MATINVENT: REINFORCEMENT LEARNING FOR 3D CRYSTAL DIFFUSION GENERATION

Junwu Chen, Jeff Guo & Philippe Schwaller
Laboratory of Artificial Chemical Intelligence (LIAC)
Institute of Chemical Sciences and Engineering
& National Centre of Competence in Research (NCCR) Catalysis
Ecole Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland
{junwu.chen,jeff.guo,philippe.schwaller}@epfl.ch

ABSTRACT

Recent advances in diffusion models have enabled increasing capabilities for inverse materials design. The key capability is achieving tailored design towards desired property profiles, with wide applications for climate change, semiconductor design, and catalysis. In this work, we present MatInvent, a reinforcement learning (RL) framework tailored to optimize diffusion models for goal-directed crystal generation. By formulating equivariant denoising as a multi-step decision-making problem, MatInvent leverages policy optimization with reward-weighted KL regularization, including experience replay and diversity filters to enhance sample efficiency and diversity. Experimental results demonstrate that MatInvent outperforms existing baselines, offering an effective strategy for crystal generation with single or multiple property optimization.

1 INTRODUCTION

Accelerating the design and discovery of new functional materials is important to address pressing challenges such as climate change (Al-Rowaili et al., 2021) and semiconductor design (Zunger, 2018; Long et al., 2024). In recent years, the release of large open-source datasets (Stuke et al., 2020; Gallarati et al., 2022; Blaskovits et al., 2024; Jain et al., 2013; Chanussot et al., 2021; Barroso-Luque et al., 2024) have facilitated *in silico* materials design, particularly inverse design (Long et al., 2024) using generative models. To date, there have been more examples of applying generative models for drug discovery (Du et al., 2024) and many existing works adapt such frameworks for the design of *organic* materials (Marques et al., 2021; Staker et al., 2022; Han et al., 2024; Li & Tabor, 2023; Ma et al., 2022; Matsuzawa et al., 2024; Westermayr et al., 2023; Yang et al., 2023; Sharma et al., 2025). An important observation amongst these works is that learnings from generative drug design can directly benefit organic materials design, leading to several case studies that have demonstrated experimental validation (Yang et al., 2023; Matsuzawa et al., 2024). However, there have been less works for *inorganic* materials with existing approaches leveraging architectures such as generative adversarial networks (GANs) (Goodfellow et al., 2014; Nouira et al., 2018a; Kim et al., 2020a;b), variational auto-encoders (Kingma, 2013; Xie et al., 2022a; Luo et al., 2023), and GFlowNets (Bengio et al., 2021; AI4Science et al., 2023). With the advancements in diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020), more recent methods have proposed to model the 3D geometry of crystals, leading to capable models showing interpolation and generalization (Jiao et al., 2023; Yang et al., 2024; Zeni et al., 2025). While these recent models have effectively leveraged large datasets for conditional pre-training (Chanussot et al., 2021; Barroso-Luque et al., 2024), global coverage of all materials is generally infeasible. In response to this, works have investigated algorithmic optimization such as using evolutionary algorithms (Allahyari & Oganov, 2020) and reinforcement learning (RL) (Karpovich et al., 2024), but do not jointly consider information on the 3D geometry.

In this work, we combine the flexibility of RL with recent advances in diffusion models to tackle inorganic materials design. Building on observations made from generative drug design, we adapt and propose new optimization techniques to improve the sample efficiency of tailored generation

while ensuring crystal diversity. **Our contribution is as follows:** (1) We demonstrate that RL can tailor crystal diffusion generation for single/multiple property optimization. (2) We adapt and implement algorithmic components that improve sample efficiency and diversity during generation. (3) We show that our method can generate out-of-distribution of pre-training data, offering a flexible and practical framework for potential discovery.

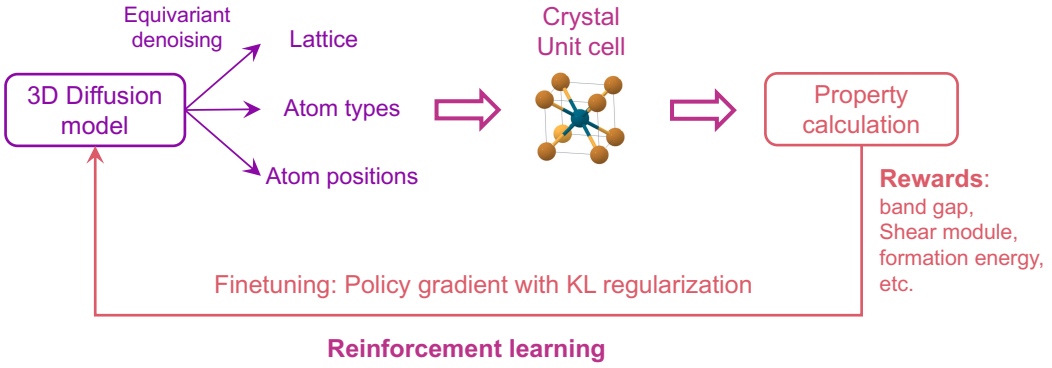

Figure 1: The schematic overview of this study.

2 Related Work

Diffusion models for *de novo* crystal generation. Generative models have revolutionized material discovery by directly generating 3D structures of stable materials, circumventing costly brute-force methods (Court et al., 2020; Nouira et al., 2018b). Traditional approaches relied on random atomic substitutions of crystal structure templates and quantum chemical calculations (Wang et al., 2021) or accelerated processes like genetic algorithms and energy prediction models (Glass et al., 2006; Pickard & Needs, 2011). Recent advancements in generative techniques, such as diffusion models Sohl-Dickstein et al. (2015); Ho et al. (2020), have proven effective. These models initially paired with Variational Autoencoders (VAEs) (Kingma, 2013) for partial variable predictions (Xie et al., 2022b; Luo et al., 2023) and later evolved to jointly diffuse atom types, atom coordinates, and lattice parameters (Jiao et al., 2023; Yang et al., 2024; Zeni et al., 2025), incorporating space group symmetries as inductive biases (Jiao et al., 2024; Lin et al., 2024; Cao et al., 2024). Other innovations include Riemannian Flow Matching (Miller et al., 2024; Sriram et al., 2024), Normalizing Flows (Wirnsberger et al., 2022), and standalone VAEs (Ren et al., 2022). Additionally, autoregressive Large Language Models (LLMs) have emerged as a parallel approach, representing materials as sequences of discretized tokens and leveraging powerful pretraining on natural language to embed rich prior knowledge (Flam-Shepherd & Aspuru-Guzik, 2023; Xiao et al., 2023; Antunes et al., 2024; Gruver et al., 2024). We note that there are other existing non-diffusion-based works that generate crystals by their composition (Pathak et al., 2020; Karpovich et al., 2024) and lattice parameters (AI4Science et al., 2023), which foregoes explicit 3D generation.

