CrystalREPA: Transferring Physical Priors from Universal MLIPs to Crystal Generative Models

Chengqian Zhang[1,2], **Yucheng Jin**[1], **Duo Zhang**[1,2], **Tiejun Li**[2,3,4,*], **Han Wang**[5,6,*]

[1]AI for Science Institute, Beijing, China;
[2]Center for Data Science, Peking University, Beijing, China;
[3]LMAM and School of Mathematical Sciences, Peking University, Beijing, China;
[4]Center for Machine Learning Research, Peking University, Beijing, China;
[5]National Key Laboratory of Computational Physics,
Institute of Applied Physics and Computational Mathematics, Beijing, China;
[6]HEDPS, CAPT, College of Engineering, Peking University, Beijing, China;

Abstract

Crystal generative models mainly learn what stable crystals look like, with little explicit supervision for what makes them stable. We reveal a substantial representation gap between state-of-the-art crystal generative models and pretrained universal machine learning interatomic potentials (MLIPs) via energy probing, and show this gap can be closed by a simple training-time alignment. We propose Crystal REPresentation Alignment (CrystalREPA), a plug-and-play framework that aligns the atom-wise hidden states of generative encoders with frozen MLIP representations through an element-aware contrastive objective, transferring stability-aware atomistic priors with marginal training overhead and no additional inference cost. Across three generative frameworks, ten MLIP teachers, and two benchmark datasets, CrystalREPA consistently improves the thermodynamic stability, structural validity, and structural fidelity of generated crystals. Equally important, we find that an MLIP's transfer effectiveness is poorly predicted by its accuracy on standard leaderboards (e.g., Matbench Discovery) but strongly predicted by the distinguishability of its atom-wise representation space, yielding a practical, accuracy-independent criterion for selecting MLIP teachers for generative transfer. Code is available at https://github.com/Chengqian-Zhang/CrystalREPA.

1 Introduction

The inverse design of crystalline materials aims to generate novel structures with desired stability and functionality, but remains challenging due to the vast combinatorial space of atomic configurations and the stringent periodic constraints of crystals [1]. Recent advances in generative modeling, particularly diffusion- and flow-based approaches, have enabled direct sampling of candidate crystal structures from data-driven distributions, marking a significant step toward automated materials discovery [2, 3, 4, 5].

Despite this progress, current crystal generative models mainly learn what stable crystals look like, rather than what makes a configuration stable. Popular benchmark datasets such as MP-20 [6] and Alex-MP-20 [3] consist primarily of stable or metastable equilibrium structures. Training on such data encourages models to reproduce plausible geometries and compositions, but exposes them to only a narrow region of the potential energy landscape. In particular, the model rarely observes

*Correspondence to: tieli@pku.edu.cn, wang_han@iapcm.ac.cn

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

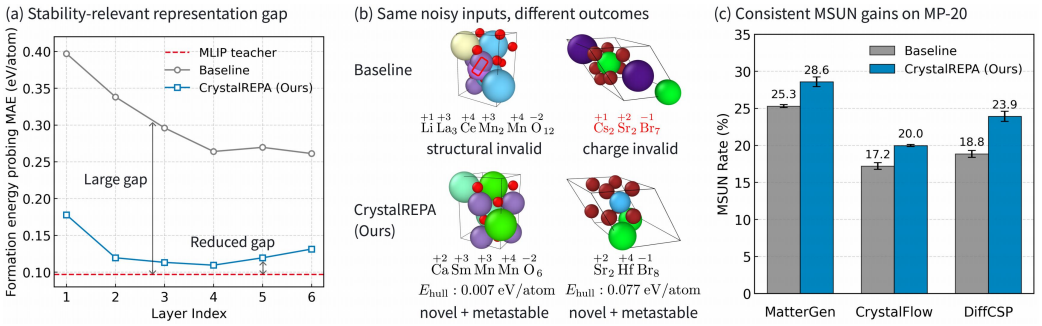


Figure 1: CrystalREPA transfers stability-aware priors from universal MLIPs to crystal generative models. (a) Formation energy probing reveals that vanilla generative representations are substantially less predictive of stability than pretrained MLIP representations, while CrystalREPA narrows this gap through representation alignment. (b) Given the same noisy initial structures, vanilla CrystalFlow can generate physically invalid crystals, whereas CrystalREPA generates novel and metastable structures. (c) CrystalREPA also yields consistent gains in MSUN (Metastable, Unique, and Novel) rate across MatterGen, CrystalFlow, and DiffCSP, indicating better discovery-oriented generation quality. Panels (a) and (b) use CrystalFlow, and panel (c) compares all three generators; all CrystalREPA results shown here use DPA-3.1-3M as the MLIP teacher.

non-equilibrium, distorted, or high-energy configurations, nor does it receive explicit energy or force labels that quantify how unfavorable such configurations are. As a result, generated crystals may appear geometrically plausible while still deviating from low-energy or physically reliable regions of configuration space.

Universal machine learning interatomic potentials (MLIPs) offer a complementary source of supervision [7]. Unlike crystal generative models, MLIPs are trained on large-scale atomistic datasets with explicit energy and force labels, covering not only relaxed structures but also perturbed, non-equilibrium, and high-energy configurations. This supervision encourages MLIPs to learn representations that are sensitive to variations in the potential energy surface and therefore relevant to stability. Empirically, this supervision difference manifests as a substantial *representation gap*: features learned by crystal generative models are far less predictive of formation energy than those of pretrained universal MLIPs. This motivates the central question of this work: *can we transfer stability-aware atomistic priors from universal MLIPs to crystal generative models without compromising generative flexibility?*

To address this question, we propose *Crystal REPresentation Alignment (CrystalREPA)*, a simple and general framework for transferring stability-aware priors from pretrained universal MLIPs to crystal generative models. Inspired by representation alignment in image generation [8], CrystalREPA aligns the atom-wise representations of generative encoders with those of pretrained universal MLIPs through an element-aware contrastive objective, encouraging the crystal generative model to learn features predictive of stability. We validate CrystalREPA across multiple popular generative frameworks, universal MLIPs, and benchmark datasets. Extensive experiments demonstrate that CrystalREPA consistently improves the thermodynamic stability, structural validity, and structural fidelity of generated crystals while preserving diversity, introducing only marginal training overhead and no extra inference cost. Beyond performance gains, we further show that the representational distinguishability of the teacher MLIPs strongly correlates with downstream generative quality, suggesting it as an important factor associated with transfer effectiveness.

Our contributions are summarized as follows:

- We make the geometry-vs-energy gap quantitative: via formation-energy probing, crystal generative model representations are far less predictive of stability than pretrained universal MLIPs.
- We propose CrystalREPA, which transfers stability-aware priors by aligning atom-wise hidden states with frozen MLIP representations via an element-aware contrastive objective, and consistently improves stability, validity, and fidelity across three generative frameworks, ten MLIP teachers, and two datasets, at marginal training cost and no inference cost.

- We find that an MLIP teacher's transfer effectiveness is predicted by its atom-wise representation distinguishability rather than leaderboard accuracy, yielding a practical teacher-selection criterion.

2 Related Work

Crystal Generative Models. Generative modeling has become a promising paradigm for inverse crystal design, with diffusion-based methods [9, 10] becoming the dominant approach for modeling periodic atomic structures [6, 11, 5, 3, 12, 13]. Representative models such as DiffCSP [5] and MatterGen [3] adopted joint diffusion frameworks that simultaneously model atomic composition, fractional coordinates, and lattice parameters, enabling end-to-end sampling of candidate materials from data distributions. Subsequent studies have incorporated crystallographic priors such as symmetry constraints to reduce the effective search space and improve structural validity [14, 15, 16, 17]. In parallel, flow-based approaches [18, 19, 20] have recently been introduced for crystal generation [21, 22, 4]. By learning continuous transport from a simple base distribution to the data distribution, these methods often offer more efficient sampling while retaining competitive generation quality. Despite these advances, most crystal generative models learn the geometry of stable crystals without supervision from the potential energy surface that governs stability.

Universal MLIPs. Universal MLIPs have recently achieved robust generalization across broad chemical spaces by training on large-scale atomistic datasets such as MPtrj [23], Alexandria [24, 25], and OMat24 [26] with explicit energy and force supervision [27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43]. Beyond their use for accurate energy prediction and molecular dynamics, pretrained MLIPs have been shown to encode atomistic representations sensitive to variations in the potential energy surface [42, 44]. However, how to exploit these pretrained representations as transferable stability-aware priors for crystal generation remains largely unexplored.

Position of our work. Our work connects crystal generative modeling with atomistic representation learning. Instead of improving generators through architectural design, symmetry constraints, or sampling procedures, we enhance their internal representations with stability-aware priors. Instead of using MLIPs for energy prediction or post-generation relaxation, we use pretrained universal MLIPs as frozen representation teachers during generative training. Building on representation alignment in image generation [8], our contribution lies in transferring energy- and force-supervised MLIP representations to crystal generative models through atom-wise, element-aware alignment.

3 Preliminaries

Our method builds on two components: a crystal generative model that produces atom-wise representations during denoising or transport, and a pretrained universal MLIP that provides atom-wise representations supervised by energy and forces. We use CrystalFlow [4] as an example to introduce the former, and then review how universal MLIPs learn representations from energy and force supervision.

3.1 Crystal generative modeling

A crystal with N atoms in the unit cell is represented by $\mathcal{M} = (\mathbf{A}, \mathbf{F}, \mathbf{L})$, where $\mathbf{A} = [\mathbf{a}_1^\top; \mathbf{a}_2^\top; \ldots; \mathbf{a}_N^\top] \in \mathbb{R}^{N\times a}$ denotes atom types encoded as categorical vectors. $\mathbf{F} = [\mathbf{f}_1^\top; \mathbf{f}_2^\top; \ldots; \mathbf{f}_N^\top] \in [0,1)^{N\times 3}$ denotes fractional coordinates, and $\mathbf{L} \in \mathbb{R}^6$ denotes a rotation-invariant lattice parameterization derived from the lattice matrix $\tilde{\mathbf{L}} = [\mathbf{l}_1^\top; \mathbf{l}_2^\top; \mathbf{l}_3^\top] \in \mathbb{R}^{3\times 3}$. Following CrystalFlow, crystal generation is formulated as learning a continuous transport map from a simple prior distribution to the data distribution.

Specifically, let $\mathcal{M}_1 = (\mathbf{A}_1, \mathbf{F}_1, \mathbf{L}_1)$ be a crystal sampled from the data distribution and let $\mathcal{M}_0 = (\mathbf{A}_0, \mathbf{F}_0, \mathbf{L}_0)$ be sampled from the corresponding priors. A continuous path $\mathcal{M}_t = (\mathbf{A}_t, \mathbf{F}_t, \mathbf{L}_t)$, with $t \in [0,1]$, interpolates between noise and data, where $t = 0$ corresponds to the prior and $t = 1$ corresponds to the clean crystal. This path induces target velocity fields $\mathbf{u}_t^A$, $\mathbf{u}_t^F$, and $\mathbf{u}_t^L$ for atom types, fractional coordinates, and lattices, respectively. The mathematical formulation of this path is provided in Appendix A.1.

CrystalFlow learns three time-dependent vector fields, $\mathbf{v}_\theta^A(\mathcal{M}_t, t)$, $\mathbf{v}_\theta^F(\mathcal{M}_t, t)$, and $\mathbf{v}_\theta^L(\mathcal{M}_t, t)$, to regress these targets. These vector fields are parameterized by a periodic $E(3)$-equivariant graph neural network with K message-passing layers, whose final atom-wise hidden states are denoted by $f^{(K)}(\mathcal{M}_t, t) \in \mathbb{R}^{N \times d_1}$. The velocity fields are predicted as

$$\mathbf{v}_\theta^A = \varphi_A(f^{(K)}(\mathcal{M}_t, t)), \quad \mathbf{v}_\theta^F = \varphi_F(f^{(K)}(\mathcal{M}_t, t)), \quad \mathbf{v}_\theta^L = \varphi_L(\frac{1}{N}\sum_i^N f_i^{(K)}(\mathcal{M}_t, t)), \quad (1)$$

where φ_A, φ_F and φ_L are prediction heads. The model is trained with a weighted regression objective

$$\mathcal{L}_{\text{gen}} = \lambda_A \mathcal{L}_A(\mathbf{v}_\theta^A, \mathbf{u}_t^A) + \lambda_F \mathcal{L}_F(\mathbf{v}_\theta^F, \mathbf{u}_t^F) + \lambda_L \mathcal{L}_L(\mathbf{v}_\theta^L, \mathbf{u}_t^L), \quad (2)$$

where $\mathcal{L}_A$, $\mathcal{L}_F$, and $\mathcal{L}_L$ are the corresponding regression losses for atom types, fractional coordinates, and lattice variables.

For our purpose, the key observation is that the generative encoder produces atom-wise representations on corrupted intermediate structure $\mathcal{M}_t$. These representations are optimized for denoising or transport, but are not supervised by signals from the potential energy surface.

3.2 Universal MLIPs

Universal MLIPs are atomistic models trained on large-scale crystal datasets with explicit energy and force supervision. Given a crystal structure $\mathcal{M} = (\mathbf{A}, \mathbf{F}, \mathbf{L})$ with energy label E and force label $\mathbf{F}_{\text{force}}$, an MLIP produces atom-wise representations through an atomistic encoder $\tilde{f}$ and predicts the total energy by aggregating atom-level contributions:

$$\hat{E} = \sum_{i=1}^N \varphi_E\big(\tilde{f}_i(\mathcal{M})\big), \quad (3)$$

where φ_E is an MLP and $\tilde{f}(\mathcal{M}) \in \mathbb{R}^{N \times d_2}$ denotes the atom-wise representations of $\mathcal{M}$. Atomic forces $\hat{\mathbf{F}}_{\text{force}} \in \mathbb{R}^{N \times 3}$ are then obtained as the negative gradient of the predicted energy with respect to atomic Cartesian coordinates:

$$\hat{\mathbf{F}}_{\text{force}} = -\nabla_{\mathbf{R}} \hat{E}, \qquad \mathbf{R} = \mathbf{F}\tilde{\mathbf{L}}. \quad (4)$$

These models are typically trained with a combination of energy and force losses:

$$\mathcal{L}_{\text{MLIP}} = \lambda_{\text{ener}} \mathcal{L}_{\text{ener}}(\hat{E}, E) + \lambda_{\text{force}} \mathcal{L}_{\text{force}}(\hat{\mathbf{F}}_{\text{force}}, \mathbf{F}_{\text{force}}). \quad (5)$$

Because accurate energy and force prediction requires sensitivity to energy-relevant structural variations, MLIP atom-wise representations provide a natural source of stability-aware supervision for guiding crystal generative models.

4 CrystalREPA

Motivation. Existing crystal generative models learn the *geometry* of stable crystals, but their training objectives never expose them to the *potential energy surface* that determines stability. Motivated by the representation gap revealed by energy probing in Fig. 1(a) and the qualitative failures in Fig. 1(b), we propose Crystal REPresentation Alignment (CrystalREPA), a simple training framework that transfers stability-aware priors from pretrained universal MLIPs to crystal generative models.

Core idea. The central idea is to align the atom-wise hidden states of a generative model on a corrupted intermediate structure $\mathcal{M}_t = (\mathbf{A}_t, \mathbf{F}_t, \mathbf{L}_t)$ with the atom-wise representations of a pretrained universal MLIP on the corresponding clean crystal $\mathcal{M} = (\mathbf{A}, \mathbf{F}, \mathbf{L})$. The clean crystal is used as the MLIP input to provide a stability-aware target for the generative path. This choice is in line with the objective of learning the transport or denoising direction from $\mathcal{M}_t$ toward the clean structure. Appendix C.6 empirically validates this design. In this way, the generative encoder is encouraged to learn features that are not only useful for generation, but also consistent with representation patterns learned from large-scale energy and force supervision.

Formally, we align rotationally invariant atom-wise representations from both the generative encoder and the MLIP teacher. Let $\tilde{\mathbf{z}} = \tilde{f}(\mathcal{M}) \in \mathbb{R}^{N \times d_2}$ denote the invariant atom-wise teacher representations extracted from a pretrained universal MLIP, which is kept frozen throughout training. Let

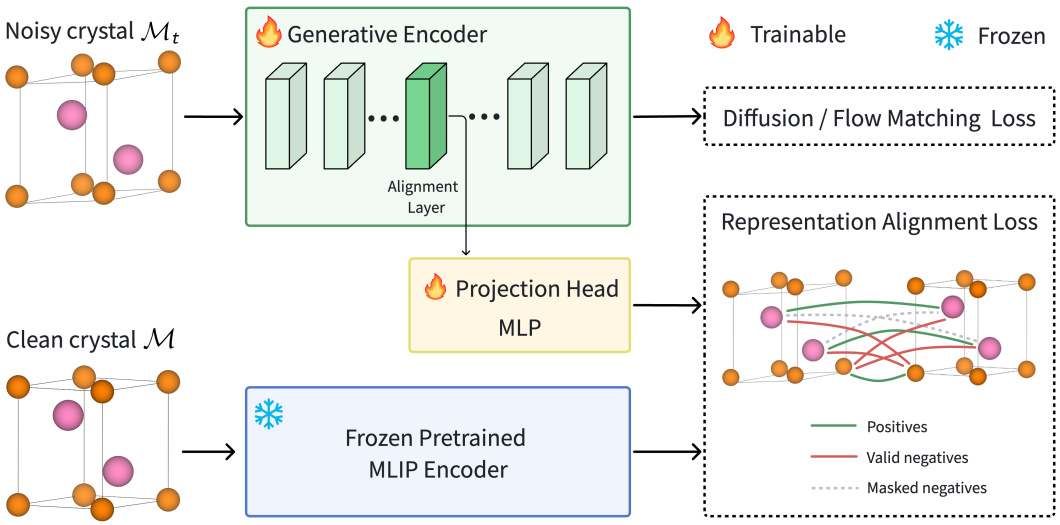


Figure 2: Overview of CrystalREPA. During training, CrystalREPA aligns atom-wise hidden states from the generative encoder with frozen MLIP representations of the corresponding clean crystal through a projection head and an element-aware contrastive loss. The projection head and MLIP teacher representations are used only during training, leaving the original sampling procedure unchanged during inference.

