

Multimodal Crystal Flow: Any-to-Any Modality Generation for Unified Crystal Modeling

Kiyoung Seong^{1,2} Sungsoo Ahn¹ Sehui Han² Changyoung Park²

Abstract

Crystal modeling spans a family of conditional and unconditional generation tasks, including crystal structure prediction (CSP) and *de novo* generation (DNG). While recent deep generative models have shown promising performance, they remain largely task-specific, lacking a unified framework that shares crystal representations across tasks. To address this limitation, we propose *Multimodal Crystal Flow (MCFlow)*, a unified multimodal flow model that realizes multiple crystal generation tasks as distinct inference trajectories via independent time variables for atom types and crystal structures. To enable multimodal flow in a standard transformer model, we introduce a composition- and symmetry-aware atom ordering with hierarchical permutation augmentation, injecting compositional and crystallographic priors without explicit structural templates. Experiments on the MP-20 and MPTS-52 benchmarks show that a single MCFlow model is competitive with task-specific baselines across CSP, DNG, and structure-conditioned atom type generation.

1. Introduction

Crystal discovery based on atomic structures is vital for developing next-generation functional materials such as battery electrolytes and catalysts (Butler et al., 2018). However, traditional approaches are computationally prohibitive due to expensive density functional theory (DFT) calculations (Glass et al., 2006; Wang et al., 2010; Pickard & Needs, 2011; Wang et al., 2012). To address this issue, prior work has introduced generative models that learn distributions of stable crystals from data under various conditional or unconditional generation schemes (Nouira et al., 2018; Noh

et al., 2019; Xie et al., 2022; Jiao et al., 2023; Miller et al., 2024). Recent methods have further incorporated crystallographic symmetry through structural templates such as Wyckoff positions (Zhu et al., 2024; Jiao et al., 2024).

At a high level, crystal modeling spans a family of generation problems across different modalities such as atom types and crystal structures. Depending on which modalities are observed, this formulation naturally covers crystal structure prediction (CSP), *de novo* generation (DNG), and structure-conditioned atom type generation. This perspective aligns with a broader trend in generative modeling toward unified frameworks where a single model supports multiple conditional generation tasks across modalities (Campbell et al., 2024; Rojas et al., 2025). However, existing crystal generative models adopt task-specific formulations, necessitating separate training for different generation settings and hindering a unified formulation that reuses the same underlying crystal representations across tasks.

In this work, we introduce *Multimodal Crystal Flow (MCFlow)*, a unified generative model that enables any-to-any modality generation across atom types and crystal structures (Figure 1). The key insight is that by decoupling flow time axes for different modalities, a single model can flexibly condition on any subset of modalities while generating the rest, without retraining or task-specific architectures. This design allows a single MCFlow model to support CSP, DNG, and atom type generation.

We instantiate MCFlow using a diffusion transformer (DiT) architecture (Peebles & Xie, 2023), following recent successes in all-atom molecular and crystalline modeling (Joshi et al., 2025; Frank et al., 2025; Yi et al., 2025; Kim et al., 2026a;b;c; Morehead et al., 2026). However, the DiT lacks inductive biases for crystallographic symmetry and permutation invariance. To address this, we introduce a composition- and symmetry-aware atom ordering and hierarchical permutation augmentation. By lexicographically sorting atoms based on electronegativity and Wyckoff positions, we enable the DiT to internalize chemical and crystallographic structure from data without relying on explicit structural templates. The hierarchical permutation augmentation then enforces invariance across and within symmetry-equivalent orbits, providing a structured inductive bias over composi-

¹Graduate School of AI, KAIST, Seoul, South Korea ²Materials Intelligence Lab, LG AI Research, Seoul, South Korea. Correspondence to: Changyoung Park <changyoung.park@lgresearch.ai>.

Proceedings of the 43rd International Conference on Machine Learning, Seoul, South Korea. PMLR 306, 2026. Copyright 2026 by the author(s).

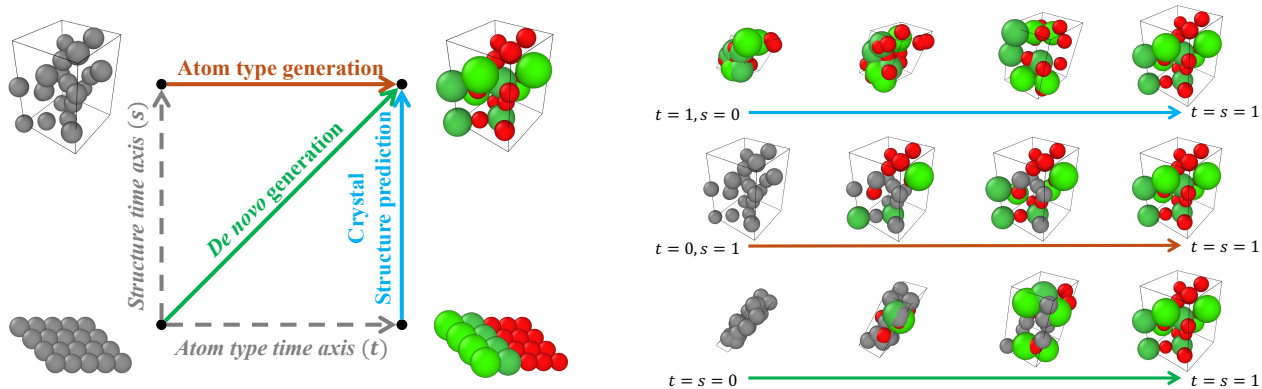


Figure 1. Overview of multimodal crystal flow with any-to-any modality generation. MCFlow trains a flow model with two independent time variables corresponding to atom types (t) and structures (s). By selecting task-specific inference trajectories in the (t, s) space, a single model performs crystal structure prediction, atom type generation, and *de novo* generation.

tionally and crystallographically equivalent atoms.

Our main contributions are as follows:

- We introduce *MCFlow*, a unified multimodal flow framework for crystal generation that enables any-to-any modality generation across atom types and crystal structures within a single model.
- We propose a composition- and symmetry-aware atom ordering with hierarchical inter- and intra-orbit permutation augmentation, enabling DiT architectures to exploit compositional and crystallographic priors from data without explicit symmetry templates.
- We demonstrate that a single MCFlow model achieves competitive performance across crystal structure prediction, *de novo* generation, and structure-conditioned atom type generation.
- Building on structure-conditioned atom type generation, we further enable atom type substitution by fixing atom types at known sites and sampling target sites from the same trained model.

Conflict of Interest Disclosure. Authors S. Han and C. Park are employed by LG AI Research, and K. Seong conducted this research during an internship at LG AI Research. MCFlow, the model proposed and evaluated in this paper, was developed at LG AI Research.

2. Related Work

Crystal generative models. Recent advances in diffusion- and flow-based generative models have significantly improved the fidelity and stability of generated crystal structures. Diffusion models (Xie et al., 2022; Jiao et al., 2023; Yang et al., 2024; Zeni et al., 2025) model crystal generation as iterative denoising of atomic coordinates in a prim-

itive unit cell. More recent work has explored flow-based models (Miller et al., 2024; Sriram et al., 2024; Luo et al., 2025), stochastic interpolants (Hoellmer et al., 2025), periodic Bayesian flows (Wu et al., 2025), and generative flow networks (Hernandez-Garcia et al., 2023). ADiT (Joshi et al., 2025) and Zatom-1 (Morehead et al., 2026) use transformer architectures for unified all-atom modeling of molecules and crystals. Large language models have also been adapted for crystal generation by fine-tuning on CIF-formatted data (Gruber et al., 2024; Mishra et al., 2026) or incorporating text guidance and hybrid LLM-diffusion pipelines (Ding et al., 2024; Park et al., 2025; Das et al., 2025; Khastagir et al., 2025). However, these models generally lack explicit inductive biases for space-group symmetry and Wyckoff structure.

Alternative approaches enforce symmetry by generating structural templates via autoregressive Wyckoff position prediction (Zhu et al., 2024; Cao et al., 2025; Kazeev et al., 2025; Kelvinius et al., 2025; Xu et al., 2025) or restricting generation to the asymmetric unit (ASU) (Jiao et al., 2024; Levy et al., 2025; Chang et al., 2025; Puny et al., 2025) conditioned on structural templates. While these methods satisfy crystallographic constraints, they are tailored to specific conditional generation settings and often rely on task-dependent structural template generation or retrieval modules. In contrast, MCFlow handles CSP, DNG, and atom type generation in a single end-to-end model by incorporating crystallographic symmetry as a soft inductive bias without structural template modules.

Any-to-any multimodal generative models. Recent diffusion- and flow-based generative frameworks increasingly adopt unified multimodal formulations to jointly model heterogeneous modalities. By decoupling modality-wise time axes, these frameworks enable flexible conditional generation, including any-to-any settings. Multiflow (Campbell et al., 2024) applies this paradigm to jointly model

amino acid sequences and structures for protein co-design, while [Rojas et al. \(2025\)](#) generalize multimodal diffusion models to arbitrary state spaces with applications to text-image domains. Despite their success, extending such frameworks to crystalline materials remains nontrivial due to periodic boundary conditions and crystallographic symmetry. MCFlow extends this multimodal generative paradigm to periodic crystal systems by decoupling atom types and crystal structures with modality-wise time axes, enabling a unified framework that supports CSP, DNG, and atom type generation without task-specific architectures or retraining.

3. Method

In this section, we present a multimodal flow framework for unified crystal generation that supports flexible conditional generation across atom types and crystal structures within a single model. The core idea is to decouple the generative processes of different modalities through modality-specific time parameterization, allowing diverse crystal generation tasks to be realized within a shared model. Building on this formulation, we introduce a composition- and symmetry-aware atom ordering together with permutation augmentation as a soft inductive bias for transformer-based architectures. Background on crystal representations and flow models is provided in [Appendix A](#).

3.1. Multimodal Flow Model for Crystals

Our goal is to generate arbitrary subsets of modalities while conditioning on the others. Standard diffusion or flow models rely on a single shared time variable, which couples all modalities to the same noise level and restricts conditional generation. We address this limitation by decoupling flow time variables to atom types and crystal structures and learning a joint generative process over a two-dimensional time space. Intuitively, this decoupling allows the model to hold one modality fixed while sampling the other, providing controllable conditional generation. Different crystal generation tasks then correspond to different inference trajectories in the (t, s) plane ([Figure 1](#)): CSP fixes $t = 1$ and integrates s from 0 to 1, atom type generation fixes $s = 1$ and integrates t , and DNG integrates (t, s) jointly along the diagonal.

Multimodal flow model. We adopt a multimodal flow model with decoupled flow time variables $t, s \in [0, 1]$ corresponding to atom types $A \in \mathbb{Z}_+^N$ and crystal structures $X = (F, L^l, L^a)$ in the primitive unit cell, which consists of fractional coordinates $F \in [0, 1]^{3 \times N}$, Niggli-reduced lattice lengths $L^l = (a, b, c) \in \mathbb{R}^{1 \times 3}$, and angles $L^a = (\alpha, \beta, \gamma) \in [60^\circ, 120^\circ]^{1 \times 3}$ ([Grosse-Kunstleve et al., 2004](#)). After training the flow model, we select an inference trajectory in a (t, s) time plane to realize different crystal discovery tasks with observed modalities fixed at 1.

To describe the multimodal flow model, we first design the joint probability path $p_{t,s}(A_t, X_s)$ defined by marginalizing modality-wise conditional paths:

$$p_{t,s}(A_t, X_s) = \int p_t(A_t | A_0, A_1) p_s(X_s | X_0, X_1) d\pi, \quad (1)$$

where $\pi(A_0, X_0, A_1, X_1) = p_0(A_0, X_0)p_1(A_1, X_1)$ is the independent coupling between joint base p_0 and data p_1 distributions over atom types and structures. By construction, this path recovers the joint base p_0 at $(t, s) = (0, 0)$ and the joint data p_1 at $(t, s) = (1, 1)$, interpolating between them through independent time evolution along each modality.

This joint path evolves through a continuous-time Markov chain (CTMC) over atom types along the t -axis and an ordinary differential equation (ODE) over crystal structures along the s -axis, expressed as:

$$p_{t+h|t,s}(\cdot | A_t, X_s) = \delta_{A_t} + h u^A(A_t, X_s, t, s) + o(h), \quad (2)$$

$$\frac{dX_s}{ds} = u^X(A_t, X_s, t, s), \quad (3)$$

where δ_{A_t} denotes the probability vector supported on the current state A_t , $u^A(A_t, X_s, t, s)$ denotes the rate vector of the CTMC over atom types in a discrete state space, $u^X(A_t, X_s, t, s)$ denotes the velocity field of the ODE over crystal structures in a continuous state space, and $o(h)$ denotes residual terms that vanish as $h \rightarrow 0$. The t -axis governs atom type updates conditioned on crystal structures, whereas the s -axis handles crystal structure updates conditioned on atom types.

Multimodal flow matching. We adopt a flow matching objective that minimizes a weighted sum of divergences between the model prediction and the modality-specific target velocity field or rate vector:

$$\mathcal{L}(\theta) = \mathbb{E}_{t,s,A_t,X_s} \left[\sum_{m \in \{A,X\}} w_m D_m(u_\theta^m, u^m) \right], \quad (4)$$

where $u_\theta^m := u_\theta^m(A_t, X_s, t, s)$ and $u^m := u^m(A_t, X_s, t, s)$, D_A is generalized KL divergence for atom types and D_X is squared Euclidean distance for crystal structures.

For tractable supervision, we match to modality-wise conditional targets $u^A(A_t | A_0, A_1)$ and $u^X(X_s | X_0, X_1)$ rather than directly to the marginal targets $u^A(A_t, X_s, t, s)$ and $u^X(A_t, X_s, t, s)$, since the marginal targets are the posterior expectation of the modality-wise conditional targets under endpoint coupling

$$u^A = \mathbb{E}_{A_0, A_1} [u^A(A_t | A_0, A_1) | A_t, X_s], \quad (5)$$

$$u^X = \mathbb{E}_{X_0, X_1} [u^X(X_s | X_0, X_1) | A_t, X_s]. \quad (6)$$

	Index	Element	Wyckoff_position	Orbit_index	X	Y	Z	
	0	Th	a	0	0.0000	0.0000	0.0000	
Inter-orbit permutation	1	Th	c	1	0.6018	0.6018	0.1945	Intra-orbit permutation
	2	Th	c	1	0.3982	0.3982	0.8055	
	3	Th	c	2	0.2003	0.2003	0.3992	
	4	Th	c	2	0.7997	0.7997	0.6008	
	5	C	b	3	0.5000	0.5000	0.5000	

Figure 2. **Composition- and symmetry-aware atom ordering with hierarchical permutation augmentation.** Atoms in the primitive unit cell are lexicographically ordered by Pauling electronegativity and Wyckoff position (denoted by letter $a, b, c, \dots$) to expose compositional and crystallographic structure. The ordering and augmentation are illustrated on a Th_5C crystal in the $R\bar{3}m$ space group. Inter-orbit permutation reorders orbits sharing the same elements and Wyckoff position, while intra-orbit permutation permutes atoms within each orbit. This ordering and hierarchical permutation augmentation provides compositional and crystallographic symmetry information to sequence-based Transformers without explicitly conditioning space group or Wyckoff positions.