Controllable generation and guidance. To enhance the controllability and quality of diffusion models in tasks such as text-to-image generation, many strategies have been explored, including fine-tuning based alignment (Ruiz et al., 2023), adapter-based conditional control (Zhang et al., 2023), and inference-time techniques using classifier (Dhariwal & Nichol, 2021) or classifier-free guidance (Ho & Salimans, 2022). In inverse material design, generative models aim to discover novel and stable materials with desired properties. These property values can be obtained through methods such as DFT calculations, prediction models, or wet-lab experiments. Jiao et al. (2023) trained a time-dependent property predictor to guide the generation process of diffusion models for single property optimization of crystals. Moreover, Zeni et al. (2025) employed a classifier-free guidance method to fine-tune unconditional diffusion models with additional adapter modules of property information, enabling material generation for single or multiple target properties. These conditional generation methods all require sufficient and diverse labeled data for the target properties. However, many material properties could be computationally expensive to obtain or lack labeled datasets.

RL fine-tuning of diffusion models. Recently, some studies have explored using reinforcement learning (RL) to optimize diffusion models for controllable generation aligned with downstream objectives. Fan & Lee (2023) introduced a method to improve pre-trained diffusion models by integrating policy gradient and GAN training. They used policy gradient with reward signals from the discriminator to update the diffusion model and improve data distribution matching. Fan et al. (2023) proposed DPOK method to better align text-to-image diffusion models to human preferences using a policy gradient algorithm with Kullback–Leibler (KL) regularization. Similarly, Black et al. (2023) designed a policy gradient-based RL approach named DDPO for fine-tuning text-to-image diffusion models, which can generalize to unseen prompts and adapt to multiple reward functions. In the context of inverse inorganic materials design, Karpovich et al. (2024) investigated policy- and value-based RL methods to optimize for single and multiple property constraints. The formulations tested were: (1) a stack recurrent neural network (RNN) based on Popova et al. (2018) where the vocabulary consists of elements and their coefficients and (2) a conditional VAE model based on Pathak et al. (2020) where materials are represented by concatenated one-hot vectors representing the elements. Our work differs in several important ways: (1) We directly model the 3D geometry of materials (by diffusing on the atom types, atom coordinates and lattice) which is important as many properties are geometry-dependent. (2) We use policy-based RL and explicitly demonstrate factors that improve sample efficiency which have been seldom discussed in existing works, and the trade-off on diversity. (3) We re-formulate diversity filters (Blaschke et al., 2020) originally proposed for generative drug design which penalizes repeated generation of scaffolds. By penalizing crystal-intrinsic properties, e.g., elemental composition, we show that this is an effective approach to generating diverse samples.

3 PRELIMINARIES

3.1 REPRESENTATION OF CRYSTAL STRUCTURES

The periodic structure of crystals arises from the repeating arrangement of atoms in 3D space, and the simplest repeating unit is defined as the unit cell. A unit cell with N atoms can be described by $\mathcal{M} = (\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{L})$, where $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_N] \in \mathbb{R}^{h \times N}$ represents the one-hot encoding of atom types, $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N] \in \mathbb{R}^{3 \times N}$ symbolizes atoms' Cartesian coordinates, and $\boldsymbol{L} = [\boldsymbol{l}_1, \boldsymbol{l}_2, \boldsymbol{l}_3] \in \mathbb{R}^{3 \times 3}$ expresses the crystal lattice matrix. The volume of a unit cell $V = |\det(\boldsymbol{L})|$ must be non-zero, meaning that $\boldsymbol{L}$ is invertible. Based on periodic boundary conditions, the atomic positions within the unit cell can also be described using fractional coordinates $\boldsymbol{F} = \boldsymbol{L}^{-1}\boldsymbol{X} = [\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_N] \in [0, 1)^{3 \times N}$, which are widely used in crystallography and crystal generation. Thus, the infinite crystal structure can be represented as

$$\left\{(\boldsymbol{a}_i', \boldsymbol{f}_i') \mid \boldsymbol{a}_i' = \boldsymbol{a}_i, \boldsymbol{f}_i' = \boldsymbol{f}_i + \boldsymbol{L}\boldsymbol{k}\mathbf{1}_N, \forall \boldsymbol{k} \in \mathbb{Z}^3\right\} \tag{1}$$

where elements of $\boldsymbol{k}$ express integer translations of the lattice and $\mathbf{1}$ is a $1 \times n$ matrix of ones to emulate broadcasting.

3.2 EQUIVARIANT DIFFUSION MODELS

A function f is considered to be equivariant to the action of a group G if $f(g \cdot s) = g \cdot f(s)$ for any $s \in \mathcal{S}$ and $g \in G$, where $\cdot$ indicates group action in the relevant space. In this work, we consider the E(3) equivariance containing translations, rotations and reflections. The equivariant diffusion models of crystal generation involve two Markov chains, a forward noising process on atom types, atom positions and lattice matrix, and a reverse denoising process learned by an equivariant graph neural network.

Diffusion on lattice $\boldsymbol{L}$ The diffusion on the continuous variable $\boldsymbol{L}$ is based on Denoising Diffusion Probabilistic Model (DDPM) (Ho et al., 2020). Specifically, in the forward process, Gaussian noises are gradually added to $\boldsymbol{L}$ according to a variance schedule $\beta_1, \ldots, \beta_T$:

$$\begin{aligned} q\left(\boldsymbol{L}_{1:T} \mid \boldsymbol{L}_0\right) &= \prod_{t=1}^{T} q\left(\boldsymbol{L}_t \mid \boldsymbol{L}_{t-1}\right), \\ q\left(\boldsymbol{L}_t \mid \boldsymbol{L}_{t-1}\right) &= \mathcal{N}\left(\boldsymbol{L}_t \mid \sqrt{1-\beta_t}\boldsymbol{L}_{t-1}, \beta_t \boldsymbol{I}\right), \end{aligned} \tag{2}$$

which can be expressed as the probability conditional on the initial state:

$$q\left(\boldsymbol{L}_t \mid \boldsymbol{L}_0\right)=\mathcal{N}\left(\boldsymbol{L}_t \mid \sqrt{\bar{\alpha}_t} \boldsymbol{L}_0,\left(1-\bar{\alpha}_t\right) \boldsymbol{I}\right), \tag{3}$$