$\mathbf{h}_t^{(k)} = f^{(k)}(\mathcal{M}_t, t) \in \mathbb{R}^{N \times d_1}$ denote the invariant atom-wise hidden states of the generative model at layer k. Since the teacher and student representations generally live in different latent spaces, we introduce a lightweight projection head φ_p to map the student features into the teacher space: $\mathbf{z} = \varphi_p(\mathbf{h}_t^{(k)}) \in \mathbb{R}^{N \times d_2}$. Here, φ_p is implemented as an MLP and is only used during training.

In practice, we compute the contrastive alignment objective over all atoms in a mini-batch rather than within each crystal independently. Let $\mathcal{I}$ denote the flattened set of atoms from all crystals in the mini-batch, and let $N_B = |\mathcal{I}|$ be the total number of atoms. Because $\mathcal{M}_t$ is constructed by corrupting the clean crystal $\mathcal{M}$ without permuting atoms, the atom ordering is preserved within each crystal. Thus, for each atom $i \in \mathcal{I}$, the projected student representation $\mathbf{z}_i$ and teacher representation $\tilde{\mathbf{z}}_i$ form a natural positive pair. We compute the cosine similarity matrix $\mathbf{S} \in \mathbb{R}^{N_B \times N_B}$ between all projected student representations and teacher representations in the mini-batch:

$$\mathbf{S}_{ij} = \frac{\mathbf{z}_i^\top \tilde{\mathbf{z}}_j}{\|\mathbf{z}_i\| \|\tilde{\mathbf{z}}_j\|}, \quad i, j \in \mathcal{I}. \tag{6}$$

A standard contrastive objective would treat all off-diagonal pairs as negatives [45]. However, this is suboptimal for crystals: atoms of the same element often share similar chemical characteristics and may therefore be naturally close in the teacher representation space. Treating such pairs as negatives introduces false-negative supervision and can undesirably distort the representation geometry. To mitigate this issue, we introduce an element-aware mask $\mathbb{M}$, constructed from the clean atom types $\mathbf{A}$, that removes same-element off-diagonal pairs from the negative set while keeping diagonal positive pairs unchanged.

$$\mathbb{M}_{ij} = \begin{cases} 0, & i \neq j \text{ and } \mathbf{A}_i = \mathbf{A}_j, \\ 1, & \text{otherwise.} \end{cases} \tag{7}$$

Based on this mask, we define the representation alignment loss as a symmetric Element-Aware InfoNCE (EA-NCE) objective in Eq. (8), where τ is a temperature hyperparameter. This objective encourages atom-wise consistency between student representations and teacher representations, while avoiding unnecessary repulsion between chemically similar atoms of the same species.

$$\mathcal{L}_{\text{REPA}} = -\frac{1}{2N_B} \sum_{i=1}^{N_B} \left(\log \frac{\exp(\mathbf{S}_{ii}/\tau)}{\sum_{j=1}^{N_B} \mathbb{M}_{ij} \exp(\mathbf{S}_{ij}/\tau)} + \log \frac{\exp(\mathbf{S}_{ii}/\tau)}{\sum_{j=1}^{N_B} \mathbb{M}_{ji} \exp(\mathbf{S}_{ji}/\tau)} \right). \tag{8}$$

Finally, we combine the alignment objective with the original generative loss. The resulting training objective contains both the original generation loss and the representation alignment loss, weighted

Table 1: Performance comparison of CrystalREPA on MP-20. We evaluate robustness across three crystal generative frameworks and 10 universal MLIP teachers. "Avg. over 10 MLIPs" reports the average metric value across all MLIP teachers. Values in parentheses denote standard errors over five independent sampling and evaluation runs.

Model	Target Rep.	Metastable%↑	MSUN%↑	Stable%↑	SUN%↑	$\bar{E}_{\text{hull}}$ ↓	RMSD↓	Valid%↑	Unique%↑	Novel%↑
Vanilla CrystalFlow [4]		41.2(0.5)	17.2(0.5)	5.1(0.3)	1.2(0.1)	0.198(0.003)	0.60(0.01)	84.9(0.6)	99.6(0.2)	**73.1**(0.4)
+CrystalREPA	Avg. over 10 MLIPs	50.5	19.1	6.2	1.5	0.178	0.51	86.1	99.7	65.0
(Ours)	DPA-3.1-3M	**55.0**(0.7)	**20.0**(0.1)	**7.0**(0.4)	1.4(0.1)	0.179(0.005)	0.46(0.01)	86.6(0.4)	99.5(0.1)	61.3(0.7)
	DPA-3.1-3M-FT	52.5(0.7)	18.6(0.5)	**7.0**(0.5)	1.5(0.2)	**0.174**(0.006)	0.55(0.01)	**87.5**(0.1)	99.6(0.1)	62.7(0.4)
	DPA-3.1-MPtrj	51.5(0.4)	19.8(0.2)	6.2(0.7)	1.7(0.2)	**0.174**(0.004)	0.53(0.00)	86.7(1.0)	99.7(0.0)	65.0(0.5)
	MACE-MPA-0	47.8(0.5)	19.6(0.5)	5.8(0.2)	1.4(0.1)	0.176(0.005)	**0.42**(0.00)	84.6(0.6)	**99.9**(0.1)	68.2(0.3)
	MACE-MP-0	49.1(0.5)	18.7(0.6)	5.9(0.3)	1.5(0.2)	0.178(0.006)	0.54(0.01)	85.4(0.5)	99.7(0.0)	65.5(0.9)
	ORB v3	50.1(0.7)	18.0(0.5)	6.2(0.2)	1.4(0.1)	**0.174**(0.006)	0.57(0.01)	86.3(0.8)	99.7(0.1)	63.8(0.5)
	ORB v2	51.5(0.4)	**20.0**(0.4)	6.2(0.5)	1.6(0.1)	0.178(0.005)	0.46(0.01)	85.6(0.7)	99.8(0.0)	64.6(0.6)
	SevenNet-Omni-i12	46.5(0.6)	18.5(0.5)	5.7(0.3)	1.4(0.1)	0.190(0.005)	0.59(0.01)	85.2(0.7)	99.7(0.0)	69.0(1.0)
	SevenNet-l3i5	49.5(0.4)	19.0(0.2)	5.7(0.2)	1.3(0.2)	0.176(0.005)	0.51(0.00)	86.6(0.6)	99.7(0.0)	65.7(0.4)
	MatterSim-v1-5M	51.3(0.7)	19.4(0.8)	6.5(0.2)	**1.8(0.1)**	0.179(0.006)	0.51(0.01)	86.2(0.8)	99.8(0.1)	64.2(0.6)
Vanilla MatterGen [3]		57.4(0.5)	25.3(0.2)	6.7(0.4)	1.9(0.2)	0.148(0.003)	0.12(0.00)	84.7(0.5)	99.6(0.0)	**64.6**(0.7)
+CrystalREPA	Avg. over 10 MLIPs	62.7	27.7	7.8	2.5	0.131	0.11	85.6	99.5	61.8
(Ours)	DPA-3.1-3M	64.3(0.7)	28.6(0.7)	7.4(0.3)	2.3(0.2)	0.129(0.004)	**0.11**(0.00)	85.8(0.4)	99.3(0.1)	60.7(0.9)
	DPA-3.1-3M-FT	63.7(0.7)	27.6(0.4)	**8.7**(0.4)	2.7(0.1)	0.125(0.003)	**0.11**(0.00)	85.5(0.6)	99.5(0.1)	60.5(0.6)
	DPA-3.1-MPtrj	62.5(0.3)	28.0(0.8)	7.3(0.2)	2.4(0.2)	0.127(0.003)	0.12(0.00)	84.5(0.2)	99.5(0.1)	62.7(0.8)
	MACE-MPA-0	61.3(0.6)	28.4(0.7)	7.8(0.1)	2.6(0.1)	0.138(0.003)	**0.11**(0.00)	85.2(0.6)	**99.7**(0.1)	63.5(0.5)
	MACE-MP-0	61.3(0.5)	27.6(0.5)	7.3(0.2)	2.4(0.2)	0.135(0.004)	**0.11**(0.00)	85.1(0.5)	99.5(0.1)	63.1(0.5)
	ORB v3	**66.2**(0.4)	**29.2**(0.4)	8.2(0.3)	2.6(0.2)	**0.119**(0.004)	**0.11**(0.00)	86.5(0.6)	99.5(0.1)	59.9(0.3)
	ORB v2	60.9(0.7)	26.8(0.3)	7.7(0.6)	2.5(0.2)	0.136(0.006)	0.14(0.00)	86.1(0.2)	99.5(0.1)	63.2(0.5)
	SevenNet-Omni-i12	59.5(0.5)	25.7(0.8)	7.3(0.3)	2.0(0.1)	0.143(0.007)	**0.11**(0.00)	85.7(0.5)	99.4(0.1)	62.8(0.7)
	SevenNet-l3i5	63.1(0.7)	27.0(0.5)	7.9(0.6)	2.5(0.2)	0.133(0.004)	**0.11**(0.00)	**86.6**(0.3)	99.6(0.1)	60.3(0.4)
	MatterSim-v1-5M	64.1(0.5)	28.5(0.6)	8.4(0.3)	**2.9**(0.2)	0.127(0.001)	**0.11**(0.00)	85.4(0.6)	99.6(0.0)	61.5(0.5)
Vanilla DiffCSP [5]		43.2(0.7)	18.8(0.5)	5.9(0.3)	2.0(0.2)	0.208(0.005)	0.36(0.01)	82.7(0.6)	98.9(0.1)	**72.0**(0.4)
+CrystalREPA	Avg. over 10 MLIPs	50.9	23.4	7.2	2.6	0.166	0.37	84.3	99.2	69.3
(Ours)	DPA-3.1-3M	50.6(0.8)	23.9(0.8)	7.5(0.7)	2.8(0.3)	0.166(0.006)	0.37(0.01)	84.1(0.4)	99.2(0.1)	70.0(1.0)
	DPA-3.1-3M-FT	52.0(0.8)	24.3(0.7)	**8.2**(0.2)	**2.9**(0.3)	0.165(0.006)	0.36(0.00)	84.5(0.6)	99.2(0.1)	68.6(0.6)
	DPA-3.1-MPtrj	50.2(0.4)	24.7(0.7)	6.7(0.1)	2.5(0.1)	0.175(0.004)	0.40(0.01)	84.9(0.5)	**99.5**(0.1)	71.4(0.6)
	MACE-MPA-0	49.9(0.8)	23.4(0.5)	6.0(0.4)	2.1(0.2)	0.169(0.004)	0.38(0.01)	84.8(0.9)	99.2(0.1)	70.4(0.6)
	MACE-MP-0	49.8(0.6)	23.4(0.4)	6.4(0.5)	2.5(0.3)	0.166(0.004)	0.37(0.01)	84.4(0.8)	99.2(0.1)	70.4(0.5)
	ORB v3	**56.5**(0.8)	**24.8**(0.8)	8.1(0.3)	2.5(0.1)	**0.147**(0.003)	**0.32**(0.00)	84.5(0.7)	99.2(0.1)	65.3(0.4)
	ORB v2	53.5(1.0)	23.2(1.1)	8.1(0.1)	2.7(0.1)	0.158(0.007)	0.36(0.01)	84.9(0.6)	99.2(0.0)	66.2(0.7)
	SevenNet-Omni-i12	45.8(0.7)	21.0(0.8)	6.6(0.5)	2.6(0.1)	0.180(0.005)	0.35(0.01)	83.2(0.4)	98.8(0.1)	**72.0**(0.5)
	SevenNet-l3i5	47.7(0.6)	22.0(0.7)	6.3(0.4)	2.6(0.2)	0.180(0.005)	0.40(0.00)	82.7(0.6)	99.3(0.2)	70.6(0.3)
	MatterSim-v1-5M	52.6(0.5)	23.7(0.3)	7.9(0.3)	**2.9**(0.2)	0.158(0.004)	0.36(0.01)	**85.4**(0.7)	99.4(0.2)	68.0(0.7)

by a hyperparameter λ_{REPA} that controls the alignment strength.

$$\mathcal{L} = \mathcal{L}_{\text{gen}} + \lambda_{\text{REPA}} \mathcal{L}_{\text{REPA}}. \tag{9}$$

Computational cost. CrystalREPA introduces marginal training overhead and no extra inference cost. In practice, teacher representations of clean crystals are precomputed offline, so extra training overhead comes only from the lightweight projection head and the alignment loss. Tab. 16 shows that CrystalREPA increases the training time by only 1.0% on average. At inference time, both the teacher encoder and the projection head are discarded, preserving the original sampling procedure.

5 Experiments

We evaluate CrystalREPA by addressing the following research questions: (1) Does CrystalREPA narrow the representation gap between generative models and universal MLIPs? (Fig. 1a); (2) Does CrystalREPA consistently improve crystal generation quality across different generative frameworks, teacher MLIPs, and benchmark datasets? (Fig. 1, Tabs. 1 and 2); (3) Does benchmark accuracy (e.g., Matbench Discovery) predict an MLIP's value as a teacher, and if not, what property does? (Fig. 3); and (4) How do individual components of CrystalREPA contribute to its overall performance? (Fig. 4).

Table 2: Performance comparison of CrystalREPA on Alex-MP-20. We use DPA-3.1-3M as the MLIP teacher and evaluate CrystalREPA across three crystal generative frameworks. Values in parentheses denote standard errors over five independent sampling and evaluation runs.

Model	Target Rep.	Metastable%↑	MSUN%↑	Stable%↑	SUN%↑	$\bar{E}_{\mathrm{hull}}$ ↓	RMSD↓	Valid%↑	Unique%↑	Novel%↑
Vanilla CrystalFlow [4]		61.8(0.8)	31.5(0.6)	**4.8**(0.1)	1.6(0.3)	0.100(0.002)	0.36(0.00)	87.7(0.3)	**100.0**(0.0)	66.3(0.6)
+CrystalREPA	DPA-3.1-3M	**63.0**(0.2)	**33.7**(0.6)	**4.8**(0.1)	**1.9**(0.1)	**0.097**(0.001)	**0.31**(0.01)	**88.1**(0.4)	99.9(0.0)	**67.9**(0.6)
Vanilla MatterGen [3]		72.4(0.2)	38.1(0.7)	6.9(0.1)	2.9(0.3)	0.079(0.001)	**0.07**(0.00)	**85.8**(0.8)	**100.0**(0.0)	62.3(0.6)
+CrystalREPA	DPA-3.1-3M	**73.0**(0.7)	**39.3**(0.5)	**7.2**(0.1)	**3.6**(0.1)	**0.075**(0.000)	**0.07**(0.00)	**85.8**(0.5)	99.9(0.0)	**63.1**(0.5)
Vanilla DiffCSP [5]		63.1(1.0)	35.0(0.5)	5.4(0.5)	1.7(0.2)	0.100(0.002)	**0.25**(0.01)	87.6(0.3)	**99.8**(0.1)	69.5(0.7)
+CrystalREPA	DPA-3.1-3M	**63.4**(0.5)	**36.3**(0.6)	**6.0**(0.3)	**2.2**(0.2)	**0.097**(0.002)	0.26(0.01)	**87.9**(0.3)	99.6(0.1)	**70.7**(0.5)

5.1 CrystalREPA narrows the representation gap

To evaluate how predictive generative representations are of formation energy, we perform a linear probing task on their learned representations. Specifically, we extract the atom-wise hidden states from the encoders of various models using clean structures ($t = 1$) from the MP-20 dataset. These hidden states are aggregated into global structural representations via mean pooling, which then serve as inputs for a ridge regression model to predict the formation energy. As an MLIP reference, we apply the same probing protocol to DPA-3.1-3M. See Appendix B.4 for details.

Fig. 1(a) shows that the representations learned by vanilla CrystalFlow exhibit a significant representation gap compared to the pretrained MLIP DPA-3.1-3M, characterized by substantially higher Mean Absolute Errors (MAE) across all encoder layers. CrystalREPA significantly narrows this gap, with the probing error in the intermediate layers approaching that of the universal MLIP, suggesting that representation alignment makes the generative representations more predictive of stability-related quantities. This improvement is also reflected qualitatively in Fig. 1(b). In the examples shown, starting from the same noisy structures, vanilla CrystalFlow can generate structurally or compositionally invalid crystals, whereas CrystalREPA produces novel and metastable structures. Together with broader probing results in Appendix C.1, these results suggest that CrystalREPA narrows the representation gap and that the improved representation quality is associated with more stable and valid generated structures. Across three generative frameworks and their CrystalREPA variants, formation-energy probing MAE is also negatively correlated with downstream Metastable% (Pearson $r = -0.62$), as detailed in Appendix C.4.

