented in Appendix N, capturing a wide range of space groups.

Ablation Studies Ablation studies on AE latent regularisation as well as CFG and self-conditioning for the DiT are provided in Appendix B.

6 Discussion and Conclusion

In this paper, we propose SymADiT, a symmetry-aware extension of ADiT, a generative framework for crystals based on latent flow matching. We construct SymADiT using a Wyckoff representation that explicitly encodes symmetry. Our results demonstrate the capability of SymADiT to generate crystals with realistic symmetry properties, which ADiT and other symmetry-agnostic methods struggle to capture, while achieving the highest number of thermodynamically stable crystals among the symmetry-aware methods evaluated. Notably, these improvements are achieved using a simple off-the-shelf Transformer architecture, without requiring additional architectural modifications, while also reducing the number of tokens relative to ADiT.

Due to computational constraints, we conducted only 100 DFT evaluations for Table 2, adding unknown variance to those reported M/S.U.N. rates. A limitation of the generative model developed here and in previous work is that, while proposing new crystals, it provides no guidance on how they could be synthesized, which leaves an open question for future work. Additionally, while ADiT was originally developed for enabling generation of both crystals and molecules, we focused only on the crystal generation aspect, and future work could include molecule generation capability together with our symmetry-aware approach to further study transfer learning between the two atomic systems.

Acknowledgements

This work was supported by the Knut and Alice Wallenberg Foundation (KAW) via the Wallenberg AI, Autonomous Systems and Software Program (WASP) and the Wallenberg Initiative Material Science for Sustainability (WISE) through the joint WASP-WISE project *Accelerated design of functional 2D materials by leveraging foundation models in a closed-loop approach*, and by Excellence Center at Linköping–Lund in Information Technology (ELLIIT). Computations were enabled by the Berzelius resource provided by the Knut and Alice Wallenberg Foundation at the National Supercomputer Centre (NSC) and by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at the National Supercomputer Centre (NSC), partially funded by the Swedish Research Council through grant agreement no. 2022-06725.

References

[1] Addis S Fuhr and Bobby G Sumpter. Deep generative models for materials discovery and machine learning-accelerated innovation. *Frontiers in Materials*, 9:865270, 2022.

[2] Hyunsoo Park, Zhenzhu Li, and Aron Walsh. Has generative artificial intelligence solved inverse materials design? *Matter*, 7(7):2355–2367, 2024.

[3] Tian Xie, Xiang Fu, Octavian-Eugen Ganea, Regina Barzilay, and Tommi S Jaakkola. Crystal diffusion variational autoencoder for periodic material generation. In *International Conference on Learning Representations (ICLR)*, 2022.

[4] 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.

[5] Benjamin Kurt Miller, Ricky TQ Chen, Anuroop Sriram, and Brandon M Wood. Flowmm: generating materials with riemannian flow matching. In *Proceedings of the 41st International Conference on Machine Learning*, pages 35664–35686, 2024.

[6] Chaitanya K Joshi, Xiang Fu, Yi-Lun Liao, Vahe Gharakhanyan, Benjamin Kurt Miller, Anuroop Sriram, and Zachary Ward Ulissi. All-atom diffusion transformers: Unified generative modelling of molecules and materials. In *International Conference on Machine Learning*, pages 28393–28417. PMLR, 2025.

[7] Xiaohan Yi, Guikun Xu, Zhong Zhang, Liu Liu, Yatao Bian, Xi Xiao, and Peilin Zhao. Crystaldit: A diffusion transformer for crystal generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, 2026.

[8] Andrei L Kholkin, Nikolay A Pertsev, and Alexander V Goltsev. Piezoelectricity and crystal symmetry. In *Piezoelectric and acoustic materials for transducer applications*, pages 17–38. Springer, 2008.

[9] 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.

[10] Rees Chang, Angela Pak, Alex Guerra, Ni Zhan, Nick Richardson, Elif Ertekin, and Ryan P Adams. Space group equivariant crystal diffusion. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[11] Filip Ekström Kelvinius, Oskar B. Andersson, Abhijith S Parackal, Dong Qian, Rickard Armiento, and Fredrik Lindsten. WyckoffDiff – a generative diffusion model for crystal symmetry. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu, editors, *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 15130–15147. PMLR, 13–19 Jul 2025.

[12] 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.

[13] Nikita Kazeev, Wei Nong, Ignat Romanov, Ruiming Zhu, Andrey E Ustyuzhanin, Shuya Yamazaki, and Kedar Hippalgaonkar. Wyckoff transformer: Generation of symmetric crystals. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu, editors, *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 29495–29526. PMLR, 13–19 Jul 2025.

[14] Rui Jiao, Wenbing Huang, Yu Liu, Deli Zhao, and Yang Liu. Space group constrained crystal generation. In *The Twelfth International Conference on Learning Representations*, 2024.

[15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[16] Minkai Xu, Alexander Powers, Ron Dror, Stefano Ermon, and Jure Leskovec. Geometric latent diffusion models for 3d molecule generation. In *International Conference on Machine Learning*. PMLR, 2023.