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

The reverse process is defined by:

$$\begin{aligned} p_\theta\left(\boldsymbol{L}_{0: T}\right) & =p\left(\boldsymbol{L}_T\right) \prod_{t=1}^T p_\theta\left(\boldsymbol{L}_{t-1} \mid \boldsymbol{L}_t\right), \\ p_\theta\left(\boldsymbol{L}_{t-1} \mid \boldsymbol{L}_t\right) & =\mathcal{N}\left(\boldsymbol{L}_{t-1} \mid \boldsymbol{\mu}_{\theta, \boldsymbol{L}}\left(\mathcal{M}_t, t\right), \sigma_t^2 \boldsymbol{I}\right), \end{aligned} \tag{4}$$

where $\boldsymbol{\mu}_{\theta, \boldsymbol{L}}\left(\mathcal{M}_t, t\right)=\frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{L}_t-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \hat{\boldsymbol{\epsilon}}_{\theta, \boldsymbol{L}}\left(\mathcal{M}_t, t\right)\right)$ and $p\left(\boldsymbol{L}_T\right)=\mathcal{N}(0, \boldsymbol{I})$. The denoising term $\hat{\boldsymbol{\epsilon}}_{\theta, \boldsymbol{L}}\left(\mathcal{M}_t, t\right) \in \mathbb{R}^{3 \times 3}$ is predicted by the equivariant graph neural network $\theta\left(\mathcal{M}_t, t\right)=\theta\left(\boldsymbol{L}_t, \boldsymbol{F}_t, \boldsymbol{A}_t, t\right)$.

For training the denoising model θ, let $\boldsymbol{L}_t=\sqrt{\bar{\alpha}_t} \boldsymbol{L}_0+\sqrt{1-\bar{\alpha}_t} \boldsymbol{\epsilon}_{\boldsymbol{L}}$ and $\boldsymbol{\epsilon}_{\boldsymbol{L}} \sim \mathcal{N}(0, \boldsymbol{I})$ according to Eq. (3). The training objective is denoted as the ℓ_2 loss between $\boldsymbol{\epsilon}_{\boldsymbol{L}}$ and $\hat{\boldsymbol{\epsilon}}_{\theta, \boldsymbol{L}}$:

$$\mathcal{L}_{\boldsymbol{L}}=\mathbb{E}_{t \sim \mathcal{U}(1, T)}\left[\left\|\boldsymbol{\epsilon}_{\boldsymbol{L}}-\hat{\boldsymbol{\epsilon}}_{\theta, \boldsymbol{L}}\left(\mathcal{M}_t, t\right)\right\|^2\right]. \tag{5}$$

Diffusion on atom types $\boldsymbol{A}$ The discrete atom types $\boldsymbol{A}$ can be simply considered as continuous variables in real space $\mathbb{R}^{h \times N}$, facilitating the DDPM-based approach for diffusion on atom types, as also shown in (Hoogeboom et al., 2022). Similar to diffusion on $\boldsymbol{L}$ (Eq. 2-5), the forward process of $\boldsymbol{A}$ is denoted as

$$q\left(\boldsymbol{A}_t \mid \boldsymbol{A}_0\right)=\mathcal{N}\left(\boldsymbol{A}_t \mid \sqrt{\bar{\alpha}_t} \boldsymbol{A}_0,\left(1-\bar{\alpha}_t\right) \boldsymbol{I}\right), \tag{6}$$

the reverse process is expressed as

$$p_\theta\left(\boldsymbol{A}_{t-1} \mid \boldsymbol{A}_t\right)=\mathcal{N}\left(\boldsymbol{A}_{t-1} \mid \boldsymbol{\mu}_{\theta, \boldsymbol{A}}\left(\mathcal{M}_t, t\right), \sigma_t^2 \boldsymbol{I}\right), \tag{7}$$

and the training objective for diffusion on $\boldsymbol{A}$ is

$$\mathcal{L}_{\boldsymbol{A}}=\mathbb{E}_{t \sim \mathcal{U}(1, T)}\left[\left\|\boldsymbol{\epsilon}_{\boldsymbol{A}}-\hat{\boldsymbol{\epsilon}}_{\theta, \boldsymbol{A}}\left(\mathcal{M}_t, t\right)\right\|^2\right]. \tag{8}$$

Diffusion on atom positions $\boldsymbol{F}$ As the domain of fractional coordinates $[0,1)^{3 \times N}$ forms a quotient space $\mathbb{R}^{3 \times N} / \mathbb{Z}^{3 \times N}$, the score matching method (Song et al., 2021) with wrapped normal distribution (Bortoli et al., 2022) is used to achieve diffusion on $\boldsymbol{F}$ (Jiao et al., 2023). The forward process is implemented by wrapped normal distribution to maintain periodic translation invariance according to:

$$q\left(\boldsymbol{F}_t \mid \boldsymbol{F}_0\right)=\mathcal{N}_W\left(\boldsymbol{F}_t \mid \boldsymbol{F}_0, \sigma_t^2 \boldsymbol{I}\right), \quad \boldsymbol{F}_t=w\left(\boldsymbol{F}_0+\sigma_t \boldsymbol{\epsilon}_{\boldsymbol{F}}\right), \tag{9}$$

where $\boldsymbol{\epsilon}_{\boldsymbol{F}} \sim \mathcal{N}(0, \boldsymbol{I})$ and $w(\cdot)$ retains the fractional part of the input. The noise scale σ_t obeys the exponential scheduler: $\sigma_0=0$ and $\sigma_t=\sigma_1\left(\frac{\sigma_T}{\sigma_1}\right)^{\frac{t-1}{T-1}}$, if $t>0$.

For the reverse process, $\boldsymbol{F}_T \sim \mathcal{U}(0,1)$ and $\boldsymbol{F}_0$ are generated using a two-step predictor-corrector sampler method (Song et al., 2021; Jiao et al., 2023) with the denoising term $\hat{\boldsymbol{\epsilon}}_{\theta, \boldsymbol{F}}\left(\mathcal{M}_t, t\right) \in \mathbb{R}^{3 \times N}$:

$$p_\theta\left(\boldsymbol{F}_{t-1} \mid \mathcal{M}_t\right)=p_P\left(\boldsymbol{F}_{t-\frac{1}{2}} \middle| \boldsymbol{L}_t, \boldsymbol{F}_t, \boldsymbol{A}_t\right) p_C\left(\boldsymbol{F}_{t-1} \mid \boldsymbol{L}_{t-1}, \boldsymbol{F}_{t-\frac{1}{2}}, \boldsymbol{A}_{t-1}\right), \tag{10}$$

where p_P, p_C are the transitions of the predictor and corrector, and more details are described in Appendix **??**.