5.2 CrystalREPA consistently improves crystal generation quality

We next evaluate whether CrystalREPA improves crystal generation quality. We evaluate CrystalREPA on MP-20 and Alex-MP-20, using the MatterGen evaluation pipeline [3]. For each method, we sample 1,024 structures, relax them with MatterSim-v1-1M [39], and report thermodynamic stability, structural fidelity, validity, and diversity metrics. Metastable% and Stable% denote the fractions of generated structures with $E_{\mathrm{hull}} \leq 0.1$ eV/atom and $E_{\mathrm{hull}} \leq 0$ eV/atom, respectively; MSUN% and SUN% further require the samples to be unique and novel. RMSD measures structural fidelity, with lower values indicating smaller structural distortion after relaxation. Detailed metric definitions are provided in Appendix B.7.

Tab. 1 shows that CrystalREPA consistently improves generation quality on MP-20 across CrystalFlow, MatterGen, and DiffCSP. Averaged over 10 MLIP teachers, CrystalREPA improves Metastable% from 41.2 to 50.5 for CrystalFlow, from 57.4 to 62.7 for MatterGen, and from 43.2 to 50.9 for DiffCSP. MSUN% also increases from 17.2 to 19.1, 25.3 to 27.7, and 18.8 to 23.4, respectively. These gains are accompanied by higher Stable% and SUN%, lower average energy above hull ($\bar{E}_{\mathrm{hull}}$), and generally improved validity, indicating that the gains are not limited to a single stability metric. The individual teacher rows further show that the gains are robust across MLIP families, including DPA, MACE, ORB, SevenNet, and MatterSim. The strongest MP-20 result is achieved by MatterGen + CrystalREPA with ORB v3, reaching 66.2 Metastable%, 29.2 MSUN%, and the lowest average energy above hull of 0.119 eV/atom.

Tab. 2 shows the same trend on the larger Alex-MP-20 dataset. The gains are smaller than on MP-20, which is expected because the Alex-MP-20 baselines are already stronger, but the direction of improvement remains consistent across all three generators. With DPA-3.1-3M as the MLIP teacher,

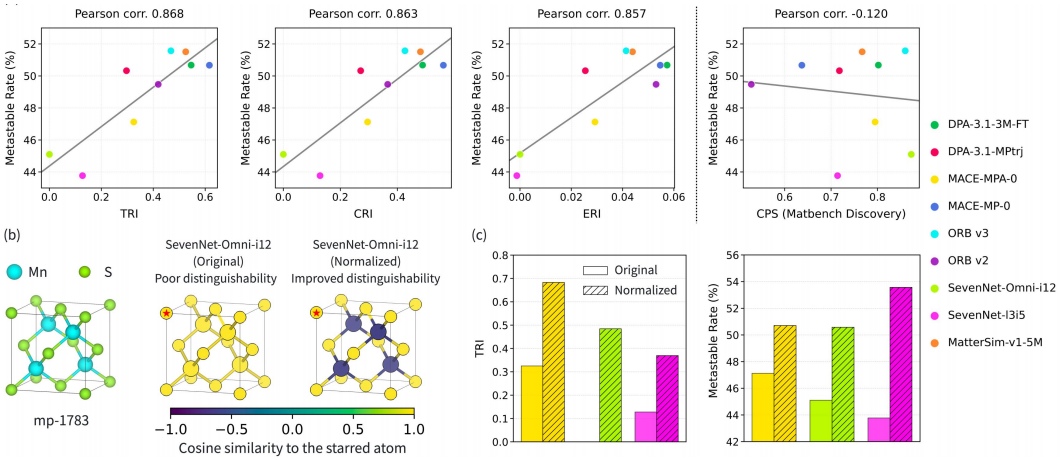


Figure 3: Representation distinguishability is associated with differences in MLIP transfer effectiveness. (a) Metastable rate correlates strongly with Chemical Resolution Index (CRI), Environment Resolution Index (ERI), and Total Resolution Index (TRI), but shows little association with the Matbench Discovery Combined Performance Score (CPS). (b) Similarity visualization shows improved atom-wise distinguishability after representation normalization. (c) Representation normalization increases distinguishability and is accompanied by consistent improvements in metastable rates. The three teacher MLIPs from left to right are MACE-MPA-0, SevenNet-Omni-i12, and SevenNet-l3i5.

CrystalREPA improves MSUN% from 31.5 to 33.7 for CrystalFlow, from 38.1 to 39.3 for MatterGen, and from 35.0 to 36.3 for DiffCSP. These results indicate that CrystalREPA remains beneficial across datasets and base generators.

Finally, Unique% remains nearly saturated around 99% across most settings. Although raw Novel% decreases in some MP-20 cases, both MSUN% and SUN%, which jointly require novelty, uniqueness, and stability, still improve. This suggests that CrystalREPA improves the yield of novel, unique, and stable structures despite a modest decrease in raw novelty in some settings.

We further assess generated crystals through DFT-based evaluation, finite-temperature molecular dynamics simulations, predicted synthesizability, and validity-aware MSUN. Across these complementary evaluations, CrystalREPA consistently improves the key metrics. Notably, on CrystalFlow, CrystalREPA improves DFT-evaluated Metastable% from 50.0 to 64.0 and MSUN% from 15.0 to 22.5. Further details are provided in Appendix C.2.

5.3 Distinguishability, not benchmark accuracy, predicts transfer effectiveness

While CrystalREPA improves over the vanilla baseline across all teachers, the magnitude of improvement varies considerably. For example, SevenNet-Omni-i12 achieves the highest Combined Performance Score (CPS) on Matbench Discovery [27] among the MLIPs considered, but yields relatively small gains in generation metrics in Tab. 1. This mismatch suggests that benchmark performance alone does not predict an MLIP's transfer effectiveness, and the answer instead lies in the structure of its atom-wise representation space.

We therefore examine transfer effectiveness from a representation perspective. Inspired by recent work showing that pairwise similarity between patch tokens among teacher representations affects transfer effectiveness in image generation [46], we quantify the distinguishability of atom-wise teacher representations using three metrics: Chemical Resolution Index (CRI), Environment Resolution Index (ERI), and Total Resolution Index (TRI). Intuitively, CRI measures distinguishability across elemental species, ERI measures distinguishability across local geometric environments, and TRI summarizes both aspects. Higher values indicate a more distinguishable atom-wise representation space. Precise definitions are provided in Appendix B.5. To avoid confounding teacher distinguishability with the element-aware negative sampling in EA-NCE, we conduct this diagnostic analysis using a cosine alignment objective rather than EA-NCE.

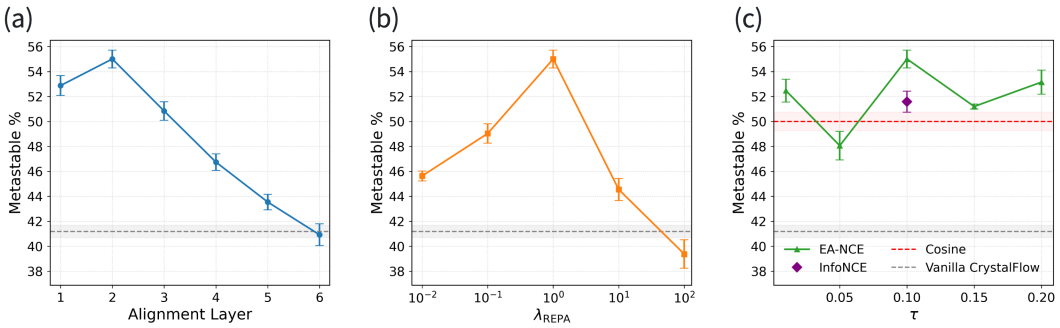


Figure 4: Ablation studies of CrystalREPA using CrystalFlow as the base generative model and DPA-3.1-3M as the MLIP teacher. We study the effects of (a) the alignment layer in the generative model, (b) the alignment weight λ_{REPA}, and (c) the choice of alignment objective.

Fig. 3(a) shows that representation distinguishability strongly correlates with generative performance across the unnormalized representations of the nine MLIP teachers with available Matbench Discovery CPS scores. The metastable rate has high positive Pearson correlations with CRI, ERI, and TRI, reaching 0.863, 0.857, and 0.868, respectively. In contrast, CPS from Matbench Discovery shows little positive association, with a correlation of -0.120. This helps explain why a teacher MLIP can perform well on conventional predictive benchmarks yet provide limited benefit for CrystalREPA if its atom-wise representation space does not clearly resolve chemical identity and local environment. We further introduce the Energy Resolution Index (EnRI) to quantify the distinguishability of MLIP representations across thermodynamic stability levels. EnRI also exhibits a strong correlation with downstream Metastable% (Pearson $r = 0.936$), as detailed in Appendix C.4.

Fig. 3(b, c) further suggests that improved teacher representation distinguishability is associated with better transfer effectiveness. We apply a simple representation normalization, defined in Appendix B.6, to frozen teacher representations of MACE-MPA-0, SevenNet-Omni-i12 and SevenNet-l3i5 before alignment. As shown in Fig. 3(b), for SevenNet-Omni-i12, normalization induces sharper contrast between sulfur atoms and manganese atoms, illustrating that the transformed representation space makes chemically distinct atoms more distinguishable. Quantitatively, as illustrated in Fig. 3(c), normalization increases TRI for all three teachers with relatively low original distinguishability and is accompanied by consistent improvements in metastable rate. For example, when using SevenNet-Omni-i12 as the MLIP teacher, TRI increases from 0.00 to 0.48, and the metastable rate improves from 45.1 to 50.6. Beyond increasing distinguishability, normalization also changes the scale and geometry of the alignment targets and may affect optimization, so improvements in distinguishability may not translate proportionally into downstream performance gains.

Overall, Fig. 3 suggests that, in our experiments, the transfer effectiveness of a teacher MLIP is better predicted by the distinguishability of its atom-wise representations than by benchmark performance, providing a practical criterion for selecting or preprocessing future teachers for crystal generation.

5.4 Ablation studies

We further conduct ablation studies to validate the design of CrystalREPA and examine its robustness to hyperparameter choices. All results in Fig. 4 use CrystalFlow as the base generative model and DPA-3.1-3M as the MLIP teacher. Overall, CrystalREPA improves over the vanilla model across a broad range of configurations, indicating that its gains do not rely on delicate hyperparameter tuning.

Alignment Layer. As shown in Fig. 4(a), aligning most encoder layers of CrystalFlow improves the metastable rate over the vanilla baseline, with the best performance achieved when aligning the hidden states after the second message-passing layer. Performance gradually declines for deeper alignment layers, suggesting that early intermediate representations provide the most effective transfer.

Effect of λ_{REPA}. Fig. 4(b) shows that CrystalREPA remains effective for λ_{REPA} from 0.01 to 10, achieves the best performance at $\lambda_{\text{REPA}} = 1$, and degrades when the alignment weight becomes excessively large. This indicates that CrystalREPA is robust to moderate choices of the loss weight.

Alignment objective. Fig. 4(c) compares different alignment objectives. The proposed EA-NCE consistently outperforms vanilla CrystalFlow across different temperatures (τ) from 0.01 to 0.2, with the best result at $\tau = 0.1$. It also performs better than standard InfoNCE [45] and the negative cosine similarity objective, validating the element-aware masking design for avoiding false negatives between same-element atoms.

Compared to conditional generation. Tab. 13 shows that energy-conditioned generation improves some stability-oriented metrics, such as $\bar{E}_{\mathrm{hull}}$, Stable%, and SUN%, but consistently reduces Valid% and therefore harms the physical plausibility of generated structures. In contrast, CrystalREPA yields more balanced improvements in Metastable%, MSUN%, RMSD, and Valid%, leading to stronger overall generation quality. Detailed results are deferred to Appendix C.5.

Effect of data overlap. We also analyze the potential effect of overlap between the MLIP pretraining data and the generator training data. Appendix C.7 shows that CrystalREPA remains beneficial with a non-overlap teacher and achieves comparable gains with a high-overlap teacher, suggesting that dataset overlap alone is unlikely to be the primary driver of the observed improvements.

Additional ablations. We further investigate SOAP-based alignment, energy-direction removal, MLIP teacher-layer selection, and direct MLIP-only optimization. These additional experiments provide complementary evidence for the contribution of energy-related MLIP representations and the importance of structural priors learned by generative models. Detailed results are provided in Appendix C.3.

6 Conclusion

In this work, we identified a clear representation gap between crystal generative models and pretrained universal MLIPs, and proposed CrystalREPA, a framework that transfers stability-aware priors to crystal generative models via atom-wise representation alignment. Across three generative frameworks, ten MLIP teachers, and two benchmark datasets, CrystalREPA consistently improves thermodynamic stability, structural validity, and structural fidelity, while preserving strong diversity, introducing only marginal training overhead, and incurring no additional inference cost. Beyond these empirical gains, we find that transfer effectiveness is predicted by the distinguishability of teacher atomistic representations rather than by benchmark accuracy, yielding a practical criterion for selecting MLIP teachers. We hope this work provides a useful step toward bridging generative modeling and atomistic representation learning, and motivates future generative models that produce structures of higher thermodynamic stability for materials discovery.

Limitation. First, the current framework aligns only atom-level representations and does not consider edge-level or higher-order structural representations. Second, our experiments are limited to crystalline materials, and we do not test whether the same alignment strategy transfers to other atomistic domains such as molecules or proteins. We leave these directions to future work.

Acknowledgments and Disclosure of Funding

We acknowledge the AI for Science Institute for providing the computational resources used in this work. The work of H.W. is supported by the National Natural Science Foundation of China (Grants No. 12525113 and No. 12561160120). Tiejun Li acknowledges the support from National Key R&D Program of China under grant 2021YFA1003301, and National Science Foundation of China under grant 12288101.

Competing interests. The authors declare no competing interests.

References

[1] Mouyang Cheng, Chu-Liang Fu, Ryotaro Okabe, Abhijatmedhi Chotrattanapituk, Artittaya Boonkird, Nguyen Tuan Hung, and Mingda Li. Artificial intelligence-driven approaches for materials design and discovery. *Nature Materials*, pages 1–17, 2026.

[2] Houssam Metni, Laura Ruple, Lauren N Walters, Luca Torresi, Jonas Teufel, Henrik Schopmans, Jona Östreicher, Yumeng Zhang, Marlen Neubert, Yuri Koide, et al. Generative models for crystalline materials. *Advanced Materials*, page e23620, 2026.

[3] Claudio Zeni, Robert Pinsler, Daniel Zügner, Andrew Fowler, Matthew Horton, Xiang Fu, Zilong Wang, Aliaksandra Shysheya, Jonathan Crabbé, Shoko Ueda, et al. A generative model for inorganic materials design. *Nature*, 639(8055):624–632, 2025.

[4] Xiaoshan Luo, Zhenyu Wang, Qingchang Wang, Xuechen Shao, Jian Lv, Lei Wang, Yanchao Wang, and Yanming Ma. Crystalflow: a flow-based generative model for crystalline materials. *Nature Communications*, 16(1):9267, 2025.

[5] Rui Jiao, Wenbing Huang, Peijia Lin, Jiaqi Han, Pin Chen, Yutong Lu, and Yang Liu. Crystal structure prediction by joint equivariant diffusion. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

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

[7] Eric C-Y Yuan, Yunsheng Liu, Junmin Chen, Peichen Zhong, Sanjeev Raja, Tobias Kreiman, Santiago Vargas, Wenbin Xu, Martin Head-Gordon, Chao Yang, et al. Foundation models for atomistic simulation of chemistry and materials. *Nature Reviews Chemistry*, pages 1–19, 2026.

[8] Sihyun Yu, Sangkyung Kwak, Huiwon Jang, Jongheon Jeong, Jonathan Huang, Jinwoo Shin, and Saining Xie. Representation alignment for generation: Training diffusion transformers is easier than you think. In *The Thirteenth International Conference on Learning Representations*, 2025.

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

[10] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

[11] Youzhi Luo, Chengkai Liu, and Shuiwang Ji. Towards symmetry-aware generation of periodic materials. *Advances in Neural Information Processing Systems*, 36:53308–53329, 2023.

[12] François R J Cornet, Federico Bergamin, Arghya Bhowmik, Juan Maria Garcia-Lastra, Jes Frellsen, and Mikkel N. Schmidt. Kinetic langevin diffusion for crystalline materials generation. In *Forty-second International Conference on Machine Learning*, 2025.

[13] Peijia Lin, Pin Chen, Rui Jiao, Qing Mo, Cen Jianhuan, Wenbing Huang, Yang Liu, Dan Huang, and Yutong Lu. Equivariant diffusion for crystal structure prediction. In *Proceedings of the 41st International Conference on Machine Learning*, pages 29890–29913, 2024.

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

[15] Daniel Levy, Siba Smarak Panigrahi, Sékou-Oumar Kaba, Qiang Zhu, Kin Long Kelvin Lee, Mikhail Galkin, Santiago Miret, and Siamak Ravanbakhsh. SymmCD: Symmetry-preserving crystal generation with diffusion models. In *The Thirteenth International Conference on Learning Representations*, 2025.

[16] Filip Ekström Kelvinius, Oskar B. Andersson, Abhijith S Parackal, Dong Qian, Rickard Armiento, and Fredrik Lindsten. Wyckoffdiff – a generative diffusion model for crystal symmetry. In *Forty-second International Conference on Machine Learning*, 2025.

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

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

[19] Michael Samuel Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations*, 2023.

[20] Xingchao Liu, Chengyue Gong, and qiang liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023.

[21] Benjamin Kurt Miller, Ricky T. Q. Chen, Anuroop Sriram, and Brandon M Wood. FlowMM: Generating materials with riemannian flow matching. In *Forty-first International Conference on Machine Learning*, 2024.

[22] Anuroop Sriram, Benjamin Kurt Miller, Ricky T. Q. Chen, and Brandon M Wood. 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.

