Generative Inverse Design of Crystal Structures via Diffusion Models with Transformers

Anonymous authors

Abstract

Recent advances in deep learning have enabled the generation of realistic data by training generative models on large datasets of text, images, and audio. While these models have demonstrated exceptional performance in generating novel and plausible data, it remains an open question whether they can effectively accelerate scientific discovery through the data generation and drive significant advancements across various scientific fields. In particular, the discovery of new inorganic materials poses a critical challenge, both scientifically and for industrial applications. However, unlike textual or image data, materials, or more specifically crystal structures, consist of multiple types of variables - including lattice vectors, atom positions, and atomic species. This complexity in data give rise to a variety of approaches for representing and generating such data. Consequently, the design choices of generative models for crystal structures remain an open question. In this study, we explore a new type of diffusion model for the generative inverse design of crystal structures, with a backbone based on a Transformer architecture. We demonstrate that our models are superior to previous methods in their versatility for generating crystal structures with desired properties. Furthermore, our empirical results suggest that the optimal conditioning methods vary depending on the dataset.

1 Introduction

The advancements in artificial intelligence, particularly in the domains of large language models and generative AI for image and audio synthesis, are having a significant impact on our social lives [OpenAI, 2023; Rombach *et al.*, 2022]. Such advancements in AI are also expected to accelerate research and development in materials science, which could potentially drive scientific discoveries and accelerate the development of materials. The discovery of novel materials, for example catalysis, battery materials, and superconducting materials, holds the potential to enable innovation in a wide range of industries [Toyao *et al.*, 2020; Chen *et al.*, 2020a].

Traditionally, the exploration of materials has required repeated try-and-error, consuming enormous amounts of time and effort. If novel and promising materials could be discovered in-silico, i.e., on computers, the exploration process could be further accelerated. Based on this concept, high-throughput virtual screenings using density functional theory (DFT) simulations or machine learning-based predictive models have been employed [Noh *et al.*, 2020]. However, such screening-based approaches have required enumerating and comprehensively simulating a vast number of candidate materials. If it were possible to selectively enumerate promising materials or directly generate materials with desired properties, the process of materials research and development would be significantly streamlined. To address these challenges, the inverse design of materials using deep generative models has emerged as a potential approach.

Materials, or more specifically crystal structures, consist of multiple types of variables including lattice vectors, atomic coordinates, and atomic species. There are several ways to represent crystal structures in a computational framework, and several approaches can be employed to generate them. Moreover, the inverse design requires generating crystal structures with desired properties rather than just generating them in a random manner, where several strategies can be employed to achieve this [Noh *et al.*, 2019; Xie *et al.*, 2021; Yang *et al.*, 2023; Zeni *et al.*, 2023]. In other words, there are various design choices for constructing the generative models for the inverse design of crystal structures.

Diffusion models are a type of generative model that have exhibited distinguished performance, particularly in the domain of image generation [Ho *et al.*, 2020; Song and Ermon, 2019]. They have also demonstrated its versatility and effectiveness in generating audio waveforms [Chen *et al.*, 2020b; Kong *et al.*, 2021], molecular and protein design [Hoogeboom *et al.*, 2022; Watson *et al.*, 2023], as well as crystal structure generation [Yang *et al.*, 2023; Jiao *et al.*, 2023; Zeni *et al.*, 2023]. In diffusion models, data is generated by progressively denoising an initial random input. For image generation tasks, variants of U-Net [Ronneberger *et al.*, 2015] based on convolutional neural networks (CNNs) has been used as the backbone for the denoising process. Recently, diffusion models replacing the U-Net with a Vision Transformers (ViTs) [Dosovitskiy *et al.*, 2021] have also been proposed, with the aim of improving scalability and enhanc-

ing the quality of generated data [Peebles and Xie, 2022; Hatamizadeh *et al.*, 2023]. This suggests that the examination of the backbone model in diffusion models can have a significant impact on the overall model performance.

In this study, we explore a new type of diffusion model for crystal structure generation, where the backbone is formulated based on a Transformer [Vaswani *et al.*, 2017] architecture. Furthermore, we explore suitable conditioning methods for the high-precision inverse design of crystal structures, and provide solid baselines for future research on generative crystal structure design techniques. Ultimately, we demonstrate that our proposed model is capable of performing inverse design with accuracy equal to or exceeding that of prior methods.