The training objective from score matching of $\boldsymbol{F}$ is

$$\mathcal{L}_{\boldsymbol{F}}=\mathbb{E}_{t \sim \mathcal{U}(1, T)}\left[\lambda_t\left\|\nabla \log q\left(\boldsymbol{F}_t \mid \boldsymbol{F}_0\right)-\hat{\boldsymbol{\epsilon}}_{\theta, \boldsymbol{F}}\left(\mathcal{M}_t, t\right)\right\|^2\right] \tag{11}$$

where $\lambda_t=\mathbb{E}_{\boldsymbol{F}_t}^{-1}\left[\left\|\nabla \log q\left(\boldsymbol{F}_t \mid \boldsymbol{F}_0\right)\right\|^2\right]$ is calculated by Monte-Carlo sampling (Appendix **??**).

3.3 MARKOV DECISION PROCESSES AND REINFORCEMENT LEARNING

A Markov decision process (MDP) formalizes sequential decision-making problems. It can be characterized by a tuple (S, A, ρ_0, P, R), where S denotes the state space, A represents the action space, ρ_0 is the initial state distribution, P specifies the transition kernel, and R defines the reward function. In every timestep t, the agent observes a state $s_t \in S$, selects an action $a_t \in A$, receives a reward $R(s_t, a_t)$, and transforms into a subsequent state $s_{t+1} \sim P(s_{t+1}|s_t, a_t)$. The agent's behavior is determined by its policy $\pi(a|s)$. As the agent interacts with the MDP, it generates trajectories of states and actions $\tau = (s_0, a_0, s_1, a_1, \ldots, s_T, a_T)$. The goal of reinforcement learning (RL) is to optimize the agent's policy π to maximize the expected cumulative reward $J_{\mathrm{RL}}(\pi)$ over sampled trajectories:

$$\mathcal{J}_{\mathrm{RL}}(\pi) = \mathbb{E}_{\tau \sim p(\tau|\pi)} \left[\sum_{t=0}^{T} R\left(s_t, a_t\right) \right] \tag{12}$$

4 REINFORCEMENT LEARNING FOR CRYSTAL DIFFUSION MODELS

This section describes our online on-policy RL algorithms to formulate the denoising process as a MDP and optimize diffusion models for crystal generation with target properties.

Given a crystal diffusion model $p_\theta(\mathcal{M}_{0:T})$, parameterized by θ and a reward function $r(\mathcal{M}_0)$ involving single or multiple target crystal properties, the denoising process can be reframed as a T-step MDP:

$$\begin{gathered} s_t = \mathcal{M}_{T-t}, \quad a_t = \mathcal{M}_{T-t-1}, \\ \rho_0\left(s_0\right) = (\mathcal{N}(0, \boldsymbol{I}), \mathcal{U}(0,1)), \quad P\left(s_{t+1} \mid s_t, a_t\right) = \delta_{a_t}, \\ \pi\left(a_t \mid s_t\right) = p_\theta\left(\mathcal{M}_{T-t-1} \mid \mathcal{M}_{T-t}\right), \\ R\left(s_t, a_t\right) = \begin{cases} r\left(s_{t+1}\right) = r\left(\mathcal{M}_0\right) & \text{if } t = T-1, \\ 0 & \text{otherwise} \end{cases} \end{gathered} \tag{13}$$

where δ_y is the Dirac delta distribution with nonzero density only at y. The initial state s_0 of a trajectory is sampled by $\boldsymbol{L}_T, \boldsymbol{A}_T \sim \mathcal{N}(0, \boldsymbol{I})$ and $\boldsymbol{F}_T \sim \mathcal{U}(0,1)$, similar to the first state $\mathcal{M}_T$ of the denoising generation. The cumulative reward of every trajectory is equal to $r\left(\mathcal{M}_0\right)$, because all intermediate rewards are 0, as only the final state $\mathcal{M}_0$ of the denoising process is meaningful for computing crystal properties and rewards. Thus, a common goal in RL fine-tuning of diffusion models is to maximize the expected reward of the generated crystals:

$$\mathcal{J}_{\mathrm{RL}}(\theta) = \mathbb{E}_{p_\theta(\mathcal{M}_0)}\left[r\left(\mathcal{M}_0\right)\right] \tag{14}$$

Based on the likelihoods and likelihood gradients, the gradient of RL objective is

$$\nabla_\theta \mathcal{J}_{\mathrm{RL}} = \mathbb{E}_{p_\theta(\mathcal{M}_{0:T})} \left[r\left(\mathcal{M}_0\right) \sum_{t=1}^{T} \nabla_\theta \log p_\theta\left(\mathcal{M}_{t-1} \mid \mathcal{M}_t\right) \right] \tag{15}$$

The risk of fine-tuning solely based on rewards related to target properties is that the diffusion model may overfit to the rewards and move too far away from the initial state (pre-trained model). To retain the broad material knowledge that the diffusion model has learned from the pre-training dataset for generating reasonable and valid crystal structures, we add the reward-weighted KL between the pre-trained and current fine-tuned models as a regularizer to the objective function according to:

$$\mathbb{E}_{p_\theta(\mathcal{M}_{0:T})} \left[\left(\lambda - r\left(\mathcal{M}_0\right)\right) \sum_{t=1}^{T} \mathrm{KL}\left(p_\theta\left(\mathcal{M}_{t-1} \mid \mathcal{M}_t\right) \| p_{\mathrm{pre}}\left(\mathcal{M}_{t-1} \mid \mathcal{M}_t\right)\right) \right], \tag{16}$$

where λ is a constant slightly larger than the maximum reward. The reward weight allows the current diffusion model to appropriately move away from the initial state (pre-trained model), thereby encouraging the model to shift its distribution to higher reward regions. And the final loss function of RL fine-tuning is:

$$L(\theta) = -\alpha r\left(\mathcal{M}_0\right) \sum_{t=1}^{T} \log p_\theta\left(\mathcal{M}_{t-1} \mid \mathcal{M}_t\right) + \beta(\lambda - r\left(\mathcal{M}_0\right)) \sum_{t=1}^{T} \mathrm{KL}\left(p_\theta\left(\mathcal{M}_{t-1} \mid \mathcal{M}_t\right) \| p_{\mathrm{pre}}\left(\mathcal{M}_{t-1} \mid \mathcal{M}_t\right)\right) \tag{17}$$

We also add experience replay (Lin, 1992) to MatInvent. Experience replay is a technique used in RL to improve the stability and efficiency of learning by storing past high-reward crystals and reusing them during training. It breaks the correlation between consecutive experiences by sampling from a buffer of previous experiences (called the replay buffer) rather than relying only on the most recent experience.