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

[24] Jonathan Schmidt, Noah Hoffmann, Hai-Chen Wang, Pedro Borlido, Pedro JMA Carriço, Tiago FT Cerqueira, Silvana Botti, and Miguel AL Marques. Machine-learning-assisted determination of the global zero-temperature phase diagram of materials. *Advanced Materials*, 35(22):2210788, 2023.

[25] Jonathan Schmidt, Tiago FT Cerqueira, Aldo H Romero, Antoine Loew, Fabian Jäger, Hai-Chen Wang, Silvana Botti, and Miguel AL Marques. Improving machine-learning models in materials science through large datasets. *Materials Today Physics*, 48:101560, 2024.

[26] Luis Barroso-Luque, Muhammed Shuaibi, Xiang Fu, Brandon M Wood, Misko Dzamba, Meng Gao, Ammar Rizvi, C Lawrence Zitnick, and Zachary W Ulissi. Open materials 2024 (omat24) inorganic materials dataset and models. *arXiv preprint arXiv:2410.12771*, 2024.

[27] Janosh Riebesell, Rhys EA Goodall, Philipp Benner, Yuan Chiang, Bowen Deng, Gerbrand Ceder, Mark Asta, Alpha A Lee, Anubhav Jain, and Kristin A Persson. A framework to evaluate machine learning crystal stability predictions. *nature machine intelligence*, 7(6):836–847, 2025.

[28] Arslan Mazitov, Filippo Bigi, Matthias Kellner, Paolo Pegolo, Davide Tisi, Guillaume Fraux, Sergey Pozdnyakov, Philip Loche, and Michele Ceriotti. Pet-mad as a lightweight universal interatomic potential for advanced materials modeling. *Nature Communications*, 16(1):10653, 2025.

[29] Xiang Fu, Brandon M Wood, Luis Barroso-Luque, Daniel S. Levine, Meng Gao, Misko Dzamba, and C. Lawrence Zitnick. Learning smooth and expressive interatomic potentials for physical property prediction. In *Forty-second International Conference on Machine Learning*, 2025.

[30] Seung Yul Lee, Hojoon Kim, Yutack Park, Dawoon Jeong, Seungwu Han, Yeonhong Park, and Jae W. Lee. FlashTP: Fused, sparsity-aware tensor product for machine learning interatomic potentials. In *Forty-second International Conference on Machine Learning*, 2025.

[31] Yuanchang Zhou, Siyu Hu, Xiangyu Zhang, Hongyu Wang, Guangming Tan, and Weile Jia. MatRIS: Toward reliable and efficient pretrained machine learning interaction potentials. In *The Fourteenth International Conference on Learning Representations*, 2026.

[32] Jaesun Kim, Jinmu You, Yutack Park, Yunsung Lim, Yujin Kang, Jisu Kim, Haekwan Jeon, Suyeon Ju, Deokgi Hong, Seung Yul Lee, et al. Optimizing cross-domain transfer for universal machine learning interatomic potentials. *Nature Communications*, 2026.

[33] Chuin Wei Tan, Marc L Descoteaux, Mit Kotak, Gabriel de Miranda Nascimento, Seán R Kavanagh, Laura Zichi, Menghang Wang, Aadit Saluja, Yizhong R Hu, Tess Smidt, et al. High-performance training and inference for deep equivariant interatomic potentials. *Digital Discovery*, 2026.

[34] Yury Lysogorskiy, Anton Bochkarev, and Ralf Drautz. Graph atomic cluster expansion for foundational machine learning interatomic potentials. *npj Computational Materials*, 2026.

[35] Benjamin Rhodes, Sander Vandenhaute, Vaidotas Šimkus, James Gin, Jonathan Godwin, Tim Duignan, and Mark Neumann. Orb-v3: atomistic simulation at scale, 2025. URL https://arxiv.org/abs/2504.06231.

[36] Anton Bochkarev, Yury Lysogorskiy, and Ralf Drautz. Graph atomic cluster expansion for semilocal interactions beyond equivariant message passing. *Physical Review X*, 14(2):021036, 2024.

[37] Duo Zhang, Anyang Peng, Chun Cai, Wentao Li, Yuanchang Zhou, Jinzhe Zeng, Mingyu Guo, Chengqian Zhang, Bowen Li, Hong Jiang, Tong Zhu, Weile Jia, Linfeng Zhang, and Han Wang. A graph neural network for the era of large atomistic models. *arXiv preprint arXiv:2506.01686*, 2025.

[38] Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M Elena, Dávid P Kovács, Janosh Riebesell, Xavier R Advincula, Mark Asta, Matthew Avaylon, William J Baldwin, et al. A foundation model for atomistic materials chemistry. *The Journal of chemical physics*, 163(18), 2025.

[39] Han Yang, Chenxi Hu, Yichi Zhou, Xixian Liu, Yu Shi, Jielan Li, Guanzhi Li, Zekun Chen, Shuizhou Chen, Claudio Zeni, et al. Mattersim: A deep learning atomistic model across elements, temperatures and pressures. *arXiv preprint arXiv:2405.04967*, 2024.

[40] Yutack Park, Jaesun Kim, Seungwoo Hwang, and Seungwu Han. Scalable parallel algorithm for graph neural network interatomic potentials in molecular dynamics simulations. *Journal of chemical theory and computation*, 20(11):4857–4868, 2024.

[41] Mark Neumann, James Gin, Benjamin Rhodes, Steven Bennett, Zhiyi Li, Hitarth Choubisa, Arthur Hussey, and Jonathan Godwin. Orb: A fast, scalable neural network potential, 2024. URL https://arxiv.org/abs/2410.22570.

[42] Duo Zhang, Xinzijian Liu, Xiangyu Zhang, Chengqian Zhang, Chun Cai, Hangrui Bi, Yiming Du, Xuejian Qin, Anyang Peng, Jiameng Huang, et al. Dpa-2: a large atomic model as a multi-task learner. *npj Computational Materials*, 10(1):293, 2024.

[43] Brandon M Wood, Misko Dzamba, Xiang Fu, Meng Gao, Muhammed Shuaibi, Luis Barroso-Luque, Kareem Abdelmaqsoud, Vahe Gharakhanyan, John R. Kitchin, Daniel S. Levine, Kyle Michel, Anuroop Sriram, Taco Cohen, Abhishek Das, Sushree Jagriti Sahoo, Ammar Rizvi, Zachary Ward Ulissi, and C. Lawrence Zitnick. UMA: A family of universal models for atoms. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[44] Chengqian Zhang, Duo Zhang, Anyang Peng, Mingyu Guo, Yuzhi Zhang, Lei Wang, Guolin Ke, Linfeng Zhang, Tiejun Li, and Han Wang. Multi-task fine-tuning enables robust out-of-distribution generalization in atomistic models. *arXiv preprint arXiv:2601.08486*, 2026.

[45] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

[46] Jaskirat Singh, Xingjian Leng, Zongze Wu, Liang Zheng, Richard Zhang, Eli Shechtman, and Saining Xie. What matters for representation alignment: Global information or spatial structure? In *The Fourteenth International Conference on Learning Representations*, 2026.

[47] Valentin De Bortoli, Emile Mathieu, Michael John Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modelling. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022.

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

[49] Alec Belsky, Mariette Hellenbrandt, Vicky Lynn Karen, and Peter Luksch. New developments in the inorganic crystal structure database (icsd): accessibility in support of materials research and design. *Structural Science*, 58(3):364–369, 2002.

[50] The openlam-v1 dataset, 2025. https://www.aissquare.com/datasets/detail?name=OpenLAM-TrainingSet-v1&id=308&pageType=datasets.

[51] Albert P Bartók, Risi Kondor, and Gábor Csányi. On representing chemical environments. *Physical Review B—Condensed Matter and Materials Physics*, 87(18):184115, 2013.

[52] James P Darby, James R Kermode, and Gábor Csányi. Compressing local atomic neighbourhood descriptors. *npj Computational Materials*, 8(1):166, 2022.

[53] Lauri Himanen, Marc O. J. Jäger, Eiaki V. Morooka, Filippo Federici Canova, Yashasvi S. Ranawat, David Z. Gao, Patrick Rinke, and Adam S. Foster. DScribe: Library of descriptors for machine learning in materials science. *Computer Physics Communications*, 247:106949, 2020. ISSN 0010-4655. doi: 10.1016/j.cpc.2019.106949.

[54] Jarno Laakso, Lauri Himanen, Henrietta Homm, Eiaki V Morooka, Marc OJ Jäger, Milica Todorović, and Patrick Rinke. Updates to the dscribe library: New descriptors and derivatives. *The Journal of Chemical Physics*, 158(23), 2023.

[55] Daniel W Davies, Keith T Butler, Adam J Jackson, Jonathan M Skelton, Kazuki Morita, and Aron Walsh. Smact: Semiconducting materials by analogy and chemical theory. *Journal of Open Source Software*, 4(38):1361, 2019.

[56] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

[57] Min Zhao, Fan Bao, Chongxuan Li, and Jun Zhu. EGSDE: Unpaired image-to-image translation via energy-guided stochastic differential equations. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022.

[58] Fan Bao, Min Zhao, Zhongkai Hao, Peiyao Li, Chongxuan Li, and Jun Zhu. Equivariant energy-guided SDE for inverse molecular design. In *The Eleventh International Conference on Learning Representations*, 2023.

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

[60] Ask Hjorth Larsen, Jens Jørgen Mortensen, Jakob Blomqvist, Ivano E Castelli, Rune Christensen, Marcin Dułak, Jesper Friis, Michael N Groves, Bjørk Hammer, Cory Hargus, et al. The atomic simulation environment—a python library for working with atoms. *Journal of Physics: Condensed Matter*, 29(27):273002, 2017.

[61] Jidon Jang, Juhwan Noh, Lan Zhou, Geun Ho Gu, John M Gregoire, and Yousung Jung. Synthesizability of materials stoichiometry using semi-supervised learning. *Matter*, 7(6): 2294–2312, 2024.

Contents

1 Introduction **1**

2 Related Work **3**

3 Preliminaries **3**

3.1 Crystal generative modeling 3

3.2 Universal MLIPs 4

4 CrystalREPA **4**

5 Experiments **6**

5.1 CrystalREPA narrows the representation gap 7

5.2 CrystalREPA consistently improves crystal generation quality 7

5.3 Distinguishability, not benchmark accuracy, predicts transfer effectiveness 8

5.4 Ablation studies 9

6 Conclusion **10**

A Formulation of CrystalREPA **16**

A.1 CrystalREPA in flow matching frameworks 16

A.2 CrystalREPA in diffusion frameworks 17

B Experiment Details **18**

B.1 Training datasets 18

B.2 Universal MLIPs 18

B.3 Hyperparameters 19

B.4 Formation energy probing 19

B.5 Definition of CRI, ERI and TRI 20

B.6 Representation normalization 21

B.7 Evaluation metrics 21

B.8 Conditional generation 21

C Additional Results **23**

C.1 Additional representation probing results 23

C.2 Additional validation of generated crystals 24

C.3 Additional ablation studies 25

C.4 Additional representation analysis 27

C.5 Comparison to conditional generation 28

C.6 Effect of MLIP teacher input 29

C.7 Effect of data overlap 29

C.8 Computational cost 30

D Broader Impact **30**

A Formulation of CrystalREPA

A.1 CrystalREPA in flow matching frameworks

Flow matching training. We use CrystalFlow as a representative flow-based crystal generative framework. We denote a crystal with N atoms in the unit cell as $\mathcal{M} = (\mathbf{A}, \mathbf{F}, \mathbf{L})$, where $\mathbf{A} = [\mathbf{a}_1^\top; \mathbf{a}_2^\top; \ldots; \mathbf{a}_N^\top] \in \mathbb{R}^{N\times a}$ denotes atom types, with each atom type mapped to an a-dimensional categorical representation. $\mathbf{F} = [\mathbf{f}_1^\top; \mathbf{f}_2^\top; \ldots; \mathbf{f}_N^\top] \in [0,1)^{N\times 3}$ denotes fractional coordinates, and $\mathbf{L} \in \mathbb{R}^6$ denotes a rotation-invariant lattice parameterization derived from the lattice matrix $\tilde{\mathbf{L}} = [\mathbf{l}_1^\top; \mathbf{l}_2^\top; \mathbf{l}_3^\top] \in \mathbb{R}^{3\times 3}$ via polar decomposition as $\tilde{\mathbf{L}}^\top = \mathbf{Q}\exp(\sum_{i=1}^6 \mathbf{L}_i \mathbf{B}_i)$. Here, $\mathbf{Q}$ is an orthogonal matrix representing rotational degrees of freedom, $\exp(\cdot)$ denotes the matrix exponential, and $\{\mathbf{B}_i\}_{i=1}^6 \in \mathbb{R}^{3\times 3}$ forms a standard basis of symmetric matrices. In CrystalFlow, crystal generation is formulated as learning a continuous transport map from a simple prior distribution to the data distribution.

Specifically, let $\mathcal{M}_1 = (\mathbf{A}_1, \mathbf{F}_1, \mathbf{L}_1)$ be a crystal sampled from the data distribution and let $\mathcal{M}_0 = (\mathbf{A}_0, \mathbf{F}_0, \mathbf{L}_0)$ be sampled from the corresponding priors. A continuous path $\mathcal{M}_t = (\mathbf{A}_t, \mathbf{F}_t, \mathbf{L}_t)$, with $t \in [0,1]$, interpolates between the prior and data, where $t = 0$ corresponds to the prior and $t = 1$ corresponds to the clean crystal. The interpolations for atom types and lattice variables are defined by linear interpolation between the prior sample and the data sample:

$$\mathbf{A}_t = t\mathbf{A}_1 + (1-t)\mathbf{A}_0, \quad \mathbf{A}_0 \sim \mathcal{N}(0, \mathbf{I}) \in \mathbb{R}^{N\times a}, \tag{10}$$

$$\mathbf{L}_t = t\mathbf{L}_1 + (1-t)\mathbf{L}_0, \quad \mathbf{L}_0 \sim \mathcal{N}\left((0,0,0,0,0,1), 0.1^2\mathbf{I}\right) \in \mathbb{R}^6. \tag{11}$$

Fractional coordinates lie on a three-dimensional torus and must respect periodic translation invariance. CrystalFlow addresses this by using a uniform prior and determining $\mathbf{F}_t$ with the minimum image convention:

$$\mathbf{F}_t = \mathbf{F}_0 + t(w(\mathbf{F}_1 - \mathbf{F}_0 - 0.5) - 0.5), \quad \mathbf{F}_0 \sim \mathcal{U}([0,1)^{N\times 3}), \tag{12}$$

where $w(\mathbf{x}) = \mathbf{x} - \lfloor \mathbf{x} \rfloor$ is the wrapping operation.

Given this probability path, the target velocity fields for atom types, fractional coordinates, and lattice variables are

$$\mathbf{u}_t^A = \mathbf{A}_1 - \mathbf{A}_0, \qquad \mathbf{u}_t^F = w(\mathbf{F}_1 - \mathbf{F}_0 - 0.5) - 0.5, \qquad \mathbf{u}_t^L = \mathbf{L}_1 - \mathbf{L}_0. \tag{13}$$

CrystalFlow learns three time-dependent vector fields, $\mathbf{v}_\theta^A(\mathcal{M}_t, t)$, $\mathbf{v}_\theta^F(\mathcal{M}_t, t)$, and $\mathbf{v}_\theta^L(\mathcal{M}_t, t)$, to regress these targets. These vector fields are parameterized by a periodic E(3)-equivariant graph neural network with K message-passing layers. Let $f^{(k)}(\mathcal{M}_t, t) \in \mathbb{R}^{N\times d_1}$ denote atom-wise hidden states after the k-th message-passing layer. The original CrystalFlow training objective is the weighted regression loss defined in Eqs. (1)–(2).

Applying CrystalREPA to flow matching models. CrystalREPA is applied to CrystalFlow by aligning the atom-wise hidden states along the flow path with MLIP representations of the corresponding clean crystal. For a corrupted structure $\mathcal{M}_t$, we extract the student representation $\mathbf{h}_t^{(k)} = f^{(k)}(\mathcal{M}_t, t) \in \mathbb{R}^{N\times d_1}$ from the k-th layer of the generative encoder. The clean crystal $\mathcal{M}_1$ is fed into a frozen pretrained MLIP encoder $\tilde{f}$ to obtain the teacher representation $\tilde{\mathbf{z}} = \tilde{f}(\mathcal{M}_1) \in \mathbb{R}^{N\times d_2}$. A lightweight projection head φ_p maps the student representation to the teacher representation space: $\mathbf{z} = \varphi_p(\mathbf{h}_t^{(k)}) \in \mathbb{R}^{N\times d_2}$. We therefore apply the element-aware contrastive loss defined in Eqs. (6)–(8). The final training objective for CrystalREPA in the flow matching framework is in Eq. (9).

Inference. At inference time, CrystalREPA does not change the original CrystalFlow sampling procedure. A random initial state $\mathcal{M}_0 = (\mathbf{A}_0, \mathbf{F}_0, \mathbf{L}_0)$ is sampled from the prior distributions defined above and then evolved from $t = 0$ to $t = 1$ using the learned vector fields. We discretize the interval $[0,1]$ into 100 steps with $\Delta t = 1/100$ and $t_m = m\Delta t$. Using the Euler solver, the updates are given by

$$\mathbf{A}_{t_{m+1}} = \mathbf{A}_{t_m} + \mathbf{v}_\theta^A(\mathcal{M}_{t_m}, t_m)\Delta t, \tag{14}$$

$$\mathbf{F}_{t_{m+1}} = \mathbf{F}_{t_m} + (1 + 5t_m)\mathbf{v}_\theta^F(\mathcal{M}_{t_m}, t_m)\Delta t, \tag{15}$$

$$\mathbf{L}_{t_{m+1}} = \mathbf{L}_{t_m} + \mathbf{v}_\theta^L(\mathcal{M}_{t_m}, t_m)\Delta t, \tag{16}$$

for $m = 0, \ldots, 99$, where $(1 + 5t_m)$ is the anti-annealing scaling factor applied to fractional coordinate updates in CrystalFlow. The MLIP teacher and projection head are discarded during inference, so CrystalREPA introduces no additional inference cost.