[17] Claudio Zeni, Robert Pinsler, Daniel Zügner, Andrew Fowler, Matthew Horton, Xiang Fu, Zilong Wang, Aliaksandra Shysheya, Jonathan Crabbé, Shoko Ueda, Roberto Sordillo, Lixin Sun, Jake Smith, Bichlien Nguyen, Hannes Schulz, Sarah Lewis, Chin-Wei Huang, Ziheng Lu, Yichi Zhou, Han Yang, Hongxia Hao, Jielan Li, Chunlei Yang, Wenjie Li, Ryota Tomioka, and Tian Xie. A generative model for inorganic materials design. *Nature*, 2025. doi: 10.1038/s41586-025-08628-5.

[18] Ralf W Grosse-Kunstleve, Nicholas K Sauter, and Paul D Adams. Numerically stable algorithms for the computation of reduced unit cells. *Foundations of Crystallography*, 60(1):1–6, 2004.

[19] Filip Ekström Kelvinius. *Deep Learning for the Atomic Scale : Graph Neural Networks and Deep Generative Models with Some Applications to Materials and Molecules*. PhD thesis, 2025.

[20] M I Aroyo, editor. *International Tables for Crystallography Volume A: Space-group Symmetry*. International Tables for Crystallography. International Union of Crystallography, 6th edition edition, December 2016. ISBN 978-0-470-97423-0.

[21] B Souvignier, H Wondratschek, M I Aroyo, G Chapuis, and A M Glazer. Space groups and their descriptions. In *International Tables for Crystallography*, pages 42–74. International Union of Crystallography, Chester, England, December 2016.

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

[23] Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured denoising diffusion models in discrete state-spaces. In *Advances in Neural Information Processing Systems*, volume 34, pages 17981–17993, 2021.

[24] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations*, 2021.

[25] Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pages 1530–1538. PMLR, 2015.

[26] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *stat*, 1050:1, 2014.

[27] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

[28] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.

[29] 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.

[30] François Rozet, Ruben Ohana, Michael McCabe, Gilles Louppe, Francois Lanusse, and Shirley Ho. Lost in latent space: An empirical study of latent diffusion models for physics emulation. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[31] William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4195–4205, 2023.

[32] Jason Yim, Brian L Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. Se (3) diffusion model with application to protein backbone generation. In *International Conference on Machine Learning*, pages 40001–40039. PMLR, 2023.

[33] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021.

[34] Philipp Höllmer, Thomas Egg, Maya Martirossyan, Eric Fuemmeler, Zeren Shui, Amit Gupta, Pawan Prakash, Adrian Roitberg, Mingjie Liu, George Karypis, Mark Transtrum, Richard Hennig, Ellad B. Tadmor, and Stefano Martiniani. Open materials generation with stochastic interpolants. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu, editors, *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 23417–23450. PMLR, 13–19 Jul 2025.

[35] Ruiming Zhu, Wei Nong, Shuya Yamazaki, and Kedar Hippalgaonkar. WyCryst: Wyckoff inorganic crystal generator framework. *Matter*, 7(10):3469–3488, October 2024. ISSN 2590-2385. doi: 10.1016/j.matt.2024.05.042.

[36] Daniel W Davies, Keith T Butler, Adam J Jackson, Andrew Morris, Jarvist M Frost, Jonathan M Skelton, and Aron Walsh. Computational screening of all stoichiometric inorganic materials. *Chem*, 1(4):617–627, 2016.

[37] 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.

[38] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136(3B):B864, 1964.

[39] Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):A1133, 1965.

[40] Stefan Elfwing, Eiji Uchibe, and Kenji Doya. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural networks*, 107:3–11, 2018.

[41] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[42] Georg Kresse and Jürgen Hafner. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B*, 47(1):558, 1993.

[43] Georg Kresse and Jürgen Hafner. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Phys. Rev. B*, 49(20):14251, 1994.

[44] Georg Kresse and Jürgen Furthmüller. Efficiency of ab initio total-energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.*, 6(1):15–50, 1996.

[45] Georg Kresse and Jürgen Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54(16):11169, 1996.

[46] John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77(18):3865, 1996.

[47] Peter E Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50(24):17953, 1994.

[48] Anubhav Jain, Geoffroy Hautier, Charles J Moore, Shyue Ping Ong, Christopher C Fischer, Tim Mueller, Kristin A Persson, and Gerbrand Ceder. A high-throughput infrastructure for density functional theory calculations. *Computational Materials Science*, 50(8):2295–2310, 2011.

[49] Anubhav Jain, Geoffroy Hautier, Shyue Ping Ong, Charles J Moore, Christopher C Fischer, Kristin A Persson, and Gerbrand Ceder. Formation enthalpies by mixing gga and gga+ u calculations. *Physical Review B—Condensed Matter and Materials Physics*, 84(4):045115, 2011.

[50] Martin Dahlqvist, Michel W Barsoum, and Johanna Rosen. Max phases–past, present, and future. *Materials Today*, 72:1–24, 2024.

[51] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL materials*, 1(1):011002, 2013.