Diversity filter (DF) We reformulate DFs originally introduced by Blaschke et al. (2020). Here, we linearly penalize non-unique crystal compositions based on the number of previous occurrences, which acts as a more lenient version of the unique DF, i.e., directly truncate the reward to 0. The score is transformed according to the number of previous occurrences (Occ) beyond an allowed tolerance (Tol) until a hard threshold is reached, referred to as the buffer (Buff):

$$\text{Filtered reward } = \begin{cases} r\left(\mathcal{M}_0\right) \times \frac{\text{Occ} - \text{Tol}}{\text{Buff} - \text{Tol}} & \text{if Tol} < \text{Occ} < \text{ Buff} \\ r\left(\mathcal{M}_0\right) & \text{if } \quad \text{Occ } \leq \text{ Tol} \\ 0 & \text{if } \quad \text{Occ } \geq \text{ Buff} \end{cases} \tag{18}$$

5 Experiments

5.1 Experimental Setup

Our crystal diffusion model has been pre-trained on the MP-20 dataset (Jain et al., 2013) and used as the prior model in all experiments. MP-20 dataset (Jain et al., 2013) extracts 45,231 stable inorganic materials from the Materials Project (Jain et al., 2013), encompassing most experimentally synthesized materials with up to 20 atoms per unit cell. In each task, the target properties of the generated crystals are computed using the pre-trained ALIGNN prediction models (Choudhary & DeCost, 2021) or PyMatGen package (Ong et al., 2013), and the final reward values are scaled to the range of 0 to 1 for stable RL fine-tuning. The sampling size of each loop is set to 32 for all experiments, and all results were obtained from 5 independent replicate experiments.

5.2 Metrics

Inspired by Guo & Schwaller (2024a), we define two metrics to evaluate sample efficiency and the diversity of samples of RL methods for goal-directed crystal generation:

Burden In an RL task, c property calculations are spent to generate n unique and valid candidate crystals above the reward threshold ξ. The metric is defined as:

$$\text{Burden} = \frac{c}{n}, \;\; n = \sum_{\mathcal{M} \in G} \mathbb{I}[R(\mathcal{M}) > \xi], \tag{19}$$

where $\mathcal{M} \in G$ are the crystals in the generated set G, and $\mathbb{I}$ is the indicator function which returns 1 if the reward $R(\mathcal{M})$ is above the threshold ξ. It directly reflects sample efficiency, as one is *always* interested in generating crystals with the desired properties. It becomes particularly important when using costly calculations for crystal property assessment. In all experiments, we measure the efficiency to generate 100 unique crystals satisfying the target property thresholds, i.e., n is set to 100.

Diversity ratio During RL fine-tuning, the generative model tends to produce crystals in specific regions, leading to reduced sample diversity. This metric is defined as the ratio between the number of unique and valid crystals generated (u) and the property calculation budget (b):

$$\text{Div. Ratio} = \frac{u}{b}, \tag{20}$$

where b is set to 3000 in all experiments. To sustain the generation of diverse, high reward samples, we use our implementation of diversity filters, inspired by Blaschke et al. (2020).

5.3 BASELINES

Hill climbing Hill climbing is an iterative optimization algorithm used to find a local maximum (or minimum) of a function. In goal-directed crystal generation, the current diffusion model (agent) first samples a batch of crystals at each iteration, and then the diffusion model is fine-tuned using only the top k crystals ranked by the final rewards from the batch. In this work, we use top ratio $k = 0.4$ and four fine-tuning epochs in each iteration of all hill climbing experiments.

REINFORCE REINFORCE (Williams, 1992) is a Monte Carlo policy gradient method used in RL. It's a type of algorithm that directly learns the optimal policy by adjusting parameters based on rewards. We implemented this method for goal-directed crystal generation.

Baseline The frozen pre-trained diffusion model continuously samples new crystals without fine-tuning, and then uses property calculations to discover candidate crystals required for each task.

5.4 SINGLE PROPERTY OPTIMIZATION

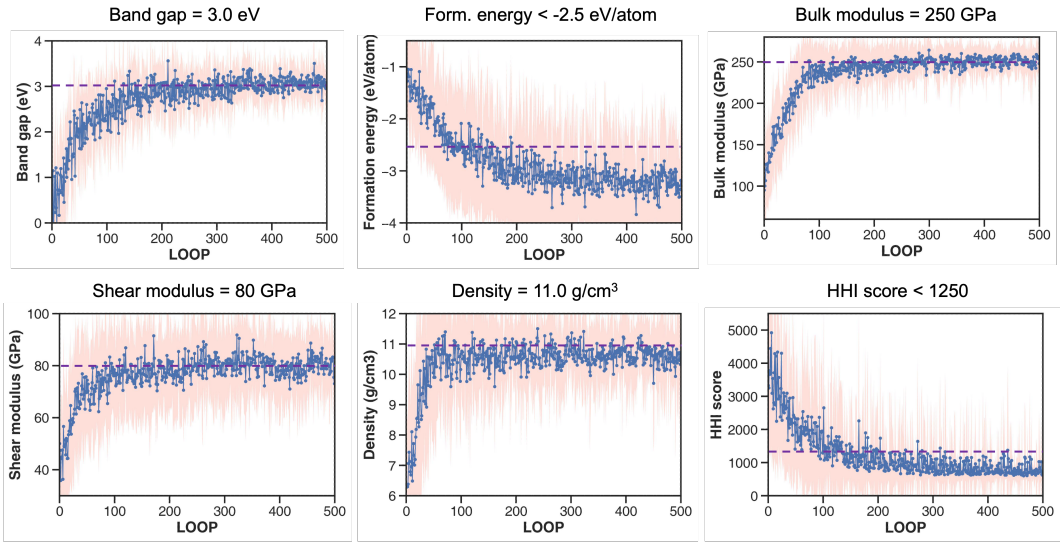

Figure 2: The RL curves of our MatInvent method in SPO tasks.

We define six single-property optimization (SPO) tasks and set boundaries for identifying candidate crystals to calculate reward thresholds and evaluation metrics:

- Band gap of 3 ± 0.25 eV;
- Formation energy (form_e) below -2.5 eV/atom;
- Bulk modulus of 250 ± 20 GPa;
- Shear modulus of 80 ± 5 GPa;
- Density of 11 ± 0.5 g/cm^3;
- Herfindahl–Hirschman index (HHI) score below 1250.