A.2 CrystalREPA in diffusion frameworks

Diffusion training. We first formulate the forward corruption process and denoising objective for diffusion-based crystal generative frameworks, including MatterGen and DiffCSP. In this section, we provide a unified formulation for applying CrystalREPA to such models, while noting that the exact diffusion parameterization of each variable is model-specific. Different from the flow matching convention used in CrystalFlow, where $t = 0$ corresponds to the prior and $t = 1$ corresponds to the clean crystal, diffusion models usually define $t = 0$ as the clean structure and $t = T$ as pure noise.

The forward diffusion of the continuous lattice variable $\mathbf{L}$ follows Denoising Diffusion Probabilistic Model (DDPM) [9]. Given the variance schedule $\beta_1, \ldots, \beta_T$, the forward process can be expressed as:

$$q(\mathbf{L}_t \mid \mathbf{L}_0) = \mathcal{N}\big(\mathbf{L}_t \mid \sqrt{\bar{\alpha}_t}\mathbf{L}_0, (1-\bar{\alpha}_t)\mathbf{I}\big), \quad \mathbf{L}_t = \sqrt{\bar{\alpha}_t}\mathbf{L}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_L, \quad \boldsymbol{\epsilon}_L \sim \mathcal{N}(0, \mathbf{I}), \tag{17}$$

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$.

Fractional coordinates $\mathbf{F}$ are diffused using score matching [10] with a wrapped normal distribution [47]. The forward process uses a wrapped normal distribution to maintain periodic invariance:

$$\begin{aligned} q(\mathbf{F}_t \mid \mathbf{F}_0) = \mathcal{N}_w\big(\mathbf{F}_t \mid \mathbf{F}_0, \sigma_t^2\mathbf{I}\big) &\propto \sum_{\mathbf{Z}\in\mathbb{Z}^{N\times 3}} \exp\left(-\frac{\|\mathbf{F}_t - \mathbf{F}_0 + \mathbf{Z}\|^2}{2\sigma_t^2}\right), \\ \mathbf{F}_t = w(\mathbf{F}_0 + \sigma_t\boldsymbol{\epsilon}_F), \quad \boldsymbol{\epsilon}_F &\sim \mathcal{N}(0, \mathbf{I}), \end{aligned} \tag{18}$$

where the noise schedule σ_t obeys the exponential scheduler: $\sigma_0 = 0$ and $\sigma_t = \sigma_1(\frac{\sigma_T}{\sigma_1})^{\frac{t-1}{T-1}}$, if $t > 0$.

The exact treatment of atom types is model-specific in diffusion-based crystal generative models. For notation simplicity, we describe the diffusion of atom types using a form similar to the lattice diffusion:

$$q(\mathbf{A}_t \mid \mathbf{A}_0) = \mathcal{N}\big(\mathbf{A}_t \mid \sqrt{\bar{\alpha}_t}\mathbf{A}_0, (1-\bar{\alpha}_t)\mathbf{I}\big), \quad \mathbf{A}_t = \sqrt{\bar{\alpha}_t}\mathbf{A}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_A, \quad \boldsymbol{\epsilon}_A \sim \mathcal{N}(0, \mathbf{I}). \tag{19}$$

According to these forward processes, the corrupted structure $\mathcal{M}_t = (\mathbf{A}_t, \mathbf{F}_t, \mathbf{L}_t)$ can be obtained using the corresponding reparameterized forms shown above. The denoising model predicts $\hat{\boldsymbol{\epsilon}}_\theta^A$, $\hat{\boldsymbol{\epsilon}}_\theta^F$, and $\hat{\boldsymbol{\epsilon}}_\theta^L$ to regress $\boldsymbol{\epsilon}_A$, $\nabla_{\mathbf{F}_t} \log q\left(\mathbf{F}_t \mid \mathbf{F}_0\right)$, and $\boldsymbol{\epsilon}_L$. The denoising model is parameterized by a graph neural network with K message-passing layers. Let $f^{(K)}(\mathcal{M}_t, t) \in \mathbb{R}^{N\times d_1}$ denote the final atom-wise hidden states. Based on these hidden states, the model predicts the denoising targets for atom types, fractional coordinates, and lattice variables:

$$\hat{\boldsymbol{\epsilon}}_\theta^A = \varphi_A(f^{(K)}(\mathcal{M}_t, t)), \quad \hat{\boldsymbol{\epsilon}}_\theta^F = \varphi_F(f^{(K)}(\mathcal{M}_t, t)), \quad \hat{\boldsymbol{\epsilon}}_\theta^L = \varphi_L(\frac{1}{N}\sum_i^N f_i^{(K)}(\mathcal{M}_t, t)), \tag{20}$$

$$\mathcal{L}_A = \mathbb{E}_{t\sim\mathcal{U}(1,T)}\left[\left\|\boldsymbol{\epsilon}_A - \hat{\boldsymbol{\epsilon}}_\theta^A(\mathcal{M}_t, t)\right\|^2\right], \tag{21}$$

$$\mathcal{L}_F = \mathbb{E}_{t\sim\mathcal{U}(1,T)}[\lambda_t \left\|\nabla_{\mathbf{F}_t} \log q\left(\mathbf{F}_t \mid \mathbf{F}_0\right) - \hat{\boldsymbol{\epsilon}}_\theta^F(\mathcal{M}_t, t)\right\|^2], \tag{22}$$

$$\mathcal{L}_L = \mathbb{E}_{t\sim\mathcal{U}(1,T)}\left[\left\|\boldsymbol{\epsilon}_L - \hat{\boldsymbol{\epsilon}}_\theta^L(\mathcal{M}_t, t)\right\|^2\right], \tag{23}$$

where φ_A, φ_F, φ_L are prediction heads, and $\lambda_t = \mathbb{E}_{\mathbf{F}_t}^{-1}\left[\|\nabla_{\mathbf{F}_t} \log q(\mathbf{F}_t \mid \mathbf{F}_0)\|^2\right]$ is approximated via Monte-Carlo sampling.

The total generative loss is:

$$\mathcal{L}_{\text{diff}} = \lambda_{\text{A}}\mathcal{L}_{\text{A}} + \lambda_{\text{F}}\mathcal{L}_{\text{F}} + \lambda_{\text{L}}\mathcal{L}_{\text{L}}. \tag{24}$$

Applying CrystalREPA to diffusion models. CrystalREPA is applied to the atom-wise hidden states produced during this denoising process. Given a clean crystal $\mathcal{M}_0 = (\mathbf{A}_0, \mathbf{F}_0, \mathbf{L}_0)$, we first construct its corrupted state $\mathcal{M}_t = (\mathbf{A}_t, \mathbf{F}_t, \mathbf{L}_t)$ using the forward diffusion process in Eqs. (17)–(19). The denoising model takes $\mathcal{M}_t$ and the timestep t as input and predicts the corresponding denoising targets for atom types, fractional coordinates, and lattices. Let $\mathbf{h}_t^{(k)} = f^{(k)}(\mathcal{M}_t, t) \in \mathbb{R}^{N\times d_1}$ denote the atom-wise hidden states from the k-th layer of the diffusion model. The frozen MLIP encoder

provides teacher representations $\tilde{\mathbf{z}} = \tilde{f}(\mathcal{M}_0) \in \mathbb{R}^{N\times d_2}$ from the corresponding clean crystal. We then project the student representations into the teacher space: $\mathbf{z} = \varphi_p(\mathbf{h}_t^{(k)}) \in \mathbb{R}^{N\times d_2}$, where φ_p is an MLP. Finally, we compute the same element-aware contrastive alignment loss $\mathcal{L}_{\mathrm{REPA}}$ as defined in Eqs. (6)–(8).

The final training objective for diffusion frameworks is therefore

$$\mathcal{L} = \mathcal{L}_{\mathrm{diff}} + \lambda_{\mathrm{REPA}}\mathcal{L}_{\mathrm{REPA}}. \tag{25}$$

Inference. Starting from the pure noise state $\mathcal{M}_T = (\mathbf{A}_T, \mathbf{F}_T, \mathbf{L}_T)$, where $\mathbf{A}_T \sim \mathcal{N}(0, \mathbf{I})$, $\mathbf{F}_T \sim \mathcal{U}(0, 1)$, $\mathbf{L}_T \sim \mathcal{N}(0, \mathbf{I})$, the reverse updates for lattice variables and atom types are defined by:

$$\mathbf{L}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{L}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\hat{\boldsymbol{\epsilon}}_\theta^L(\mathcal{M}_t, t)\right) + \sqrt{\frac{\beta_t(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_L, \tag{26}$$

$$\mathbf{A}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{A}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\hat{\boldsymbol{\epsilon}}_\theta^A(\mathcal{M}_t, t)\right) + \sqrt{\frac{\beta_t(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_A. \tag{27}$$

Fractional coordinates are generated using a two-step predictor-corrector sampler method [10]:

$$\mathbf{F}_{t-\frac{1}{2}} = w\left(\mathbf{F}_t + \left(\sigma_t^2 - \sigma_{t-1}^2\right)\hat{\boldsymbol{\epsilon}}_\theta^F(\mathcal{M}_t, t) + \frac{\sigma_{t-1}\sqrt{\sigma_t^2 - \sigma_{t-1}^2}}{\sigma_t}\boldsymbol{\epsilon}_F\right), \tag{28}$$

$$\mathbf{F}_{t-1} = w\left(\mathbf{F}_{t-\frac{1}{2}} + d_t\hat{\boldsymbol{\epsilon}}_\theta^F(\mathbf{A}_{t-1}, \mathbf{F}_{t-\frac{1}{2}}, \mathbf{L}_{t-1}, t-1) + \sqrt{2d_t}\boldsymbol{\epsilon}'_F\right), \quad d_t = \frac{\gamma\sigma_{t-1}}{\sigma_1}, \tag{29}$$

where $\boldsymbol{\epsilon}_L$, $\boldsymbol{\epsilon}_A$, $\boldsymbol{\epsilon}_F$, $\boldsymbol{\epsilon}'_F \sim \mathcal{N}(0, \mathbf{I})$, γ is the step size of Langevin dynamics. As in the flow matching case, CrystalREPA does not modify the original diffusion sampling procedure. For MatterGen and DiffCSP, we follow the original sampling settings and use $T = 1000$ reverse diffusion steps during inference.

B Experiment Details

B.1 Training datasets

MP-20. MP-20, originally introduced by Xie *et al.* [6], is curated from the Materials Project [48] (Creative Commons Attribution 4.0 International License) and contains 45,231 materials with diverse crystal structures and chemical compositions, spanning 89 elements and unit cells containing 1 to 20 atoms. To ensure experimental fidelity, the dataset is restricted to materials originating from the Inorganic Crystal Structure Database (ICSD) [49], which covers the majority of experimentally reported crystals within this size range. It further retains only materials with energy above hull below 0.08 eV/atom and formation energy below 2 eV/atom, and all structures are relaxed using DFT. We adopt the original random 60/20/20 split for training, validation, and testing following Xie *et al.* [6].

Alex-MP-20. Alex-MP-20, used in MatterGen [3], is a large-scale crystal dataset derived from the Materials Project [48] and Alexandria [24, 25] (Creative Commons Attribution 4.0 International License) datasets, containing 607,683 structures with up to 20 atoms in the unit cell. It is constructed by retaining structures with DFT-computed energy above hull below 0.1 eV/atom, while manually excluding structures from well-explored chemical systems, yielding a large and diverse set of stable and metastable crystals.

B.2 Universal MLIPs

We use 10 pretrained universal MLIPs as frozen representation teachers, spanning different model families, pretraining data sources, model sizes, and representation dimensions. These teachers include DPA, MACE, ORB, SevenNet, and MatterSim models, as summarized in Tab. 3. Their pretraining data cover several large-scale atomistic datasets, including MPtrj [23], Alexandria [24, 25], sAlex [26], OMat24 [26], OpenLAM [50], COSMOSDataset [32], and MatterSim [39]. DPA-3.1-3M-FT denotes the DPA-3.1-3M checkpoint further fine-tuned on MPtrj.

Table 3: Universal MLIP teachers considered in this work. We summarize the atom-wise embedding dimension used for representation alignment, the number of model parameters, the approximate scale of the pretraining data, and the main pretraining data sources.

Model	Embedding dim.	#Params	Pretraining scale	Pretraining data sources
DPA-3.1-3M [37]	128	3.27M	163M	OpenLAM
DPA-3.1-3M-FT [37]	128	3.27M	163M	OpenLAM
DPA-3.1-MPtrj [37]	128	3.27M	1.58M	MPtrj
MACE-MPA-0 [38]	256	9.06M	12M	MPtrj, sAlex
MACE-MP-0 [38]	256	4.69M	1.58M	MPtrj
ORB v3 [35]	256	25.5M	133M	MPtrj, Alex, OMat24
ORB v2 [41]	256	25.2M	32.1M	MPtrj, Alex
SevenNet-Omni-i12 [32]	64	54.9M	243M	COSMOSDataset
SevenNet-l3i5 [40]	64	1.17M	1.58M	MPtrj
MatterSim-v1-5M [39]	256	4.55M	17M	MatterSim

Table 4: Training hyperparameters for CrystalREPA across crystal generative models and datasets.

Category	Hyperparameter	CrystalFlow		MatterGen		DiffCSP	
		MP-20	Alex-MP-20	MP-20	Alex-MP-20	MP-20	Alex-MP-20
Training	# Params	20.9M	20.9M	44.6M	44.6M	12.3M	12.3M
	Number of GNN layers	6	6	4	4	6	6
	Atom-wise embedding dim.	512	512	512	512	512	512
	Optimizer	Adam	Adam	Adam	Adam	Adam	Adam
	Learning rate scheduling	ReduceLROnPlateau		ReduceLROnPlateau		ReduceLROnPlateau	
	Maximum learning rate	1e-03	1e-03	1e-04	1e-04	1e-04	1e-04
	Minimum learning rate	1e-04	1e-04	1e-06	1e-06	1e-05	1e-05
	Gradient accumulation steps	1	1	4	4	1	1
	Training epochs	10,000	5,000	1,900	1,200	10,000	3,000
	Batch size (per GPU)	256	128	128	32	256	256
	Number of GPUs	1	4	1	4	1	4
	Atom type loss prefactor λ_A	10	10	1	1	20	20
	Coordinate loss prefactor λ_F	10	10	0.1	0.1	1	1
	Lattice loss prefactor λ_L	1	1	1	1	1	1
CrystalREPA	Alignment loss	EA-NCE	EA-NCE	EA-NCE	EA-NCE	EA-NCE	EA-NCE
	Alignment loss prefactor λ_{REPA}	1	1	1	1	1	1
	Temperature τ	0.1	0.1	0.1	0.1	0.1	0.1
	Alignment layer	2	2	2	2	2	2

B.3 Hyperparameters

Tab. 4 summarizes the training hyperparameters used for CrystalREPA across different crystal generative models and training datasets. All reported experiments are conducted using 80GB NVIDIA A800 GPUs, with the number of GPUs specified in the table. The CrystalREPA projection head is implemented as a lightweight residual MLP, consisting of one residual block with a linear layer and a SiLU activation function, followed by a linear projection to the MLIP representation dimension. CrystalREPA only affects training and leaves the sampling procedure unchanged during inference; therefore, all inference settings are identical to those of the corresponding baseline models.

B.4 Formation energy probing

We evaluate representation quality by probing formation energy per atom. For generative models, we extract atom-wise hidden states from each encoder layer on clean structures at $t = 1.0$. For each crystal, let N be the number of atoms in the unit cell. The k-th layer atom-wise representations $\{\mathbf{h}_i^{(k)}\}_{i=1}^N$ are aggregated by mean pooling:

$$\mathbf{h}_{\mathrm{global}}^{(k)} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{h}_i^{(k)}. \tag{30}$$

For MLIP teachers, we apply the same mean-pooling procedure to their atom-wise representations. Before fitting the regressor, each feature dimension is standardized using the mean and standard deviation computed on the training split. We then train a ridge regression model on the MP-20 training split, select the regularization strength from $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 100\}$ using the validation split,

and report the mean absolute error (MAE) on the test split. The probing results are shown in Fig. 1(a) and Appendix Fig. 5.

B.5 Definition of CRI, ERI and TRI

To quantify the distinguishability of atom-wise representations learned by different teacher MLIPs, we define three resolution metrics: the *Chemical Resolution Index* (CRI), the *Environment Resolution Index* (ERI), and the *Total Resolution Index* (TRI). These metrics are designed to assess whether the representation space distinguishes chemical identities and reflects similarities in local geometric environments, which we hypothesize to be important for effective generative transfer.

