MULTI-PROPERTY FEYNMAN-KAC STEERING OF CRYSTAL DIFFUSION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Pretrained diffusion models provide powerful priors for generating new materials, including inorganic crystals, but generating useful materials requires satisfying specific chemical and physical properties, from conductivity to band gaps. Existing approaches either retrain models for new target properties or steer only the atom coordinates and lattice during generation, leaving composition unchanged despite its fundamental role in determining physical properties. We present **FK-ReMask**, an inference-time, gradient-free method that steers both the continuous and discrete channels of a pretrained crystal diffusion model. During generation, FK-ReMask maintains multiple denoising trajectories and replicates promising candidates using Feynman–Kac particle selection. However, selection alone is insufficient: in the atom-type channel, committed elements cannot be revised, so cloning propagates existing compositions and limits the population's ability to explore new ones. We address this with an atom re-masking mechanism, which reopens committed sites during the generation process and allows alternative compositions to emerge. Across four electronic, photovoltaic, and mechanical objectives, our method improves the yield of SUN (stable, unique, novel) and property-on-target materials. We further demonstrate joint steering of three photovoltaic properties, showing that multiple physical objectives can be specified entirely at inference time without retraining.

1 INTRODUCTION

Designing new crystal materials with desired properties holds huge promise to advance diverse technologies, from photovoltaic cells, through carbon capture and catalysts, to efficient batteries. Diffusion models provide a natural framework for crystal generation (Zeni et al., 2025), but they typically learn to model the distribution of known materials they were trained on, rather than satisfy desired property requirements. For example, a photovoltaic absorber may require a suitable band gap together with favorable optical and electronic properties, while remaining stable and novel. Designing methods that can generate such materials largely remains an open problem.

Current diffusion generative approaches target desired properties in two main ways. First, during training, using classifier-free guidance (CFG), input conditioning (Zeni et al., 2025), or reinforcement-learning fine-tuning (Chen et al., 2025; Guo, 2026). Second, at inference time, through training-free guidance (steering) (de Lambilly et al., 2026). Current training-based approaches modify the model for target seen in training, and quite disappointingly, tend to generalize poorly to combinations of multiple properties.Training-free guidance avoids retraining cost by steering with gradients at generation time, but they are limited to differentiable objective. They act only on continuous channels (atom positions and lattice parameters) because a score-based gradient has no straightforward counterpart in the atom-type discrete channel. These approaches therefore do not guide the material composition (atom element type) despite being a fundamental determinant of material properties.

Here we describe **FK-ReMask**, an inference-time, gradient-free framework for multi-property control of pretrained crystal diffusion models. It has two important components: FK selection, and Atom remasking. For the first, FK-ReMask uses reward-based Feynman–Kac particle selection over a frozen pretrained generator frozen. When generating a crystal, we maintain a small population of parallel denoising trajectories for parallel exploration, periodically scoring their intermediate struc-

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

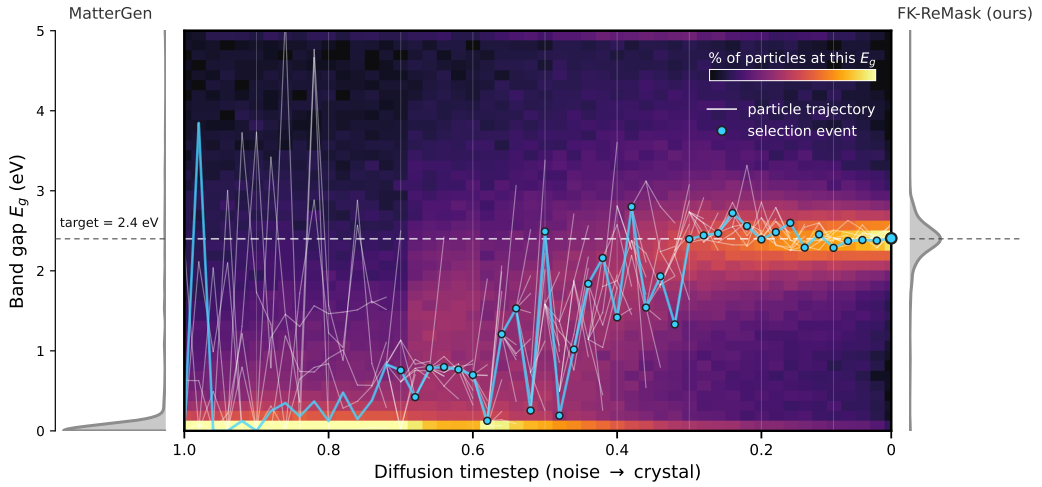


Figure 1: **A snapshot of guiding material generation towards a desired band gap, with FK sampling.** The figure traces the distribution of (predicted) band-gap as it changes through the diffusion denoising process, from random noise $t = 1$ to finished crystals $t = 0$. Light blue line traces one particle, with selection events marked with circles; thin white lines trace other particle lineages. The heatmap gives the distribution of the 4096 live particles at each timestep.; the dashed line marks the target 2.4 eV . Once FK selection begins ($t = 0.7$) the population shifts towards the target line. Side panels show the BG distribution for unsteered MatterGen (left) and FK-ReMask (right).

tures using external property predictors, and cloning higher-reward trajectories. Since rewards are computed on the full material state, this selection jointly steers the continuous and discrete channels. Multiple properties can be targeted by combining their reward terms, without updating the generator or requiring additional property-labeled data.