2 Preliminary

Crystal structure is characterized by the periodic arrangement of atoms in three-dimensional space. The crystal structure $\boldsymbol{M}$ can be defined using a repeating unit called unit cell. When a unit cell contains N atoms, the crystal structure $\boldsymbol{M}$ can be represented as $\boldsymbol{M} = (\boldsymbol{L}, \boldsymbol{X}, \boldsymbol{A})$, where $\boldsymbol{L} = [\boldsymbol{l}_1, \boldsymbol{l}_2, \boldsymbol{l}_3] \in \mathbb{R}^{3\times3}$ is the lattice matrix containing the lattice vectors of the unit cell, $\boldsymbol{X} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_N] \in \mathbb{R}^{3\times N}$ is the atomic coordinates in the Cartesian coordinate system, and $\boldsymbol{A} = [a_1, ...a_N] \in \mathbb{Z}^N$ is the corresponding atomic species. $\boldsymbol{f}_i = \boldsymbol{L}^{-1}\boldsymbol{x}_i \in [0, 1)^{3\times1}$ is called the fractional coordinate, which is advantageous for representing atomic positions in crystal structures considering their periodic nature. Therefore, the crystal structure can also be represented as $\boldsymbol{M} = (\boldsymbol{L}, \boldsymbol{F}, \boldsymbol{A})$, where $\boldsymbol{F} = [\boldsymbol{f}_1, ..., \boldsymbol{f}_N] \in [0, 1)^{3\times N}$ denotes the fractional coordinate matrix. A conceptual diagram of a crystal structure represented in two dimensions for simplicity is shown in Figure 1

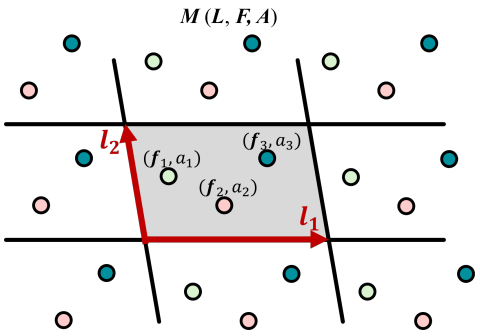

Figure 1: A conceptual diagram of a crystal structure represented in 2D for intuitive understanding. The red arrows represent the lattice vectors of the unit cell, while colored dots represent atoms, with the colors corresponding to different atomic species. The gray region highlights the area of the unit cell.

3 Related Work

3.1 Diffusion Models for Structured Data

Diffusion models are a class of generative models originally proposed for image generation, with two main formulation based on Denoising Diffusion Probabilistic Models (DDPMs) [Ho *et al.*, 2020] and Noise Conditional Score Networks (NCSCs) [Song and Ermon, 2019], where data is generated from the noise with the same dimension. Diffusion models have also been applied to other structured data, such as text [Hoogeboom *et al.*, 2021], point cloud [Luo and Hu, 2021] and graphs [Vignac *et al.*, 2023].

When applying to crystal structure generation, it is necessary to jointly model continuous variables, such as lattice vectors and atomic coordinates, alongside discrete variables, like atom types. Furthermore, to take into account the translational, rotational, and periodic invariance of crystal structures, it is necessary to consider the rotational equivariance of lattice vectors and the periodic invariance of fractional coordinates [Jiao *et al.*, 2023], which requires careful design of the denoising model architecture.

3.2 Transformers for Recognition and Generation

Transformer is a neural network architecture that was originally proposed for natural language processing (NLP) tasks [Vaswani *et al.*, 2017]. The self-attention mechanism in Transformers effectively captures long-range dependencies and efficiently encodes input sequences, leading to its success across various domains beyond NLP.

One prominent example of their application is in the field of computer vision. Vision Transformers (ViTs), which incorporate attention mechanisms to capture dependencies between image patches, have demonstrated superior performance compared to CNN-based models in image recognition tasks [Dosovitskiy *et al.*, 2021]. Due to this advantage, ViTs have recently been increasingly employed as backbone models for diffusion models in image generation tasks [Peebles and Xie, 2022; Gao *et al.*, 2023; Hatamizadeh *et al.*, 2023].

Furthermore, the application of Transformer-based models have advanced in domains related to materials science. For instance, self-attention mechanisms have been utilized to model the complex relationship between atoms within materials. This approach has enabled the effective representation learning of the 3D structures of molecules [Ying *et al.*, 2021] and materials [Yan *et al.*, 2022; Yan *et al.*, 2024; Taniai *et al.*, 2024], facilitating the accurate prediction of materials properties. Our study aims to leverage the power of self-attention mechanisms to model inter-atomic interactions, which is then used as the backbone of a diffusion model for crystal structure generation.