Under these divergences, the conditional flow matching objective induces the same gradient with respect to the model parameters as matching the marginal target implied by this posterior expectation (Lipman et al., 2024). We construct conditional target velocity fields or rate vector as follows: straight line flows between prior and target samples for lattice parameter, geodesic flows between prior and target samples on the torus manifold $\mathbb{T}^{3 \times N}$ for fractional coordinates (Miller et al., 2024) to enforce periodicity, and mixture paths between two point masses of prior and target samples for atom types; full expressions are provided in Appendix B.

Parameterization and base distributions. We adopt clean data prediction across all modalities instead of predicting noise or velocity targets, motivated by the manifold structure of data distribution (Li & He, 2025). For fractional coordinates, we predict fractional coordinates F_θ and project them onto the torus $\mathbb{T}^{3 \times N}$ via $\text{proj}(F) = F \pmod{1}$. We use a uniform base distribution on $\mathbb{T}^{3 \times N}$ for fractional coordinates. For lattice length and angles, we predict directly L_θ^l and L_θ^a , using LogNormal priors for lengths to ensure positivity (Miller et al., 2024) and uniform priors for angles. For atom types, we predict atom type logits A_θ with a mask-token point mass δ_M as base distribution.

Diffusion transformer. For all-atom modeling of crystals, we adopt a Diffusion Transformer (DiT) backbone (Peebles & Xie, 2023) with modality-specific time conditioning and prediction heads tailored to the proposed multimodal flow. This choice follows recent successes of transformer architectures in all-atom molecular and crystalline modeling (Joshi et al., 2025; Frank et al., 2025; Morehead et al., 2026), providing a scalable backbone that uniformly handles heterogeneous modalities through attention. We embed all modalities and add a fixed sinusoidal positional embedding PosEmb (Vaswani et al., 2017) indexed by the atom position in the proposed atom ordering (Section 3.2) to obtain a

token embedding:

$$z = \text{Emb}(A) + \text{Lin}(F) + \text{Lin}(L^l) + \text{Lin}(L^a) + \text{PosEmb}, \quad (7)$$

where $\text{Lin}(\cdot)$ denotes modality-specific linear layer and lattice embeddings are broadcasted to all atom tokens. The joint conditioning vector for learnable layer normalization is a sum of modality-wise time embeddings

$$c = \text{TimeEmb}(t) + \text{TimeEmb}(s), \quad (8)$$

where $\text{TimeEmb}(\cdot)$ denotes modality-specific time embedding layer. We feed the multimodal embedding vector z and the conditioning vector c into the DiT backbone, $h = \text{DiT}(z, c)$. From the token-wise representations, we linearly predict atom type logits and fractional coordinates,

$$A_\theta = \text{Lin}(h), \quad F_\theta = \text{Lin}(h). \quad (9)$$

Global lattice parameters are predicted via mean pooling.

$$L_\theta^l = \text{Lin}(\bar{h}), \quad L_\theta^a = \text{Lin}(\bar{h}), \quad \bar{h} = \frac{1}{N} \sum_{i=1}^N h_i. \quad (10)$$

The final velocity field and rate vector are constructed from these outputs as provided in Appendix B. Training and inference algorithms are listed in Appendix C. Since crystals have variable numbers of atoms, we pad sequences within a batch and use attention masks; the pooling for lattice heads is computed over non-padded atom tokens only.

Inference-time guidance. To improve sample fidelity in our multimodal flow framework, we adopt noisy guidance (NG; Rojas et al., 2025), originally introduced for diffusion score models, and apply it to the velocity field of multimodal flow as an inference-time technique for conditional generation. This follows from the probability flow ODE (Song et al., 2020), where the score function and the velocity field are deterministically related. For crystal structure prediction,

the guided velocity field is given by

$$u_{\text{ng}}^X(A_1, X_s, 1, s) = (1 - \omega) u^X(A_\sigma, X_s, \sigma, s) + \omega u^X(A_1, X_s, 1, s). \quad (11)$$

where σ controls the corruption level of the conditioning modality and ω denotes the guidance scale, biasing inference trajectories toward the regions favored by the conditional velocity field. The noisy guidance procedure is summarized in Algorithm 3, with additional ablation and sensitivity analyses provided in Appendix E.

3.2. Composition- and Symmetry-Aware Atom Ordering and Hierarchical Augmentation

This subsection introduces an atom ordering and augmentation strategy for the DiT architecture, designed to support multimodal flow models by incorporating compositional and crystallographic symmetry information.

Prior work has proposed ordering atoms in crystal structures according to their Wyckoff positions for an autoregressive transformer model, leveraging crystallographic symmetry (Cao et al., 2025). While effective, directly applying Wyckoff-based ordering to a DiT-based multimodal flow model suffers from two limitations. First, in crystal structure prediction (CSP), where only the composition is given, the absence of structural symmetry information at inference time leads to ambiguity in defining the initial atom ordering for the DiT. Second, this ordering does not explicitly encode chemical information, which can limit the model’s ability to learn compositionally valid crystals.

Composition- and symmetry-aware atom ordering. To address these limitations, we propose a lexicographical atom ordering that jointly incorporates compositional and crystallographic information. As illustrated in Figure 2, we first partition atoms into groups $O = \{g \cdot f \mid g \in G\}$ under the action of the space group G , referred to as *orbits*. Each orbit represents compositionally and symmetry-equivalent atomic sites sharing the same atom type and Wyckoff position.

We then impose a lexicographical ordering over these orbits, sorting them first by the Pauling electronegativity of their atom type and then by Wyckoff position, as defined in the International Tables for Crystallography (Prince, 2004). Electronegativity provides a chemical signal that correlates with bonding preferences and charge-transfer tendencies, while Wyckoff positions capture crystallographic symmetry with earlier letters indicating higher site symmetry. This ordering injects a soft inductive bias by arranging atoms in a symmetry-aligned and contiguous sequence, enabling the DiT to learn symmetry without explicit conditioning on space group or Wyckoff positions.

Augmentation via inter- and intra-orbit permutation. While the proposed ordering captures compositional and

Table 1. Crystal structure prediction on MP-20 and MPTS-52. MCFlow achieves the best results on MPTS-52 across all metrics and is competitive on MP-20, while achieving consistently lower RMSE. We report match rate (MR) and RMSE for single-sample (1) and multi-sample (20) generation. S/B/L denote small/base/large model sizes. Noisy guidance is applied only for single-sample generation. **Bold** indicates the best value within each column across all models and sampling budgets; underline indicates the best value within the single-sample group.

Model	MP-20		MPTS-52	
	MR (%) $\uparrow$	RMSE $\downarrow$	MR (%) $\uparrow$	RMSE $\downarrow$
CDVAE (1)	33.90	0.1045	5.34	0.2106
DiffCSP (1)	51.49	0.0631	12.19	0.1786
FlowMM (1)	61.39	0.0566	17.54	0.1726
CrystalFlow (1)	62.02	0.0710	21.00	0.1613
OMatG (1)	63.75	0.0720	25.15	0.1931
DiffCSP (20)	77.93	0.0492	34.02	0.1749
CrystalFlow (20)	78.34	0.0577	37.81	0.1584
MCFlow-S (1)	56.10	0.1269	22.88	0.1845
MCFlow-B (1)	63.14	0.0611	26.46	0.1572
MCFlow-L (1)	<u>64.08</u>	<u>0.0561</u>	<u>27.16</u>	<u>0.1401</u>
MCFlow-S (20)	77.84	0.0939	38.91	0.1717
MCFlow-B (20)	77.12	0.0439	41.40	0.1465
MCFlow-L (20)	76.08	0.0383	41.45	0.1296

symmetry information, permutation ambiguity remains among tied orbits and atoms within the same orbit. We address this by introducing a hierarchical permutation augmentation scheme that preserves compositional and crystallographic equivalence. For each group of orbits $\mathcal{W}^i = (O_1^i, \dots, O_m^i)$ sharing the same atom type and Wyckoff position in a primitive unit cell, we apply inter-orbit permutation across orbits followed by intra-orbit permutation within each orbit. This reduces the permutation space from $N!$ to $\prod_i |\mathcal{W}^i|! \prod_j |O_j^i|!$, with a quantitative analysis provided in Appendix D. We additionally apply a global modulo translation to account for translational invariance.

4. Experiments

We aim to answer two research questions:

- Can a single MCFlow model achieve competitive performance across crystal structure prediction (CSP), *de novo* generation (DNG), and structure-conditioned atom type generation?
- Do the proposed composition- and symmetry-aware atom ordering and symmetry-preserving hierarchical permutation augmentations enable a transformer-based model to capture crystallographic constraints without relying on explicit symmetry templates?

To this end, we evaluate MCFlow on all three tasks using the same trained model and conduct ablation studies to assess

Table 2. De novo generation evaluated on LeMat-GenBench (Betala et al., 2025) with MP-20 training. MCFlow achieves competitive or superior performance across most metrics both with and without MLIP pre-relaxation. Reported values are averaged over an ensemble of three MLIPs (MACE-MP, UMA, Orb-v3) with LeMat-Bulk as the reference set, with all percentages computed over the 2,500 submitted structures following the LeMat-GenBench convention. Standard deviations are computed across the 2,500 generated structures, averaged over the three MLIPs. SUN and MSUN values are reported as percentage (count). Results are organized into two groups: models with and without MLIP pre-relaxation; MCFlow uses Orb-v3 for pre-relaxation. **Bold** and underline indicate the best and second-best results within each group.

Model	# Params	Valid (%) $\uparrow$	Unique (%) $\uparrow$	Novel (%) $\uparrow$	Energy-based			Stability		Metastability	
					E_f (eV) $\downarrow$	E_{hull} (eV) $\downarrow$	RMSD (Å) $\downarrow$	Stable (%) $\uparrow$	SUN (%) $\uparrow$	Meta (%) $\uparrow$	MSUN (%) $\uparrow$
<i>With pre-relaxation</i>											
MatterGen	47M	95.7	95.1	70.5	-0.70 ± 0.79	0.18 ± 0.18	0.39 ± 0.50	2.0	0.2 (6)	33.4	<u>15.0</u>
PLaID++	7B	96.0	77.8	24.2	-0.50 ± 0.44	0.09 ± 0.16	0.13 ± 0.29	12.4	1.0 (26)	60.7	7.6
WyFormer	8B	93.4	93.0	<u>66.4</u>	-0.43 ± 0.95	0.50 ± 0.51	0.81 ± 0.98	0.5	0.1 (3)	15.7	1.9
WyFormer-DFT	8B	95.2	95.0	<u>66.4</u>	-0.67 ± 0.91	0.27 ± 0.36	0.42 ± 0.60	3.7	0.4 (9)	24.8	7.8
MCFlow-S	34M	97.2	96.3	52.2	-0.85 ± 0.84	0.10 ± 0.12	0.16 ± 0.27	11.7	0.7 (17)	49.5	18.9
MCFlow-B	134M	<u>97.7</u>	<u>95.5</u>	25.4	<u>-0.91 ± 0.85</u>	<u>0.05 ± 0.10</u>	<u>0.08 ± 0.18</u>	<u>17.6</u>	<u>0.7 (17)</u>	<u>64.3</u>	11.9
MCFlow-L	464M	98.6	95.2	18.6	<u>-0.93 ± 0.87</u>	0.04 ± 0.08	0.06 ± 0.15	18.8	0.5 (13)	68.3	9.3
<i>Without pre-relaxation</i>											
ADiT	180M	90.6	87.8	26.0	-0.73 ± 0.93	0.33 ± 0.45	0.38 ± 0.40	0.4	0.0 (0)	36.5	1.0
Crystal-GFN	1M	51.7	51.7	51.7	-1.30 ± 2.63	2.09 ± 2.38	1.87 ± 0.97	0.0	0.0 (0)	0.0	0.0
CrystalFormer	5M	69.9	69.4	31.8	-0.17 ± 1.48	0.70 ± 1.33	0.66 ± 1.05	1.4	0.0 (0)	28.8	3.1
DiffCSP	12M	<u>95.7</u>	<u>94.8</u>	66.2	-0.64 ± 0.96	0.28 ± 0.56	0.59 ± 0.63	2.3	0.1 (3)	29.8	8.5
DiffCSP++	12M	95.3	95.1	<u>62.0</u>	-0.52 ± 0.87	0.41 ± 0.44	0.69 ± 0.82	1.0	0.2 (4)	26.4	<u>5.0</u>
LLaMat2-CIF	7B	84.4	81.4	30.0	-0.47 ± 1.01	0.44 ± 0.71	0.54 ± 0.71	0.7	0.1 (3)	34.7	2.1
LLaMat3-CIF	8B	15.4	15.2	10.5	0.74 ± 2.69	1.71 ± 2.48	1.01 ± 1.10	0.1	0.0 (0)	2.1	0.2
SymmCD	12M	73.4	73.0	47.0	-0.02 ± 1.22	0.88 ± 1.07	0.87 ± 1.09	1.4	0.1 (2)	18.6	2.4
MCFlow-S	34M	86.5	85.8	48.2	-0.42 ± 0.95	0.49 ± 0.59	0.64 ± 0.61	1.2	0.1 (2)	30.7	2.5
MCFlow-B	134M	93.3	91.2	25.1	-0.76 ± 0.88	<u>0.18 ± 0.35</u>	<u>0.29 ± 0.44</u>	<u>4.6</u>	0.2 (4)	<u>60.4</u>	3.2
MCFlow-L	464M	96.0	92.9	19.4	<u>-0.84 ± 0.88</u>	0.13 ± 0.27	0.23 ± 0.37	4.9	0.1 (2)	69.7	3.7

the contribution of each component. Detailed experimental configurations and metrics are in Appendices F and G.

Dataset. Our main evaluation uses two inorganic crystal benchmarks that support all three tasks (CSP, DNG, ATG): *MP-20* (Xie et al., 2022) and *MPTS-52* (Baird et al., 2024). *MP-20* comprises all materials from the Materials Project database (Jain et al., 2013) as of July 2021 that contain no more than 20 atoms per unit cell and reside within 0.08 eV/atom of the convex hull energy. We also test on *MPTS-52*, a dataset consisting of structures with up to 52 atoms per unit cell. For both datasets, we follow a consistent 60/20/20 ratio for training, validation, and test sets. *MPTS-52* dataset is partitioned into time slices based on their initial publication year to reflect a chronological discovery.