Given a pooled set of N atoms collected from randomly sampled crystals, let $\mathbf{z} \in \mathbb{R}^{N\times d_2}$ be the atom-wise representations extracted from a pretrained universal MLIP. We first compute the cosine similarity matrix

$$\mathbf{S}_{ij} = \frac{\mathbf{z}_i^\top \mathbf{z}_j}{\|\mathbf{z}_i\|\|\mathbf{z}_j\|}. \tag{31}$$

Chemical Resolution Index (CRI). CRI measures how well the representation separates atoms of different elemental species. We partition all off-diagonal atom pairs into two sets: intra-element pairs, where $\mathbf{A}_i = \mathbf{A}_j$, and inter-element pairs, where $\mathbf{A}_i \neq \mathbf{A}_j$. We then define

$$\mathrm{CRI} = \frac{1}{|\mathcal{P}_{\mathrm{intra}}|} \sum_{(i,j)\in\mathcal{P}_{\mathrm{intra}}} \mathbf{S}_{ij} - \frac{1}{|\mathcal{P}_{\mathrm{inter}}|} \sum_{(i,j)\in\mathcal{P}_{\mathrm{inter}}} \mathbf{S}_{ij}, \tag{32}$$

where $\mathcal{P}_{\mathrm{intra}} = \{(i,j)|i \neq j, \mathbf{A}_i = \mathbf{A}_j\}$ and $\mathcal{P}_{\mathrm{inter}} = \{(i,j)|i \neq j, \mathbf{A}_i \neq \mathbf{A}_j\}$. A larger CRI indicates that atoms from the same element cluster more tightly in the representation space while atoms from different elements are clearly separated.

Environment Resolution Index (ERI). ERI measures whether the learned representation is sensitive to local geometric environments. As a reference for local-environment similarity, we use element-agnostic SOAP descriptors [51, 52]. Specifically, let $\mathbf{z}^{\mathrm{soap}} \in \mathbb{R}^{N\times d_3}$ denote the corresponding element-agnostic SOAP descriptors computed using the DScribe package [53, 54]. We then compute the cosine similarity matrix

$$\mathbf{S}^{\mathrm{soap}}_{ij} = \frac{(\mathbf{z}^{\mathrm{soap}}_i)^\top \mathbf{z}^{\mathrm{soap}}_j}{\|\mathbf{z}^{\mathrm{soap}}_i\|\|\mathbf{z}^{\mathrm{soap}}_j\|}. \tag{33}$$

We next compute the mean SOAP similarity over all off-diagonal pairs:

$$\bar{\mathbf{S}}^{\mathrm{soap}} = \frac{1}{N(N-1)} \sum_{i\neq j} \mathbf{S}^{\mathrm{soap}}_{ij}. \tag{34}$$

Using this mean value as a parameter-free threshold, we divide atom pairs into two groups (pairs exactly at the mean threshold are ignored):

$$\mathcal{P}_{\mathrm{high}} = \{(i,j)|i \neq j, \mathbf{S}^{\mathrm{soap}}_{ij} > \bar{\mathbf{S}}^{\mathrm{soap}}\}, \quad \mathcal{P}_{\mathrm{low}} = \{(i,j)|i \neq j, \mathbf{S}^{\mathrm{soap}}_{ij} < \bar{\mathbf{S}}^{\mathrm{soap}}\}. \tag{35}$$

We then define

$$\mathrm{ERI} = \frac{1}{|\mathcal{P}_{\mathrm{high}}|} \sum_{(i,j)\in\mathcal{P}_{\mathrm{high}}} \mathbf{S}_{ij} - \frac{1}{|\mathcal{P}_{\mathrm{low}}|} \sum_{(i,j)\in\mathcal{P}_{\mathrm{low}}} \mathbf{S}_{ij}. \tag{36}$$

A larger ERI indicates that the learned representation assigns higher similarity to atoms with more similar local geometric environments, and lower similarity to atoms in dissimilar environments.

Total Resolution Index (TRI). Finally, we combine the above two aspects into a single overall distinguishability score:

$$\mathrm{TRI} = \mathrm{CRI} + \mathrm{ERI}. \tag{37}$$

A larger TRI indicates that the representation is simultaneously discriminative with respect to both chemical species and local structural environment. We use the unweighted sum because both terms are defined as differences of cosine similarities and therefore lie on comparable scales. In practice, we pool atom-wise representations from 1,000 sampled crystal structures and compute CRI, ERI, and TRI over all resulting atom pairs.

B.6 Representation normalization

For the diagnostic experiments in Fig. 3, we apply a simple dimension-wise normalization to the frozen MLIP teacher representations before using them as alignment targets. This normalization is computed once over all atom-wise teacher representations from all structures in the corresponding training split. Specifically, let $\{\tilde{\mathbf{z}}_i\}_{i=1}^{N_{\text{tot}}}$ denote the flattened set of teacher representations over all atoms in the training split, where N_{tot} is the total number of atoms. We compute the dimension-wise mean and standard deviation as

$$\boldsymbol{\mu} = \frac{1}{N_{\text{tot}}} \sum_{i=1}^{N_{\text{tot}}} \tilde{\mathbf{z}}_i, \qquad \boldsymbol{\sigma} = \sqrt{\frac{1}{N_{\text{tot}}} \sum_{i=1}^{N_{\text{tot}}} (\tilde{\mathbf{z}}_i - \boldsymbol{\mu})^2}. \tag{38}$$

The normalized teacher representation is then given by

$$\tilde{\mathbf{z}}_i^{\text{norm}} = \frac{\tilde{\mathbf{z}}_i - \boldsymbol{\mu}}{\boldsymbol{\sigma} + \epsilon}, \tag{39}$$

where $\epsilon = 10^{-6}$ is used for numerical stability. The normalized representations are kept fixed and used as alignment targets.

B.7 Evaluation metrics

We follow the MatterGen evaluation protocol [3] to evaluate generated crystal structures. For each generative model, we sample 1,024 candidate structures and relax them using MatterSim-v1-1M [39]. Generated structures containing elements not supported by the MatterGen evaluation pipeline are excluded from metric computation. For each relaxed structure, we compute the energy above the convex hull, denoted as E_{hull}, where the reference convex hull is constructed from 845,997 structures from the Materials Project [48] and Alexandria [24, 25] datasets. A generated structure is considered metastable if $E_{\text{hull}} \leq 0.1$ eV/atom and stable if $E_{\text{hull}} \leq 0$ eV/atom. We report *Metastable%* and *Stable%* as the fractions of generated structures satisfying these two criteria, respectively. We also report the average energy above hull, denoted as $\bar{E}_{\text{hull}}$, over all generated structures.

As a proxy for structural fidelity, we report the root mean squared displacement (*RMSD*) between each generated structure before relaxation and its corresponding relaxed structure. A lower RMSD indicates smaller relaxation-induced structural changes, suggesting that the generated structure is closer to a local equilibrium configuration.

We report *Valid%* as the fraction of generated structures that pass both structural and compositional validity checks. For structural validity, we reject structures with interatomic distances smaller than 0.5 Å or unit cell volumes smaller than 0.1 Å^3. Compositional validity is evaluated using SMACT [55], which checks whether the generated composition can satisfy charge neutrality. A generated structure is counted as valid only when both checks are satisfied.

We report *Unique%* as the fraction of generated structures that do not match any other generated structure in the same batch. Following MatterGen, uniqueness is computed among structures with the same reduced chemical formula using an ordered-disordered structure matcher. We report *Novel%* as the fraction of generated structures that do not match any structure in the reference set used to construct the convex hull, which contains 845,997 entries. This definition does not necessarily guarantee novelty with respect to the complete pretraining corpora of all MLIP teachers.

Following prior work, we further report *MSUN%* and *SUN%*, which measure the yield of generated structures that simultaneously satisfy multiple discovery-oriented criteria. Specifically, MSUN requires a structure to be metastable, unique, and novel, while SUN requires a structure to be stable, unique, and novel. We report MSUN% and SUN% as the corresponding fractions among all generated samples.

B.8 Conditional generation

B.8.1 CrystalFlow

For conditional CrystalFlow, we augment the time-dependent vector fields with an external condition c, resulting in $\mathbf{v}_\theta^A(\mathcal{M}_t, t, c)$, $\mathbf{v}_\theta^F(\mathcal{M}_t, t, c)$, and $\mathbf{v}_\theta^L(\mathcal{M}_t, t, c)$. The original training objective in

Eqs. (1)–(2) is modified as

$$\mathcal{L}_{\text{gen}}^{\text{cond}} = \lambda_A \mathcal{L}_A(\mathbf{v}_\theta^A(\mathcal{M}_t, t, c), \mathbf{u}_t^A) + \lambda_F \mathcal{L}_F(\mathbf{v}_\theta^F(\mathcal{M}_t, t, c), \mathbf{u}_t^F) + \lambda_L \mathcal{L}_L(\mathbf{v}_\theta^L(\mathcal{M}_t, t, c), \mathbf{u}_t^L). \quad (40)$$

In the conditional generation experiments in Appendix C.5, we use energy above hull as the external condition. During training, c is set to the energy above hull label $c(\mathcal{M})$ of the corresponding clean structure. All other training hyperparameters are kept the same as those of the vanilla CrystalFlow baseline.

At inference time, we set the target condition to $c_{\text{target}} = 0$ eV/atom and keep CrystalFlow's original Euler sampler unchanged. The only modification is to replace the velocity field terms in Eqs. (14)–(16), namely $\mathbf{v}_\theta^A(\mathcal{M}_{t_m}, t_m)$, $\mathbf{v}_\theta^F(\mathcal{M}_{t_m}, t_m)$, and $\mathbf{v}_\theta^L(\mathcal{M}_{t_m}, t_m)$, with their conditional counterparts $\mathbf{v}_\theta^A(\mathcal{M}_{t_m}, t_m, c_{\text{target}})$, $\mathbf{v}_\theta^F(\mathcal{M}_{t_m}, t_m, c_{\text{target}})$, and $\mathbf{v}_\theta^L(\mathcal{M}_{t_m}, t_m, c_{\text{target}})$, respectively. All other sampling steps follow the original CrystalFlow sampler.

B.8.2 MatterGen

For conditional MatterGen, we follow its original property-guided generation strategy based on adapter fine-tuning and classifier-free guidance [56]. MatterGen fine-tunes the diffusion score network with condition-dependent adapter modules, while using the same denoising objective as the unconditional model. In our experiments, we fine-tune the adapter modules for 200 epochs.

In the conditional generation experiments in Appendix C.5, we use energy above hull as the external condition. Following the diffusion convention in Appendix A.2, where $\mathcal{M}_0$ denotes the clean structure, each $\mathcal{M}_0$ is paired with its energy above hull label $c(\mathcal{M}_0)$ during adapter fine-tuning. Following classifier-free guidance, the condition input is randomly replaced by a null condition c_{null}. Accordingly, the denoising objective is modified as

$$\begin{aligned}\mathcal{L}_{\text{diff}}^{\text{cond}} = \lambda_A \mathcal{L}_A(\hat{\boldsymbol{\epsilon}}_\theta^A(\mathcal{M}_t, t, \bar{c}), \boldsymbol{\epsilon}_A) + \lambda_F \mathcal{L}_F(\hat{\boldsymbol{\epsilon}}_\theta^F(\mathcal{M}_t, t, \bar{c}), \nabla_{\mathbf{F}_t} \log q\left(\mathbf{F}_t \mid \mathbf{F}_0\right)) \\ + \lambda_L \mathcal{L}_L(\hat{\boldsymbol{\epsilon}}_\theta^L(\mathcal{M}_t, t, \bar{c}), \boldsymbol{\epsilon}_L),\end{aligned} \quad (41)$$

where $\bar{c} \in \{c(\mathcal{M}_0), c_{\text{null}}\}$ denotes either the true condition or the null condition. This allows the same network to provide both conditional predictions $\hat{\boldsymbol{\epsilon}}_\theta(\mathcal{M}_t, t, c)$ and unconditional predictions $\hat{\boldsymbol{\epsilon}}_\theta(\mathcal{M}_t, t, c_{\text{null}})$ for classifier-free guidance.

At inference time, we set the target condition to $c_{\text{target}} = 0$ eV/atom. The conditional and unconditional denoising predictions are combined as

$$\hat{\boldsymbol{\epsilon}}_\theta^{A,\gamma}(\mathcal{M}_t, t, c_{\text{target}}) = \gamma \hat{\boldsymbol{\epsilon}}_\theta^A(\mathcal{M}_t, t, c_{\text{target}}) + (1-\gamma)\hat{\boldsymbol{\epsilon}}_\theta^A(\mathcal{M}_t, t, c_{\text{null}}), \quad (42)$$

$$\hat{\boldsymbol{\epsilon}}_\theta^{F,\gamma}(\mathcal{M}_t, t, c_{\text{target}}) = \gamma \hat{\boldsymbol{\epsilon}}_\theta^F(\mathcal{M}_t, t, c_{\text{target}}) + (1-\gamma)\hat{\boldsymbol{\epsilon}}_\theta^F(\mathcal{M}_t, t, c_{\text{null}}), \quad (43)$$

$$\hat{\boldsymbol{\epsilon}}_\theta^{L,\gamma}(\mathcal{M}_t, t, c_{\text{target}}) = \gamma \hat{\boldsymbol{\epsilon}}_\theta^L(\mathcal{M}_t, t, c_{\text{target}}) + (1-\gamma)\hat{\boldsymbol{\epsilon}}_\theta^L(\mathcal{M}_t, t, c_{\text{null}}), \quad (44)$$

where γ is the guidance factor. Following the original MatterGen setting, we set $\gamma = 2.0$.

We keep MatterGen's original reverse diffusion sampler unchanged. The only modification is to replace the corresponding denoising prediction terms in Eqs. (26)–(29), $\hat{\boldsymbol{\epsilon}}_\theta^A$, $\hat{\boldsymbol{\epsilon}}_\theta^F$, and $\hat{\boldsymbol{\epsilon}}_\theta^L$, with the classifier-free guided predictions defined in Eqs. (42)–(44), $\hat{\boldsymbol{\epsilon}}_\theta^{A,\gamma}$, $\hat{\boldsymbol{\epsilon}}_\theta^{F,\gamma}$, and $\hat{\boldsymbol{\epsilon}}_\theta^{L,\gamma}$, respectively. All other sampling steps follow the original MatterGen sampler.

B.8.3 DiffCSP

For conditional DiffCSP, we follow its original energy-guided sampling strategy for property optimization [57, 58], which uses formation energy per atom as the guidance target. Different from CrystalFlow and MatterGen, DiffCSP does not introduce the target property as an explicit input to the denoising network. Instead, it trains an additional time-dependent property predictor $E_\psi(\mathcal{M}_t, t)$, which takes an intermediate noisy structure $\mathcal{M}_t = (\mathbf{A}_t, \mathbf{F}_t, \mathbf{L}_t)$ as input and is supervised by the property label of the corresponding clean structure.

In the conditional generation experiments in Appendix C.5, the predictor $E_\psi(\mathcal{M}_t, t)$ is trained to predict the formation energy label $c(\mathcal{M}_0)$ of the corresponding clean structure. At inference time, we evaluate E_ψ on the current intermediate structure and use its gradients with respect to the current

variables to guide the reverse diffusion process toward lower-energy structures. The sampling process changes from Eqs. (26)–(28) to:

$$\mathbf{L}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{L}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\hat{\boldsymbol{\epsilon}}_\theta^L\right) - s\frac{\beta_t(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\nabla_{\mathbf{L}_t}E_\psi + \sqrt{\frac{\beta_t(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_L, \tag{45}$$

$$\mathbf{A}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{A}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\hat{\boldsymbol{\epsilon}}_\theta^A\right) - s\frac{\beta_t(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\nabla_{\mathbf{A}_t}E_\psi + \sqrt{\frac{\beta_t(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_A, \tag{46}$$

$$\mathbf{F}_{t-\frac{1}{2}} = w\left(\mathbf{F}_t + \left(\sigma_t^2 - \sigma_{t-1}^2\right)\hat{\boldsymbol{\epsilon}}_\theta^F - s\frac{\sigma_{t-1}^2(\sigma_t^2-\sigma_{t-1}^2)}{\sigma_t^2}\nabla_{\mathbf{F}_t}E_\psi + \frac{\sigma_{t-1}\sqrt{\sigma_t^2-\sigma_{t-1}^2}}{\sigma_t}\boldsymbol{\epsilon}_F\right), \tag{47}$$

where s is the guidance scale. The subsequent corrector step for fractional coordinates remains unchanged. Following DiffCSP, we set $s = 50$.