3.3 Generative Models for Crystal Generation

Generative models for crystal structures are essentially models designed to create representations of crystal structures $\boldsymbol{M} = (\boldsymbol{L}, \boldsymbol{F}, \boldsymbol{A})$. In general, these models can be designed based on two aspects: how crystal structures are represented in a computational framework and how those representations are generated. Typical invertible crystal structure representations include 2D arrays containing crystallographic information on $\boldsymbol{M}$ [Ren *et al.*, 2022], voxel images [Hoffmann *et al.*, 2019; Court *et al.*, 2020; Noh *et al.*, 2019], and graphs [Xie *et al.*, 2021; Luo *et al.*, 2023; Zeni *et al.*, 2023]. Regarding the generation methods, generative frameworks originally developed for image generation, such as Variational Autoencoders (VAEs) [Hoffmann *et al.*, 2019; Court *et al.*, 2020; Ren *et al.*, 2022; Xie *et al.*, 2021; Luo *et al.*, 2023; Noh *et al.*,

2019], Generative Adversarial Networks [Nouira *et al.*, 2018; Zhao *et al.*, 2021], and diffusion models [Xie *et al.*, 2021; Luo *et al.*, 2023; Zeni *et al.*, 2023; Yang *et al.*, 2023; Jiao *et al.*, 2023], have been widely adopted.

Many previous studies have focused on generating crystal structures for a limited range of crystal systems, such as cubic crystals, or for systems with a limited number of elements, such as binary or ternary compounds. However, recently, versatile generative models capable of generating diverse and plausible crystal structures across various crystal systems and elemental compositions have been developed. [Xie *et al.*, 2021; Luo *et al.*, 2023; AI4Science *et al.*, 2023; Yang *et al.*, 2023; Jiao *et al.*, 2023]. Xie *et al*. proposed CDVAE, a model that combines VAE and diffusion models, where crystal structures are represented as graphs. They demonstrated that the model is capable of generating diverse and reasonable crystal structures by learning the Cartesian coordinate scores using a graph neural network (GNN) as the backbone. Zeni *et al*. proposed a diffusion model that jointly generates $\boldsymbol{L}$, $\boldsymbol{X}$, and $\boldsymbol{A}$, which uses graphs as a representation of crystal structures, and the backbone model is based on a GNN [Zeni *et al.*, 2023]. A diffusion generative model by Yang *et al*. uses a well designed representation called UniMat, and its backbone is based on a U-Net [Yang *et al.*, 2023].

In our diffusion model, atoms are treated as a point cloud in fractional space, and the backbone model is formulated based on a Transformer.

4 Methodology

In this section, we introduce our proposed model for the generative inverse design of crystal structures. We first introduce the joint diffusion framework in Sec. 4.1. In Sec. 4.2, an overview of the base architecture of the transformer model is provided, and in Sec. 4.3, two approaches of conditional generation for the generative inverse design of crystal structures are described.

4.1 Joint Diffusion Framework

As a general concept, a diffusion model defines two Markov processes: a fixed forward diffusion process that gradually adds noise to the original data $\boldsymbol{M}_0 = (\boldsymbol{L}_0, \boldsymbol{F}_0, \boldsymbol{A}_0)$ over T steps, from $t = 1$ to $t = T$, and a learned generative process that removes the noise from the prior $\boldsymbol{M}_T = (\boldsymbol{L}_T, \boldsymbol{F}_T, \boldsymbol{A}_T)$.

Diffusion on Lattices

For lattice vectors with continuous variables, the forward diffusion process can be defined as follows [Jiao *et al.*, 2023], according to DDPM [Ho *et al.*, 2020]:

$$q(\boldsymbol{L}_t|\boldsymbol{L}_{t-1}) = \mathcal{N}(\boldsymbol{L}_t; \sqrt{1-\beta_t}\boldsymbol{L}_{t-1}, \beta_t\mathbf{I}). \quad (1)$$

Here, $\beta_t \in [0, 1]$ is the predefined noise schedule, and $\mathbf{I}$ is the identity matrix. By applying the Markov property, $\boldsymbol{L}_t$ can be directly derived from $\boldsymbol{M}_0$ as:

$$q(\boldsymbol{L}_t|\boldsymbol{M}_0) = \mathcal{N}(\boldsymbol{L}_t; \sqrt{\overline{\alpha}_t}\boldsymbol{L}_0, (1-\overline{\alpha}_t)\mathbf{I}), \quad (2)$$

where $\overline{\alpha}_t = \prod_{s=1}^{t} \alpha_s$ and $\alpha_t = 1-\beta_t$. By the reparametrization trick, $\boldsymbol{L}_t$ can be written as $\boldsymbol{L}_t = \sqrt{\overline{\alpha}_t}\boldsymbol{L}_0 + \sqrt{1-\overline{\alpha}_t}\boldsymbol{\epsilon}_{\boldsymbol{L}}$, where $\boldsymbol{\epsilon}_{\boldsymbol{L}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