Setup. MCFlow is based on a Diffusion Transformer (DiT) backbone (Peebles & Xie, 2023) for all-atom crystal modeling. We evaluate three model scales: small (34M parameters, hidden dimension 384, 6 attention heads, 12 layers), base (134M parameters, 768 dimension, 12 heads, 12 layers), and large (464M parameters, 1024 dimension, 16 heads, 24 layers), denoted as MCFlow-S, -B, and -L, respectively. Throughout this paper, MCFlow refers to the base model unless its scale is explicitly stated. Detailed model, training, and inference configurations are provided in Appendix F. Computational cost analysis is provided in Table H.1.

4.1. Crystal Structure Prediction

CSP predicts stable crystal structures given a chemical composition: the number of atoms and their types.

Metrics. Following the metric calculation of Jiao et al. (2023); Miller et al. (2024), we evaluate the accuracy of the generated samples against the ground truth structures of the test set using the StructureMatcher (Ong et al., 2013). We report the Match Rate (MR), defined as the fraction of valid candidates that match their ground truth counterparts. For matched pairs, we also calculate the Root Mean Square Error (RMSE) to assess the precision of the predicted atomic locations within the unit cell. We additionally report METRe (Martirosyan et al., 2025) on a polymorph-aware *MP-20* split, which accounts for compositions admitting multiple stable structures (Table H.2).

Baselines. We compare MCFlow against baselines: CD-VAE (Xie et al., 2022), DiffCSP (Jiao et al., 2023), FlowMM (Miller et al., 2024), CrystalFlow (Luo et al., 2025), and OMatG (Hoellmer et al., 2025), for which we report the best-performing LinearODE variant. We exclude FlowLLM (Sriram et al., 2024), SymmCD (Levy et al., 2025), and ADiT (Joshi et al., 2025), as they do not support CSP. We also exclude DiffCSP++ (Jiao et al., 2024), SGEquiDiff (Chang et al., 2025), and SGFM (Puny et al., 2025), which rely on structural templates extracted from the target structure or retrieved via CSPML (Kusaba et al., 2022). For a given composition, multiple templates often exist, while one template leads to a nearly unique stable structure (Kazeev et al., 2025), making template-conditioned gen-

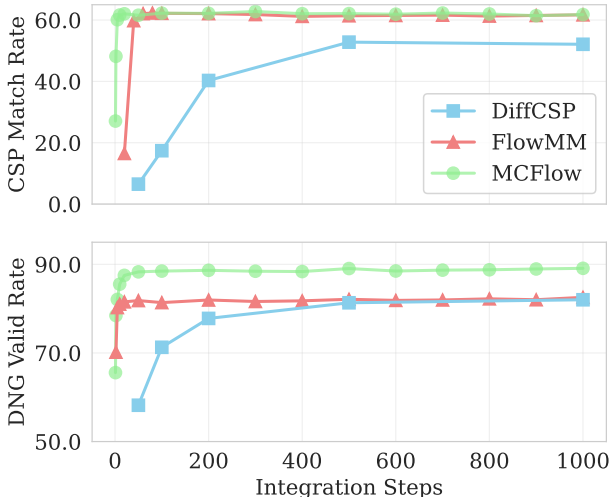


Figure 3. **Effect of the number of integration steps on performance.** MCFlow reaches competitive crystal structure prediction (CSP) match rate and *de novo* generation (DNG) validity with far fewer integration steps than FlowMM and DiffCSP.

eration a different problem setting from full CSP. We further compare our method under target-template filtering against these target template-conditioned methods (Table H.3).

Results. MCFlow achieves competitive performance on MP-20 and improved performance on MPTS-52, which contains larger unit cells (Table 1). The proposed ordering and augmentation drive these gains (Tables 5 and H.4); the larger margin on MPTS-52 reflects the DiT’s scaling advantage on longer sequences. MCFlow saturates significantly earlier than FlowMM and DiffCSP in CSP match rate (Figure 3). We also observe consistent scaling behavior across model sizes, as detailed in Figure H.1; on MP-20, multi-sample match rate saturates around the base model while RMSE continues to improve with scale, whereas MPTS-52 benefits from scale across both metrics. Noisy guidance (NG) further improves single-sample generation accuracy, as demonstrated by the ablation study in Appendix E. Further qualitative samples illustrating structural diversity are provided in Appendix I. MCFlow also attains the highest polymorph-aware match rate (Table H.2) and leads under both filtered and unfiltered match rate evaluations (Table H.5).

4.2. De Novo Generation

In the DNG task, the model generates atom types, lattice parameters, and fractional coordinates at once.

Metrics. We evaluate DNG using the standardized LeMat-GenBench (Betala et al., 2025) protocol, which provides a unified evaluation pipeline across models. All metrics are computed for 2,500 sampled crystals using an ensemble of three MLIPs—MACE-MP (Batatia et al., 2025),

UMA (Wood et al., 2025), Orb-v3 (Rhodes et al., 2025)—with LeMat-Bulk (Siron et al., 2025) as the reference set. We report validity, uniqueness, novelty, formation energy (E_f), energy above hull (E_{hull}), relaxation RMSD, stability ($E_{\text{hull}} \leq 0$ eV/atom), metastability ($0 < E_{\text{hull}} \leq 0.1$ eV/atom, disjoint from stability), S.U.N., and MS.U.N. rates (definitions in Appendix G).

Baselines. We compare MCFlow against representative crystal generative models on LeMat-GenBench, including MatterGen (Zeni et al., 2025), PlaID++ (Xu et al., 2025), WyFormer (Kazeev et al., 2025), DiffCSP (Jiao et al., 2023), DiffCSP++ (Jiao et al., 2024), ADiT (Joshi et al., 2025), LLaMat-2/3 (Mishra et al., 2026), SymmCD (Levy et al., 2025), CrystalFormer (Cao et al., 2025), and CrystalGFN (Hernandez-Garcia et al., 2023).

Results. MCFlow achieves strong performance on LeMat-GenBench across both evaluation groups (Table 2). MCFlow leads on most metrics with and without MLIP pre-relaxation, particularly on stability-related ones (stability, metastability, E_{hull}), which we attribute to the proposed ordering combined with the DiT backbone. Under pre-relaxation, we further observe a stability–novelty trade-off across model scales: larger variants produce more stable but less novel samples, and MCFlow-S consequently leads MS.U.N. by retaining sample diversity. Notably, MCFlow better captures crystallographic symmetry distributions over space groups and Wyckoff properties, whereas FlowMM and ADiT exhibit mode collapse toward lower-symmetry groups such as P1 (Figure 4). These results indicate that MCFlow internalizes crystallographic symmetry from data without explicit symmetry constraints.

MCFlow also produces geometrically stable samples. The distribution profiles in Figure 5 demonstrate that MCFlow samples are closer to local energy minima as a population, exhibiting shorter relaxation trajectories and lower E_{hull} values than the baselines. MCFlow achieves competitive performance with a relatively small number of integration steps (Figure 3). MCFlow maintains high structural and compositional validity on the larger MPTS-52 dataset (Table H.6). While different DNG metrics exhibit varying trends across model sizes, validity shows consistent improvement, as detailed in Figure H.1. Additional qualitative samples are provided in Appendix I.

4.3. Atom Type Generation

We consider the atom type generation task, abbreviated ATG, where we generate atom types conditioned on a crystal structure to demonstrate that the proposed multimodal flow model naturally supports structure-to-composition generation using the same model. This setting reflects a common case in materials datasets, where nearly identical crystal structures can support multiple chemically valid composi-

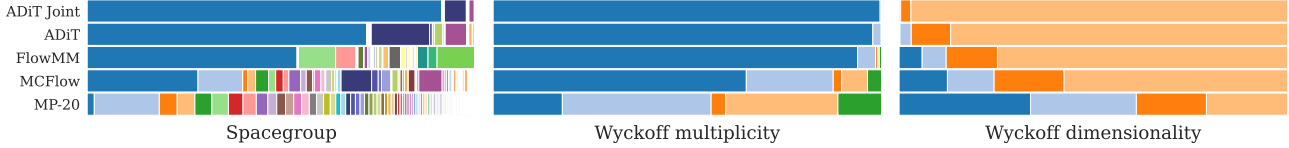


Figure 4. Distributions of crystallographic symmetry properties. MCFlow better captures crystallographic symmetry distributions than FlowMM and ADiT. From left to right, panels show the distributions of space groups (P1, Fm $\bar{3}$ m, C2/m, P6 $\bar{3}$ /mmc, I4/mmm, Pnma, R $\bar{3}$ m, Cmc, Pm $\bar{3}$ m, P4/mmm, P $\bar{1}$, ...), Wyckoff multiplicity (1, 2, 3, 4, 6), and Wyckoff dimensionality (0, 1, 2, 3). Space groups and Wyckoff positions are determined using the SpaceGroupAnalyzer module in pymatgen with distance of 0.1 Å.

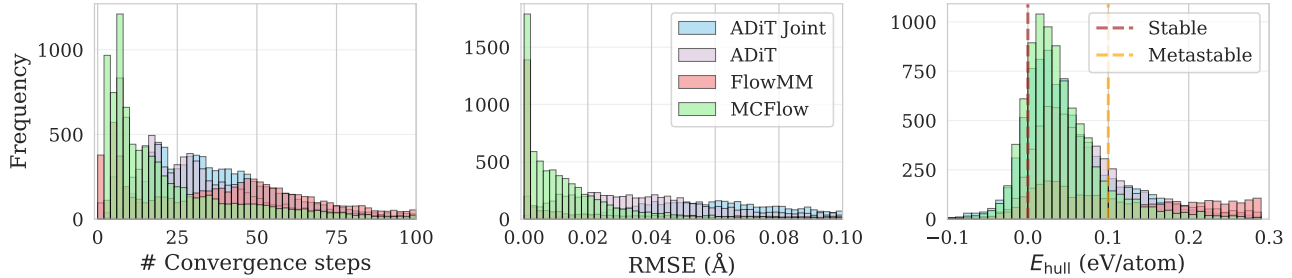


Figure 5. Distributions of relaxed structures. MCFlow samples are closer to local energy minima than FlowMM and ADiT, with smaller relaxation RMSE, fewer convergence steps, and lower E_{hull} . From left to right, plots show distributions of convergence steps, RMSE between initial and relaxed structures, and E_{hull} along CHGNet (Deng et al., 2023) geometry optimization trajectories. The convergence rates are 99.73/60.79/87.65/99.54% for MCFlow/FlowMM/ADiT/ADiT Joint.

Table 3. Atom type generation on MP-20 and MPTS-52. MCFlow shows higher compositional validity (Comp) and MR than the baselines when sampling multiple candidates per structure. Comp follows Xie et al. (2022); details in Appendix G.

Model	MP-20		MPTS-52	
	Comp (%) $\uparrow$	MR (%) $\uparrow$	Comp (%) $\uparrow$	MR (%) $\uparrow$
CGCNN	87.06	4.71	82.08	3.50
FlowMM (1)	83.39	2.40	76.12	0.88
MCFlow-S (1)	88.97	1.57	83.14	1.01
MCFlow-B (1)	90.17	3.12	83.56	1.20
MCFlow-L (1)	90.07	3.32	83.67	1.24
MCFlow-S (20)	88.71	15.34	83.22	7.22
MCFlow-B (20)	90.22	14.73	84.17	5.72
MCFlow-L (20)	90.23	15.30	84.25	7.20

tions. As a result, it is particularly relevant for exploring atomic substitution patterns within a fixed structure. To our knowledge, structure-conditioned ATG has not been studied as a stand-alone generative task in prior crystal generative modeling work; we therefore adapt the deterministic property-prediction model CGCNN (Xie & Grossman, 2018) and the generative model FlowMM (Miller et al., 2024) to the ATG task as reference points for compositional validity and accuracy.

Results. MCFlow generates compositionally valid and diverse atom types when conditioned on a fixed crystal structure. Compared to the baselines, MCFlow exhibits greater diversity across multiple samples per structure while maintaining higher compositional validity (Table 3). Further

qualitative samples illustrating compositional diversity in atom type generation are provided in Appendix I. These results illustrate that a single fixed crystal structure can be compatible with multiple valid compositions.

Atom type substitution. Building on atom type generation, MCFlow can perform atom type substitution: at each ATG step, we sample atom types from the CTMC posterior and overwrite the atom types at the known sites with given types, so only the masked sites are sampled. Across four structurally and functionally diverse MP-20 test-set materials—a cubic Pm $\bar{3}$ m perovskite oxide (AcGaO $_3$), an orthorhombic Pnma silicide intermetallic for hydrogen storage (Zr $_4$ Fe $_4$ Si $_4$), a cubic Fm $\bar{3}$ m full Heusler spintronic half-metal (Mn $_2$ VGa), and a cubic Fd $\bar{3}$ m spinel photocatalyst (Ca $_2$ In $_4$ O $_8$)—masking one element and drawing 100 samples yields chemically meaningful top substitutions (Table 4): the top-3 predicted elements mostly correspond to known Materials Project entries that share the same space group as the original structure. This indicates that MCFlow has learned physically meaningful atom substitution patterns from data, without substitution rules.

4.4. Ablation Studies

We investigate the impact of composition- and symmetry-aware atom ordering and inter- and intra-orbit permutation augmentation to quantify component contributions. We evaluate these strategies using Match Rate and RMSE for CSP, and structural and compositional validity for DNG.

Table 4. Atom type substitutions across four MP-20 test-set materials. MCFlow recovers chemically meaningful substitutions in different applications: the top-3 predicted elements mostly correspond to known materials sharing the same space group as the original structure. For each case, we mask one element type and draw 100 samples using inference-time guidance built on atom type generation. The four parent structures, all drawn from the MP-20 test set, are AcGaO_3 (mp-1183053, cubic $\text{Pm}\bar{3}\text{m}$ perovskite oxide), $\text{Zr}_4\text{Fe}_4\text{Si}_4$ (mp-1102452, orthorhombic Pnma silicide intermetallic for hydrogen storage), Mn_2VGa (mp-21235, cubic $\text{Fm}\bar{3}\text{m}$ full Heusler spintronic half-metal), and $\text{Ca}_2\text{In}_4\text{O}_8$ (mp-22766, cubic $\text{Fd}\bar{3}\text{m}$ spinel photocatalyst). A blank in the MP ID column means no matching Materials Project entry exists; a “—” in the Split column means the entry is not in MP-20.