Selection alone, however, creates a challenge in the discrete atom channel. The pretrained generator fixes atom types once they are committed, so cloning high-reward particles progressively concentrates the population around previously selected compositions. As a second key component of our method, we therefore use an atom re-masking mechanism, which probabilistically reopens committed sites, allowing selected lineages to explore alternative chemical formulas during the generation process. Figure 1 shows an example of band-gap steering with FK-ReMask: once selection begins, the population shifts toward the target value.

We test our method on four single-property objectives and a joint three-objective photovoltaic absorber specification. The latter combines complementary requirements on optical performance, carrier transport, and electronic structure, demonstrating a realistic setting for target multiple material properties simultaneously.

Our contributions are: (1) FK-ReMask is the first training-free steering method for crystal diffusion models that jointly guides both the continuous channel of the material (coordinates and lattice) and its discrete (element composition) channel. (2) We introduce a novel use of atom re-masking to mitigate compositional collapse induced by particle cloning, enabling continued exploration of alternative compositions during the generation process. (3) Across single- and multi-property targeting tasks, FK-ReMask increases the fraction of generated crystals that are simultaneously stable, unique, novel and on the target property.

2 Related work

Particle-based diffusion steering. These methods steer generation by selecting among a population of trajectories. Interacting-particle methods provide the foundation for Feynman–Kac and sequential Monte Carlo approaches (Del Moral, 2004), which have been adapted to diffusion models for conditional sampling and inverse problems (Trippe et al., 2023; Cardoso et al., 2024; Wu et al., 2023).

More recent work uses particle populations explicitly for reward-based steering. (Singhal et al., 2025) formulate Feynman–Kac steering with trajectory-level potentials, and (Didi et al., 2026) apply it to atomistic binder design. Hasan et al. (2026) use SMC steering in masked discrete diffusion in text generation, where they report population collapse under particle cloning and address it by suspending resampling during the final stage of generation. This is in contrast to our approach which purpose to uses re-masking mechanism to remedy this diversity collapse.

Property targeting in crystal diffusion models. Most existing approaches reach property targets by modifying or fine-tuning the generator, which is costly in time and property-labeled data, and limits flexibility when targeting new properties or combinations of them. MatterGen(Zeni et al., 2025) fine-tunes adapter modules and conditions through classifier-free guidance, while exploring up to two joint property optimization. MatInvent (Chen et al., 2025) fine-tunes the same backbone with reinforcement learning, reaching up to three simultaneous objectives. As with other training-based methods, changing the target value requires a new fine-tuning run. (Guo, 2026) avoid jointly labeled multi-property data by fine-tuning one flow model per property, then composing the resulting correction fields at sampling time. New combinations then require no retraining, but each new property still requires its own fine-tuning run.

Recent work has also explored inference-time targeting. de Lambilly et al. (2026) steer MatterGen with Universal Guidance, a form of training-free guidance that requires a differentiable objective. Their updates act on atom positions and lattice parameters and leave the discrete atom channel unguided, so they can reshape geometry but not composition. ALM (Edamadaka et al., 2026) couples a language model to MatterGen through a learned conditioning bridge that injects the text condition into the diffusion process, and introduces T2C-FK, an N-particle SMC sampler that resamples trajectories according to rewards evaluated on clean estimates of partially denoised structures. However, they only targrt one physical property, formation energy, while keeping it's composition fixed.

Discrete diffusion and revision operators. D3PM (Austin et al., 2021) introduces the absorbing corruption used by the discrete atom channel: a site remains masked until it commits to an element, after which its identity cannot be changed. Wang et al. (2025) observe that masked discrete diffusion therefore cannot revise a generated token "even when it introduces an error", and derive ReMDM, a principled re-masking backward process whose operator we adopt. *Remask, Don't Replace* (Yao, 2026), *Revise, Don't Freeze* (Yu et al., 2026) and ProSeCo (Schiff et al., 2026) similarly enable previously generated tokens to be revised in diffusion language models. However, these methods use re-masking to enable revision within a single generation trajectory, while we use this mechanism to preserve compositional diversity within a group of trajectories.

3 PRELIMINARIES

Representating crystal materials. A crystal is represented as a $M = (A, X, L)$, where $A = (a_1, \ldots, a_N)$ is the list of atom types at the N sites, X contains fractional coordinates and L the lattice.

Score-based crystal diffusion models. It uses a score network s_θ to jointly denoise the continuous variables X and L with VE-SDE and VP-SDE, respectively, and the discrete atom types A with an absorbing D3PM process. (Austin et al., 2021) . This network takes as input crystal M_t and returns continuous scores for X and L, together with logits over clean atom types at each site. A softmax over the non-mask vocabulary, applied independently at each site, gives $\hat{p}_\theta(a \mid M_t)$, the predicted probability that a site's clean element is a. We keep s_θ frozen and never differentiate through it.

4 OUR METHOD: FK-REMASK

FK-ReMask combines two components for inference-time control of a frozen crystal diffusion model. **(1) Feynman–Kac Selection** reallocates particles toward higher-reward denoising paths. **(2) Atom Re-Masking Mechanism** reopens committed atom types, allowing the model to explore compositional alternatives during the generation. Together, they steer both continuous and discrete material channels. Figure 2 illustrates the method and Algorithm 1 gives the full sampler.