[52] Matthew K. Horton, Patrick Huck, Ruo Xi Yang, Jason M. Munro, Shyam Dwaraknath, Alex M. Ganose, Ryan S. Kingsbury, Mingjian Wen, Jimmy X. Shen, Tyler S. Mathis, Aaron D. Kaplan, Karlo Berket, Janosh Riebesell, Janine George, Andrew S. Rosen, Evan W. C. Spotte-Smith, Matthew J. McDermott, Orion A. Cohen, Alex Dunn, Matthew C. Kuner, Gian-Marco Rignanese, Guido Petretto, David Waroquiers, Sinead M. Griffin, Jeffrey B. Neaton, Daryl C. Chrzan, Mark Asta, Geoffroy Hautier, Shreyas Cholia, Gerbrand Ceder, Shyue Ping Ong, Anubhav Jain, and Kristin A. Persson. Accelerated data-driven materials science with the materials project. *Nature Materials*, 24(10):1522–1532, 2025. ISSN 1476-4660. doi: 10.1038/s41563-025-02272-0.

[53] Koichi Momma and Fujio Izumi. Vesta 3 for three-dimensional visualization of crystal, volumetric and morphology data. *Applied Crystallography*, 44(6):1272–1276, 2011.

A Example of the Representation

Here we showcase an example of how our representation can be used to create a crystal $\mathbf{C} = (G, \mathbf{W}, \mathbf{A}, \mathbf{F}, \ell)$ using the crystal library `PyXtal` [29]. We take the classic example of NaCl, which belongs to space group 225. Na and Cl occupy two highly symmetric orbits in space group 225, and have Wyckoff positions "4a" and "4b", respectively, which results in 8 atoms once expanded to the full conventional unit cell. The representatives of these orbits are completely fixed by symmetry: $(0, 0, 0)$ for Na at "4a", and $(0.5, 0.5, 0.5)$ for Cl at "4b" (i.e., both representatives have 0 DOF), meaning that no prediction of $\mathbf{F}$ is required. Furthermore, the unit cell in space group 225 is a cube, meaning that all sides have the same length (in the case of NaCl, 5.65 Å), with angles all being fixed at $90°$. Again, this implies that the autoencoder only gets supervised with one number for the unit cell, namely the length of the cube (a in ℓ). `PyXtal` combines the free fractional and free unit cell parameters into a tuple $\mathbf{X}$, which in this case would be $\mathbf{X} = (5.65)$. Given all of this, the tuple needed to build the crystal via `PyXtal` would look like

$$\mathbf{C} = (G, \mathbf{W}, \mathbf{A}, \mathbf{F}, \ell) = (G, \mathbf{W}, \mathbf{X}, \mathbf{A}) = [225,\ (4a, 4b),\ (5.65),\ (\text{Na}, \text{Cl})].$$

This can later be expanded to the full crystal using pymatgen [37].

Generation From a generative standpoint, the process of generating NaCl would begin by sampling space group 225 from the empirical training distribution. Next, Wyckoff positions and atomic species are sampled, with the constraint that zero-degree-of-freedom (0 DOF) Wyckoff positions cannot be selected more than once (e.g., assigning "4a" to multiple orbits is disallowed). Since both "4a" and "4b" are selected, positional predictions are not utilized (no free parameters). Similarly, for lattice parameter prediction, only one degree of freedom is available (length of cube); therefore, only the first prediction is retained, and all subsequent predictions of ℓ are discarded.

B Ablation Studies

We conduct extensive ablation studies to justify our choice of training settings.

To assess latent-space regularisation, we compare saturation-based regularisation with KL regularisation at different scales, as well as using no latent space regularisation at all. Reconstruction performance for stage 1 is evaluated using the match rate (MR) and root mean square distance (RMSD) on the MP20 test set. Overall, we found that using a downweighted KL divergence penalty did not have any significant effect on performance, which is why we decided to follow [30] and use a saturation function instead. Moreover, we note that we achieve an approximately 7% increase in match rate compared to ADiT's reconstruction performance, as well as a lower RMSD [6]. Results are presented in Table 3.

Table 3: **Regularisation ablation.** Effect of different regularisation methods on match rate (MR) and RMSD.

Reg. method	MR ↑	RMSD ↓
No regularisation	97.92	0.029
Saturation	97.90	**0.019**
KL scale = 0.1	79.98	0.104
KL scale = 0.01	95.35	0.040
KL scale = 0.001	97.79	0.022
KL scale = 0.0001	**98.93**	**0.019**

To assess the effect of classifier-free guidance (CFG) during generation, we vary the CFG scale and evaluate generation performance using uniqueness and novelty. Higher values indicate better performance. Lower CFG scales (focusing less on the space group conditioning) achieve higher novelty at the cost of validity metrics. Overall, we found that CFG scales of 1.5 and 2.0 work well in our experiments, placing greater emphasis on conditioning. Results are presented in Table 4 .

Table 4: **Classifier-free guidance ablation.** Effect of the CFG scale on structural validity, compositional validity, uniqueness, and novelty.