Rank	AcGaO_3 (mask Ac)				$\text{Zr}_4\text{Fe}_4\text{Si}_4$ (mask Fe)				Mn_2VGa (mask V)				$\text{Ca}_2\text{In}_4\text{O}_8$ (mask Ca)			
	Element	Freq (%)	MP ID	Split	Element	Freq (%)	MP ID	Split	Element	Freq (%)	MP ID	Split	Element	Freq (%)	MP ID	Split
1	Ac	23	1183053	Test	Co	40	1102565	Test	Mn	23	999550	—	Mn	32	35162	Train
2	Yb	22	1187561	—	Cu	32	22522	Train	V	16	21235	Test	Ca	19	22766	Test
3	Sr	10	—	—	Ni	19	1103272	Val	Nb	15	1185972	Train	Cd	12	19803	Val

Table 5. Ablation results on MP-20 and MPTS-52. Removing the proposed hierarchical permutation augmentation and atom ordering leads to consistent performance degradation across tasks. Rows denote cumulative ablations from the full MCFlow model by removing inter-orbit permutation (— Inter), intra-orbit permutation (— Intra), and positional embedding (— PosEmb). MR and RMSE are reported for CSP, while Struct and Comp denote structural and compositional validity for DNG; details in Appendix G.

Model	MP-20				MPTS-52			
	MR	RMSE	Struct	Comp	MR	RMSE	Struct	Comp
MCFlow	63.14	0.0611	99.44	89.85	26.46	0.1572	98.34	85.57
— Inter	60.41	0.0772	99.48	89.64	22.80	0.1927	98.12	85.20
— Intra	58.54	0.1023	99.35	89.30	18.80	0.2231	97.70	84.83
— PosEmb	54.83	0.0979	99.48	85.30	16.44	0.2105	98.20	75.20

Results. Lexicographical ordering based on electronegativity and Wyckoff position provides compositional and crystallographic priors that are exploited by the DiT model, as shown in Table 5. The PosEmb ablation removes explicit positional information from the sequence representation to assess the contribution of the proposed atom ordering. The resulting performance degradation indicates that the ordering encodes compositional and symmetry information (Table 5). Inter- and intra-orbit permutation augmentation further improves performance by promoting robustness to symmetry-preserving permutations within this ordering. Furthermore, electronegativity-based ordering consistently outperforms atomic numbers, and the proposed hierarchical augmentation significantly outperforms full $N!$ random permutation (Table H.4). We also show that unified training matches CSP-only training (Table H.7), and that decoupled time axes are critical for CSP: coupling $t=s$ substantially degrades CSP performance, and DiffCSP/FlowMM adapted to CSP fail to effectively condition on fixed compositions due to their coupled noise schedules (Table H.8).

5. Conclusion

We introduced MCFlow, a unified multimodal flow framework that formulates crystal generation tasks as different inference trajectories through a single joint generative model. By decoupling flow time axes for atom types and crystal

structures, MCFlow enables flexible conditional generation, allowing crystal structure prediction, *de novo* generation, and atom type generation to be realized within a single model without task-specific architectures or retraining.

To instantiate this unified framework in an all-atom transformer architecture, we proposed a composition- and symmetry-aware atom ordering together with hierarchical permutation augmentation, providing a soft inductive bias that allows composition and crystallographic symmetry to be learned directly from data without explicit templates. Experiments demonstrate that a single MCFlow model achieves competitive performance across multiple crystal generation tasks, while reducing collapse toward low-symmetry space groups compared to existing baselines.

Limitations. The composition- and symmetry-aware ordering exploits crystallographic symmetry and is therefore inapplicable to amorphous or disordered systems, which would require alternative inductive biases not relying on well-defined space groups. We further observe a stability–novelty trade-off across model scales (Table 2): larger variants yield more stable but less novel samples, biasing generation toward known regions of materials space.

A natural next step is extending the any-to-any formulation to material property modalities, enabling property-conditioned generation within the same framework. A complementary direction is integrating MCFlow into agent-based, closed-loop materials discovery (Takahara et al., 2025), where its flexible generation modes can propose candidates under different conditioning settings.

Impact Statement

MCFlow offers a computational tool for discovering next-generation materials, with applications including solid-state electrolytes, green-hydrogen catalysts, and advanced semiconductors. By enabling any-to-any modality generation in a single model, it supports diverse materials generation tasks without task-specific architectures.

References

- Baird, S. G., Sayeed, H. M., and Riebesell, J. sparks-baird/matbench-genmetrics. <https://github.com/sparks-baird/matbench-genmetrics>, 2024. [Accessed 03-05-2024].
- Batatia, I., Benner, P., Chiang, Y., Elena, A. M., Kovács, D. P., Riebesell, J., et al. A foundation model for atomistic materials chemistry. *The Journal of Chemical Physics*, 163(18):184110, 2025.
- Betala, S., Gleason, S. P., Ramlaoui, A., Xu, A., Channing, G., Levy, D., Fourrier, C., Kazeev, N., Joshi, C. K., Kaba, S.-O., Therrien, F., Hernandez-Garcia, A., Mercado, R., Krishnan, N. M. A., and Duval, A. LeMat-GenBench: A unified evaluation framework for crystal generative models. In *NeurIPS Workshop on AI for Accelerated Materials Design (AI4Mat)*, 2025.
- Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., and Walsh, A. Machine learning for molecular and materials science. *Nature*, 559(7715):547–555, 2018.
- Campbell, A., Yim, J., Barzilay, R., Rainforth, T., and Jaakkola, T. Generative flows on discrete state-spaces: Enabling multimodal flows with applications to protein co-design. In *International Conference on Machine Learning*, pp. 5453–5512. PMLR, 2024.
- Cao, Z., Luo, X., Lv, J., and Wang, L. Space group informed transformer for crystalline materials generation. *Science Bulletin*, 2025.
- Chang, R., Pak, A., Guerra, A., Zhan, N., Richardson, N., Ertekin, E., and Adams, R. P. Space group equivariant crystal diffusion. In *Advances in Neural Information Processing Systems*, 2025.
- Chen, R. T. and Lipman, Y. Riemannian flow matching on general geometries. In *The Twelfth International Conference on Learning Representations*, 2024. URL <https://openreview.net/forum?id=g7ohDlTITL>.
- Chen, R. T., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.
- Das, K., Khastagir, S., Goyal, P., Lee, S.-C., Bhattacharjee, S., and Ganguly, N. Periodic materials generation using text-guided joint diffusion model. In *International Conference on Learning Representations*, 2025.
- Deng, B., Zhong, P., Jun, K., Riebesell, J., Han, K., Bartel, C. J., and Ceder, G. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5(9):1031–1041, 2023.
- Ding, Q., Miret, S., and Liu, B. MatExpert: Decomposing materials discovery by mimicking human experts. *arXiv preprint arXiv:2410.21317*, 2024.
- Frank, J. T., Ripken, W., Lied, G., MÄßler, K.-R., Unke, O. T., and Chmiela, S. Sampling 3d molecular conformers with diffusion transformers. *arXiv preprint arXiv:2506.15378*, 2025.
- Glass, C. W., Oganov, A. R., and Hansen, N. USPEX—evolutionary crystal structure prediction. *Computer physics communications*, 175(11-12):713–720, 2006.
- Grosse-Kunstleve, R. W., Sauter, N. K., and Adams, P. D. Numerically stable algorithms for the computation of reduced unit cells. *Acta Crystallographica Section A: Foundations of Crystallography*, 60(1):1–6, 2004.
- Gruver, N., Sriram, A., Madotto, A., Wilson, A. G., Zitnick, C. L., and Ulissi, Z. Fine-tuned language models generate stable inorganic materials as text. In *International Conference on Learning Representations*, 2024.
- Hernandez-Garcia, A., Duval, A., Volokhova, A., Bengio, Y., Sharma, D., Carrier, P. L., Benabed, Y., Koziarski, M., Schmidt, V., Rignanese, G.-M., De Breuck, P.-P., and Clancy, P. Crystal-GFN: Sampling crystals with desirable properties and constraints. *arXiv preprint arXiv:2310.04925*, 2023.
- Hillier, H. Crystallography and cohomology of groups. *The American Mathematical Monthly*, 93(10):765–779, 1986.
- Hoellmer, P., Egg, T., Martirosyan, M. M., Fuemmeler, E., Shui, Z., Gupta, A., Prakash, P., Roitberg, A., Liu, M., Karypis, G., Transtrum, M., Hennig, R. G., Tadmor, E. B., and Martiniani, S. Open materials generation with stochastic interpolants. In *International Conference on Machine Learning*, 2025. URL <https://arxiv.org/abs/2502.02582>.
- Holderrieth, P., Havasi, M., Yim, J., Shaul, N., Gat, I., Jaakkola, T., Karrer, B., Chen, R. T., and Lipman, Y. Generator matching: Generative modeling with arbitrary markov processes. *arXiv preprint arXiv:2410.20587*, 2024.
- Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL materials*, 1(1):011002, 2013.
- Jiao, R., Huang, W., Lin, P., Han, J., Chen, P., Lu, Y., and Liu, Y. Crystal structure prediction by joint equivariant diffusion. In *Advances in Neural Information Processing Systems*, 2023. URL <https://arxiv.org/abs/2309.04475>.

- Jiao, R., Huang, W., Liu, Y., Zhao, D., and Liu, Y. Space group constrained crystal generation. In *International Conference on Learning Representations*, 2024. URL <https://arxiv.org/abs/2402.03992>.
- Joshi, C. K., Fu, X., Liao, Y.-L., Gharakhanyan, V., Miller, B. K., Sriram, A., and Ulissi, Z. W. All-atom diffusion transformers: Unified generative modelling of molecules and materials. In *Forty-second International Conference on Machine Learning*, 2025.
- Kazeev, N., Nong, W., Romanov, I., Zhu, R., Ustyuzhanin, A. E., Yamazaki, S., and Hippalgaonkar, K. Wyckoff transformer: Generation of symmetric crystals. In *Forty-second International Conference on Machine Learning*, 2025.
- Kelvinus, F. E., Andersson, O. B., Parackal, A. S., Qian, D., Armiento, R., and Lindsten, F. Wyckoffdiff—a generative diffusion model for crystal symmetry. In *Forty-second International Conference on Machine Learning*, 2025.
- Khastagir, S., Das, K., Goyal, P., Lee, S.-C., Bhattacharjee, S., and Ganguly, N. LLM meets diffusion: A hybrid framework for crystal material generation. In *Advances in Neural Information Processing Systems*, 2025.
- Kim, M., Kim, N., Kim, H., and Ahn, S. CatFlow: Co-generation of slab-adsorbate systems via flow matching. *arXiv preprint arXiv:2602.05372*, 2026a.
- Kim, N., Kim, H., Yu, S., Kim, M., Kim, S., and Ahn, S. AtomMOF: All-atom flow matching for MOF-adsorbate structure prediction. *arXiv preprint arXiv:2602.07351*, 2026b.
- Kim, N., Kim, S., and Ahn, S. Flexible MOF generation with torsion-aware flow matching. *Advances in Neural Information Processing Systems*, 38:13790–13820, 2026c.
- Kusaba, M., Liu, C., and Yoshida, R. Crystal structure prediction with machine learning-based element substitution. *Computational Materials Science*, 211:111496, 2022.
- Levy, D., Panigrahi, S. S., Kaba, S.-O., Zhu, Q., Lee, K. L. K., Galkin, M., Miret, S., and Ravanbakhsh, S. Symmcd: Symmetry-preserving crystal generation with diffusion models. In *International Conference on Learning Representations*, 2025. URL <https://arxiv.org/abs/2502.03638>.
- Li, T. and He, K. Back to basics: Let denoising generative models denoise. *arXiv preprint arXiv:2511.13720*, 2025.
- Lipman, Y., Chen, R. T., Ben-Hamu, H., Nickel, M., and Le, M. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2022.
- Lipman, Y., Havasi, M., Holderrieth, P., Shaul, N., Le, M., Karrer, B., Chen, R. T. Q., Lopez-Paz, D., Ben-Hamu, H., and Gat, I. Flow matching guide and code, 2024. URL <https://arxiv.org/abs/2412.06264>.
- Liu, Q. Rectified flow: A marginal preserving approach to optimal transport, 2022. URL <https://arxiv.org/abs/2209.14577>.
- Loshchilov, I. and Hutter, F. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.
- Luo, X., Wang, Z., Wang, Q., Shao, X., Lv, J., Wang, L., Wang, Y., and Ma, Y. Crystalflow: A flow-based generative model for crystalline materials. *Nature Communications*, 16(1):9267, 2025.
- Martirosyan, M. M., Egg, T., Hoellmer, P., Karypis, G., Transtrum, M., Roitberg, A., Liu, M., Hennig, R. G., Tadmor, E. B., and Martiniani, S. All that structure matches does not glitter. In *Advances in Neural Information Processing Systems*, 2025.
- Miller, B. K., Chen, R. T. Q., Sriram, A., and Wood, B. M. Flowmm: Generating materials with riemannian flow matching. In *International Conference on Machine Learning*, 2024. URL <https://arxiv.org/abs/2406.04713>.
- Mishra, V., Singh, S., Ahlawat, D., Zaki, M., Bihani, V., Grover, H. S., Mishra, B., Miret, S., Mausam, and Krishnan, N. M. A. A family of large language models for materials research with insights into model adaptability in continued pretraining. *Nature Machine Intelligence*, 8: 435–448, 2026.
- Morehead, A., Cretu, M., Panescu, A., Anand, R., Weiler, M., Perez, T., Blau, S., Farrell, S., Bhimji, W., Jain, A., Sahasrabudhe, H., Lio, P., Jaakkola, T., Gomez-Bombarelli, R., Ying, R., Erichson, N. B., and Mahoney, M. W. Zatom-1: A multimodal flow foundation model for 3D molecules and materials. *arXiv preprint arXiv:2602.22251*, 2026.
- Noh, J., Kim, J., Stein, H. S., Sanchez-Lengeling, B., Gregoire, J. M., Aspuru-Guzik, A., and Jung, Y. Inverse design of solid-state materials via a continuous representation. *Matter*, 1(5):1370–1384, 2019.
- Nouira, A., Sokolovska, N., and Crivello, J.-C. Crystalgan: learning to discover crystallographic structures with generative adversarial networks. *arXiv preprint arXiv:1810.11203*, 2018.
- Ong, S. P., Richards, W. D., Jain, A., Hautier, G., Kocher, M., Cholia, S., Gunter, D., Chevrier, V. L., Persson, K. A., and Ceder, G. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013.