These tasks encompass the electronic, stability, mechanical, and physical properties of materials. As shown in the Figure 2, our MatInvent method can iteratively optimize the crystal diffusion model during the RL process and achieve goal-oriented generation in all SPO tasks. As shown in Table 1, our MatInvent method outperforms other approaches in sample efficiency and generates more desirable crystals under the budget. In most tasks, the Burden metric of MatInvent is less than 10, indicating that fewer than 1,000 property evaluations are required to obtain 100 valid and unique target crystals. As shown in Table 2, MatInvent demonstrates a high diversity ratio comparable to the baseline across all tasks and significantly outperforms the hill climbing and REINFORCE methods,

benefiting from KL regularization and diversity filters. This result shows that MatInvent can generate high reward samples while *maintaining* diversity. It is worth noting that the hill climbing method can occasionally exhibit excellent performance but is unstable (with large standard deviations). This is not unexpected, as hill climbing is a greedy algorithm that can get stuck in local optima without the ability to backtrack, leading to a continuous reduction in the diversity of the generated crystals Guo & Schwaller (2024b). Moreover, in each iteration of hill climbing, the diffusion model is fine-tuned on the generated crystal structure without DFT optimization, which could cause the diffusion model to move too far away from the initial pre-trained distribution, resulting in the generation of numerous invalid structures or even model collapse.

Table 1: Burden (↓) results of different methods on SPO tasks. Best results are bolded. MatInvent consistently outperforms the baselines across all tasks.

Tasks	Baseline	Hill climbing	REINFORCE	MatInvent (Our)
Band gap	75.2 ± 10.6	48.7 ± 11.2	33.2 ± 5.6	$\mathbf{14.4 \pm 1.3}$
form_e	15.7 ± 2.8	8.1 ± 3.4	11.3 ± 4.8	$\mathbf{4.3 \pm 2.1}$
Bulk modulus	168.5 ± 18.7	51.9 ± 31.0	36.7 ± 9.8	$\mathbf{11.5 \pm 1.5}$
Shear modulus	67.8 ± 9.8	12.9 ± 3.3	27.6 ± 4.2	$\mathbf{9.4 \pm 1.1}$
Density	39.1 ± 6.2	11.2 ± 2.4	13.6 ± 1.8	$\mathbf{8.7 \pm 0.9}$
HHI	38.4 ± 5.4	27.5 ± 5.7	18.1 ± 2.3	$\mathbf{7.7 \pm 0.8}$

Table 2: Diversity ratio (↑) results of different methods on SPO tasks.

Tasks	Baseline	Hill climbing	REINFORCE	MatInvent (Our)
Band gap	0.965 ± 0.002	0.767 ± 0.176	0.893 ± 0.076	0.970 ± 0.005
form_e	0.964 ± 0.002	0.807 ± 0.159	0.871 ± 0.083	0.969 ± 0.009
Bulk modulus	0.966 ± 0.003	0.704 ± 0.280	0.856 ± 0.106	0.961 ± 0.015
Shear modulus	0.966 ± 0.004	0.749 ± 0.188	0.877 ± 0.086	0.965 ± 0.013
Density	0.963 ± 0.002	0.752 ± 0.303	0.821 ± 0.116	0.957 ± 0.015
HHI	0.961 ± 0.003	0.761 ± 0.233	0.884 ± 0.067	0.958 ± 0.008

5.5 Multiple Property Optimization

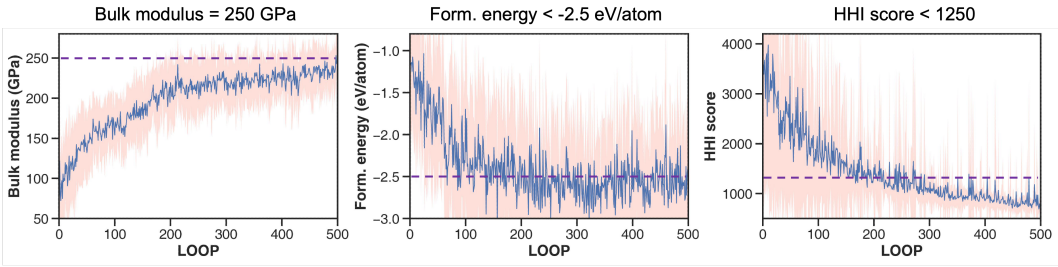

Figure 3: The RL curves of our MatInvent method in the T4 MPO task.

We define four multiple property optimization (MPO) tasks and set boundaries for identifying candidate crystals:

- (T1) band gap of 3 ± 0.25 eV and formation energy below -2.5 eV/atom;
- (T2) band gap of 3 ± 0.25 eV and bulk modulus of 250 ± 20 GPa;
- (T3) bulk modulus of 250 ± 20 GPa and HHI score below 1250;
- (T4) bulk modulus of 250 ± 20 GPa, formation energy below -2.5 eV/atom, and HHI score below 1250.

As shown in Figure 3, our MatInvent method can iteratively optimize the crystal diffusion model and achieve simultaneous optimization of each property in the MPO task. As shown in Table 3, the MPO task is more challenging than the SPO task and elicits a higher burden. It is difficult for diffusion models pre-trained on the MP-20 dataset to generate crystals with multiple target properties, even with the expense of extensive property calculations. MatInvent shows lower burden and higher sampling efficiency than other methods on all MPO tasks. Moreover, MatInvent maintains a high diversity ratio comparable to the baseline, and outperforming other approaches (Table 4). All results indicate that our MatInvent is a high-efficiency RL algorithm for crystal diffusion models in single/multi-property optimization tasks, without the need for a large amount of pre-existing labeled data on target properties.

Table 3: Burden (↓) results of different methods on MPO tasks. The best results are bolded.

Tasks	Baseline	Hill climbing	REINFORCE	MatInvent (Our)
T1	83.2 ± 11.1	41.6 ± 13.2	43.7 ± 8.7	$\mathbf{15.8 \pm 1.8}$
T2	> 1000	56.2 ± 23.4	83.0 ± 12.4	$\mathbf{51.9 \pm 6.1}$
T3	> 1000	48.9 ± 21.8	72.6 ± 15.8	$\mathbf{44.1 \pm 3.7}$
T4	> 1000	67.1 ± 39.7	88.4 ± 14.1	$\mathbf{47.4 \pm 4.2}$

Table 4: Diversity ratio (↑) results of different methods on MPO tasks.

Tasks	Baseline	Hill climbing	REINFORCE	MatInvent (Our)
T1	0.966 ± 0.002	0.836 ± 0.133	0.920 ± 0.056	0.959 ± 0.004
T2	0.965 ± 0.003	0.841 ± 0.089	0.889 ± 0.074	0.947 ± 0.006
T3	0.963 ± 0.003	0.811 ± 0.107	0.914 ± 0.064	0.966 ± 0.007
T4	0.964 ± 0.002	0.897 ± 0.096	0.908 ± 0.077	0.951 ± 0.009

6 CONCLUSION

In this work, we introduced MatInvent, a reinforcement learning framework for optimizing diffusion models to generate crystals with desired properties. By reformulating equivariant denoising as a multi-step decision process and incorporating policy optimization with reward-weighted KL regularization, our approach enables efficient goal-directed crystal generation. The experimental results demonstrate that MatInvent significantly outperforms baseline methods across both single and multiple property optimization tasks, achieving up to 10x improvement in sample efficiency while maintaining high sample diversity. This improved performance can be attributed to the combination of experience replay and diversity filters, which help balance exploration and exploitation during the optimization process.