CFG	Struct. (%) ↑	Comp. (%) ↑	U (%) ↑	N (%) ↑
0.0	70.03	65.83	**99.73**	**96.50**
0.5	92.48	82.62	97.12	74.90
1.0	93.20	83.25	82.40	48.20
1.5	96.86	88.18	90.72	54.40
2.0	96.50	88.23	87.42	54.60
2.5	**97.14**	88.27	87.69	52.40
3.0	97.05	**88.41**	86.34	54.00

To assess the effect of self-conditioning during training, we conduct an ablation in which the self-conditioning probability is set to zero. We train an otherwise identical model using the same experimental setup described in Table 1 in the paper. Without self-conditioning, higher uniqueness is achieved, although with lower novelty. Results are presented in Table 5 .

Table 5: **Self-conditioning ablation.** Effect of self-conditioning on structural validity, compositional validity, uniqueness, and novelty.

Self-cond.	Struct. (%) ↑	Comp. (%) ↑	U (%) ↑	N (%) ↑
Yes	96.50	88.23	87.42	**54.60**
No	**97.85**	**88.27**	**89.20**	50.40

C Network Details

Autoencoder The MLPs in the encoder (Algorithm 1) for the fractional coordinates $\mathbf{F}$ and the unit cell ℓ are two-layer feedforward networks, using a SiLU [40] nonlinearity between layers. We concatenated $\mathbf{F}$ and ℓ with their corresponding mask $M_{\mathbf{F}}$ and M_ℓ before passing to the MLP. For the atom types $\mathbf{A}$ and space group G, we use a typical `nn.Embedding` from PyTorch [41]. For the Wyckoff positions $\mathbf{W}$, since Wyckoff letters and multiplicities (e.g., `4a`, `4b`) are not unique across different space groups, we encode each position by concatenating its space group identifier with the Wyckoff label (e.g., `225_4a`). Again, we use an `nn.Embedding` layer to encode this. The Linear layers in the decoder (Algorithm 2) are single-layer neural networks for the discrete predictions ($\mathbf{W}$ and $\mathbf{A}$), and similar two-layer networks as the encoder for the continuous predictions ($\mathbf{F}$ and ℓ).

During training, the free parameters of $\mathbf{F}$ (constrained by G and $\mathbf{W}$, see Wyckoff Positions) and ℓ (constrained by G, see Unit Cell) are indicated by binary masks $M_{\mathbf{F}} \in \{0,1\}^3$ and $M_\ell \in \{0,1\}^6$. For Wyckoff positions, invalid entries are masked by assigning their logits to $-\infty$ prior to the softmax operation.

D Latent Space Visualisation of Autoencoder

We project the learned crystal representations to two dimensions using t-SNE, preceded by PCA for dimensionality reduction and denoising. Figure 3 shows this for 30,000 samples for the MP20 dataset, with the colours indicating which space group each crystal belongs to. We observe that the latent representations form distinct clusters according to the crystals' space groups, with similar space groups located closer to one another. Still, we observe some mixing, which could be due to other information that the crystals share (e.g. similar Wyckoff or unit cell information). We observe that removing the space group conditioning during the latent flow matching part results in fewer valid crystals (see Appendix E) and hypothesize that this additional information helps guide the noisy latents toward the correct cluster.

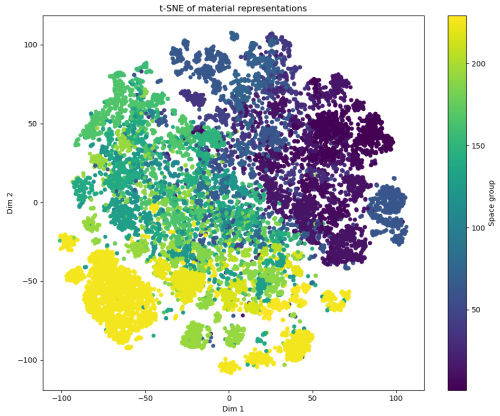
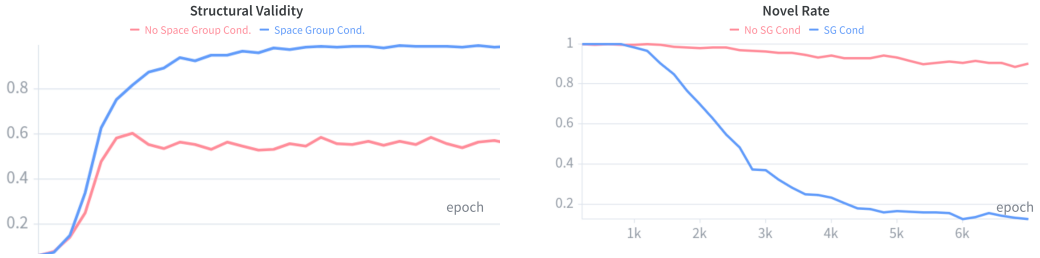


Figure 3: t-SNE and PCA plot of latent embeddings from a trained autoencoder on the MP20 dataset.

We evaluated several strategies for encoding local information in Algorithm 1. In particular, we experimented with concatenating and then down-projecting local, global, and combined representations. Ultimately, we found that a simple additive combination yielded the best performance in terms of generating novel and stable crystals.