- Park, H., Onwuli, A., and Walsh, A. Exploration of crystal chemical space using text-guided generative artificial intelligence. *Nature Communications*, 16(1), 2025.
- Peebles, W. and Xie, S. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4195–4205, 2023.
- Pickard, C. J. and Needs, R. Ab initio random structure searching. *Journal of Physics: Condensed Matter*, 23(5): 053201, 2011.
- Prince, E. (ed.). *International Tables for Crystallography, Volume C: Mathematical, physical and chemical tables*. Springer Science & Business Media, 2004.
- Puny, O., Lipman, Y., and Miller, B. K. Space group conditional flow matching. *arXiv preprint arXiv:2509.23822*, 2025.
- Rhodes, B., Vandenhaute, S., Šimkus, V., Gin, J., Godwin, J., Duignan, T., and Neumann, M. Orb-v3: Atomistic simulation at scale. *arXiv preprint arXiv:2504.06231*, 2025.
- Rojas, K., Zhu, Y., Zhu, S., Ye, F. X., and Tao, M. Diffuse everything: Multimodal diffusion models on arbitrary state spaces. In *Forty-second International Conference on Machine Learning*, 2025.
- Siron, M., Djafar, I., Ramlaoui, A., du Fayette, E., Rossello, A., Fako, E., McDermott, M., Therrien, F., Barroso-Luque, L., Cipcigan, F., Schwaller, P., Wolf, T., and Duval, A. LeMat-Bulk: Aggregating, and de-duplicating quantum chemistry materials databases. *arXiv preprint arXiv:2511.05178*, 2025.
- Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2020.
- Sriram, A., Miller, B. K., Chen, R. T., and Wood, B. M. FlowLLM: Flow matching for material generation with large language models as base distributions. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.
- Takahara, I., Mizoguchi, T., and Liu, B. Accelerated inorganic materials design with generative AI agents. *Cell Reports Physical Science*, 6(12), 2025.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- Wang, Y., Lv, J., Zhu, L., and Ma, Y. Crystal structure prediction via particle-swarm optimization. *Physical Review B*, 82(9):094116, 2010.
- Wang, Y., Lv, J., Zhu, L., and Ma, Y. Calypso: A method for crystal structure prediction. *Computer Physics Communications*, 183(10):2063–2070, 2012.
- Wood, B. M., Dzamba, M., Fu, X., Gao, M., Shuaibi, M., Barroso-Luque, L., Abdelmaqsoud, K., Gharakhanyan, V., Kitchin, J. R., Levine, D. S., Michel, K., Sriram, A., Cohen, T., Das, A., Rizvi, A., Sahoo, S. J., Ulissi, Z. W., and Zitnick, C. L. UMA: A family of universal models for atoms. In *Advances in Neural Information Processing Systems*, 2025.
- Wu, H., Song, Y., Gong, J., Cao, Z., Ouyang, Y., Zhang, J., Zhou, H., Ma, W.-Y., and Liu, J. A periodic Bayesian flow for material generation. In *International Conference on Learning Representations*, 2025.
- Wyckoff, R. W. G. *The analytical expression of the results of the theory of space groups*. Number 318 in Carnegie Institution of Washington Publication. Carnegie Institution of Washington, 1930.
- Xie, T. and Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters*, 120(14):145301, 2018.
- Xie, T., Fu, X., Ganea, O.-E., Barzilay, R., and Jaakkola, T. Crystal diffusion variational autoencoder for periodic material generation. In *International Conference on Learning Representations*, 2022.
- Xu, A., Desai, R., Wang, L., Hope, G., and Ritz, E. PLalD++: A preference aligned language model for targeted inorganic materials design. *arXiv preprint arXiv:2509.07150*, 2025.
- Yang, M., Cho, K., Merchant, A., Abbeel, P., Schuurmans, D., Mordatch, I., and Cubuk, E. D. Scalable diffusion for materials generation. In *International Conference on Learning Representations*, 2024.
- Yi, X., Xu, G., Xiao, X., Zhang, Z., Liu, L., Bian, Y., and Zhao, P. Crystaldit: A diffusion transformer for crystal generation. *arXiv preprint arXiv:2508.16614*, 2025.
- Zeni, C., Pinsler, R., Zügner, D., Fowler, A., Horton, M., Fu, X., Shysheya, S., Crabbé, J., Sun, L., Smith, J., et al. MatterGen: a generative model for inorganic materials design. *Nature*, 639:624–632, 2025. URL <https://arxiv.org/abs/2312.03687>.
- Zhu, R., Nong, W., Yamazaki, S., and Hippalgaonkar, K. Wycryst: Wyckoff inorganic crystal generator framework. *Matter*, 7(10):3469–3488, 2024.

A. Notations and Backgrounds

Table A.1. Summary of Notations

Symbol	Description
C	Infinite periodic crystal
N	Number of atoms in the primitive unit cell
A	Atom types sequence ($[a_i]_{i=1}^N \in \mathcal{A}^{1 \times N}$)
F	Fractional coordinates ($[f_i]_{i=1}^N \in [0, 1)^{3 \times N}$)
X	Cartesian coordinates ($X = LF$)
L	Lattice matrix
L^l	Lattice lengths ($\{a, b, c\} \in \mathbb{R}^{1 \times 3}$)
L^a	Lattice angles ($\{\alpha, \beta, \gamma\} \in [60^\circ, 120^\circ]^{1 \times 3}$)
G	Space group
O	Orbits ($\{g \cdot f \mid g \in G\}$)
G_f	Site symmetry group ($\{g \mid g \cdot f = f\}$)
$\mathcal{W}$	Wyckoff position
$\mathcal{S}$	Discrete state space (e.g., atom type space $\mathcal{A}$)
$u(x_t, t)$	Velocity field (continuous) or rate vector (discrete)
p_t	Probability path at time t
t, s	Flow time coordinates (t : atom types, s : structure)
π	Coupling distribution between p_0 and p_1
δ_{x_t}	Probability vector with $\delta_{x_t}(x_t) = 1$ on discrete space
$\mathbb{T}^{3 \times N}$	Torus manifold for fractional coordinates ($\mathbb{R}^{3 \times N} / \mathbb{Z}^{3 \times N}$)
$\text{proj}(\cdot)$	Projection onto the torus ($x \pmod{1}$)

We first formalize periodic crystals and symmetry (space groups, Wyckoff positions), then review flow-based generative models on various state spaces. Table A.1 summarizes the key notations used throughout this paper.

A.1. Crystal Representation

We represent crystals as infinite periodic arrangements of primitive unit cells. The unit cell is represented by atom types $A = [a_i]_{i=1}^N \in \mathbb{Z}^{1 \times N}$, fractional coordinates $F = [f_i]_{i=1}^N \in [0, 1)^{3 \times N}$, and a lattice matrix $L = [l_i]_{i=1}^3 \in \mathbb{R}^{3 \times 3}$. While there are many possible choices of primitive unit cells, we use Niggli reduction (Grosse-Kunstleve et al., 2004) to fix a canonical representative (up to numerical tolerance). The lattice matrix L can be also represented as roto-translation invariant lattice parameters: three side lengths $L^l = \{a, b, c\} \in \mathbb{R}^{1 \times 3}$ and three internal angles $L^a = \{\alpha, \beta, \gamma\} \in [60^\circ, 120^\circ]^{1 \times 3}$ of the unit cell. Given this lattice representation L , the 3D Cartesian coordinates $X = [x_i]_{i=1}^N \in \mathbb{R}^{3 \times N}$ are obtained by the linear combination of lattice vectors with fractional coefficients: $X = LF$. The infinite periodic crystal C is then defined as:

$$C = \{(a_i, x_i + Lk) \mid i \in \{1, \dots, N\}, \quad k \in \mathbb{Z}^3\} \quad (12)$$

where k_j denotes the integral translation along the j -th lattice vector l_j .

A.2. Crystal Symmetry

Space groups (Hiller, 1986) characterize the invariances of infinite periodic structures.

Space groups. A space group G is a group of isometries that tile $\mathbb{R}^3$ with repeating unit cells, i.e., a set of operation g under which the transformed crystal C is identical to the original crystal where operation $g = (R, t) \in G$ consists of an orthogonal matrix $R \in O(3)$ and a translation $t \in \mathbb{R}^3$. Its action on a Cartesian coordinate $x \in \mathbb{R}^3$ denotes $g \cdot x := Rx + t$. When expressed in fractional coordinates f with $x = Lf$, we use the corresponding induced action

$$g \cdot f := L^{-1}(R(Lf) + t) \pmod{1}, \quad (13)$$

where $\pmod{1}$ is applied component-wise, identifying fractional coordinates that differ by lattice translations. A space group operation first acts on atomic positions in Cartesian space and then maps the result back into the unit cell via fractional coordinates modulo lattice translations.

A crystal C is said to be invariant under space group G if for every $g \in G$, there exists a permutation $\sigma_g \in S_N$ of the atom indices such that the transformed atomic positions coincide with the original ones up to lattice translations and a permutation of atom indices preserving atom types. This permutation σ_g rearranges indices changed by the symmetry operation g while preserving their atom types. Every crystal has space group symmetry which is one of the 230 distinct space groups, a unique way a crystal's atoms can be arranged periodically.

Wyckoff positions. Given a space group G , Wyckoff positions (Wyckoff, 1930) describe crystal symmetry at the level of atomic sites. For a fractional coordinate f , a set of equivalent atoms generated by group G , called *orbit*, is

$$G \cdot f := \{g \cdot f \mid g \in G\},$$

Site symmetry group of f is a subgroup of the space group consisting of operations that leave the site invariant (up to lattice translations):

$$G_f := \{g \in G \mid g \cdot f = f\}.$$

We can classify site symmetry groups based on conjugate relation, which is equivalent relation, up to basis change by space group, between G_f and $G_{f'}$ defined by $G_f = gG_{f'}g^{-1}$ for some $g \in G$. The conjugate relation induces an equivalence class $[G_f]$ of site symmetry groups

$$[G_f] := \{gG_fg^{-1} \mid g \in G\}.$$

Wyckoff position is a set of orbits whose site symmetry groups have conjugate relation:

$$\mathcal{W}_{[G_f]} := \{G \cdot f' \mid G_{f'} \in [G_f]\}.$$

The size of its orbit $|G \cdot f|$ is called the *Wyckoff multiplicity*. Wyckoff positions are conventionally labeled by letters $a, b, c, \dots$, called the *Wyckoff letter*. The letter is assigned according to the site symmetry following the IUCr convention (Prince, 2004). Ordering atoms by Wyckoff position encodes crystallographic symmetry information that is useful for transformer models.

A.3. Flow Generative Model

We provide a self-contained background on flow-based models and flow matching in various state spaces. For a more comprehensive guide and practical implementation details of flow matching, we refer the reader to Lipman et al. (2024). Flow models naturally extend to multimodal settings by combining continuous and discrete state spaces under a unified generator matching framework (Holderieth et al., 2024).

Flow-based generative model. Flow-based generative model (Chen et al., 2018), or simply flow model, transports a base distribution $p_0(x_0) = p_{\text{base}}$ to data distribution $p_1(x_1) = p_{\text{data}}$ by solving an ordinary differential equation (ODE) characterized by a time-dependent velocity field $u(x_t, t)$ for $t \in [0, 1]$:

$$\frac{dx_t}{dt} = u(x_t, t), \quad x_0 \sim p_0. \quad (14)$$

Flow models construct and learn $u(x_t, t)$ which corresponds to probability path $\{p_t\}_{t \in [0, 1]}$ evolving from $p_0 = p_{\text{base}}$ to $p_1 = p_{\text{data}}$.

Flow matching. Flow matching (FM) is a simulation-free training method to learn the constructed velocity field $u(x_t, t)$ (Lipman et al., 2022; Liu, 2022). Given a neural network $u_\theta(x_t, t)$, the objective is to minimize the squared distance

$$\mathcal{L}_{\text{FM}}(\theta) = \mathbb{E}_{t, x_t} [\|u_\theta(x_t, t) - u(x_t, t)\|^2], \quad (15)$$

where $t \sim \mathcal{U}[0, 1]$ and $x_t \sim p_t$. Lipman et al. (2022) defines the probability path as a mixture of tractable conditional probability paths $p_t(\cdot | x_0, x_1)$:

$$p_t(x) = \int p_t(x | x_0, x_1) d\pi(x_0, x_1), \quad (16)$$

where $\pi(x_0, x_1)$ is a independent coupling distribution over the base and data distributions, i.e., $\pi(x_0, x_1) = p_0(x_0)p_1(x_1)$. Lipman et al. (2022) introduces conditional flow matching (CFM) which minimizes the distance between $u_\theta(x_t, t)$ with tractable conditional velocity field $u(x_t, t | x_0, x_1)$:

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{t, x_0, x_1, x_t} [\|u_\theta(x_t, t) - u(x_t, t | x_0, x_1)\|^2], \quad (17)$$

where pairs (x_0, x_1) are sampled from the coupling distribution $\pi(x_0, x_1)$, and conditional flow x_t is sampled from conditional probability path $p_t(\cdot | x_0, x_1)$. The conditional flow matching objective induces the same gradient with respect to θ as the marginal flow matching objective, i.e., $\nabla_\theta \mathcal{L}_{\text{FM}}(\theta) = \nabla_\theta \mathcal{L}_{\text{CFM}}(\theta)$

A simple and commonly used conditional flow is linear interpolation $x_t = (1 - t)x_0 + tx_1$ between two endpoints x_0 and x_1 , corresponding to the conditional velocity field:

$$u(x_t, t | x_0, x_1) = \frac{x_1 - x_0}{1 - t}. \quad (18)$$

We can construct the marginal velocity field $u(x_t, t)$, constructed via posterior expectation over π :

$$u(x_t, t) = \mathbb{E}_{x_0, x_1} [u(x_t, t | x_0, x_1) | x_t], \quad (19)$$

ensuring transports from p_0 to p_1 . The CFM objective extends the squared distance loss to a Bregman divergence $D_\Phi(u, v) = \Phi(u) - \Phi(v) - \langle \nabla \Phi(v), u - v \rangle$:

$$\mathbb{E}_{t, x_0, x_1, x_t} [D(u_\theta(x_t, t), u(x_t, t | x_0, x_1))], \quad (20)$$

where Φ is a convex function. Bregman divergences are used since they preserve the gradient equivalence between conditional and marginal flow matching objectives. This formulation allows different choices of Φ and recovers the squared distance loss $D(u, v) = \|u - v\|^2$ when $\Phi(u) = \|u\|^2$.

Riemannian flow model. The above flow model assumes that data lie in Euclidean space. When data distribution is on Riemannian manifold $\mathcal{M}$, Riemannian flow model (Chen & Lipman, 2024) transports base distribution on the manifold to the data distribution with velocity field $u(x_t, t)$ in tangent space $T_{x_t}\mathcal{M}$. The construction $u(x_t, t)$ from the conditional velocity field $u(x_t, t | x_0, x_1)$ is obtained by replacing straight-line paths with geodesics: $x_t = \exp_{x_0}(t \log_{x_0}(x_1))$ where $\exp_x$ and $\log_x$ denote the Riemannian exponential and logarithmic maps. The corresponding conditional velocity field is

$$u(x_t, t | x_0, x_1) = \frac{\log_{x_t}(x_1)}{1 - t} \in T_{x_t}\mathcal{M}. \quad (21)$$

Flow matching regresses $u_\theta(x_t, t)$ onto the target $u(x_t, t | x_0, x_1)$ in the tangent space. A commonly used Bregman divergence is $D(u, v) = \|u - v\|_g^2$ for a Riemannian metric g on the manifold $\mathcal{M}$.