Our work bridges an important gap between recent advances in crystal diffusion models and the practical needs of materials discovery. While existing diffusion models have shown promise in generating stable crystal structures, they often struggle with targeted generation of materials having specific desired properties. MatInvent addresses this limitation by providing a flexible framework that can be applied to both single and multiple property optimization scenarios, even in cases where labeled training data is scarce. The framework's ability to generate out-of-distribution samples while maintaining physical validity suggests its potential for discovering novel materials that lie beyond the scope of the pre-training dataset.

REFERENCES

Mila AI4Science, Alex Hernandez-Garcia, Alexandre Duval, Alexandra Volokhova, Yoshua Bengio, Divya Sharma, Pierre Luc Carrier, Yasmine Benabed, Michał Koziarski, and Victor Schmidt. Crystal-gfn: sampling crystals with desirable properties and constraints. *arXiv preprint arXiv:2310.04925*, 2023.

Fayez Nasir Al-Rowaili, Umer Zahid, Sagheer Onaizi, Mazen Khaled, Aqil Jamal, and Eid M AL-Mutairi. A review for metal-organic frameworks (mofs) utilization in capture and conversion of carbon dioxide into valuable products. *Journal of CO2 Utilization*, 53:101715, 2021.

Zahed Allahyari and Artem R Oganov. Coevolutionary search for optimal materials in the space of all possible compounds. *npj Computational Materials*, 6(1):55, 2020.

Luis M Antunes, Keith T Butler, and Ricardo Grau-Crespo. Crystal structure generation with autoregressive large language modeling. *Nature Communications*, 15(1):1–16, 2024.

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.

Emmanuel Bengio, Moksh Jain, Maksym Korablyov, Doina Precup, and Yoshua Bengio. Flow network based generative models for non-iterative diverse candidate generation. *Advances in Neural Information Processing Systems*, 34:27381–27394, 2021.

Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. *arXiv preprint arXiv:2305.13301*, 2023.

Thomas Blaschke, Ola Engkvist, Jürgen Bajorath, and Hongming Chen. Memory-assisted reinforcement learning for diverse molecular de novo design. *Journal of cheminformatics*, 12(1):68, 2020.

J Terence Blaskovits, Ruben Laplaza, Sergi Vela, and Clémence Corminboeuf. Data-driven discovery of organic electronic materials enabled by hybrid top-down/bottom-up design. *Advanced Materials*, 36(2):2305602, 2024.

Valentin De Bortoli, Emile Mathieu, MJ Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modelling. In *Conference on Neural Information Processing Systems*, 2022.

Zhendong Cao, Xiaoshan Luo, Jian Lv, and Lei Wang. Space group informed transformer for crystalline materials generation. *arXiv preprint arXiv:2403.15734*, 2024.

Lowik Chanussot, Abhishek Das, Siddharth Goyal, Thibaut Lavril, Muhammed Shuaibi, Morgane Riviere, Kevin Tran, Javier Heras-Domingo, Caleb Ho, Weihua Hu, et al. Open catalyst 2020 (oc20) dataset and community challenges. *Acs Catalysis*, 11(10):6059–6072, 2021.

Kamal Choudhary and Brian DeCost. Atomistic line graph neural network for improved materials property predictions. *npj Computational Materials*, 7(1):185, 2021.

Callum J Court, Batuhan Yildirim, Apoorv Jain, and Jacqueline M Cole. 3-d inorganic crystal structure generation and property prediction via representation learning. *Journal of Chemical Information and Modeling*, 60(10):4518–4535, 2020.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. In *Conference on Neural Information Processing Systems*, 2021.

Yuanqi Du, Arian R Jamasb, Jeff Guo, Tianfan Fu, Charles Harris, Yingheng Wang, Chenru Duan, Pietro Liò, Philippe Schwaller, and Tom L Blundell. Machine learning-aided generative molecular design. *Nature Machine Intelligence*, pp. 1–16, 2024.

Ying Fan and Kangwook Lee. Optimizing ddpm sampling with shortcut fine-tuning. In *International Conference on Machine Learning*, 2023.

Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. Reinforcement learning for fine-tuning text-to-image diffusion models. In *Conference on Neural Information Processing Systems*, 2023.

Daniel Flam-Shepherd and Alán Aspuru-Guzik. Language models can generate molecules, materials, and protein binding sites directly in three dimensions as xyz, cif, and pdb files. *arXiv preprint arXiv:2305.05708*, 2023.

Simone Gallarati, Puck van Gerwen, Ruben Laplaza, Sergi Vela, Alberto Fabrizio, and Clemence Corminboeuf. Oscar: an extensive repository of chemically and functionally diverse organocatalysts. *Chemical Science*, 13(46):13782–13794, 2022.

Colin W Glass, Artem R Oganov, and Nikolaus Hansen. Uspex—evolutionary crystal structure prediction. *Computer physics communications*, 175(11-12):713–720, 2006.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

Nate Gruver, Anuroop Sriram, Andrea Madotto, Andrew Gordon Wilson, C Lawrence Zitnick, and Zachary Ulissi. Fine-tuned language models generate stable inorganic materials as text. In *International Conference on Learning Representations*, 2024.

Jeff Guo and Philippe Schwaller. Beam enumeration: probabilistic explainability for sample efficient self-conditioned molecular design. In *International Conference on Learning Representations*, 2024a.

Jeff Guo and Philippe Schwaller. Saturn: Sample-efficient generative molecular design using memory manipulation. *arXiv preprint arXiv:2405.17066*, 2024b.

Minhi Han, Joonyoung F Joung, Minseok Jeong, Dong Hoon Choi, and Sungnam Park. Generative deep learning-based efficient design of organic molecules with tailored properties. *ACS Central Science*, 2024.

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

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Conference on Neural Information Processing Systems*, 2020.

Emiel Hoogeboom, Vıctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. In *International conference on machine learning*, pp. 8867–8887. PMLR, 2022.

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.

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

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

Christopher Karpovich, Elton Pan, and Elsa A Olivetti. Deep reinforcement learning for inverse inorganic materials design. *npj Computational Materials*, 10(1):287, 2024.

Baekjun Kim, Sangwon Lee, and Jihan Kim. Inverse design of porous materials using artificial neural networks. *Science advances*, 6(1):eaax9324, 2020a.