E Space Group Conditioning

In this section, we examine the role of space group conditioning in our framework. We modify Algorithm 1 by removing the space group encoding to train a new autoencoder. We then train two latent flow matching models: one with space group conditioning and one without. The results are shown in Figure 4. Without conditioning, the model achieves high novelty but low structural validity. In contrast, incorporating space group conditioning significantly improves structural validity, albeit at the cost of reduced novelty. This highlights the importance of space group conditioning in our representation, in contrast to ADiT, which uses the same conditioning but is only able to generate mostly one space group (P1). Furthermore, this also suggests an interesting avenue for future research: identifying additional sources of information that may be beneficial to condition on.



(a) Structural Validity (b) Novelty

Figure 4: Comparison of structural validity and novelty for DiT models with and without space group conditioning.

F DFT Details

First-principles structural optimisations were conducted in the framework of density functional theory (DFT) [38, 39] using the Vienna *Ab initio* Simulation Package (VASP) [42–45]. The exchange-correlation potential was treated using the generalized gradient approximation (GGA) functional in conjunction with the Perdew, Burke, and Ernzerhof (PBE) method [46]. The projector augmented wave method [47] was applied. All input parameters, including plane-wave cutoffs, convergence criteria, and k-point densities, were selected using pymatgen's MPRelaxSet [37]. For systems where

the tetrahedron smearing method led to convergence issues, Gaussian smearing with a smearing width of 0.05 was employed. Thermodynamic stability was assessed by constructing the convex hull relative to all competing phases reported in the Materials Project MP20 and MPTS52 datasets. To enable direct comparison with Materials Project data, the MP energy correction scheme was employed for anions as well as for combining GGA and GGA+U calculations [48, 49]. The convex hull analysis was performed using the pymatgen library [37]. The convex hull was constructed without incorporating the energies of the newly predicted structures. Accordingly, phases lying on or below the hull are considered thermodynamically stable, while those with energies up to 0.1 eV/atom above the hull are classified as metastable. The reported energy difference with respect to the convex hull is equivalent to the formation enthalpy (ΔH_{cp}) defined relative to the set of competing phases, as commonly used in the materials science literature [50].

G Evaluation Metrics

Similar to prior works ADiT [6], we follow the evaluation protocol established by [28], where we sample $10,000$ crystals and compute validity, stability, uniqueness, and novelty, as defined as follows:

- Structural Validity: % of crystals with all pairwise distances >= 0.5 and crystal volume >= 0.1.
- Compositional Validity: % of crystal compositions with charge neutrality and electronegativity balance according to SMACT [36]. As noted in the main paper, only approximately 90% of the MP20 crystals pass this check; therefore, it should not be relied upon heavily.
- Uniqueness: % of generated crystals that are not duplicates of each other. Comparisons were performed using StructureMatcher from pymatgen [37].
- Novelty: % of generated crystals that are not duplicates of the training data. Comparisons were performed using StructureMatcher from pymatgen [37].
- Stability: % of crystals with ΔH_{cp} <0.0 eV/atom and number of unique elements >= 2. (We also report metastability as ΔH_{cp} <0.1 eV/atom and number of unique elements >= 2.)
- U.N. Rate: % of structurally valid crystals that are both unique and novel
- S.U.N.: # of structurally valid crystals that are stable (See Appendix F), unique and novel.
- M.S.U.N.: # of crystals that are metastable (See Appendix F), unique and novel.
- P1 rate: the fraction of the 10,000 generated crystals classified as P1, as determined by pymatgen [37].

H Licenses

Code We have used the ADiT codebase [12] as starting point for our implementation. This code is available under an MIT license.

Training data We used datasets derived from Materials Project [51, 52], which provide their data under a CC-BY 4.0 license. [13].

Data from compared methods Generated structures for compared methods, except ADiT, SymmCD, SGEquiDiff, and OMatG were provided by Yi et al. [7], available online[14] under a MIT license. For ADiT, SymmCD, SGEquiDiff, and OMatG, we trained models from scratch to generate structures ourselves using official public implementations, without any changes.

I Space group Distribution of Generated Structures

To showcase clearly the diversity of generated crystal symmetries, and further support the results in Table 1, we compare the space group distributions produced by each method to the MP20 dataset.

[12] https://github.com/facebookresearch/all-atom-diffusion-transformer

[13] https://creativecommons.org/licenses/by/4.0/

[14] https://huggingface.co/xiaohan-yi/CrystalDiT

Results can be seen in Figure 5. The figure shows the fraction of generated structures assigned to each of the ten most frequent space groups across the evaluated methods, with all remaining space groups grouped into *Other*. We note that SymADiT best follows the distribution of the reference MP20 dataset. This suggests that, for some competing symmetry-aware methods, the final generated structures may not always preserve the space group initially sampled during generation, thereby altering the resulting space-group distribution.