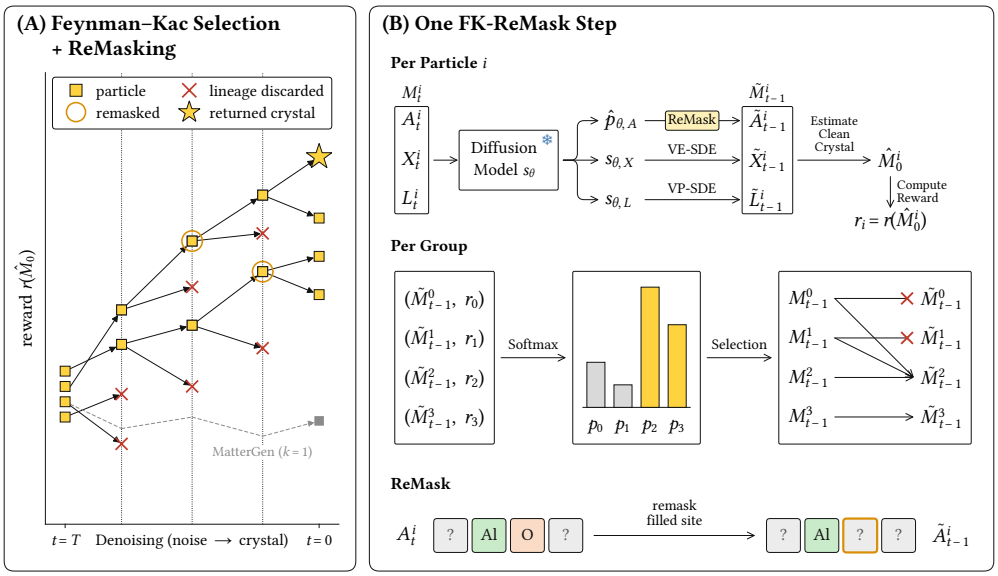


Figure 2: **Overview of FK-ReMask. (A)** Four particles are denoised in parallel ($k = 4$ shown). At each selection event (dotted lines), low-reward particles are discarded (red crosses) and high-reward particles are cloned. Re-masking lets a clone reopen some of its atom sites (circles), so clones of the same parent can diverge. The best particle at $t = 0$ is returned (star). The dashed line is a single unsteered MatterGen sample. **(B)** One FK-ReMask step. *Per particle:* the frozen diffusion model updates the atom types with re-masking and the positions and lattice with their SDEs. The reward is computed on the estimated clean crystal $\hat{M}_0^i$. *Per group:* a softmax over the rewards gives selection probabilities p_i, and the k slots are redrawn from them. *ReMask:* a filled atom site is potentially returned to the mask state.

4.1 Feynman–Kac selection

We formulate inference-time steering as a Feynman–Kac interacting particle system (Del Moral, 2004). When generating a crystal, we maintain a population of k particles which are parallel evolving under the same frozen reverse dynamics. Steering toward predefined property targets occurs through periodic selection: Every E reverse steps within a selection window, each particle's clean crystal is estimated by TWEEDIE, which applies the Tweedie formula (Efron, 2011) to positions and lattice, and by IMPUTEMASKED, which fills each masked atom site with $\arg\max_a \hat{p}_\theta(a \mid M_t)$, its most likely clean element. Sites that already hold an element are left unchanged. The k slots are then redrawn with probability proportional to the potential $G_t^{(i)} = \exp\left(r(\hat{M}_0^{(i)})/\tau\right)$, where τ sets how sharply the group concentrates.

This process repeatedly allocates more particles to promising regions of the generation process while retaining multiple trajectories for exploration. The reward never sees a noisy state, which is what lets it be any off-the-shelf predictor trained on finished crystals. Selection occurs only every E steps rather than at every step, which lets a selected lineage evolve between these events. At $t = 0$ all the particles are rewarded using their clean finished crystals, and the one with the highest reward is returned.

4.2 Atom re-masking

Selection only samples from crystals that the population already holds, limiting exploration. The continuous channels in our model keep exploring: every reverse step draws fresh Gaussian noise per particle, so positions and lattices go on moving. The atom-type channel does not. It is absorbing, so once a site commits to an element no later step can change it, and selection over composition reduces to choosing among formulas committed earlier in the trajectory. That is a real limit, since the properties we target depend on composition as much as on geometry.

Algorithm 1 FK-ReMask. Guided generation of a single crystal. Blue text marks our changes to standard FK. Every particle update applies to all $i \in [k]$.

Input: frozen score network s_θ, reward r, particles $i = 1, ..., k$, atom count N, reverse steps T, selection interval E, temperature τ, re-mask rate σ, re-masking T_{rm} and selection T_{sel} windows.

$M_T^i \sim \text{prior}(\cdot \mid N)$ ▷ k particles, N atoms each
for $t = T, \dots, 1$ **do**
  $X_{t-1}^i, L_{t-1}^i \leftarrow \textsc{ContinuousDenoise}\big(s_\theta, M_t^i\big)$
  $A_{t-1}^i \leftarrow \textsc{DiscreteDenoise}\big(s_\theta, T_{\text{rm}}, M_t^i\big)$ ▷ Eq. 1
  $M_{t-1}^i = (A_{t-1}^i, X_{t-1}^i, L_{t-1}^i)$
  if $t \in T_{\text{sel}}$ **and** $t \bmod E = 0$ **then**
    $\hat{X}_0^i, \hat{L}_0^i \leftarrow \textsc{Tweedie}\big(s_\theta, M_{t-1}^i\big)$
    $\hat{A}_0^i \leftarrow \textsc{ImputeMasked}\big(s_\theta, M_{t-1}^i\big)$
    $\hat{M}_0^i = (\hat{A}_0^i, \hat{X}_0^i, \hat{L}_0^i)$ ▷ clean crystal estimation
    $w \leftarrow \text{softmax}\big(r(\hat{M}_0)/\tau\big)$ ▷ over the group
    $\text{pa}(i) \sim \text{Cat}(w)$ ▷ k draws with replacement
    $M_{t-1}^i \leftarrow M_{t-1}^{\text{pa}(i)}$
return $M_{\text{final}} = \arg\max_i r(M_0^i)$ ▷ on finished crystals

Resampling every committed site would restore that freedom, but it would discard the composition built so far and leave a masked fraction the frozen model was never trained on. We reopen only part of it instead, controlled by a re-masking rate σ, and compensate so that the schedule is preserved.