In the backward process, the lattice vectors are represented using a Gaussian distribution $\mathcal{N}(\boldsymbol{L}_{t-1}|\boldsymbol{\mu}_\theta(\boldsymbol{M}_t), \Sigma_t\mathbf{I})$, where $\boldsymbol{\mu}_\theta(\boldsymbol{M}_t) = \frac{1}{\sqrt{\alpha_t}}(\boldsymbol{L}_t - \frac{\beta_t}{\sqrt{1-\overline{\alpha}_t}}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{L}}(\boldsymbol{M}_t, t))$ and $\Sigma_t = \beta_t\frac{1-\overline{\alpha}_{t-1}}{1-\overline{\alpha}_t}$. The neural network is trained to predict $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{L}}$, given $\boldsymbol{M}_t$ and t, with the loss function

$$\mathcal{L}_{\boldsymbol{L}} = \mathbb{E}_{\boldsymbol{\epsilon}_{\boldsymbol{L}}\sim\mathcal{N}(\mathbf{0},\mathbf{I}),t\sim\mathcal{U}(1,T)}[||\boldsymbol{\epsilon}_{\boldsymbol{L}} - \hat{\boldsymbol{\epsilon}}_{\boldsymbol{L}}(\boldsymbol{M}_t, t)||_2^2]. \quad (3)$$

Diffusion on Coordinates

While fractional coordinates $\boldsymbol{F} \in [0, 1)^{3\times N}$ are convenient for handling periodicity, applying DDPM with Gaussian functions is not appropriate. Therefore, score-matching based models [Song and Ermon, 2019; Song *et al.*, 2019; Song and Ermon, 2020] have been adopted together with wrapped normal (WN) distribution [Jiao *et al.*, 2023; Zeni *et al.*, 2023]. In this case, we assume that $\boldsymbol{F}_t$ follows the perturbed distribution with a predefined noise schedule σ_t as follows:

$$q(\boldsymbol{F}_t|\boldsymbol{F}_0) = \mathcal{N}_w(\boldsymbol{F}_t; \boldsymbol{F}_0, \sigma_t^2\mathbf{I}), \quad (4)$$

where $\mathcal{N}_w$ denotes a WN distribution, and σ_t is defined using a hyper parameter σ_T as $\sigma_0 = 0$ and $\sigma_t = \sigma_1(\frac{\sigma_T}{\sigma_1})^{\frac{t-1}{T-1}}$ for $t > 0$.

In the backward process, the output of neural network $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{F}}$ estimates the score of perturbed data distribution, where the neural network is trained with the objective:

$$\mathcal{L}_{\boldsymbol{F}} = \mathbb{E}_{\boldsymbol{F}_t\sim q(\boldsymbol{F}_t|\boldsymbol{F}_0),t\sim\mathcal{U}(1,T)}[\lambda_t||\nabla_{\boldsymbol{F}_t}\log q(\boldsymbol{F}_t|\boldsymbol{F}_0) - \hat{\boldsymbol{\epsilon}}_{\boldsymbol{F}}(\boldsymbol{M}_t, t)||_2^2]. \quad (5)$$

Here, $\lambda_t = \mathbb{E}^{-1}_{\boldsymbol{F}_t}[||\nabla_{\boldsymbol{F}_t}\log q(\boldsymbol{F}_t|\boldsymbol{F}_0)||_2^2]$ is approximated via Monte Carlo Sampling, and the data is generated by the ancestral predictor with the Langevin corrector, as detailed in [Jiao *et al.*, 2023].

Diffusion on Species

There are several possible approaches to modelling the diffusion of atomic species; however, in this study, we consider them as categorical data and apply discrete denoising diffusion probabilistic models (D3PMs) [Austin *et al.*, 2023; Hoogeboom *et al.*, 2021], as employed in the previous works [Guan *et al.*, 2023; Peng *et al.*, 2023; Zeni *et al.*, 2023].

In D3PM, the forward diffusion process is formulated using categorical distribution with probability vector $\boldsymbol{p}$ as:

$$q(\boldsymbol{a}_{i,t}|\boldsymbol{a}_{i,t-1}) = \mathrm{Cat}(\boldsymbol{a}_{i,t}; \boldsymbol{p} = \boldsymbol{a}_{i,t-1}\boldsymbol{Q}_t), \quad (6)$$

where $\boldsymbol{a}_{i,t} \in \{0, 1\}^K$ is the one-hot row-vector representation of atomic species a_i at timestep t, $\boldsymbol{Q}_t$ is the transition matrix, and K is the number of classes. The Markov property allows for directly computing $\boldsymbol{a}_{i,t}$ from $\boldsymbol{a}_{i,0}$ as follows:

$$q(\boldsymbol{a}_{i,t}|\boldsymbol{a}_{i,0}) = \mathrm{Cat}(\boldsymbol{a}_{i,t}; \boldsymbol{p} = \boldsymbol{a}_{i,0}\overline{\boldsymbol{Q}}_{i,t}), \quad (7)$$

with $\overline{\boldsymbol{Q}}_t = \boldsymbol{Q}_1\boldsymbol{Q}_2...\boldsymbol{Q}_t$. In this study, we define transition matrix as $\boldsymbol{Q}_t = (1-\beta_t)\mathbf{I} + \beta_t/K\,\mathbb{1}\mathbb{1}^\top$ so that atomic species follow a uniform distribution at $t = T$, which was selected based on experiments.