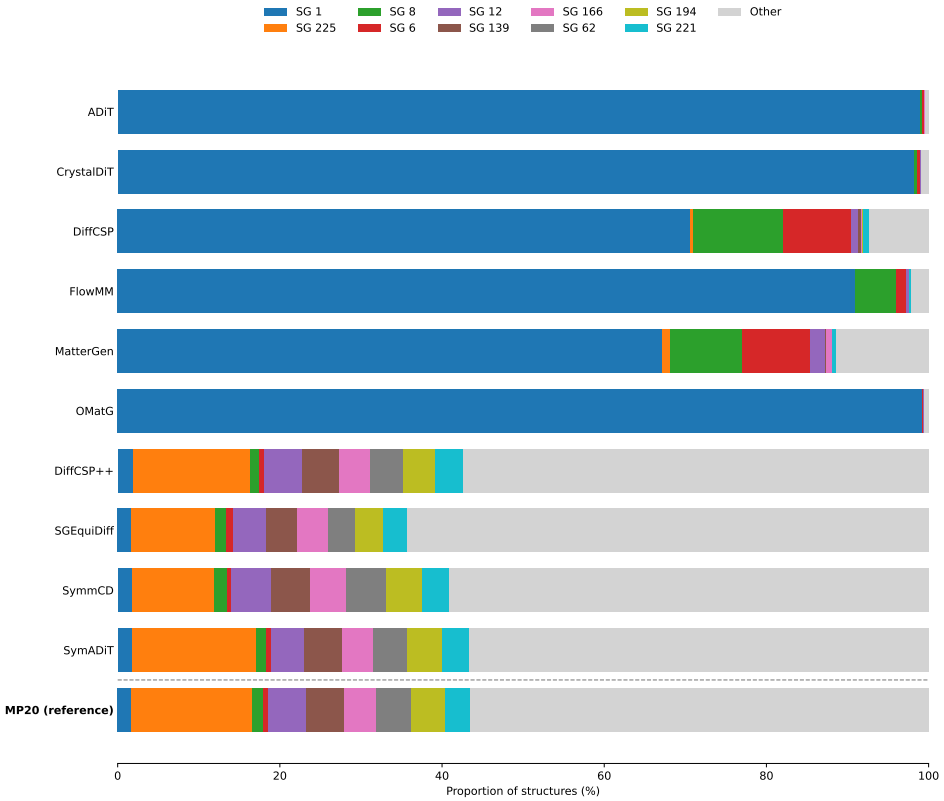


Figure 5: Space group distributions of structures generated by each method in relation to the MP20 dataset. Each horizontal bar represents the full set of generated structures for a given method, with the width of each colored segment indicating the fraction assigned to the corresponding space group. Colors denote the ten most frequent space groups across all evaluated methods, while all remaining space groups are grouped into *Other*.

J Novel Crystal Statistics

In this section, we present some statistical analyses of the novel structures generated by each model, based on 1,000 samples drawn from the set of structurally valid and unique crystals, and compare them to the MP20 training set.

Table 6 showcases compositions in terms of the number of unique elements per structure. Overall, all methods exhibit similar statistics, with mean values close to the training distribution, indicating that compositional complexity is largely preserved across models. SymADiT matches best with the training distribution.

Table 7 reports the occurrence of rare-earth elements across generated structures. RE denotes the number of generated structures that contain at least one rare-earth element, while %RE indicates the corresponding percentage relative to the total number of structures. Mean RE measures the average number of distinct rare-earth elements per structure, providing a finer-grained view of how heavily such elements are incorporated. We observe that most methods produce comparable percentages to the training data, suggesting that rare-earth inclusion is largely consistent across approaches. Notably, CrystalDiT exhibits a higher proportion and mean count of rare-earth elements, which may explain

its increased novelty and could also contribute to its high M/S.U.N. metrics, since fewer comparable crystals are available in the convex hull calculations.

Table 8 reveals a pronounced difference in structural complexity across methods. Symmetry-agnostic models generate structures with substantially higher orbit counts, while symmetry-aware methods produce distributions that closely align with the training data, suggesting more physically plausible crystal structures. This can be explained by the fact that in P1 structures, which symmetry-agnostic methods predominantly sample from, each atom forms its own symmetry orbit, leading to inflated orbit counts.

Table 6: **Composition complexity.** Statistics of the number of unique element types per generated structure.

Method	Sym.	#Struct.	Mean	Std.	Min	Max
SymmCD	Yes	927	3.522	1.241	1	9
DiffCSP++	Yes	898	3.477	1.278	1	10
CrystalDiT	No	651	3.353	1.016	1	7
DiffCSP	No	911	3.334	1.066	1	8
MatterGen	No	918	3.215	0.874	1	7
SGEquiDiff	Yes	895	3.179	0.938	1	8
FlowMM	No	909	3.164	0.801	1	6
ADiT	No	486	3.156	0.765	1	5
SymADiT	Yes	546	3.147	0.764	1	6
MP20 (train)	Yes	27138	3.014	0.726	1	7

Table 7: **Rare-earth composition.** Statistics of rare-earth element occurrence in generated structures.

Method	Sym.	#Struct.	#RE	%RE	Mean RE
SGEquiDiff	Yes	895	307	34.30	0.406
SymmCD	Yes	927	301	32.47	0.392
DiffCSP++	Yes	898	374	41.65	0.490
SymADiT	Yes	546	215	39.38	0.430
ADiT	No	486	192	39.51	0.422
CrystalDiT	No	651	310	47.62	0.922
MatterGen	No	918	362	39.43	0.460
FlowMM	No	909	390	42.90	0.470
DiffCSP	No	911	288	31.61	0.384
MP20 (train)	Yes	27138	11018	40.60	0.440

Table 8: **Structural complexity.** Statistics of the number of symmetry orbits per generated structure.