Formally, let γ_t be the fraction of sites committed at time t, so that $\gamma_{t-1} > \gamma_t$ at the next, less noisy step, and write δ_x for the point mass on x. Under the stock absorbing kernel a committed site is frozen and a masked site unmasks with probability $(\gamma_{t-1} - \gamma_t)/(1 - \gamma_t)$. We replace it with an **atom re-masking** kernel, applied independently to each of the N sites:

$$q_\sigma\big(a_{t-1} \mid a_t\big) = \begin{cases} (1-\sigma_t)\,\delta_{a_t} \;+\; \sigma_t\,\delta_{[\text{MASK}]}, & a_t \neq [\text{MASK}], \\ \dfrac{\gamma_{t-1} - (1-\sigma_t)\,\gamma_t}{1-\gamma_t}\,\hat{p}_\theta \;+\; \dfrac{1-\gamma_{t-1}-\sigma_t\,\gamma_t}{1-\gamma_t}\,\delta_{[\text{MASK}]}, & a_t = [\text{MASK}], \end{cases} \tag{1}$$

with the rate gated to the window and capped so the mask coefficient stays non-negative,

$$\sigma_t \;=\; \min\Big\{\, \sigma \cdot \mathbf{1}\big[\, t \in T_{\text{rm}} \,\big],\;\; \tfrac{1-\gamma_{t-1}}{\gamma_t} \,\Big\}, \qquad \text{so } \sigma_t = 0 \text{ outside } T_{\text{rm}}. \tag{2}$$

The unmasking rate in the second branch is the stock rate plus $\sigma_t\gamma_t/(1-\gamma_t)$, exactly the mass returned to the mask state spread over the $1-\gamma_t$ masked sites, so the committed fraction is unchanged in expectation: $\gamma_t(1-\sigma_t) + (\gamma_{t-1} - (1-\sigma_t)\gamma_t) = \gamma_{t-1}$. Because unmasking is conditioned on the surrounding structure, variation enters without leaving the model's own distribution.

We adopt this operator from ReMDM (Wang et al., 2025) but repurpose it: they re-mask to correct decoding errors within one trajectory, whereas we use it against the compositional collapse that cloning induces across trajectories (Figure 3).

4.3 Reward Design

For a target property, let v_{pred} denote the predicted value and v_{target} the desired target. Our default reward is the negative normalised target distance, $r = -|v_{\text{pred}} - v_{\text{target}}|/s$. The rescaling factor s is computed once from samples generated by the unsteered model and then fixed throughout steering, $s = \text{sd}\big(|v - v_{\text{target}}|\big)$, putting reward terms for different properties on a comparable scale. For multiple objectives, we sum the per-objective rewards, $r = \sum_{j=1}^{P} \lambda_j r_j$; each r_j uses its own target and scale values.

5 Experimental Setup and Details

Experimental Design. We evaluate FK-ReMask against baseline methods across four material properties, first in single-objective settings (§6.1) and then by simultaneously targeting three photovoltaic properties (§6.2). We use MatterGen as our pretrained unconditional crystal diffusion model.

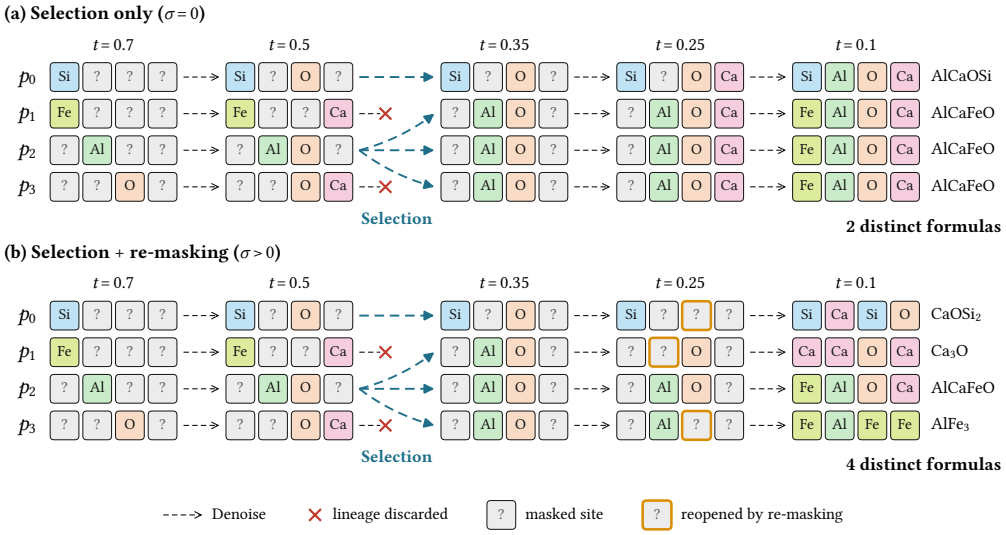


Figure 3: **Illustration of Atom re-masking under selection.** Four particles go through denoising, selection, and remasking.Cells denote the element composition at atom sites. **(a)** Without remasking, $\sigma = 0$, the selection step at $t = 0.35$ allocate winning particles to more slots. This reduces the diversity of the final generated compositions.. **(b)** With remasking, $\sigma > 0$, committed sites reopen inside the re-mask ($t = 0.25$) window yielding four distinct formulas.