The backward process is modeled using categorical distribution $\mathrm{Cat}(\boldsymbol{a}_{i,t-1}; \boldsymbol{p} = \frac{\boldsymbol{a}_{i,t}\boldsymbol{Q}_t^\top \odot \hat{\boldsymbol{a}}_{i,0}\overline{\boldsymbol{Q}}_{t-1}}{\hat{\boldsymbol{a}}_{i,0}\overline{\boldsymbol{Q}}_t\boldsymbol{a}_{i,t}^\top})$, wherein $\hat{\boldsymbol{a}}_{i,0} = \hat{\boldsymbol{a}}_{i,0}(\boldsymbol{M}_t, t)$ is predicted by the neural network, which is trained with the loss function:

$$\mathcal{L}_{\boldsymbol{a}} = \mathbb{E}_{\boldsymbol{a}_{i,t}\sim q(\boldsymbol{a}_{i,t}|\boldsymbol{a}_{i,0}),t\sim\mathcal{U}(1,T)}[q(\boldsymbol{a}_{i,t-1}|\boldsymbol{M}_t, \boldsymbol{M}_0)||p_\theta(\boldsymbol{a}_{i,t-1}|\boldsymbol{M}_t)]. \quad (8)$$

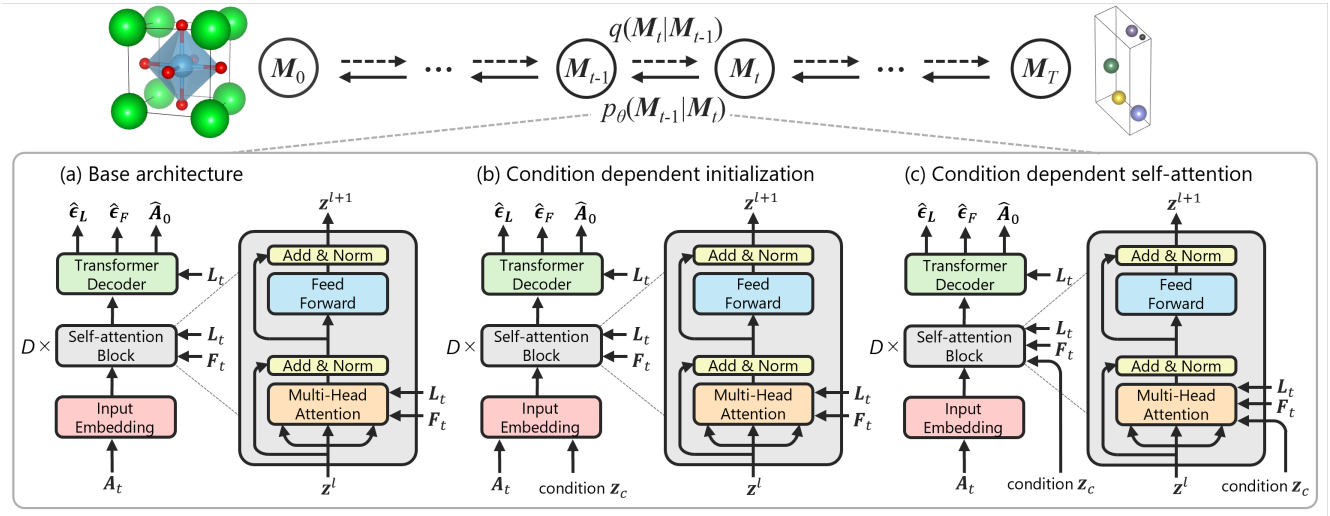

Figure 2: (a) Overview of the base model architecture. (b) Overview of the conditional model with condition-dependent initialization method. (c) Overview of the conditional model with condition-dependent self-attention method. The models of crystal structures in this figure were visualized using VESTA [Momma and Izumi, 2008].

4.2 Backbone Denoising Model

The overview of the base model architecture is shown in Figure 2 (a). Under the joint diffusion processes defined in Sec. 4.1, we construct a model that receives $\boldsymbol{M}_t = (\boldsymbol{L}_t, \boldsymbol{F}_t, \boldsymbol{A}_t)$ as input and outputs $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{L}}$, $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{F}}$, and $\hat{\boldsymbol{A}}_0$. In the following, we describe several key components of the model.