C Additional Results

C.1 Additional representation probing results

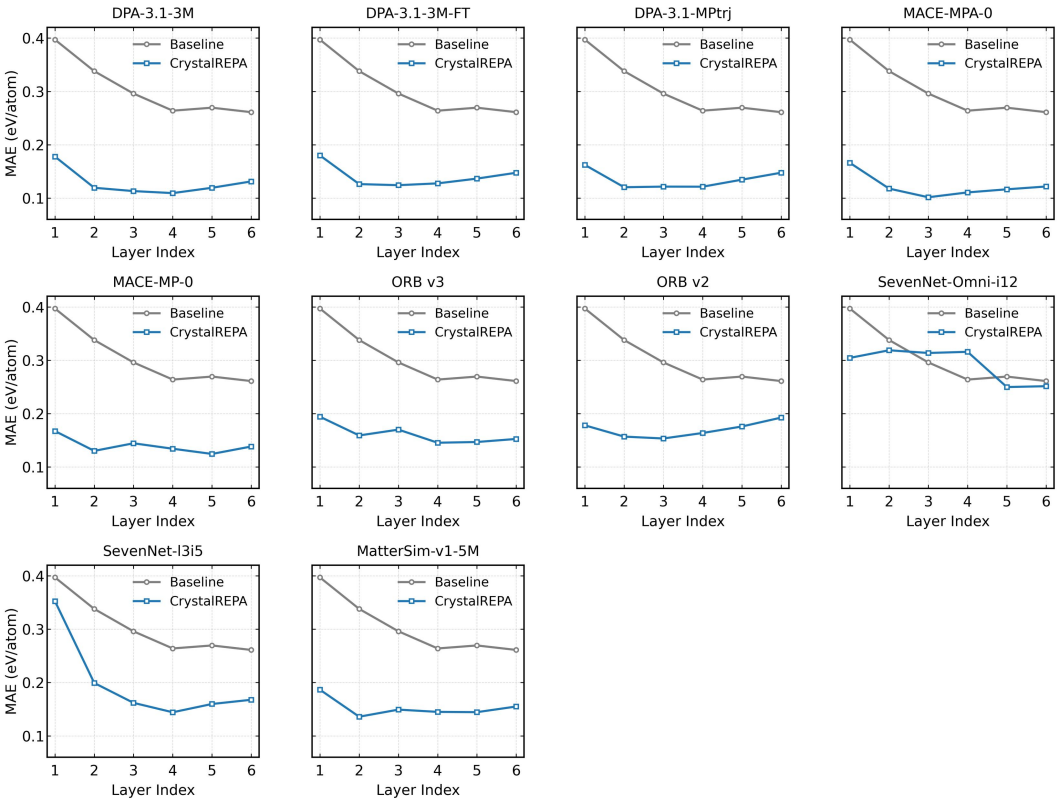


Figure 5: Additional formation energy probing results across different MLIP teachers on MP-20.

Fig. 5 provides additional formation energy probing results for CrystalREPA trained with different MLIP teachers, following the same protocol as in Appendix B.4. Across most teachers, CrystalREPA reduces the probing MAE compared with vanilla CrystalFlow, especially in early and intermediate layers, suggesting that representation alignment generally improves the formation-energy predictiveness of the generative representations. One exception is SevenNet-Omni-i12, which shows little probing improvement, consistent with its relatively weak transfer effectiveness in generation experiments.

Table 5: DFT-based evaluation of CrystalFlow and CrystalREPA using DPA-3.1-3M as the MLIP teacher. Each method is evaluated on 200 generated structures.

Model	Metastable%↑	MSUN%↑	Stable%↑	SUN%↑	$\bar{E}_{\text{hull}}$ ↓	RMSD↓	Valid%↑	Unique%↑	Novel%↑
Vanilla CrystalFlow	50.0	15.0	**11.0**	**1.5**	0.149	0.51	89.5	**100.0**	**61.5**
+ CrystalREPA	**64.0**	**22.5**	**11.0**	**1.5**	**0.114**	**0.44**	**90.0**	**100.0**	55.5

Table 6: Finite-temperature MD evaluation of CrystalFlow and CrystalREPA using DPA-3.1-3M as the MLIP teacher.

Model	MD-SR (%)↑	Max RMSD (Å)↓	Max AD (Å)↓
Vanilla CrystalFlow	89.8	0.198	0.553
+ CrystalREPA	**93.2**	**0.157**	**0.452**

C.2 Additional validation of generated crystals

DFT-based evaluation. To further validate the stability of generated crystals beyond MLIP-based evaluation, we perform DFT calculations on 200 structures generated by vanilla CrystalFlow and 200 structures generated by CrystalFlow with CrystalREPA, using DPA-3.1-3M as the MLIP teacher. All calculations are performed using VASP 5.4.4 with the projector-augmented-wave (PAW) method and the PBE exchange–correlation functional, following Materials Project-compatible settings, including GGA+U corrections where applicable [59]. We use a plane-wave cutoff of 520 eV and Γ-centered k-point meshes with a reciprocal-space density of 64 Å^{-3}. Atomic positions and cell parameters, including cell shape and volume, are jointly relaxed until the residual forces are below 0.02 eV/Å. All relaxations are initialized directly from the generated structures without prior MLIP relaxation.

As shown in Tab. 5, CrystalREPA improves Metastable% from 50.0 to 64.0 and MSUN% from 15.0 to 22.5, while reducing $\bar{E}_{\text{hull}}$ from 0.149 to 0.114 eV/atom and RMSD from 0.51 to 0.44. These results confirm that the improvements in metastability and structural fidelity persist under DFT-based evaluation.

Finite-temperature molecular dynamics. We further evaluate the dynamical stability of generated crystals through NVT molecular dynamics simulations using the Atomic Simulation Environment (ASE) package [60] with MatterSim-v1.0.0-1M. Each structure is expanded to a supercell with all lattice dimensions of at least 10 Å. Simulations are performed at 100 K using a Langevin thermostat with a timestep of 1 fs. Each trajectory runs for 25 ps, consisting of 5 ps of equilibration and 20 ps of production dynamics.

A trajectory is considered dynamically stable if it completes successfully and satisfies all the following criteria: a minimum interatomic distance above 0.75 Å, a maximum trajectory RMSD below 1.0 Å, an absolute production-stage energy trend below 0.01 eV/atom/ps, and a mean production-stage temperature within 20 K of the target temperature. We evaluate 500 generated structures from each of vanilla CrystalFlow and CrystalFlow with CrystalREPA (using DPA-3.1-3M as the MLIP teacher), and report three metrics. The MD stability rate (MD-SR) is the percentage of structures satisfying all stability criteria. Max RMSD is the largest root mean square atomic displacement from the starting structure observed during a trajectory. Max AD is the largest displacement of any individual atom during a trajectory. Max RMSD and Max AD are reported as medians over the 500 evaluated structures.

As shown in Tab. 6, CrystalREPA increases MD-SR from 89.8% to 93.2% compared with vanilla CrystalFlow. It also reduces the median Max RMSD from 0.198 to 0.157 Å and the median Max AD from 0.553 to 0.452 Å, indicating smaller atomic displacements during finite-temperature dynamics.

Predicted synthesizability. For synthesizability, we adopt the predictor introduced in prior work [61]. The predictor is trained under a positive-unlabeled learning framework to estimate the synthesizability of an inorganic material from its elemental stoichiometry. It reports an output probability between 0 and 1, which we use as the synthesizability score, with higher values indicating a greater predicted likelihood of experimental synthesis. We evaluate the same 1,024 generated

Table 7: Predicted synthesizability scores of generated crystals using DPA-3.1-3M as the MLIP teacher. Higher scores indicate greater predicted synthesizability. Values in parentheses denote standard errors over five independent sampling and evaluation runs.

Model	Vanilla	+ CrystalREPA
CrystalFlow	0.64(0.00)	**0.67**(0.00)
MatterGen	0.68(0.00)	**0.69**(0.00)
DiffCSP	0.62(0.00)	**0.68**(0.01)

Table 8: Comparison of SOAP-based and MLIP-based representation alignment using CrystalFlow. All models are trained for 3,000 epochs, with cosine similarity used as the alignment loss for SOAP- and MLIP-based alignment.

Model	Target Rep.	Metastable%↑	MSUN%↑	Stable%↑	SUN%↑	$\bar{E}_{\mathrm{hull}}$ ↓	RMSD↓	Valid%↑	Unique%↑	Novel%↑
Vanilla CrystalFlow	–	34.8	15.5	4.1	1.3	0.224	0.64	84.5	**99.9**	**78.1**
+ CrystalREPA	SOAP	36.4	16.8	3.7	1.3	0.211	0.60	82.7	99.8	77.1
+ CrystalREPA	Avg. over 10 MLIPs	**44.0**	**18.2**	**5.1**	**1.4**	**0.196**	**0.57**	**84.8**	99.8	70.6

structures per model as in Tab. 1. As shown in Tab. 7, CrystalREPA consistently improves the average synthesizability scores across all three generative frameworks.

Validity-aware MSUN. Following the standard MatterGen evaluation protocol [3], MSUN% measures the fraction of generated structures that are simultaneously metastable, unique, and novel, without explicitly requiring structural and compositional validity. We additionally evaluate Valid-MSUN%, which requires generated structures to satisfy both the MSUN criteria and the validity checks described in Appendix B.7. For CrystalFlow with DPA-3.1-3M as the MLIP teacher, CrystalREPA improves Valid-MSUN% from 14.7% to 17.2%, consistent with the improvement in MSUN% from 17.2% to 20.0%. This indicates that the improvement in MSUN% is not primarily driven by invalid structures.

C.3 Additional ablation studies

SOAP alignment control. To further investigate the contribution of stability-aware information learned by MLIPs, we compare MLIP-based alignment with alignment using SOAP-based atomic representations [51]. Unlike MLIP representations, SOAP descriptors characterize local atomic environments without being learned from energy or force supervision. Standard SOAP scales poorly with chemical diversity: MP-20 contains 89 elements, which would produce a descriptor with 1,776,796 dimensions and is therefore impractical as an alignment target. We instead use an element-agnostic SOAP descriptor [52] with 252 dimensions to encode local geometry and concatenate an 89-dimensional one-hot element encoding, yielding a compact 341-dimensional target representation.

We conduct the comparison using CrystalFlow. All models are trained for 3,000 epochs, with cosine similarity used as the alignment loss for SOAP- and MLIP-based alignment. Notably, this SOAP-based representation has a TRI of 1.15, substantially higher than the maximum TRI of 0.62 among the evaluated MLIP representations.

As shown in Tab. 8, SOAP alignment improves Metastable%, MSUN%, and $\bar{E}_{\mathrm{hull}}$ over vanilla CrystalFlow, confirming that geometric and elemental information is useful. Despite its substantially higher TRI, SOAP alignment still underperforms MLIP-based alignment on the main stability metrics, indicating that geometric and elemental distinguishability, as captured by this SOAP target, is not sufficient to explain the full improvement.

Energy-direction ablation. The SOAP comparison shows that MLIP-based alignment achieves stronger stability improvements than alignment with geometric and elemental descriptors. To further investigate the role of energy-related information, we conduct a diagnostic experiment that removes a learned formation-energy direction from the MLIP teacher representations.

For each crystal i, we first mean-pool its atom-wise DPA-3.1-3M representations to obtain $\bar{\mathbf{z}}_i = \frac{1}{n_i}\sum_{a=1}^{n_i}\mathbf{z}_{i,a}$, where n_i is the number of atoms in crystal i. Following the probing setup in Fig. 1(a),

Table 9: Effect of removing the learned formation-energy direction from DPA-3.1-3M representations. All variants are trained for 3,000 epochs using CrystalFlow.

Teacher representation	Metastable%↑	MSUN%↑	Stable%↑	SUN%↑	$\bar{E}_{\text{hull}}$ ↓	RMSD↓	Valid%↑	Unique%↑	Novel%↑
Original DPA-3.1-3M	**47.9**	**19.3**	**6.0**	**1.6**	**0.178**	**0.52**	85.6	**99.8**	68.0
w/o learned energy direction	43.1	18.0	5.0	1.3	0.199	0.63	85.0	**99.8**	**71.4**
w/o random direction	47.0	19.1	5.7	1.4	0.179	0.59	**86.0**	99.6	68.6

Table 10: Effect of the MLIP teacher representation layer on CrystalREPA using CrystalFlow and DPA-3.1-3M. All models are trained for 3,000 epochs. Layer 16 is the final MLIP representation layer.

Model	MLIP Layer	Metastable%↑	MSUN%↑	Stable%↑	SUN%↑	$\bar{E}_{\text{hull}}$ ↓	RMSD↓	Valid%↑	Unique%↑	Novel%↑
Vanilla CrystalFlow	–	34.8(0.3)	15.5(0.4)	4.1(0.3)	1.3(0.1)	0.224(0.004)	0.64(0.01)	84.5(0.4)	**99.9**(0.0)	**78.1**(0.6)
+ CrystalREPA	1	40.5(0.7)	18.2(0.3)	4.3(0.2)	1.1(0.2)	0.191(0.004)	0.63(0.00)	**85.6**(0.5)	99.7(0.0)	73.8(0.8)
+ CrystalREPA	4	40.8(0.3)	17.1(0.3)	4.4(0.2)	1.2(0.1)	0.199(0.001)	**0.52**(0.01)	84.3(0.3)	99.7(0.1)	73.2(0.6)
+ CrystalREPA	8	44.1(0.3)	18.5(0.2)	5.1(0.2)	1.2(0.1)	0.184(0.005)	0.56(0.00)	85.1(0.3)	99.7(0.0)	71.0(0.1)
+ CrystalREPA	12	45.7(0.6)	18.8(0.5)	5.3(0.2)	1.3(0.1)	**0.176**(0.005)	0.58(0.00)	85.4(0.5)	99.8(0.1)	69.9(0.2)
+ CrystalREPA	16	**47.9**(0.4)	**19.3**(0.5)	**6.0**(0.2)	**1.6**(0.1)	0.178(0.005)	**0.52**(0.01)	**85.6**(0.2)	99.8(0.1)	68.0(0.6)

we fit a ridge regression probe to predict the formation energy per atom. Specifically, we solve

$$(\hat{\mathbf{w}}, \hat{b}) = \arg\min_{\mathbf{w},b} \sum_{i=1}^{N} \left(E_{\text{form},i} - \mathbf{w}^\top \bar{\mathbf{z}}_i - b\right)^2 + \alpha \|\mathbf{w}\|_2^2. \tag{48}$$

The normalized regression coefficient $\mathbf{u} = \hat{\mathbf{w}}/\|\hat{\mathbf{w}}\|_2$ defines a unit direction associated with formation energy. For each atom-wise teacher representation, we remove its component along this direction using

$$\mathbf{z}_{i,a}^{\perp} = \mathbf{z}_{i,a} - (\mathbf{z}_{i,a}^\top \mathbf{u})\mathbf{u}. \tag{49}$$

We then retrain CrystalREPA on CrystalFlow using the residual representations $\mathbf{z}_{i,a}^{\perp}$ as the alignment targets. As a dimension-matched control, we repeat the experiment after removing one randomly sampled unit direction. All variants are trained for 3,000 epochs under the same settings.

As shown in Tab. 9, removing the learned energy direction substantially degrades the main stability metrics, reducing Metastable% from 47.9 to 43.1 and MSUN% from 19.3 to 18.0, while increasing $\bar{E}_{\text{hull}}$ from 0.178 to 0.199 eV/atom. In contrast, removing a randomly sampled direction largely preserves the original performance, achieving 47.0 Metastable% and 19.1 MSUN%. This dimension-matched control suggests that the degradation is specifically associated with removing the learned energy direction, supporting a contribution from energy-related information to downstream generative performance.

MLIP teacher-layer ablation. While Fig. 4(a) examines the alignment layer in the generative model, we further investigate the choice of representation layer in the MLIP teacher. In our main experiments, we use the final representation layer of the MLIP teacher for alignment. To examine whether intermediate MLIP representations may be more effective, we conduct a teacher-layer ablation using CrystalFlow as the generative backbone and DPA-3.1-3M as the teacher MLIP. All models are trained for 3,000 epochs.

As shown in Tab. 10, all tested teacher representation layers improve Metastable% and MSUN% over vanilla CrystalFlow. Deeper teacher representation layers generally provide stronger stability improvements, with the final representation layer achieving the best performance in Metastable%, MSUN%, Stable%, and SUN%. These results support our use of the final MLIP representation layer as a strong default, while also indicating that useful transferable information is present across multiple teacher depths.

MLIP-only optimization baseline. Having examined the role of MLIP representations in CrystalREPA, we further investigate whether direct MLIP-based optimization can achieve comparable generation quality. To this end, we construct an MLIP-only optimization baseline using MatterSim-v1-5M. For each structure, we retain the composition and cell volume of a structure generated by CrystalREPA on the DiffCSP framework, using MatterSim-v1-5M as the teacher. Retaining the cell

Table 11: Comparison of MLIP-only optimization, vanilla DiffCSP, and CrystalREPA using MatterSim-v1-5M. The MLIP-only baseline optimizes randomly initialized structures while retaining the compositions and cell volumes of CrystalREPA-generated structures.

Model	Metastable%↑	MSUN%↑	Stable%↑	SUN%↑	$\bar{E}_{\text{hull}}$ ↓	RMSD↓	Valid%↑	Unique%↑	Novel%↑
MLIP-only optimization	18.1	10.1	1.8	1.6	0.439	2.04	82.5	**100.0**	**90.5**
Vanilla DiffCSP	43.2	18.8	5.9	2.0	0.208	**0.36**	82.7	98.9	72.0
+ CrystalREPA	**52.6**	**23.7**	**7.9**	**2.9**	**0.158**	**0.36**	**85.4**	99.4	68.0

volume prevents physically unreasonable densities, such as overly compressed cells. We then randomly initialize the atomic coordinates and lattice shape, rescale the randomized lattice to the retained volume, and jointly optimize the atomic coordinates and lattice parameters using MatterSim-v1-5M.

As shown in Tab. 11, direct MLIP optimization from random initialization performs substantially worse than both vanilla DiffCSP and CrystalREPA in Metastable%, MSUN%, Stable%, and SUN%, while also yielding a higher average energy above hull and substantially larger RMSD. Although the MLIP-only baseline achieves higher Unique% and Novel%, this likely reflects the broad structural diversity introduced by random initialization rather than useful crystal generation, as most resulting structures are not metastable. These results indicate that MLIP optimization alone cannot replace the structural prior learned by the generative model.

C.4 Additional representation analysis

Energy Resolution Index (EnRI). To further examine the relationship between thermodynamic stability resolution and transfer effectiveness, we introduce the Energy Resolution Index (EnRI). EnRI is a structure-level metric that measures how well crystal representations distinguish different thermodynamic stability levels. We use energy above the hull, E_{hull}, as a proxy for thermodynamic stability. Since E_{hull} is defined at the crystal level, we mean-pool the atom-wise MLIP representations to obtain one global representation per crystal.