Material Properties. We consider three photovoltaic properties and one mechanical property. **(1) Band gap (BG)** determines the spectral range available for absorption and constrains the achievable voltage, making it useful for identifying materials with suitable electronic structure for solar-energy conversion. **(2) Solar-cell efficiency (SLME)** accounts for the absorption spectrum and the distinction between fundamental and direct-allowed gaps (Yu & Zunger, 2012). **(3) Hole effective mass** captures carrier transport, as large effective masses can hinder carrier extraction. Together, BG, SLME, and hole effective mass capture complementary requirements for a practical solar absorber. **(4) Bulk modulus** characterizes resistance to volumetric compression, providing a target distinct from the electronic and optical properties.

Table 1: Properties, target values, and success criteria.

Material property	v_{target}	**On-target criteria**
Band Gap (E_g)	2.4 eV (§6.1), 1.3 eV (§6.2)	$\lvert E_g - v_{\text{target}} \rvert \leq 0.1$ eV
Solar-Cell Efficiency (SLME)	maximise (33.7 % is the ceiling)	$E_g \geq 0.5$ eV $\wedge$ SLME ≥ 25 %
Bulk Modulus (K)	150 GPa	$\lvert K - 150 \rvert \leq 15$ GPa
Hole Effective Mass (m_h^*)	minimise (0 is the floor)	$E_g \geq 0.5$ eV $\wedge$ $0 \leq m_h^* \leq 0.3$

Metrics. We report the percentages of generated structures that are valid, unique, novel, stable ($E_{\text{hull}} \leq 0.1$ eV/atom against the MP2020 convex hull) and SUN (structures that are simultaneously stable, unique, and novel). We further evaluate how well the generation met target property, using three : (1) on-target - the fraction of materials that satisfy the specified property target; (2) stable & on-target; and (3) SUN & target. Target values and property-specific criteria are given in Table 1.

Evaluation Protocol. We followed the MatterGen evaluation pipeline. For each method in each experiment, we generate $n = 512$ structures, relax them, and calculate the evaluation metrics on the relaxed structures. Structures that the evaluator cannot process are retained as failures rather than removed from the evaluation set. The on-target metric is computed using off-the-shelf pretrained property predictors, see Appendix A for details.

Baselines. We compare FK-ReMask with three baselines: **(1) MatterGen unsteered** samples from the frozen pretrained model without conditioning. **(2) MatterGen with CFG** using the released classifier-free-guidance adapter, fine-tuned on property-labelled data, available only for for band gap and bulk modulus. **(3) Best-of-N**. For each of n outputs, we generate k independent samples from the unsteered MatterGen model and select the best after relaxation, matching FK's kn trajectories.

Hyperparameters. We use $k = 8$ particles per generated crystal. Particle in a group are initialized from the prior with a shared atom count N. Selection occurs every $E = 20$ reverse steps at temperature $\tau = 0.2$ within a window $T_{\text{sel}} = [0.7, 0]$, giving 35 events on a uniform grid. Atom re-masking is applied at every reverse step within a window $T_{\text{rm}} = [0.60, 0.05]$, each committed site returning to the mask state with rate $\sigma = 0.1$.

6 RESULTS

We first evaluate steering generation using a single target property, repeated with 4 properties, then steer towards multiple properties jointly.

6.1 OPTIMIZING A SINGLE REQUESTED PROPERTY

Table 2 compares FK-ReMask to the baselines on four target properties. Our method achieves the highest SUN & on-target yield on all four, improving over the strongest baseline by 18.6 percentage points on band gap, 26.0 on SLME, 19.9 on hole effective mass, and 13.3 on bulk modulus. Figure 4 shows the distributions of predicted values for all methods and target properties.

Table 2: Single-objective results across four properties. Bold indicates the best value for each metric within each property.

property	method	valid	unique	novel	stable	SUN	on target	stable & on target	SUN & on target
Band Gap (E_g)	MatterGen	**85.9**	**99.8**	59.0	74.0	38.5	1.0	0.8±0.4	0.2±0.2
	+ CFG	**85.9**	98.6	67.6	68.8	37.5	6.4	4.3±0.9	2.0±0.6
	+ Best-of-N	84.2	**99.8**	64.3	72.1	38.9	8.4	5.7±1.0	3.1±0.8
	+ **FK-ReMask (ours)**	80.8	96.5	**68.5**	**82.0**	**50.9**	**46.3**	**40.0**±2.2	**21.7**±1.8
Solar-Cell Efficiency (SLME)	MatterGen	**85.9**	**99.8**	59.0	74.0	38.5	3.1	2.9±0.7	2.1±0.6
	+ Best-of-N	85.4	99.4	62.7	73.0	39.1	22.9	18.4±1.7	11.3±1.4
	+ **FK-ReMask (ours)**	85.4	98.8	**67.0**	**76.2**	**44.5**	**79.1**	**60.7**±2.2	**37.3**±2.1
Hole Effective Mass(m_h^*)	MatterGen	**85.9**	**99.8**	59.0	74.0	38.5	6.8	5.5±1.0	3.9±0.9
	+ Best-of-N	85.0	99.4	**63.3**	70.9	37.5	33.8	23.6±1.9	15.8±1.6
	+ **FK-ReMask (ours)**	81.8	94.7	60.9	**87.7**	**50.0**	**68.4**	**61.1**±2.2	**35.7**±2.1
Bulk Modulus (K)	MatterGen	85.9	**99.8**	**59.0**	**74.0**	**38.5**	4.9	2.9±0.7	1.4±0.5
	+ CFG	82.4	99.4	56.2	62.9	25.2	54.9	34.6±2.1	13.7±1.5
	+ Best-of-N	**86.3**	98.8	54.9	65.2	26.2	28.9	18.4±1.7	7.4±1.2
	+ **FK-ReMask (ours)**	75.0	97.9	51.6	70.1	27.3	**95.9**	**68.0**±2.1	**27.0**±2.0