Input Embedding

As shown in Figure 2, initial input embedding in a unit cell, $\boldsymbol{z}^0 = (\boldsymbol{z}_1^0, ..., \boldsymbol{z}_N^0)$, which is fed into the self-attention block, is generated based on time-dependent atomic species $\boldsymbol{A}_t = (\boldsymbol{a}_{1,t}, ..., \boldsymbol{a}_{N,t})$. We generate d-dimensional species embedding $\boldsymbol{z}_{\boldsymbol{a}_i} = \mathrm{MLP}(\boldsymbol{a}_{i,t}) \in \mathbb{R}^d$ and set $\boldsymbol{z}_i^0 = \boldsymbol{z}_{\boldsymbol{a}_i}$, where MLP denotes a simple multi-layer perceptron.

Self-Attention Block

The Self-Attention block, which serves as a core component of the proposed model, receives $\boldsymbol{z}^l = (\boldsymbol{z}_0^l, ..., \boldsymbol{z}_N^l)$, $\boldsymbol{L}_t$, and $\boldsymbol{F}_t$ as inputs, and outputs an updated $\boldsymbol{z}^{l+1} = (\boldsymbol{z}_0^{l+1}, ..., \boldsymbol{z}_N^{l+1})$, where d denotes block index $l = 0, ..., D-1$. Following the original Transformer [Vaswani *et al.*, 2017], $\boldsymbol{z}^l$ is updated as follows:

$$\hat{\boldsymbol{z}} = \mathrm{LN}(\mathrm{MHA}(\boldsymbol{z}^l, \boldsymbol{L}_t, \boldsymbol{F}_t) + \boldsymbol{z}^l), \tag{9}$$

$$\boldsymbol{z}^{l+1} = \mathrm{LN}(\mathrm{MLP}(\hat{\boldsymbol{z}}) + \hat{\boldsymbol{z}}), \tag{10}$$

where MHA and LN denote multi-head attention layer and layer normalization [Ba *et al.*, 2016], respectively.

In this work, the attention mechanism in the MHA layer relies on self-attention with relative position representations [Shaw *et al.*, 2018] to better incorporate relative positional relationships between atoms. Specifically, input sequences $\boldsymbol{z} = (\boldsymbol{z}_0, ..., \boldsymbol{z}_N)$ are transformed to $\boldsymbol{z}' = (\boldsymbol{z}'_0, ..., \boldsymbol{z}'_N)$ according to the following equation:

$$\boldsymbol{z}'_i = \frac{1}{Z_i} \sum_{j=1}^{N} \exp(\boldsymbol{q}_i^\top \boldsymbol{k}_j / \sqrt{d} + \alpha_{ij})(\boldsymbol{v}_j + \boldsymbol{\beta}_{ij}). \tag{11}$$

Here, $Z_i = \sum_{j=1}^{N} \exp(\boldsymbol{q}_i^\top \boldsymbol{k}_j / \sqrt{d} + \alpha_{ij})$, and $\boldsymbol{k}_i = \boldsymbol{z}_i \boldsymbol{W}^K$, $\boldsymbol{q}_i = \boldsymbol{z}_i \boldsymbol{W}^Q$, $\boldsymbol{v}_i = \boldsymbol{z}_i \boldsymbol{W}^V$, where $\boldsymbol{W}^Q$, $\boldsymbol{W}^K$, $\boldsymbol{W}^V$ are parameter matrices for query, key, and value, respectively.

The terms $\alpha_{ij} \in \mathbb{R}$ and $\boldsymbol{\beta}_{ij} \in \mathbb{R}^d$ serve as biases to incorporate the relative positional relationship between atoms i and j. In this study, we consider generating these terms from $\boldsymbol{F}_t$ and $\boldsymbol{L}_t$. Specifically, to incorporate information about the relative positions of atom i and j in fractional space, a Fourier transformation $\psi_{\mathrm{FT}}(\boldsymbol{f}_j - \boldsymbol{f}_i)$, which is first proposed in [Jiao *et al.*, 2023] for creating periodic-invariant message in GNN, is utilized alongside $\boldsymbol{L}_t$ to get α_{ij} and $\boldsymbol{\beta}_{ij}$ as follows:

$$\alpha_{ij} = \mathrm{MLP}(\psi_{\mathrm{FT}}(\boldsymbol{f}_j - \boldsymbol{f}_i), \boldsymbol{L}_t^\top \boldsymbol{L}_t), \tag{12}$$

$$\boldsymbol{\beta}_{ij} = \mathrm{MLP}(\psi_{\mathrm{FT}}(\boldsymbol{f}_j - \boldsymbol{f}_i), \boldsymbol{L}_t^\top \boldsymbol{L}_t). \tag{13}$$

Here, the ψ_{FT} is defined as:

$$\psi_{\mathrm{FT}}(\boldsymbol{f})[n,k] = \begin{cases} \sin(2\pi m f_n) & \text{if } k = 2m \\ \cos(2\pi m f_n) & \text{if } k = 2m+1, \end{cases} \tag{14}$$

where n is an index that runs over the dimension of 3D coordinates, and k is an index that runs over the dimensions of the embedding vector. In plactice, $\psi_{\mathrm{FT}}(\boldsymbol{f}_j - \boldsymbol{f}_i)$ and $\boldsymbol{L}_t^\top \boldsymbol{L}_t$ are flattened and concatenated before being fed into MLP.