Method	Sym.	#Struct.	Mean	Std.	Min	Max
ADiT	No	486	12.366	5.074	1	20
CrystalDiT	No	651	10.931	5.675	1	20
DiffCSP	No	911	10.692	5.593	1	20
FlowMM	No	909	10.411	5.327	1	20
MatterGen	No	918	9.584	4.925	1	20
SymADiT	Yes	546	5.328	3.260	1	20
SymmCD	Yes	927	5.028	3.284	1	20
DiffCSP++	Yes	898	5.027	3.203	1	20
SGEquiDiff	Yes	895	4.667	2.761	1	20
MP20 (train)	Yes	27138	4.758	3.012	1	20

K Statistical Analysis

Table 9 summarizes the outcomes of the 500-sample evaluation and reports the corresponding Wilson score confidence intervals. SymADiT achieves the highest number of M/S.U.N. crystals among the symmetry-aware approaches, showing its strong empirical performance so far in the crystal generation pipeline.

The 95% Wilson confidence intervals for the M/S.U.N. rates of SymADiT overlap with those of most competing methods. As a result, the observed differences cannot generally be considered statistically distinguishable, and the confidence-interval analysis alone does not provide sufficient evidence that SymADiT consistently outperforms or underperforms the evaluated models.

To further assess statistical significance, we conducted pairwise two-proportion z-tests on the S.U.N. outcomes, using SymADiT as the reference method. The improvements over SymmCD, DiffCSP, CrystalDiT, MatterGen, and FlowMM are statistically significant ($p < 0.01$). The improvement over DiffCSP++ is also notable, although it falls short of the conventional significance threshold ($p = 0.061$). In contrast, no statistically significant difference is detected between SymADiT and its non-symmetry-aware counterpart, ADiT ($p = 0.145$).

Table 9: Number of S.U.N. and M.S.U.N. crystals, with Wilson 95% confidence intervals (CI). The z- and p-values are obtained from pairwise two-proportion tests on S.U.N. using SymADiT as the reference.

Method	# S.U.N. ↑	# M.S.U.N. ↑	S.U.N. W-CI	M.S.U.N. W-CI	z	p
SGEquiDiff	51	89	7.8–13.2%	14.7–21.4%	-3.14	0.002
SymmCD	49	78	7.5–12.7%	12.7–19.0%	-3.34	< 0.001
DiffCSP++	64	105	10.2–16.0%	17.7–24.8%	-1.87	0.061
ADiT	103	170	17.3–24.4%	30.0–38.3%	$+1.46$	0.145
CrystalDiT	47	152	7.1–12.3%	26.5–34.6%	-3.55	< 0.001
MatterGen	41	126	6.1–10.9%	21.6–29.2%	-4.19	< 0.001
FlowMM	36	106	5.2–9.8%	17.8–25.0%	-4.75	< 0.001
DiffCSP	53	107	8.2–13.6%	18.0–25.2%	-2.93	0.003
OMatG	87	134	14.3–21.0%	23.1–30.8%	$+0.17$	0.867
SymADiT	85	136	14.0–20.5%	23.5–31.3%	0.00	–

L A Discussion regarding the Novelty Metric

In particular, generating a structure that is classified as novel can, in principle, be relatively straightforward: sufficiently perturbing the atomic configuration of a known structure may cause it to no longer match any structure in the training set. Whether such a structure is physically realistic, however, is a separate question, for which stability provides an important indication. This distinction is particularly relevant for the M/S.U.N. metric, which is also commonly reported in the literature. Typically, M/S.U.N. is reported as a conditional fraction among novel structures. Consequently, a high novelty rate can increase the pool of structures entering this metric without necessarily implying that these structures are physically plausible. Such candidates may ultimately fail the subsequent DFT-based stability evaluation.

For this reason, we believe it is more important to report the absolute number of evaluated structures that ultimately satisfy the stability criteria, rather than a conditional fraction. A high novelty rate alone can otherwise give an overly favourable impression if many of the novel structures are physically implausible and subsequently rejected by DFT.

This does not diminish the value of novelty as an evaluation metric. In particular, novelty is useful when combined with computationally inexpensive screening methods, where it can serve as an effective precursor to more expensive DFT calculations by identifying candidates that are distinct from the training data.

Our point is therefore not that novelty should be disregarded, but rather that it should be *interpreted jointly with stability and with the number of structures that successfully pass the full evaluation pipeline*. A high novelty rate by itself does not necessarily imply that a model generates a larger number of physically plausible new crystals.

M Comparison of Inference Speed between SymADiT and ADiT

Comparison of inference speeds between models is not a trivial task – as it is unclear what constitutes a well defined measurement (e.g., time to a single forward pass vs time to a fully generated CIF file). Furthermore, each method applies different post-processing steps. The exception in this work is the comparison between SymADiT and ADiT. The two generation pipelines are similar in how they simulate the ODE and apply the decoder, enabling a meaningful comparison. This experiment will also demonstrate the inference-speed benefits of reducing the number of tokens, specifically by using symmetry orbits as Transformer tokens rather than individual atoms.