Given N crystals, let $\mathbf{z}_i \in \mathbb{R}^d$ denote the mean-pooled representation of crystal i. We partition the crystals into $K = 10$ quantile bins according to $E_{\text{hull},i}$, with approximately equal numbers of crystals per bin. We then compute the pairwise cosine similarity as

$$S_{ij} = \frac{\mathbf{z}_i^\top \mathbf{z}_j}{\|\mathbf{z}_i\|_2 \|\mathbf{z}_j\|_2}. \tag{50}$$

Let $\mathcal{B}_k$ denote the set of crystals in the k-th E_{hull} bin. The average within-bin similarity is

$$S_{\text{within}} = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{|\mathcal{B}_k|(|\mathcal{B}_k| - 1)} \sum_{\substack{i,j \in \mathcal{B}_k \\ i \neq j}} S_{ij}. \tag{51}$$

The average between-bin similarity is

$$S_{\text{between}} = \frac{2}{K(K-1)} \sum_{1 \leq k < l \leq K} \frac{1}{|\mathcal{B}_k||\mathcal{B}_l|} \sum_{\substack{i \in \mathcal{B}_k \\ j \in \mathcal{B}_l}} S_{ij}. \tag{52}$$

We define

$$\text{EnRI} = S_{\text{within}} - S_{\text{between}}. \tag{53}$$

A higher EnRI indicates that crystals within similar E_{hull} ranges have more similar representations than crystals from different E_{hull} ranges.

We compute EnRI for all ten MLIP teachers and correlate it with Metastable% of the corresponding CrystalFlow models trained with CrystalREPA. As shown in Fig. 6, EnRI exhibits a strong positive Pearson correlation of 0.936 with downstream Metastable%. Unlike ERI, which evaluates distinguishability with respect to geometry-derived SOAP similarities, EnRI evaluates distinguishability across thermodynamic stability levels without relying on SOAP. These results suggest that representations that better distinguish crystals across thermodynamic stability levels tend to yield stronger transfer performance.

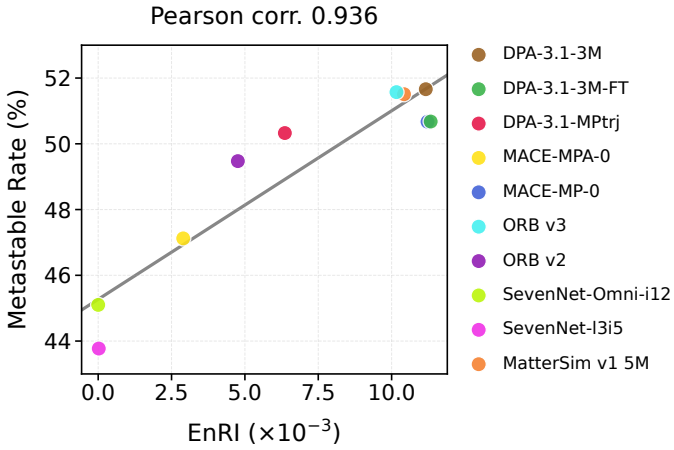


Figure 6: Correlation between the Energy Resolution Index (EnRI) and downstream Metastable% across ten MLIP teachers using CrystalFlow with CrystalREPA. Each point represents an MLIP teacher, and the line indicates a linear fit.

Table 12: Relationship between formation-energy probing performance and generative stability across three frameworks. CrystalREPA results are averaged over ten MLIP teachers.

Generative framework	Probe MAE, baseline↓	Probe MAE, +CrystalREPA↓	Metastable%, baseline↑	Metastable%, +CrystalREPA↑
CrystalFlow	0.338	0.158	41.2	50.5
DiffCSP	0.228	0.100	43.2	50.9
MatterGen	0.128	0.091	57.4	62.7

Formation-energy probing and generative performance. We further examine the relationship between formation-energy probing performance and generative stability across three generative frameworks and their CrystalREPA variants. Following the linear probing protocol in Appendix B.4, we report the formation-energy probing MAE at the second encoder layer, where CrystalREPA applies representation alignment. For each framework with CrystalREPA, the reported probing MAE and Metastable% are averaged over ten MLIP teachers.

As shown in Tab. 12, all three generative frameworks exhibit lower probing MAE and higher Metastable% after applying CrystalREPA. Across all 33 evaluated configurations, comprising three baseline models and thirty CrystalREPA variants, probing MAE has a Pearson correlation of -0.62 with Metastable%. These results further highlight the relationship between stability-related representation quality and downstream generative performance. Since these configurations differ in architecture, teacher, and training setup, this correlation should be interpreted as an association.

C.5 Comparison to conditional generation

A natural baseline for improving crystal stability is conditional generation, in which the generator is explicitly conditioned on a stability-related target property. Here, we compare CrystalREPA against this baseline using energy above hull (E_{hull}) as the conditioning signal. Implementation details for the conditional baselines are provided in Appendix B.8.

As shown in Tab. 13, conditional generation improves some stability-related metrics over vanilla generation, but its gains are not consistently better than those of CrystalREPA. While conditional generation often improves Stable%, SUN% and $\bar{E}_{\mathrm{hull}}$, it consistently reduces Valid%, indicating a trade-off between target-driven stability optimization and basic structural plausibility. By contrast, our proposed CrystalREPA yields more consistent gains on broader quality metrics. CrystalREPA improves not only post hoc energy-related metrics, but also structural validity, fidelity, and robustness across different generative frameworks.

CrystalREPA also has a practical advantage at inference time. It preserves the original sampling procedure and introduces no additional inference overhead. In contrast, conditional generation

Table 13: Comparison between CrystalREPA and conditional generation using external energy-related signal. For CrystalREPA, "average" reports the mean result across the 10 universal MLIP teachers used in this work, while "best" reports the best result among these teachers for each metric. **Bold** and underline denote the best and second-best results, respectively.

Model	Metastable%↑	MSUN%↑	Stable%↑	SUN%↑	$\bar{E}_{\mathrm{hull}}$ ↓	RMSD↓	Valid%↑	Unique%↑	Novel%↑
Vanilla CrystalFlow	41.2	17.2	5.1	1.2	0.198	0.60	84.9	99.6	**73.1**
Conditional Generation	40.8	16.2	6.4	1.7	0.177	0.62	81.9	99.7	72.6
CrystalREPA (average)	50.5	19.1	6.2	1.5	0.178	0.51	86.1	99.7	65.0
CrystalREPA (best)	**55.0**	**20.0**	**7.0**	**1.8**	**0.174**	**0.42**	**87.5**	**99.9**	69.0
Vanilla MatterGen	57.4	25.3	6.7	1.9	0.148	0.12	84.7	99.6	**64.6**
Conditional Generation	60.9	27.2	8.4	2.5	**0.119**	0.12	83.0	99.5	63.4
CrystalREPA (average)	62.7	27.7	7.8	2.5	0.131	**0.11**	85.6	99.5	61.8
CrystalREPA (best)	**66.2**	**29.2**	**8.7**	**2.9**	**0.119**	**0.11**	**86.6**	**99.7**	63.5
Vanilla DiffCSP	43.2	18.8	5.9	2.0	0.208	0.36	82.7	98.9	**72.0**
Conditional Generation	50.9	19.7	7.1	2.7	0.169	**0.31**	82.2	86.9	68.8
CrystalREPA (average)	50.9	23.4	7.2	2.6	0.166	0.37	84.3	99.2	69.3
CrystalREPA (best)	**56.5**	**24.8**	**8.2**	**2.9**	**0.147**	0.32	**85.4**	**99.5**	**72.0**

Table 14: Ablation on the MLIP teacher input for representation alignment. Using the clean structure $\mathcal{M}$ as the MLIP input yields better stability-oriented generation quality than using the noisy structure $\mathcal{M}_t$.

MLIP input	Metastable%↑	MSUN%↑	Stable%↑	SUN%↑	$\bar{E}_{\mathrm{hull}}$ ↓	RMSD↓	Valid%↑	Unique%↑	Novel%↑
None	35.8	12.9	5.1	1.4	0.278	0.4	86.3	**99.9**	**73.8**
Noisy structure $\mathcal{M}_t$	41.6	14.6	5.6	1.8	0.237	**0.32**	**87.8**	99.6	69.2
Clean structure $\mathcal{M}$	**50.5**	**16.5**	**7.9**	**2.6**	**0.204**	0.34	85.8	99.6	62.0

typically requires extra inference overhead, such as classifier-free guidance in MatterGen and energy-guidance strategy in DiffCSP.

More importantly, CrystalREPA and conditional generation should be viewed as complementary rather than mutually exclusive. CrystalREPA improves the physical quality of the learned generator itself, whereas conditional generation offers explicit controllability toward user-specified objectives. Exploring their combination is therefore a promising direction for future goal-directed crystal design.

C.6 Effect of MLIP teacher input

CrystalREPA uses the clean crystal $\mathcal{M}$, rather than the corrupted intermediate structure $\mathcal{M}_t$, as the input to the MLIP teacher. The generative encoder computes representations from $\mathcal{M}_t$ to predict the transport or denoising direction toward the clean structure, and the MLIP representation of the clean crystal provides auxiliary supervision consistent with this objective. This design is also computationally efficient, since MLIP representations of clean crystals are fixed and can be precomputed offline, whereas using $\mathcal{M}_t$ would require online MLIP forward passes for time-dependent corrupted structures during training.

We conduct this ablation on MP-20 using DiffCSP++ [14] as the base generative model and DPA-3.1-3M [37] as the MLIP teacher, with a batch size of 32 and 900,000 training steps. We compare this design with the natural alternative that feeds $\mathcal{M}_t$ into the MLIP teacher. As shown in Tab. 14, using the clean structure $\mathcal{M}$ yields stronger stability-oriented generation quality than using the noisy structure $\mathcal{M}_t$, improving Metastable%, MSUN%, Stable%, and SUN%, while reducing $\bar{E}_{\mathrm{hull}}$. These results support our choice of using the clean structure as the MLIP input and suggest that MLIP representations of clean crystals provide a more effective and efficient supervision signal for stability-oriented representation alignment.

C.7 Effect of data overlap

A potential concern is that the improvement from CrystalREPA may be driven by overlap between the MLIP teacher pretraining data and the crystal generator training data, rather than by transferable physical priors. To examine this issue, we construct a more conservative training split from Alex-MP-20, denoted as Alex-MP-clean. Specifically, we remove all structures in the Alex-MP-20 training split whose reduced formulas appear in MPtrj. This criterion is stricter than exact structure matching,

Table 15: Effect of data overlap between the MLIP teacher pretraining data and the crystal generator training data. We construct Alex-MP-clean by removing Alex-MP-20 training structures whose reduced formulas appear in MPtrj. DPA-3.1-3M still has substantial reduced formula overlap with Alex-MP-clean through broader pretraining sources such as OMat24, whereas DPA-3.1-MPtrj has no reduced formula overlap with Alex-MP-clean under this filtering criterion. All models are trained for 900 epochs.

Model	Target Rep.	Overlap	Metastable%↑	MSUN%↑	Stable%↑	SUN%↑	$\bar{E}_{\mathrm{hull}}$ ↓	RMSD↓	Valid%↑	Unique%↑	Novel%↑
Vanilla CrystalFlow [4]			57.0	30.6	4.3	1.8	0.110	0.38	87.3	**100.0**	70.7
+ CrystalREPA	DPA-3.1-3M	✓	58.5	**34.0**	4.3	**1.9**	0.107	**0.37**	88.5	**100.0**	**72.8**
(Ours)	DPA-3.1-MPtrj	×	**61.6**	31.7	**5.0**	**1.9**	**0.102**	0.44	**88.6**	**100.0**	67.2

because it removes not only identical structures but also structures with the same reduced formula. The resulting Alex-MP-clean split contains 523,090 structures, compared with 607,683 structures in the original Alex-MP-20 training split.

This filtering removes reduced formula overlap with the pretraining data of DPA-3.1-MPtrj. In contrast, DPA-3.1-3M is trained on broader data sources beyond MPtrj and therefore still has substantial remaining overlap with Alex-MP-clean. In particular, 492,438 out of the 523,090 structures in Alex-MP-clean have reduced formulas that appear in the OMat24 training set, which is part of the DPA-3.1-3M pretraining data. Therefore, DPA-3.1-3M represents a teacher with substantial remaining reduced formula overlap, while DPA-3.1-MPtrj serves as a non-overlap teacher under the same reduced formula criterion.

As shown in Tab. 15, CrystalREPA remains beneficial on Alex-MP-clean even when using DPA-3.1-MPtrj as the teacher. Compared with vanilla CrystalFlow, CrystalREPA with DPA-3.1-MPtrj improves Metastable% from 57.0 to 61.6 and Stable% from 4.3 to 5.0, while reducing $\bar{E}_{\mathrm{hull}}$ from 0.110 eV/atom to 0.102 eV/atom. MSUN% and SUN% also increase slightly, although RMSD and raw Novel% are mixed. These results indicate that reduced formula overlap with the teacher pretraining data is not necessary for CrystalREPA to improve generation quality.

Moreover, the improvements obtained with DPA-3.1-MPtrj are comparable to those obtained with DPA-3.1-3M, despite the much larger remaining overlap of DPA-3.1-3M with Alex-MP-clean. This suggests that training data overlap alone is unlikely to be the primary driver of the observed gains. Instead, the results are more consistent with our interpretation that CrystalREPA benefits from transferable stability-aware representations learned by pretrained MLIPs.

C.8 Computational cost

CrystalREPA introduces only marginal overhead during training and no additional cost at inference time. In our implementation, the atom-wise teacher representations of clean crystals are precomputed offline and stored with the corresponding training structures. Therefore, training does not require online forward passes through the pretrained MLIP, and the additional cost comes only from the lightweight projection head and the representation alignment loss. At inference time, both the teacher encoder and the projection head are removed, so the original sampling procedure is unchanged.

Tab. 16 reports the computational cost across three base crystal generative models on MP-20 and Alex-MP-20. Across the six settings in Tab. 16, CrystalREPA increases the training time per epoch by only 1.0% on average. The slightly lower training time observed for DiffCSP on Alex-MP-20 is likely due to normal runtime variation, as the measured difference is small. These results show that CrystalREPA is a lightweight training-time enhancement while preserving the inference cost of the base generative models.

D Broader Impact

This work focuses on a methodological improvement for crystal generative modeling. Its potential societal impact is expected to be indirect, mainly through downstream use in computational materials discovery. By improving the quality of generated crystal candidates, CrystalREPA may provide better initial candidates for downstream computational screening. However, generated structures should

Table 16: Computational cost comparison between baseline crystal generative models and CrystalREPA using DPA-3.1-3M as the teacher MLIP. All measurements are conducted on 80GB NVIDIA A800 GPUs. Training time follows the settings in Tab. 4, including the corresponding number of GPUs for each model and dataset. Inference time is measured with a unified batch size of 256 for fair comparison.

Data	Method	Train / epoch ↓	Infer / batch↓	Inference steps
MP-20	CrystalFlow	26.6s	7.6s	100
	+ CrystalREPA	27.1s	7.6s	100
	MatterGen	104.4s	1315.0s	1000
	+ CrystalREPA	105.9s	1315.0s	1000
	DiffCSP	26.6s	97.8s	1000
	+ CrystalREPA	26.9s	97.8s	1000
Alex-MP-20	CrystalFlow	177.8s	7.6s	100
	+ CrystalREPA	177.9s	7.6s	100
	MatterGen	837.8s	1315.0s	1000
	+ CrystalREPA	853.8s	1315.0s	1000
	DiffCSP	175.5s	97.8s	1000
	+ CrystalREPA	174.4s	97.8s	1000

undergo further computational or experimental validation before practical use. We do not identify immediate negative societal impacts specific to the proposed method.

NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

   Justification: The claims made in the abstract and introduction are empirically verified in Section 5.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

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

   Answer: [Yes]

   Justification: We discuss our limitations in Section 6.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

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

   Answer: [N/A]

Justification: This work does not present theoretical results or formal proofs.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

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

   Answer: [Yes]

   Justification: We disclose all information needed for reproduction in Appendix B.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

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

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

Answer: [Yes]

Justification: We provide publicly available code and instructions for reproducing the main CrystalREPA experiments at https://github.com/Chengqian-Zhang/CrystalREPA.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

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

   Answer: [Yes]

   Justification: We provide experimental details in Appendices B and C.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

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

   Answer: [Yes]

   Justification: We report standard errors over five independent sampling and evaluation runs for the generation results in Table 1, Table 2 and Figure 4.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

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

   Answer: [Yes]

   Justification: We provide the information on the computer resources in Appendix B.3 and Appendix C.8.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

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

   Answer: [Yes]

   Justification: We have read and obey the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

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

    Answer: [Yes]

    Justification: We discuss the broader impact of this work in the Appendix D.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

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

    Answer: [N/A]

    Justification: This work does not release high-risk data or models. It focuses on crystal generation for materials discovery using existing benchmark datasets.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

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

    Answer: [Yes]

    Justification: We cite the original papers for the datasets, models, and evaluation tools used in this work.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

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

    Answer: [Yes]

    Justification: We provide publicly available code and documentation for reproducing the main CrystalREPA experiments at https://github.com/Chengqian-Zhang/CrystalREPA.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

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

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

    Answer: [N/A]

    Justification: This work does not involve crowdsourcing experiments or research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

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

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

    Answer: [N/A]

    Justification: This work does not involve crowdsourcing experiments or research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

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

    Answer: [N/A]

    Justification: LLMs were used only for writing assistance and did not affect the core methodology, experiments, or evaluation.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.