6.2 MULTIPLE PROPERTIES

We next evaluated steering methods in the task of jointly targeting three properties: band gap, SLME, and hole effective mass. For FK-ReMask this was achieved by summing their rewards. Table 3 reports the results. The unsteered model rarely visits the region where all three conditions hold. Most samples are metallic: 79.5% have no band gap, disqualifying them from both the efficiency and hole-mass criteria. The 20.5% that do open a gap have a median of 1.92 eV, substantially above the 1.3 eV target, and 93% contain an electronegative anion, compounds in which holes are heavy and a gap that wide leaves most of the solar spectrum unabsorbed. Only 2 of 512 structures are stable, novel, and satisfy all three conditions. Best-of-N yields 6 that are stable and on target, but 3 of those already exist. In contrast, FK-ReMask yields 26, of which 14 are novel (Figure 5 displays three of them).

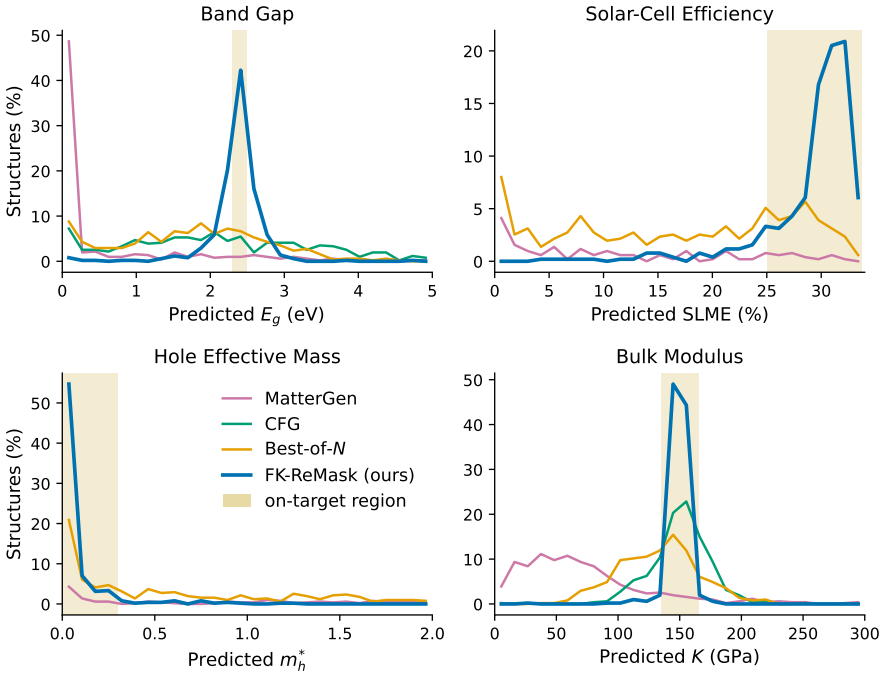


Figure 4: **Distribution of target properties**. Each curve is a histogram over crystals generated by each method. Solar-cell efficiency and hole effective mass are plotted for gapped structures only ($E_g \geq 0.5\,\text{eV}$), because the predictor returns a value for a metal that carries no physical meaning.

Table 3: **Multi-properties results**. Jointly targeting three photovoltaic objectives. *all* denotes the conjunction of the three on-target conditions. Bold indicates the best value for each metric.

						on target					
method	valid	unique	novel	stable	SUN	E_g	SLME	m_h^*	all	stable & all	SUN & all
MatterGen	**85.9**	99.8	59.0	74.0	38.5	0.8	3.1	6.8	0.4	$0.4_{\pm 0.3}$	$0.4_{\pm 0.3}$
+ Best-of-N	85.5	**99.8**	64.8	70.5	38.3	11.7	14.8	18.8	1.4	$1.2_{\pm 0.5}$	$0.6_{\pm 0.3}$
+ **FK-ReMask (ours)**	81.6	99.2	**65.4**	**76.6**	**44.9**	**17.0**	**43.2**	**46.1**	**5.7**	$\mathbf{5.1}_{\pm 1.0}$	$\mathbf{2.7}_{\pm 0.7}$

6.3 Selection is not enough

The results above show that selection with atom re-masking improves targeting, but leave open whether re-masking is necessary. To isolate this effect, we compare two otherwise identical band-gap steering experiments targeting 2.4 eV: one with re-masking ($\sigma = 0.1$) and one with selection alone ($\sigma = 0$). Along each denoising trajectory, we measure two quantities: (1) the number of distinct committed compositions among the eight particles. This measures the chemical alternatives actually represented in the particle population, rather than those introduced by imputing unresolved sites. (2) The mean band-gap reward across the particles, computed from the complete estimated structures, which tracks the quality of the current search population.