Transformer Decoder

After updating through the D-layer self-attention block, the transformer decoder receives $\boldsymbol{z}^D = (\boldsymbol{z}_0^D, ..., \boldsymbol{z}_N^D)$ and $\boldsymbol{L}_t$, and then predicts $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{L}}$, $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{F}}$, and $\hat{\boldsymbol{A}}_0$. $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{F}}$ and $\hat{\boldsymbol{A}}_0$ are obtained straightforwardly from $\boldsymbol{z}^D$ through the equations $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{f}_i} = \mathrm{MLP}(\boldsymbol{z}_i^D)$ and $\hat{\boldsymbol{a}}_{i,0} = \mathrm{MLP}(\boldsymbol{z}_i^D)$. $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{L}}$ is obtained by the linear transformation of $\boldsymbol{L}_t$ as $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{L}} = \boldsymbol{L}_t \varphi(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{z}_i^D)$, where φ denote an MLP that outputs a 3×3 matrix. We note that this procedure follows the steps performed in the read-out of GNN [Jiao *et al.*, 2023], and enables prediction of $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{L}}$ that is equivariant to the rotation of $\boldsymbol{L}_t$.

Method	Perov-5			Carbon-24			MP-20		
	SR5	SR10	SR15	SR5	SR10	SR15	SR5	SR10	SR15
CDVAE [Xie *et al.*, 2021]	0.52	0.65	0.79	0.00	0.06	0.06	0.78	0.86	0.90
SyMat [Luo *et al.*, 2023]	0.73	0.80	0.87	0.06	0.13	0.13	**0.92**	**0.97**	**0.97**
MODEL-CDI	<u>0.93</u>	<u>0.97</u>	<u>0.98</u>	<u>0.43</u>	<u>0.56</u>	<u>0.57</u>	<u>0.91</u>	<u>0.93</u>	<u>0.95</u>
MODEL-CDS	**0.97**	**0.99**	**1.00**	**0.56**	**0.64**	**0.64**	0.63	0.75	0.84

Table 1: Property optimization performance, where SR stands for success rate. The highest SR and the second highest SR achieved among the four models are emphasized with bold and underline, respectively. Performance of CDVAE and SyMat were obtained from the original literatures.

4.3 Conditioning Methods

In this study, we perform conditional generation, using physical property values as condition, for the generative inverse design of crystal structures. We provide the model with the time step t and property values as conditions, where time step t is transformed into a feature vector $\boldsymbol{z}_t$ using sinusoidal positional encoding [Vaswani *et al.*, 2017; Ho *et al.*, 2020], while property values are mapped to a vector $\boldsymbol{z}_{\text{prop}}$ via linear projection. The condition embedding is constructed from $\boldsymbol{z}_t$ and $\boldsymbol{z}_{\text{prop}}$ as follows:

$$\boldsymbol{z}_c = \text{MLP}(\boldsymbol{z}_t \oplus \boldsymbol{z}_{\text{prop}}), \tag{15}$$

where $\oplus$ denotes the concatenation of two vectors. As described below, we investigate two conditional generation approaches based on how $\boldsymbol{z}_c$ is fed into the model.

Condition Dependent Initialization (CDI)

The first approach is to make the input embedding condition-dependent, as shown in Figure 2 (b). In this case, input embedding is set using not only the species embedding $\boldsymbol{z}_{\boldsymbol{a}_i}$ but also the condition embedding $\boldsymbol{z}_c$ as follows:

$$\boldsymbol{z}_i^0 = \text{MLP}(\boldsymbol{z}_{\boldsymbol{a}_i} \oplus \boldsymbol{z}_c). \tag{16}$$

The model conditioned with condition-dependent initialization is referred to as MODEL-CDI.