Discrete flow model. To extend flow models on continuous space to flow model on discrete state space $\mathcal{S}$, we convert solving ODE with the velocity field $u(x_t, t)$ into transition probability vector $p_{t+h|t}(\cdot | x_t)$ from current state x_t for small increment $h > 0$, called continuous-time Markov chain (CTMC):

$$p_{t+h|t}(\cdot | x_t) = \delta_{x_t} + h u(x_t, t) + o(h), \quad (22)$$

where rate vector $u(x_t, t)$ governs CTMC and δ_{x_t} is the probability vector for current state x_t which satisfies $\delta_{x_t}(x_t) = 1$ and $\delta_{x_t}(x) = 0$ if $x \neq x_t$. Discrete flow model aims to learn the rate vector $u(x_t, t)$ which transports the base probability vector p_0 to the data probability vector p_1 through probability vector path $\{p_t\}_{t \in [0, 1]}$. We use masked flow model where the base probability vector $p_0 = \delta_M$ for the masked state M .

Discrete flow matching. Analogous to the continuous case, we match the parameterized and target rate vectors, $u_\theta(x_t, t)$ and $u(x_t, t)$, respectively, given the current state x_t . A commonly used Bregman divergence is the generalized KL divergence $D(u, v) = \sum_j u_j \log(u_j/v_j) - \sum_j u_j + \sum_j v_j$ and the FM objective has the same form with Equation (20) where marginal rate vector are constructed by conditional rate vector as in Equation (19). Discrete flow model also extends to CFM objective for tractability. A commonly used conditional probability vector path is the independent mixture path between the base state $x_0 = M$ and the target state x_1 sampled from p_1 :

$$p_t(\cdot | x_0, x_1) = (1 - t)\delta_{x_0} + t\delta_{x_1}, \quad (23)$$

which is the discrete analog of linear interpolation. The corresponding conditional rate vector is given by:

$$u(x_t, t | x_0, x_1) = \frac{\delta_{x_0}(x_t)\delta_{x_1}}{1 - t}. \quad (24)$$

where $x_0 = M$. While continuous and discrete flow matching are defined on different frameworks, ODE and CTMC, they can be unified into generator matching on continuous-time Markov process (Holderrieth et al., 2024). This unifying perspective enables joint generative processes in which continuous and discrete dynamics evolve simultaneously over different components of the state space.

Table B.1. Configurations for Multimodal Flow Matching. We leverage the flexibility of the flow matching framework to design components for each modality. The Base dist. column specifies the prior distribution p_0 , selected to match the support of each modality. The Cond. target column defines the conditional target vector fields or rate vectors derived from data pairs, utilizing linear interpolation for Euclidean space, geodesic paths for the torus, and independent mixture paths for discrete atom types. The Parameterization column describes how to approximate the target velocity field from the clean data prediction ($F_\theta, L_\theta^l, L_\theta^a$) or the target categorical posterior distribution via atom type logit prediction A_θ . Finally, the Bregman div. column lists the specific divergence metric minimized during training to match the parameterized and target velocity field or rate vector. Since the torus is flat and locally Euclidean, the Bregman divergence on fractional coordinates is the squared Euclidean distance in the tangent space.

Modality (m)	Base dist. p_0	Cond. target	Parameterization u_θ^m	Bregman div. $D_m(u, v)$
Atom type (A)	δ_M	$\frac{\delta_{A_0}(A_t)\delta_{A_1}}{1-t}$	$\frac{\delta_{A_0}(A_t)\text{Cat}(\text{softmax}(A_\theta))}{1-t}$	$\sum_j u_j \log\left(\frac{u_j}{v_j}\right) - \sum_j u_j + \sum_j v_j$
Frac. coord. (F)	$\mathcal{U}[0, 1]^{3 \times N}$	$\frac{\log_{F_s}(F_1)}{1-s}$	$\frac{\log_{F_s}(\text{proj}(F_\theta))}{1-s}$	$\ u - v\ ^2$
Length (L^l)	$\text{LogNorm}(\mu, \sigma)$	$\frac{L_1^l - L_s^l}{1-s}$	$\frac{L_\theta^l - L_s^l}{1-s}$	$\ u - v\ ^2$
Angle (L^a)	$\mathcal{U}[60^\circ, 120^\circ]^{1 \times 3}$	$\frac{L_1^a - L_s^a}{1-s}$	$\frac{L_\theta^a - L_s^a}{1-s}$	$\ u - v\ ^2$

B. Multimodal Flow Details

We provide detailed configurations for each modality in Table B.1.

Riemannian flow on the torus for fractional coordinates. We model fractional coordinates F on the torus manifold $\mathbb{T}^{3 \times N}$ to enforce the periodicity of crystal structures. The corresponding Riemannian exponential and logarithmic maps are given by

$$\exp_f(v) = (f + v) \pmod{1}, \quad \log_f(f') = \frac{1}{2\pi} \text{atan2}(\sin(2\pi(f' - f)), \cos(2\pi(f' - f))). \quad (25)$$

This Riemannian formulation ensures that generated fractional coordinates remain within the fundamental domain $[0, 1)$ by construction.

Parameterization details. We use x_1 -prediction across all modalities and convert the predicted clean data into velocity fields or rate vectors according to the underlying state space of each modality. For lattice lengths and angles in Euclidean space, the network outputs L_θ^l and L_θ^a , which are converted to conditional velocity targets via $(L_\theta^l - L_s^l)/(1-s)$ and $(L_\theta^a - L_s^a)/(1-s)$, respectively. For fractional coordinates on the torus $\mathbb{T}^{3 \times N}$, the network predicts F_θ , which is first projected component-wise onto the torus via $\text{proj}(F) = F \pmod{1}$. The velocity field is then obtained by applying the logarithmic map on the torus, $\log_{F_s}(\text{proj}(F_\theta))/(1-s)$, ensuring that the resulting flow respects the periodic topology of fractional coordinates. For discrete atom types, the network predicts logits A_θ corresponding to the target atom types A_1 . These logits define a posterior categorical distribution $p_\theta = \text{Cat}(\text{softmax}(A_\theta))$, which induces the rate vector of the CTMC as $\delta_{A_0}(A_t)p_\theta/(1-t)$. To mitigate numerical instability near the terminal times, we clip the flow times t and s to 0.9 during training, following Campbell et al. (2024).

Base distribution details. For lattice lengths, the parameters of the LogNormal prior are fitted to training data statistics via maximum likelihood. During inference, the number of atoms N is sampled from the empirical distribution of the training set, which determines the dimensionality of the base distribution.

C. Algorithm

Algorithm 1 MCFlow preprocessing and training

- 1: **Input:** dataset $\mathcal{D}$, clean data (and logit) prediction network f_θ , joint base dist. p_0 , weights $\{w_m\}_{m \in \{A, F, L^a, L^l\}}$.
- 2: **Preprocessing:**
- 3: Group atoms into orbits for each crystal in $\mathcal{D}$.
- 4: Sort orbits lexicographically by electronegativity and Wyckoff position for each crystal.
- Training:**
- 5: Sample $(A_1, X_1) \sim \mathcal{D}$ where $X_1 = (F_1, L_1^l, L_1^a)$.
- 6: Apply inter- and intra-orbit permutations.
- 7: Apply global modulo translation.
- 8: Sample $t, s \sim \mathcal{U}[0, 1]$ and $(A_0, F_0, L_0^l, L_0^a) \sim p_0(A_0, X_0)$.
- 9: Sample (or evaluate) intermediates from endpoint-conditioned paths:

$$A_t \sim \text{Cat}((1-t)\delta_{A_0} + t\delta_{A_1}), \quad F_s \leftarrow \exp_{F_0}(s \log_{F_0}(F_1)), \quad L_s^l \leftarrow (1-s)L_0^l + sL_1^l, \quad L_s^a \leftarrow (1-s)L_0^a + sL_1^a.$$

- 10: Construct modality-wise conditional target velocity field or rate vector:

$$u^A = \frac{\delta_{A_0}(A_t)\delta_{A_1}}{1-t}, \quad u^F \leftarrow \frac{\log_{F_s}(F_1)}{1-s}, \quad u^{L^l} \leftarrow \frac{L_1^l - L_s^l}{1-s}, \quad u^{L^a} \leftarrow \frac{L_1^a - L_s^a}{1-s}.$$

- 11: Predict $(A_\theta, F_\theta, L_\theta^l, L_\theta^a) \leftarrow f_\theta(A_t, X_s, t, s)$.
- 12: Compute rate vector for atom types and velocity fields for crystal structure:

$$u_\theta^A \leftarrow \frac{\delta_{A_0}(A_t)\text{Cat}(\text{softmax}(A_\theta))}{1-t}, \quad u_\theta^F \leftarrow \frac{\log_{F_s}(\text{proj}(F_\theta))}{1-s}, \quad u_\theta^{L^l} \leftarrow \frac{L_\theta^l - L_s^l}{1-s}, \quad u_\theta^{L^a} \leftarrow \frac{L_\theta^a - L_s^a}{1-s}$$

- 13: Compute the multimodal flow matching objective:

$$\widehat{\mathcal{L}}(\theta) = \sum_{m \in \{A, F, L^a, L^l\}} w_m D_m(u_\theta^m, u^m),$$

- 14: Update parameters θ using AdamW with learning rate η .
-

Algorithm 2 MCFlow any-to-any modality generation (DNG/CSP/ATG) during inference

```

1: Input: trained model MCFlow $_{\theta}$ , timesteps  $K$ , empirical atom-count distribution  $p_{\text{emp}}(N)$ .
2: Input: task  $\tau \in \{\text{DNG}, \text{CSP}, \text{ATG}\}$  with atom types  $A_1$  (CSP); structure  $X_1$  (ATG).
3: Output: a sample  $(A, X)$ .

4: Set task-specific trajectory  $\gamma$  and conditioning set  $\mathcal{C}$ .
5: if  $\tau = \text{DNG}$ :  $\gamma(\lambda) = (\lambda, \lambda)$ ,  $\mathcal{C} \leftarrow \emptyset$ .
6: if  $\tau = \text{CSP}$ :  $\gamma(\lambda) = (1, \lambda)$ ,  $\mathcal{C} \leftarrow \{A\}$ .
7: if  $\tau = \text{ATG}$ :  $\gamma(\lambda) = (\lambda, 1)$ ,  $\mathcal{C} \leftarrow \{X\}$ .

8: Fix observed modality:  $A \leftarrow A_1$  if  $\tau = \text{CSP}$ ;  $X \leftarrow X_1$  if  $\tau = \text{ATG}$ .
9: Set  $N \leftarrow \text{sample } p_{\text{emp}}(N)$  if  $\tau = \text{DNG}$ ; otherwise  $N = |A_1|$  (CSP) or  $N = |X_1|$  (ATG).
10: If  $A \notin \mathcal{C}$ , sample  $A \sim p_0(A)$ .
11: If  $X \notin \mathcal{C}$ , sample  $X \sim p_0(X)$ .

12: for  $k = 0$  to  $K - 1$  do
13:    $(t_k, s_k) \leftarrow \gamma(k/K)$ ;  $(t_{k+1}, s_{k+1}) \leftarrow \gamma((k+1)/K)$ 
14:    $\Delta t \leftarrow t_{k+1} - t_k$ ;  $\Delta s \leftarrow s_{k+1} - s_k$ 
15:    $(u_{\theta}^A, u_{\theta}^X) \leftarrow \text{MCFlow}_{\theta}(A, X, t_k, s_k)$ 
16:   If  $A \notin \mathcal{C}$ ,  $A \leftarrow \text{CTMCStep}(A, u_{\theta}^A, \Delta t)$ .
17:   If  $X \notin \mathcal{C}$ ,  $X \leftarrow \text{ODEStep}(X, u_{\theta}^X, \Delta s)$ .
18: end for
19: return  $(A, X)$ .
```

Algorithm 3 MCFlow inference with noisy guidance (CSP/ATG)

```

1: Input: timesteps  $K$ , guidance scale  $\omega$ , noise level  $\sigma$ .
2: Input: task  $\tau \in \{\text{CSP}, \text{ATG}\}$  with atom types  $A_1$  (CSP); structure  $X_1$  (ATG).
3: Output: a guided sample  $(A, X)$ .

4: Set task-specific trajectory  $\gamma$  and conditioning set  $\mathcal{C}$ .
5: if  $\tau = \text{CSP}$ :  $\gamma(\lambda) = (1, \lambda)$ ,  $\mathcal{C} \leftarrow \{A\}$ .
6: if  $\tau = \text{ATG}$ :  $\gamma(\lambda) = (\lambda, 1)$ ,  $\mathcal{C} \leftarrow \{X\}$ .

7: Fix observed modality:  $A \leftarrow A_1$  if  $\tau = \text{CSP}$ ;  $X \leftarrow X_1$  if  $\tau = \text{ATG}$ .
8: Set  $N \leftarrow |A_1|$  (CSP) or  $N \leftarrow |X_1|$  (ATG).

9: if  $\tau = \text{CSP}$ : sample  $X \sim p_0(X)$ .
10: if  $\tau = \text{ATG}$ : sample  $A \sim p_0(A)$ .

11: for  $k = 0$  to  $K - 1$  do
12:    $(t_k, s_k) \leftarrow \gamma(k/K)$ ;  $(t_{k+1}, s_{k+1}) \leftarrow \gamma((k+1)/K)$ 
13:    $\Delta t \leftarrow t_{k+1} - t_k$ ;  $\Delta s \leftarrow s_{k+1} - s_k$ 
14:   if  $\tau = \text{CSP}$ :
15:     Sample  $A_0 \sim p_0(A)$  and  $A_{\sigma} \sim \text{Cat}((1 - \sigma)\delta_{A_0} + \sigma\delta_{A_1})$ 
16:      $u_{\text{cond}}^X \leftarrow u_{\theta}^X(A_1, X, 1, s_k)$ ;  $u_{\text{corr}}^X \leftarrow u_{\theta}^X(A_{\sigma}, X, \sigma, s_k)$ 
17:      $u_{\text{ng}}^X \leftarrow (1 - \omega)u_{\text{corr}}^X + \omega u_{\text{cond}}^X$ 
18:      $X \leftarrow \text{ODEStep}(X, u_{\text{ng}}^X, \Delta s)$ 
19:   if  $\tau = \text{ATG}$ :
20:     Sample  $X_0 \sim p_0(X)$  and  $X_{\sigma} \sim p_{\sigma}(X_{\sigma} | X_0, X_1)$ 
21:      $u_{\text{cond}}^A \leftarrow u_{\theta}^A(A, X_1, t_k, 1)$ ;  $u_{\text{corr}}^A \leftarrow u_{\theta}^A(A, X_{\sigma}, t_k, \sigma)$ 
22:      $u_{\text{ng}}^A \leftarrow (1 - \omega)u_{\text{corr}}^A + \omega u_{\text{cond}}^A$ 
23:      $A \leftarrow \text{CTMCStep}(A, u_{\text{ng}}^A, \Delta t)$ 
24:   end for
25: return  $(A, X)$ .
```

D. Ordering and Augmentation Analysis

Permutation space reduction analysis. To quantify the impact of our hierarchical permutation augmentation, we evaluate search space complexity in the MP-20 training set. For each crystal structure, we calculate the size of the reduced permutation space using the formula $\prod_i |\mathcal{W}^i|! \prod_j |O_j^i|!$, where i denotes groups sharing the same element and Wyckoff position, and j denotes individual orbits within those groups. We then group these values by the number of atoms in the unit cell (N) and calculate the minimum, median, and maximum values for each group to capture the distribution across different symmetry.