As shown in Figure 6(a), in both setups particles diverge freely prior to selection, reaching near the maximum of distinct compositions. But the initial selection events trigger a severe collapse in compositional diversity. Without re-masking, the population remains collapsed. In contrast, activating re-masking within $t \in [0.6, 0.05]$ allows committed sites to be iteratively revised, sustaining substantially greater compositional diversity across the window. This improves generation quality, raising the terminal mean reward (Figure 6(b)). Table 4 shows the same effect on the three-objective solar experiment from §6.2: without re-masking, all three targeting rates fall, and the SUN&on-target yield drops from 2.7% to 0.2%.

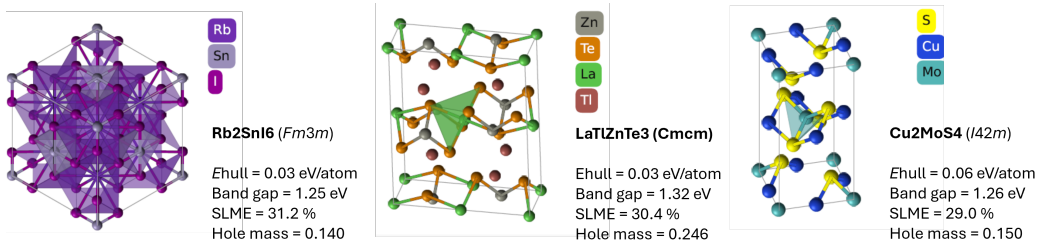


Figure 5: **Three example crystals generated by FK-ReMask.** Property values were predicted using pretrained ML predictors. Structures are higlighy diverse and meet the desired specification.

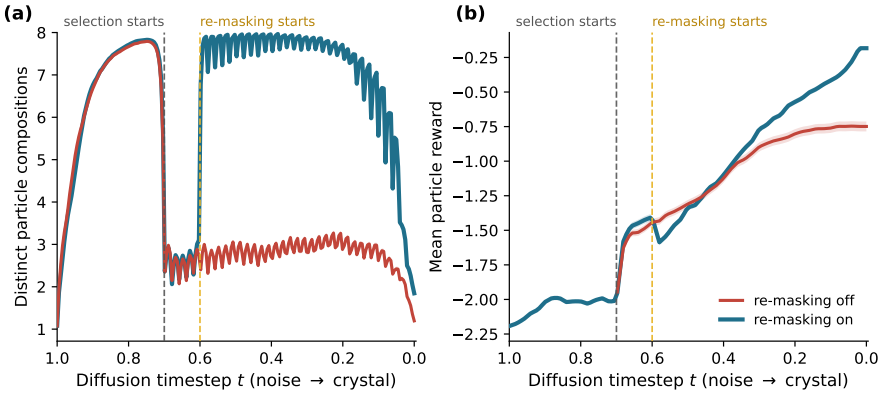


Figure 6: **Band-gap steering with and without re-masking.** Vertical dashed lines mark where selection and re-masking begin. **(a)** Selection collapses the population, but with re-masking on, diversity is restored. **(b)** Re-masking reaches higher mean particle reward.

Table 4: Re-masking on and off for the three-objective arm.

						on target					
	valid	unique	novel	stable	SUN	E_g	SLME	m_h^*	all	stable & all	SUN & all
$\sigma = 0$ (selection only)	**85.0**	**99.4**	64.5	70.1	37.7	9.4	26.0	27.0	1.4	1.0$\pm$0.4	0.2$\pm$0.2
$\sigma = 0.1$ (ours)	81.6	99.2	**65.4**	**76.6**	**44.9**	**17.0**	**43.2**	**46.1**	**5.7**	**5.1**$\pm$1.0	**2.7**$\pm$0.7

7 Limitations

FK-ReMask is designed as a fast, training-free method, that guides generation during inference. The method has several limitations. First, when large datasets of materials with desired target properties are available, finetuning models may end up performing better. Second, FK-ReMask repeatedly calls property predictors in the loop, thereby depending on available light-weight but accurate predictors. Also, real chemical validation often requires complex synthesis protocols that may be hard to achieve for novel materials. Incorporating high-fidelity evaluators that are also aware of synthesizability during denoising could also provide more useful generated materials.

8 Conclusion

We introduced FK-ReMask, a training-free, inference-time method for steering a frozen crystal generator across its continuous and discrete channels towards desired material properties. FK-ReMask Increases dramatically the fraction of generated crystals that meet the desired specification. For example 3$\times$ higher SUN and on target band-gap compared with baselines. Our approach introduces two key components into diffusion-based material generation: FK selection and Atom remasking.

REPRODUCIBILITY STATEMENT

Implementation details are provided in the designated section. The Code and trained models will be released publicly after completing the ongoing work.

USE OF GENERATIVE AI

We used generative AI tools to assist with method implementation, experiment design and methodological feedback, hypothesis refinement, result interpretation, formulation of mathematical claims and derivations, and data cleaning and reformatting. We did not use generative AI to generate synthetic datasets, translate content, or perform qualitative or thematic data analysis. We also used generative AI to assist with scientific figures, software development, paper drafting and editing, literature review and discovery, reference formatting, paper structure, method naming, experimental parameter suggestions, and brainstorming. All AI-assisted work was reviewed by the authors. Quantitative results were verified against stored evaluation artifacts and recomputed independently, and claims about prior work were checked against primary sources. The authors take full responsibility for the final content, claims, and artifacts.