Sungwon Kim, Juhwan Noh, Geun Ho Gu, Alan Aspuru-Guzik, and Yousung Jung. Generative adversarial networks for crystal structure prediction. *ACS central science*, 6(8):1412–1420, 2020b.

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

Cheng-Han Li and Daniel P Tabor. Generative organic electronic molecular design informed by quantum chemistry. *Chemical Science*, 14(40):11045–11055, 2023.

Long-Ji Lin. Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine learning*, 8:293–321, 1992.

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 *International Conference on Machine Learning*, 2024.

Teng Long, Yixuan Zhang, and Hongbin Zhang. Generative deep learning for the inverse design of materials. *arXiv preprint arXiv:2409.19124*, 2024.

Youzhi Luo, Chengkai Liu, and Shuiwang Ji. Towards symmetry-aware generation of periodic materials. In *Conference on Neural Information Processing Systems*, 2023.

Ruimin Ma, Hanfeng Zhang, and Tengfei Luo. Exploring high thermal conductivity amorphous polymers using reinforcement learning. *ACS Applied Materials & Interfaces*, 14(13):15587–15598, 2022.

Gabriel Marques, Karl Leswing, Tim Robertson, David Giesen, Mathew D Halls, Alexander Goldberg, Kyle Marshall, Joshua Staker, Tsuguo Morisato, Hiroyuki Maeshima, et al. De novo design of molecules with low hole reorganization energy based on a quarter-million molecule dft screen. *The Journal of Physical Chemistry A*, 125(33):7331–7343, 2021.

Nobuyuki N Matsuzawa, Hiroyuki Maeshima, Keisuke Hayashi, Tatsuhito Ando, Mohammad Atif Faiz Afzal, Kyle Marshall, Benjamin J Coscia, Andrea R Browning, Alexander Goldberg, Mathew D Halls, et al. Exploring molecules with low viscosity: Using physics-based simulations and de novo design by applying reinforcement learning. *Chemistry of Materials*, 36(23): 11706–11716, 2024.

Benjamin Kurt Miller, Ricky TQ Chen, Anuroop Sriram, and Brandon M Wood. Flowmm: Generating materials with riemannian flow matching. In *International Conference on Machine Learning*, 2024.

Asma Nouira, Nataliya Sokolovska, and Jean-Claude Crivello. Crystalgan: learning to discover crystallographic structures with generative adversarial networks. *arXiv preprint arXiv:1810.11203*, 2018a.

Asma Nouira, Nataliya Sokolovska, and Jean-Claude Crivello. Crystalgan: learning to discover crystallographic structures with generative adversarial networks. *arXiv preprint arXiv:1810.11203*, 2018b.

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

Yashaswi Pathak, Karandeep Singh Juneja, Girish Varma, Masahiro Ehara, and U Deva Priyakumar. Deep learning enabled inorganic material generator. *Physical Chemistry Chemical Physics*, 22 (46):26935–26943, 2020.

Chris J Pickard and RJ Needs. Ab initio random structure searching. *Journal of Physics: Condensed Matter*, 23(5):053201, 2011.

Mariya Popova, Olexandr Isayev, and Alexander Tropsha. Deep reinforcement learning for de novo drug design. *Science advances*, 4(7):eaap7885, 2018.

Zekun Ren, Siyu Isaac Parker Tian, Juhwan Noh, Felipe Oviedo, Guangzong Xing, Jiali Li, Qiaohao Liang, Ruiming Zhu, Armin G Aberle, Shijing Sun, et al. An invertible crystallographic representation for general inverse design of inorganic crystals with targeted properties. *Matter*, 5(1): 314–335, 2022.

Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22500–22510, 2023.

Vidushi Sharma, Andy Tek, Khanh Nguyen, Max Giammona, Murtaza Zohair, Linda Sundberg, and Young-Hye La. Improving electrolyte performance for target cathode loading using an interpretable data-driven approach. *Cell Reports Physical Science*, 6(1), 2025.

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

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.

Anuroop Sriram, Benjamin Kurt Miller, Ricky TQ Chen, and Brandon M Wood. Flowllm: Flow matching for material generation with large language models as base distributions. In *Conference on Neural Information Processing Systems*, 2024.

Joshua Staker, Kyle Marshall, Karl Leswing, Tim Robertson, Mathew D Halls, Alexander Goldberg, Tsuguo Morisato, Hiroyuki Maeshima, Tatsuhito Ando, Hideyuki Arai, et al. De novo design of molecules with low hole reorganization energy based on a quarter-million molecule dft screen: Part 2. *The Journal of Physical Chemistry A*, 126(34):5837–5852, 2022.

Annika Stuke, Christian Kunkel, Dorothea Golze, Milica Todorović, Johannes T Margraf, Karsten Reuter, Patrick Rinke, and Harald Oberhofer. Atomic structures and orbital energies of 61,489 crystal-forming organic molecules. *Scientific data*, 7(1):58, 2020.

Hai-Chen Wang, Silvana Botti, and Miguel AL Marques. Predicting stable crystalline compounds using chemical similarity. *npj Computational Materials*, 7(1):12, 2021.

Julia Westermayr, Joe Gilkes, Rhyan Barrett, and Reinhard J Maurer. High-throughput property-driven generative design of functional organic molecules. *Nature Computational Science*, 3(2): 139–148, 2023.

Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Mach Learn*, 8(3):229–256, May 1992. ISSN 1573-0565. doi: 10.1007/BF00992696. URL https://doi.org/10.1007/BF00992696.

Peter Wirnsberger, George Papamakarios, Borja Ibarz, Sébastien Racaniere, Andrew J Ballard, Alexander Pritzel, and Charles Blundell. Normalizing flows for atomic solids. *Machine Learning: Science and Technology*, 3(2):025009, 2022.

Hang Xiao, Rong Li, Xiaoyang Shi, Yan Chen, Liangliang Zhu, Xi Chen, and Lei Wang. An invertible, invariant crystal representation for inverse design of solid-state materials using generative deep learning. *Nature Communications*, 14(1):7027, 2023.

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

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

Sherry Yang, KwangHwan Cho, Amil Merchant, Pieter Abbeel, Dale Schuurmans, Igor Mordatch, and Ekin Dogus Cubuk. Scalable diffusion for materials generation. In *International Conference on Learning Representations*, 2024.

Zhenze Yang, Weike Ye, Xiangyun Lei, Daniel Schweigert, Ha-Kyung Kwon, and Arash Khajeh. De novo design of polymer electrolytes with high conductivity using gpt-based and diffusion-based generative models. *arXiv preprint arXiv:2312.06470*, 2023.

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*, pp. 1–3, 2025.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3836–3847, October 2023.

Alex Zunger. Inverse design in search of materials with target functionalities. *Nature Reviews Chemistry*, 2(4):0121, 2018.