Figure D.1 illustrates the size of the permutation space as a function of N . While the naive permutation space grows factorially as $N!$, our proposed ordering combined with the hierarchical augmentation constraints drastically restricts the residual freedom. As shown in the figure, for a unit cell with $N = 20$ atoms, the median reduced space size is approximately $10^{7.2}$, whereas the full permutation space is $20! \approx 10^{18.4}$. This reduction of over 11 orders of magnitude provides a strong inductive bias, enabling the Diffusion Transformer to effectively learn crystallographic symmetry and compositional priors directly from data without explicit structural templates.

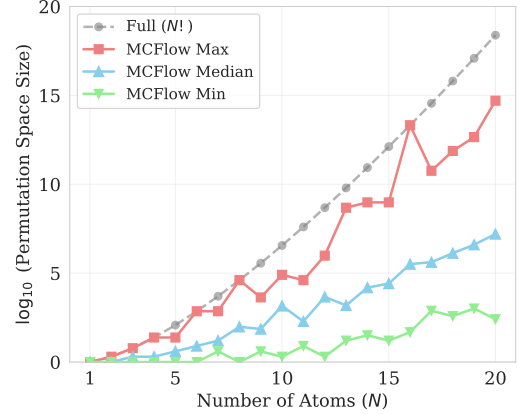


Figure D.1. **Permutation space reduction.** Logarithmic scale comparison of full permutation space $N!$ and reduced space $\prod_i |\mathcal{W}^i|! \prod_j |O_j^i|!$ across unit cell sizes $N = 1$ to 20.

Table E.1. **Ablation study of noisy guidance on CSP and atom type generation (ATG).** Noisy guidance improves structural fidelity in CSP and yields competitive performance in ATG across datasets. Single-sample generation per composition or structure. **Bold** indicates the best performance within each model size group.

Model	NG	CSP				ATG			
		MP-20		MPTS-52		MP-20		MPTS-52	
		MR (%) $\uparrow$	RMSE $\downarrow$	MR (%) $\uparrow$	RMSE $\downarrow$	Comp (%) $\uparrow$	MR (%) $\uparrow$	Comp (%) $\uparrow$	MR (%) $\uparrow$
MCFlow-S	✗	54.40	0.1339	22.58	0.1880	88.89	1.53	83.19	0.98
MCFlow-S	✓	56.10	0.1269	22.88	0.1845	88.97	1.57	83.14	1.01
MCFlow-B	✗	62.11	0.0624	25.32	0.1640	90.05	2.69	84.29	1.09
MCFlow-B	✓	63.14	0.0611	26.46	0.1572	90.17	3.12	83.56	1.20
MCFlow-L	✗	63.86	0.0546	26.63	0.1499	90.17	3.21	84.67	1.33
MCFlow-L	✓	64.08	0.0561	27.16	0.1401	90.07	3.32	83.67	1.24

E. Noisy Guidance

Noisy guidance for atom type generation. For atom type generation, we apply noisy guidance at the logit level of the categorical distribution, which induces guided rate vector for the corresponding CTMC.

Ablation study. We perform an ablation study on single-sample CSP to evaluate the effect of noisy guidance. Table E.1 shows that applying noisy guidance generally improves the match rate and RMSE on CSP across MP-20 and MPTS-52, indicating enhanced structural fidelity.

Sensitivity to noise level and guidance scale. Figure E.1 analyzes the sensitivity of noisy guidance to the atom-type noise level σ and guidance scale ω on single-sample CSP. We observe that introducing a small amount of noise ($\sigma > 0$) consistently improves both match rate and RMSE compared to the noiseless setting. Performance peaks at low noise levels $\sigma = 0.1$, while larger noise gradually degrades structural accuracy. While the method remains robust across a range of guidance scales, we observe that $\omega = 2.0$ consistently yields strong performance.

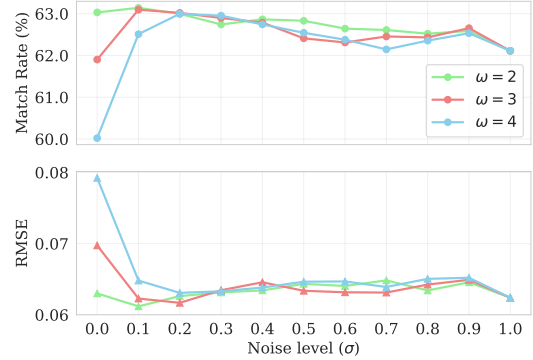


Figure E.1. **Sensitivity analysis of noisy guidance.** Single-sample CSP performance on MP-20 depending on the atom type noise level σ and guidance scale ω .

F. Model, Training, and Inference Configurations

We provide model architecture (Table F.1), training (Table F.2), and inference (Table F.3) configurations.

Table F.1. MCFlow model configurations.

Model scale	Small (S)	Base (B)	Large (L)
# Parameters	34M	134M	464M
Hidden dimension	384	768	1024
Transformer layers	12	12	24
Attention heads	6	12	16

Training settings. We train all models for a maximum of 10,000 epochs using the AdamW optimizer (Loshchilov & Hutter, 2019) with a constant learning rate of 1×10^{-4} and no weight decay. To balance learning dynamics across different modalities, we set the loss weights to 0.5 for atom types, 2.0 for fractional coordinates, and 1.0 for lattice parameters. The batch size is set to 256. For computational efficiency, all models are trained using Bfloat16 (BF16) mixed precision. To ensure training stability, we apply gradient clipping with a maximum norm of 10.0 and clip flow times t, s to 0.9 to mitigate numerical instability near the boundaries (Campbell et al., 2024). Additionally, we maintain an exponential moving average (EMA) of the model weights with a decay rate of 0.9999. The final model is selected by taking the latest epoch among the top three checkpoints ranked by total validity (structural and compositional) on the validation set. The LogNormal prior parameters for lattice lengths are fitted to the training data after dividing each length by $N^{1/3}$ (Table F.2).

Table F.2. MCFlow training configurations.

Hyperparameter	Value
Optimizer	AdamW
Learning rate	1×10^{-4}
Weight decay	0.0
Batch size	256
Max epochs	10,000
Precision	BF16 mixed precision
Gradient clipping	max norm 10.0
EMA decay	0.9999
Flow time clipping	$t, s \leq 0.9$
Loss weights ($A/F/L^l/L^a$)	0.5 / 2.0 / 1.0 / 1.0
LogNormal prior (μ, σ) MP-20	(0.847, 0.973, 1.251), (0.243, 0.216, 0.289)
LogNormal prior (μ, σ) MPTS-52	(0.788, 0.972, 1.255), (0.249, 0.206, 0.292)

Inference settings. During sampling, we solve the probability flow ODE for continuous variables and simulate the CTMC for discrete variables using 500 integration steps to generate crystals. To improve sample fidelity in single-sample CSP and structure-conditioned atom type generation, we apply noisy guidance with a guidance scale $\omega = 2.0$ and a noise level $\sigma = 0.1$. In contrast, for 20-sample generation, we disable noisy guidance to better capture the multimodal nature of the distribution. The number of atoms is sampled from the empirical distribution of atom counts in the training set.

Table F.3. MCFlow inference configurations.

Hyperparameter	Value
Integration steps	500
Guidance scale ω	2.0 (single-sample)
Noise level σ	0.1 (single-sample)

Computational resources. All experiments for the Small and Base models were conducted on a single NVIDIA GeForce RTX 3090 GPU, while experiments for the Large model were conducted on a single NVIDIA A40 GPU. Training the Base model takes approximately 2 days on MP-20 and 4 days on MPTS-52.

G. Evaluation Metrics

Crystal structure prediction. We compare generated structures against ground-truth test structures using `StructureMatcher` from `pymatgen` (Ong et al., 2013). The Match Rate (MR) is the fraction of generated candidates that match their corresponding ground-truth structures under symmetry-aware matching with tolerances `stol`=0.5, `ltol`=0.3, and `angle_tol`=10°. For matched pairs only, we compute the Root Mean Square Error (RMSE) of atomic positions, normalized by $(V/N)^{1/3}$, where V is the unit cell volume and N is the number of atoms.

De novo generation. We evaluate *de novo* generation using LeMat-GenBench (Betala et al., 2025), a standardized benchmark that provides a unified evaluation pipeline. We sample 2,500 structures and apply the LeMat-GenBench validity filter, which checks CIF readability, interatomic distances ($d_{ij} \geq 0.5(0.7 + r_i + r_j)$ Å using `pymatgen` atomic radii), density and lattice-parameter bounds, space-group determinability, and charge neutrality (oxidation-state analysis, tolerance 0.1). Validity is reported as the pass rate.

Thermodynamic stability is assessed via E_{hull} using an ensemble of three MLIPs (MACE-MP (Batatia et al., 2025), UMA (Wood et al., 2025), Orb-v3 (Rhodes et al., 2025)). Each MLIP relaxes the structure and constructs its own self-consistent convex hull from LeMat-Bulk (Siron et al., 2025), avoiding mixed MLIP-DFT inconsistencies. We report the ensemble mean and standard deviation of E_{hull} , formation energy (E_f), and relaxation RMSD; structures with $E_{\text{hull}} \leq 0$ are *stable* and those with $0 < E_{\text{hull}} \leq 0.1$ eV/atom are *metastable*.

Novelty and uniqueness use direct structure comparison via `pymatgen`’s `StructureMatcher` with a stricter `ltol`=0.1 than the CSP matching above (`ltol`=0.3), following the LeMat-GenBench convention. S.U.N. and MS.U.N. apply a stable $\rightarrow$ unique $\rightarrow$ novel funnel to the submitted set, with MS.U.N. restricted to the strict metastable range above. All count-based percentages (Valid, Unique, Novel, Stable, Metastable, S.U.N., MS.U.N.) are reported as count over the 2,500 submitted structures, following the LeMat-GenBench leaderboard convention; this caps each percentage above by the validity rate.

Atom type generation. We report compositional validity (Comp) following Xie et al. (2022): a generated composition is valid if its atomic species satisfy SMACT-based charge-neutrality and electronegativity rules. Match Rate is computed as in CSP, following Jiao et al. (2023).

Ablation study. Ablation experiments (Tables 5, H.4, H.7 and H.8) report MR and RMSE for CSP following Jiao et al. (2023), and structural and compositional validity for DNG following Xie et al. (2022): structural validity is the fraction of structures with all interatomic distances above 0.5 Å, and compositional validity is the fraction of compositions satisfying SMACT-based charge-neutrality and electronegativity rules.

Polymorph-aware evaluation. We additionally report METRe and cRMSE on a polymorph-aware MP-20 split, both introduced by Martirosyan et al. (2025). Standard match rate uses one-to-one pairing between generated and reference structures, which fails when a composition admits multiple stable polymorphs. METRe (Match Everyone to REference) instead compares every generated structure against every reference and counts a reference as matched whenever at least one generated structure falls within tolerance; the METRe rate is the fraction of reference structures that find at least one match. The corrected RMSE (cRMSE) penalizes unmatched references by assigning them an error equal to the matching tolerance `stol` (with `stol` = 0.5, the value used in our CSP evaluation):

$$\text{cRMSE} = \frac{\sum_{i \in \text{matched}} \text{RMSE}_i + \text{stol} \cdot (N_{\text{test}} - N_{\text{matched}})}{N_{\text{test}}}, \quad (26)$$

combining coverage (METRe) and precision (RMSE) into a single score.

H. Additional Results

We provide computational cost analysis (Table H.1), polymorph-aware METRe evaluation (Table H.2), template-conditioned CSP results (Table H.3), atom ordering and permutation ablations (Table H.4), scaling behavior (Figure H.1), filtered vs. unfiltered match rate (Table H.5), DNG on MPTS-52 (Table H.6), unified vs. task-specific training (Table H.7), and decoupled time axes ablation and comparison with adapted DNG baselines (Table H.8).

Table H.1. Computational cost comparison. MCFlow-B achieves faster per-step training and inference than the baselines despite more parameters, because Transformer dense operations are more hardware-efficient than E(3)-equivariant GNN message passing. Combined with faster metric saturation (Figure 3), MCFlow achieves lower total inference cost. We report wall-clock training time per epoch and inference time per integration step, measured on a single NVIDIA RTX 3090.

Model	Params	Batch	Train (1 Epoch)	Infer (1 Step)
DiffCSP	12.3M	256	26.0s	111.1ms
FlowMM	28.3M	256	52.11s	240.7ms
MCFlow-S	33.7M	256	8.54s	23.9ms
MCFlow-B	134.0M	256	22.27s	68.7ms
MCFlow-L	464.6M	128	77.68s	113.5ms

Table H.2. Polymorph-aware CSP evaluation using METRe (Martirosyan et al., 2025). MCFlow achieves the highest METRe under polymorph and original splits, indicating robustness to evaluation methodology. METRe (Match Everyone to REference) computes the fraction of reference structures that find at least one matching generated structure; this corrects the standard match rate, whose one-to-one pairing underestimates models that correctly generate multiple stable polymorphs per composition.

Split	Model	METRe (%) $\uparrow$	RMSE $\downarrow$	cRMSE $\downarrow$
Polymorph	DiffCSP	53.1	0.084	0.279
	FlowMM	65.2	0.079	0.226
	OMatG	70.5	0.056	0.187
	MCFlow-B	70.7	0.068	0.195
Original	DiffCSP	58.8	0.064	0.244
	FlowMM	67.0	0.067	0.210
	OMatG	66.0	0.058	0.208
	MCFlow-S	62.3	0.137	0.274
	MCFlow-B	69.7	0.070	0.200
	MCFlow-L	70.9	0.060	0.190

Table H.3. Template-conditioned crystal structure prediction on MP-20 and MPTS-52. MCFlow outperforms the template-conditioned baselines DiffCSP++ and SGFM on both datasets. Evaluation under structural templates extracted from the target structure; for MCFlow, MR and RMSE are computed on generated samples matching the target template.