REFERENCES

Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured denoising diffusion models in discrete state-spaces. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. arXiv:2107.03006.

Gabriel Cardoso, Yazid Janati El Idrissi, Sylvain Le Corff, and Eric Moulines. Monte carlo guided denoising diffusion models for Bayesian linear inverse problems. In *International Conference on Learning Representations (ICLR)*, 2024. arXiv:2308.07983.

Junwu Chen, Jeff Guo, Edvin Fako, and Philippe Schwaller. Accelerating inverse materials design using generative diffusion models with reinforcement learning. *arXiv preprint arXiv:2511.03112*, 2025.

Auguste de Lambilly, Vladimir Baturin, David Portehault, Guillaume Lambard, Nataliya Sokolovska, Florence d'Alché Buc, and Jean-Claude Crivello. Finetuning-free diffusion model with adaptive constraint guidance for inorganic crystal structure generation. *arXiv preprint arXiv:2604.13354*, 2026.

Pierre Del Moral. *Feynman–Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Probability and Its Applications. Springer, 2004. doi: 10.1007/978-1-4684-9393-1.

Kieran Didi, Zuobai Zhang, Guoqing Zhou, Danny Reidenbach, Zhonglin Cao, Sooyoung Cha, Tomas Geffner, Christian Dallago, Jian Tang, Michael M. Bronstein, Martin Steinegger, Emine Kucukbenli, Arash Vahdat, and Karsten Kreis. Scaling atomistic protein binder design with generative pretraining and test-time compute. In *International Conference on Learning Representations (ICLR)*, 2026. arXiv:2603.27950.

Sathya Edamadaka, Krithik Ramesh, Ju Li, and Rafael Gómez-Bombarelli. Atomistic language models understand and generate materials. *arXiv preprint arXiv:2606.21395*, 2026.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011. doi: 10.1198/jasa.2011.tm11181.

Junjie Guo. Train separately, compose at sampling: Multi-property crystal generation with orthogonal flow guidance. In *International Conference on Learning Representations (ICLR)*, 2026.

Mohsin Hasan, Viktor Ohanesian, Artem Gazizov, Yoshua Bengio, Alán Aspuru-Guzik, Roberto Bondesan, Marta Skreta, and Kirill Neklyudov. Discrete feynman-kac correctors. *arXiv preprint arXiv:2601.10403*, 2026.

Yair Schiff, Omer Belhasin, Roy Uziel, Guanghan Wang, Marianne Arriola, Gilad Turok, Ran Zilberstein, Michael Elad, and Volodymyr Kuleshov. Learn from your mistakes: Self-correcting masked diffusion models. *arXiv preprint arXiv:2602.11590*, 2026.

Raghav Singhal, Zachary Horvitz, Ryan Teehan, Mengye Ren, Zhou Yu, Kathleen McKeown, and Rajesh Ranganath. A general framework for inference-time scaling and steering of diffusion models. In *International Conference on Machine Learning (ICML)*, 2025. arXiv:2501.06848.

Brian L. Trippe, Jason Yim, Doug Tischer, David Baker, Tamara Broderick, Regina Barzilay, and Tommi Jaakkola. Diffusion probabilistic modeling of protein backbones in 3D for the motif-scaffolding problem. In *International Conference on Learning Representations (ICLR)*, 2023. arXiv:2206.04119.

Guanghan Wang, Yair Schiff, Subham Sekhar Sahoo, and Volodymyr Kuleshov. Remasking discrete diffusion models with inference-time scaling. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. arXiv:2503.00307.

Luhuan Wu, Brian L. Trippe, Christian A. Naesseth, David M. Blei, and John P. Cunningham. Practical and asymptotically exact conditional sampling in diffusion models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. arXiv:2306.17775.

Lin Yao. Remask, don't replace: Token-to-mask refinement in diffusion language models. *arXiv preprint arXiv:2604.18738*, 2026.

Liping Yu and Alex Zunger. Identification of potential photovoltaic absorbers based on first-principles spectroscopic screening of materials. *Physical Review Letters*, 108(6):068701, 2012.

Longxuan Yu, Shaorong Zhang, Yu Fu, Hui Liu, Yue Dong, and Greg Ver Steeg. Revise, don't freeze: Sampler-matched training for self-correcting masked diffusion language models. *arXiv preprint arXiv:2606.01026*, 2026.

Claudio Zeni et al. A generative model for inorganic materials design. *Nature*, 639:624–632, 2025. doi: 10.1038/s41586-025-08628-5.

APPENDIX

A PROPERTY PREDICTORS AND CFG CHECKPOINTS

Every reward is one forward call to an off-the-shelf pretrained ALIGNN model from the public zoo; none is trained or fine-tuned by us. Table 5 names them. The two released MatterGen classifier-free-guidance adapters used as baselines are `dft_band_gap` for the band gap and `ml_bulk_modulus` for bulk modulus; no adapter is released for SLME or hole effective mass.

Table 5: The predictor behind each objective of Table 1.

property	predictor
Band Gap (E_g)	`mp_gappbe_alignn`
Solar-Cell Efficiency (SLME)	`jv_slme_alignn`
Bulk Modulus (K)	`jv_bulk_modulus_kv_alignn`
Hole Effective Mass (m_h^*)	`jv_avg_hole_mass_alignn`