For both models, we generate 10,000 crystals on a single NVIDIA H200 GPU using a maximum batch size of 1,024 and 500 ODE steps. The models also have comparable parameter counts. To make the timing comparison as consistent as possible, we exclude model-specific sanity checks and filtering, such as implausible lattice parameters or atomic overlaps, since these differ between the pipelines. For ADiT, we evaluate only periodic crystal generation and exclude molecular generation.

We decided to divide the evaluation of inference time into two parts:

1. Model inference. We measure the time required to produce the 10,000 decoded outputs needed to construct all crystals. This includes the iterative ODE sampling procedure and the VAE decoder at the end.
2. Post-processing. We measure the time required to convert the 10,000 decoded model outputs into explicit pymatgen crystal structures. Here, the two models require different post-processing steps because their outputs represent different quantities. For example, SymADiT predicts Wyckoff positions and needs to post-process through Pyxtal, whereas ADiT does not.

Table 10: **Generation time.** Model inference, post-processing, and total generation time for 10,000 generated crystals. All times are reported in seconds.

Method	Model inference (s)	Post-processing (s)	Total time (s)
SymADiT	263.9	90.8	354.7
ADiT	849.4	2.3	851.7

Results are presented in Table 10. For model inference, SymADiT achieves 3.22x speedup in inference time. The improvement is consistent with SymADiT representing crystals using symmetry orbits rather than individual atoms, which on average results in fewer tokens and therefore reduces the computational cost of the iterative denoising process.

For post-processing, ADiT is much faster, as it can go straight-forwardly from the model output to a pymatgen crystal structure, while in our implementation we rely on converting the model output from ASU to full unit cell via PyXtal, which is slower.

Despite the longer post-processing time, SymADiT still achieves a 2.40× inference speed-up over ADiT.

N Visual Results

In Figure 6, we present S.U.N and M.S.U.N crystals from SymADiT. The generated structures show a wide range of space groups and compositions. Visualisation was done using the VESTA software [53].

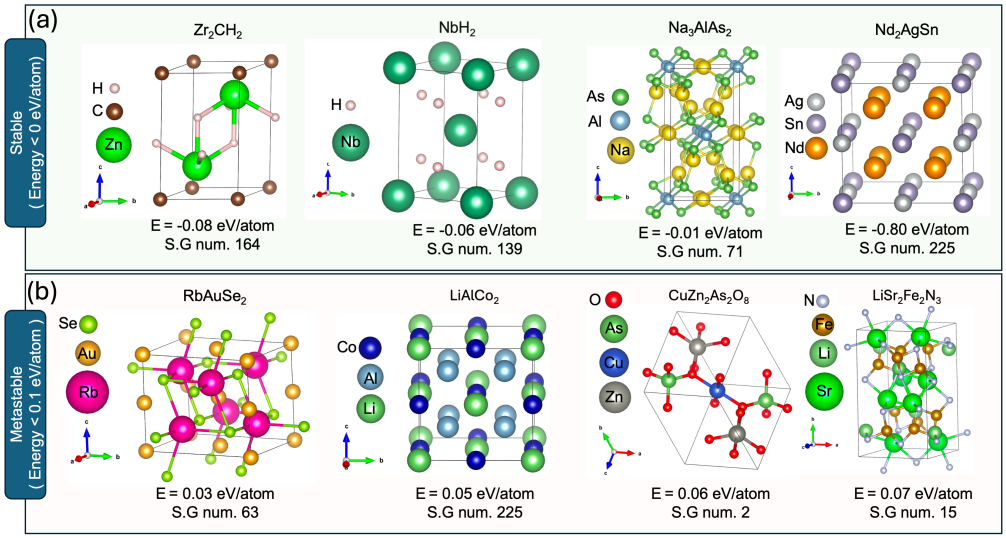


Figure 6: Schematic illustration of crystals identified by SymADiT: (a) thermodynamically stable structures (S.U.N) with $\Delta H_{cp} < 0$, and (b) metastable structures M/S.U.N. with $0 \leq \Delta H_{cp} < 0.1$ eV/atom.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claims in the abstract and introduction are consistent with the proposed method and supported by the experimental results presented in the paper. Some limitations is discussed in Section 6.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Some limitations are discussed in Section 6.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [N/A]

   Justification: The paper does not include any theoretical results or formal proofs.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Training settings and (publicly available) datasets are detailed in the paper. We plan to release after acceptance the code, models, and the exact data splits for reproducibility.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We plan to release our code once accepted. All data is publicly available.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: All training and test details are detailed. We used the reported splits of the publicly available datasets, and will release the split for the third dataset.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: Due to computational resources, we were not able to run our models and multiple models over different seeds. The same goes with running more DFT calculations, which we discuss in Section 6.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

Justification: We include which GPUs we have used for training and generation, along with other important training details such as batch size.

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have reviwed the code of ethics.

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: This paper presents work whose goal is to advance the field of machine Learning in material science. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: This is not a concern with the used datasets and the released models.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Citations are provided, and we discuss licenses in the appendix.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: We will release our code and models, and details around this once accepted.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: We did not involve such experiments.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: We did not involve such experiments.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: We used LLMs for improving the writing and producing code for visualizations.