Model	MP-20		MPTS-52	
	MR (%) $\uparrow$	RMSE $\downarrow$	MR (%) $\uparrow$	RMSE $\downarrow$
DiffCSP++	80.27	0.0295	46.29	0.0896
SGFM	82.74	0.0288	51.79	0.0827
MCFlow-S	79.21	0.0238	73.91	0.0415
MCFlow-B	85.31	0.0210	77.61	0.0361
MCFlow-L	85.96	0.0201	74.63	0.0386

Table H.4. Impact of atom ordering and permutation strategies on CSP and DNG. Electronegativity ordering outperforms atomic number ordering due to the direct correspondence with cation-anion roles exploited by the DiT. Full $N!$ random permutation substantially degrades performance, as the permutation space overwhelms learning capacity. All variants use MCFlow-B on MP-20.

Ordering	CSP		DNG	
	MR (%) $\uparrow$	RMSE $\downarrow$	Struct (%) $\uparrow$	Comp (%) $\uparrow$
Atomic Number	59.81	0.0951	99.30	88.09
Rand. Perm.	54.25	0.1184	99.45	86.12
Electronegativity	63.14	0.0611	99.44	89.85

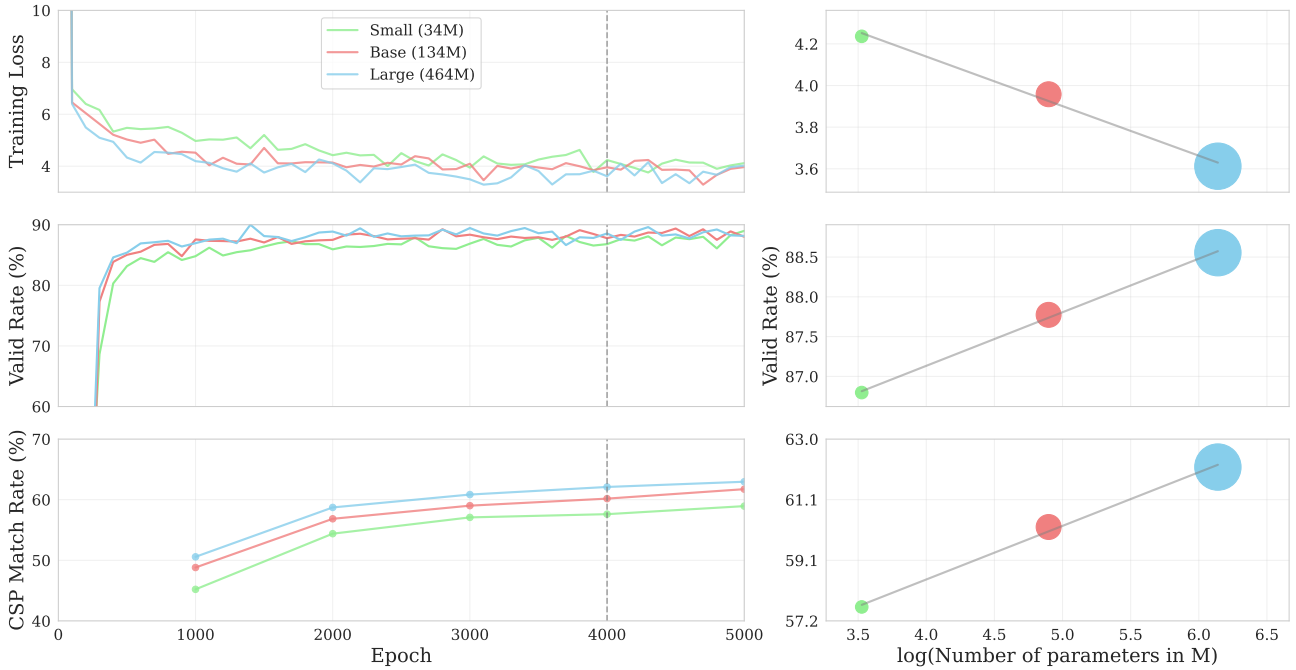


Figure H.1. Scaling behavior of MCFlow across model sizes. Larger MCFlow variants consistently improve training loss, DNG validity, and CSP match rate. Training loss, DNG valid rate and CSP match rate over training epochs for small, base, and large models (left). Correlation between model size and training loss, validity, and CSP match rate at epoch 4000 (right).

Table H.5. **Filtered and unfiltered CSP match rates on MP-20 and MPTS-52.** MCFlow-L consistently outperforms all baselines under both evaluation schemes. The unfiltered MR is included to capture physically real structures ($\sim 10\%$ in MP-20) that fail SMACT validity rules. Performance is reported as filtered / unfiltered MR using single-sample generation with noisy guidance. OMatG corresponds to the LinearODE variant.

Model	MP-20		MPTS-52	
	MR (%) $\uparrow$	RMSE $\downarrow$	MR (%) $\uparrow$	RMSE $\downarrow$
DiffCSP	52.51 / 57.82	0.0600 / 0.0627	14.29 / 15.79	0.1489 / 0.1533
FlowMM	59.98 / 66.22	0.0629 / 0.0661	20.28 / 22.29	0.1486 / 0.1541
OMatG	63.75 / 69.83	0.0720 / 0.0741	25.15 / 27.38	0.1931 / 0.1970
MCFlow-S	56.10 / 61.78	0.1269 / 0.1321	22.88 / 24.74	0.1846 / 0.1920
MCFlow-B	63.14 / 69.23	0.0611 / 0.0663	26.46 / 28.77	0.1577 / 0.1610
MCFlow-L	64.08 / 70.38	0.0561 / 0.0592	27.16 / 29.58	0.1401 / 0.1497

Table H.6. **De novo generation results on the MPTS-52 benchmark.** MCFlow achieves the highest compositional validity and competitive structural validity on MPTS-52. The MPTS-52 dataset contains 100.00% structurally valid and 88.24% compositionally valid structures. **Bold** and underline indicate the best and second-best results, respectively.

Model	Structural Validity (%) $\uparrow$	Compositional Validity (%) $\uparrow$
DiffCSP	67.47	55.80
SymmCD	87.11	78.18
DiffCSP++	99.87	77.52
SGEquivDiff	97.79	79.83
MCFlow-S	96.73	84.62
MCFlow-B	<u>98.34</u>	85.57
MCFlow-L	97.34	<u>85.21</u>

Table H.7. **CSP-only vs. unified training on MP-20.** CSP-only ($t = 1$ during training) and unified training yield nearly identical MR, indicating no destructive interference. The unified model additionally supports DNG, ATG, and masked ATG without retraining, and noisy guidance (NG) provides a further gain only available through decoupled time training.

Method	MR (%) $\uparrow$	RMSE $\downarrow$
CSP-only	61.92	0.0638
MCFlow-B w/o NG	62.11	0.0624
MCFlow-B	63.14	0.0611

Table H.8. **Decoupled time axes ablation and comparison with adapted DNG baselines on MP-20 CSP.** Decoupled time axes are essential: coupling $t=s$ substantially degrades CSP performance. DNG-trained DiffCSP and FlowMM, adapted to CSP by overwriting composition with ground-truth values at every integration step, also achieve substantially lower MR due to their coupled noise schedules.

Method	MR (%) $\uparrow$	RMSE $\downarrow$
DiffCSP DNG	3.76	0.3396
FlowMM DNG	27.01	0.1947
MCFlow-B coupled ($t=s$)	45.58	0.1499
MCFlow-B	63.14	0.0611

I. Visualizations of Samples

We visualize qualitative samples from MCFlow across the three tasks: crystal structure prediction (Figure I.1), atom type generation (Figure I.2), and *de novo* generation (Figure I.3).

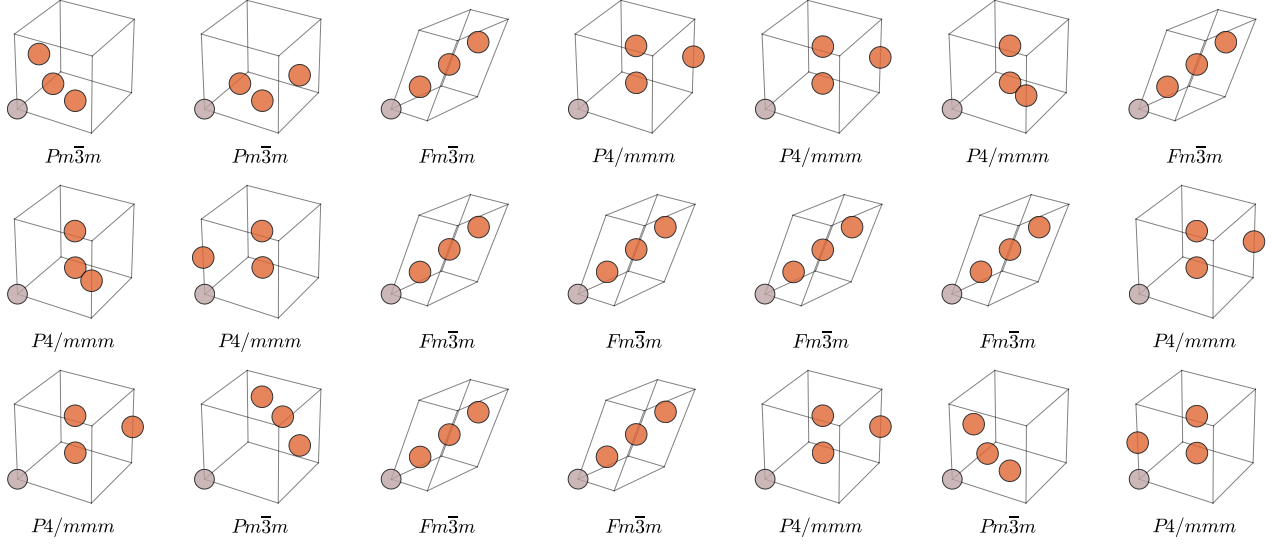


Figure I.1. **Crystal structure prediction.** MCFlow explores multiple crystal structures with different space groups conditioned on a single composition, illustrated for the alloy AlFe_3 (top-left, target crystal).

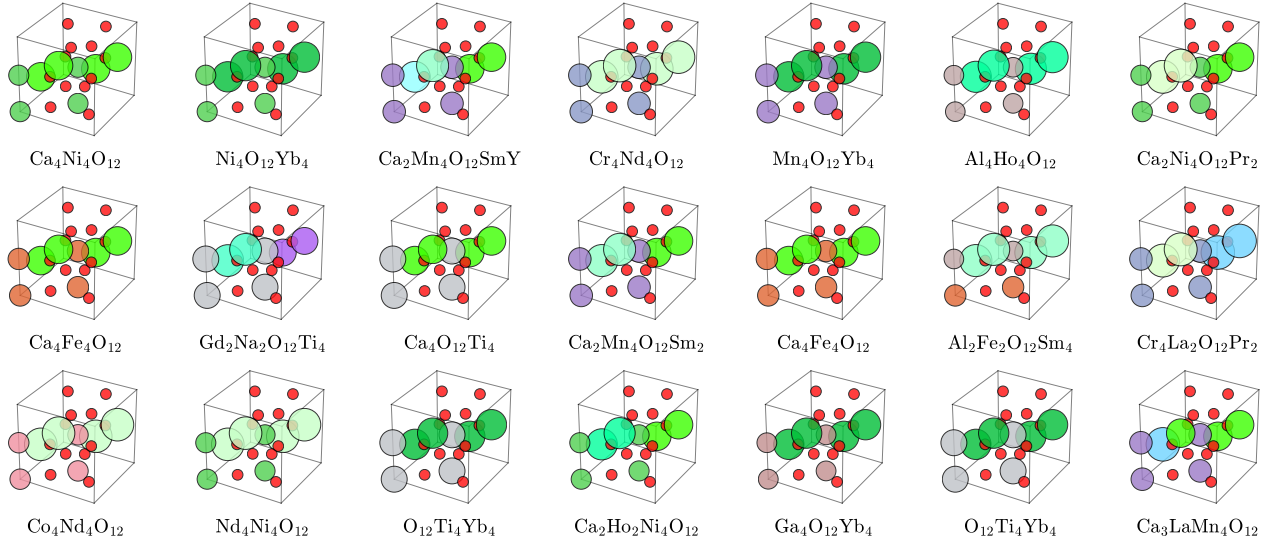


Figure I.2. **Atom type generation.** MCFlow captures multimodal compositional distributions conditioned on a structure, generating diverse compositions for a given perovskite structure of $\text{Ca}_4\text{Ni}_4\text{O}_{12}$ (top-left, target crystal).

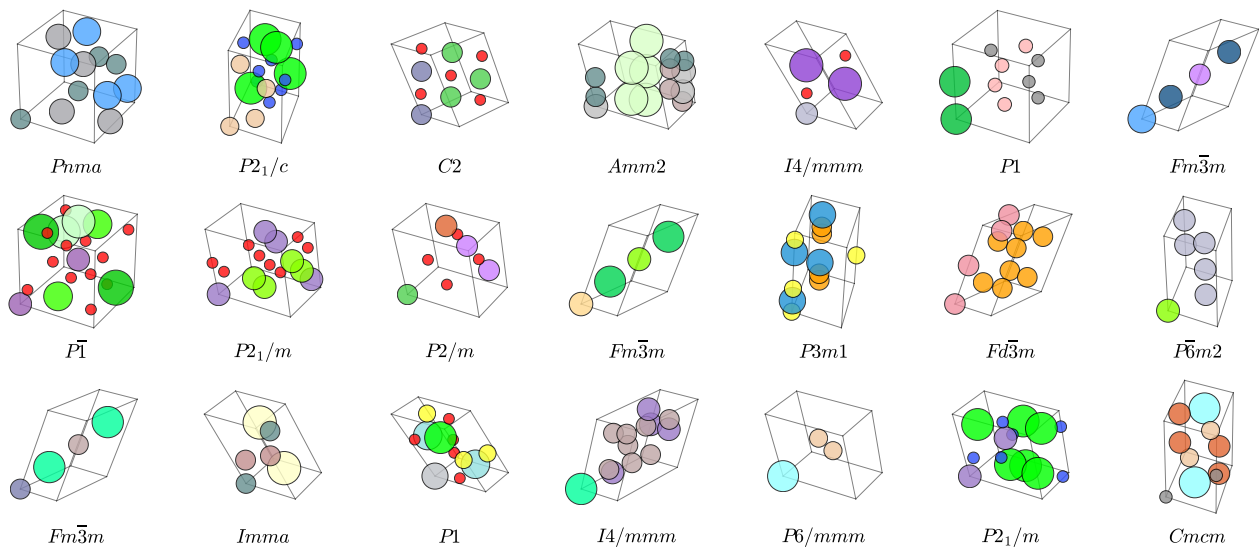


Figure I.3. **De novo generation.** MCFlow jointly generates diverse atom types and crystal structures.