Condition Dependent Self-attention (CDS)

Figure 2 (c) shows the second approach, where $\boldsymbol{z}_c$ is fed into every self-attention block. Similar to the Time-dependent Self-attention mechanism proposed in DiffiT [Hatamizadeh *et al.*, 2023], the attention block takes $\boldsymbol{z}_c$ as an additional input, and the query, key, value are made condition-dependent as follows:

$$\boldsymbol{q}_i = \boldsymbol{z}_i \boldsymbol{W}^Q + \boldsymbol{z}_c \boldsymbol{W}^{Qc}, \tag{17}$$

$$\boldsymbol{k}_i = \boldsymbol{z}_i \boldsymbol{W}^K + \boldsymbol{z}_c \boldsymbol{W}^{Kc}, \tag{18}$$

$$\boldsymbol{v}_i = \boldsymbol{z}_i \boldsymbol{W}^V + \boldsymbol{z}_c \boldsymbol{W}^{Vc}, \tag{19}$$

where $\boldsymbol{W}^{Qc}$, $\boldsymbol{W}^{Kc}$, $\boldsymbol{W}^{Vc}$ are the conditional projection matrices for query, key, and value, respectively. The model conditioned with this condition-dependent self-attention is referred to as MODEL-CDS.

5 Experiments

In this section, we evaluate the performance of our proposed models on the task of property optimization. Through the comparison with existing methods, we show that our proposed models are effective for generative inverse design of crystal structures.

5.1 Experimental Setup

Task definition

Property optimization is a task that aims to generate crystal structures that possess desired physical properties when target property values are provided.

Datasets

To evaluate the performance of proposed models across diverse materials systems, we conduct the assessments on three datasets with different compositions and crystal systems, following Xie *et al.* [Xie *et al.*, 2021]. **Perov-5** [Castelli *et al.*, 2012a; Castelli *et al.*, 2012b] is a dataset of crystal structures derived from cubic perovskite, containing 18,928 crystal structures and their corresponding property values. **Carbon-24** [Pickard, 2020] contains 10,153 structures consisting solely of carbon atoms, with each structure including between 6 and 24 atoms. **MP-20** [Jain *et al.*, 2013] is a collection of stable crystal structures from the Materials Project database, each containing 20 atoms or fewer. MP-20 includes a total of 45,231 structures and and is the most diverse dataset in terms of both composition and crystal systems. Following [Xie *et al.*, 2021], each dataset was split into a ratio of 6:2:2 for training, validation, and testing, respectively.

Metrics

In the task of property optimization, we measure the ability of generative models to generate crystal structures with low formation energies, in other words, to generate stable crystal structures. In this research, we perform conditional generation of crystal structures by providing the minimum values of the physical properties in the training dataset as conditional values, and calculate the success rate (SR). The success rate is calculated as the proportion of generated crystal structures whose physical property values fall within the top 5% (SR5), 10% (SR10), and 15% (SR15) of the target values. The property values were calculated using pre-trained GNN model by Xie *et al.* [Xie *et al.*, 2021].

Baselines

We compare our models with two existing methods: CDVAE [Xie *et al.*, 2021] and SyMat [Luo *et al.*, 2023], both of which are crystal structure generation models that combine VAE and score-based diffusion models, and they can be applied to the property optimization. Other diffusion models, such as MatterGen [Zeni *et al.*, 2023] and UniMat [Yang *et al.*, 2023], were not compared because source codes are not open at present and were not assessed on the same metrics.

5.2 Results

The performance of property optimization is reported in Table 1. For both Perov-5 and Carbon-24, our proposed models (MODEL-CDI and MODEL-CDS) demonstrated superior performance compared to the previous methods (CDVAE and SyMat). Notably, for Perov-5, MODEL-CDS yielded SRs of nearly 1.0, while for Carbon-24, the SR5, which was close to 0.0 with previous methods, increased to above 0.5. On the other hand, for MP-20, using MODEL-CDI achieved SRs as high as that of SyMat, while MODEL-CDS resulted in a slightly lower SRs. The fact that the proposed models demonstrated overall good performance suggests that Transformers can also be applied as the backbone in diffusion models for crystal structure generation. In addition, as shown in Table 1, MODEL-CDI demonstrated SRs close to the best across all datasets. This indicates that MODEL-CDI is a versatile model capable of being applied to the task of generating crystal structures with desired properties across a wide variety of datasets.

It is interesting to note that when comparing the performance of MODEL-CDI and MODEL-CDS, MODEL-CDS demonstrated higher performance on Perov-5 and Carbon-24, while MODEL-CDI exhibited better performance on MP-20. From the perspective of the crystal structure distribution, crystal structures in Perov-5 has a high degree of freedom in composition, while that in Carbon-24 has a high degree of freedom in crystal systems. MP-20, on the other hand, has high degrees of freedom in both composition and crystal systems. Therefore, it is conceivable that the factors determining the property values, particularly the formation energy in this case, vary depending on the dataset. The difference in performance between MODEL-CDI and MODEL-CDS in this study is considered to reflect this differences of characteristics of datasets, indicating that the optimal conditioning method varies depending on the dataset. The optimal method for conditional generation may also vary depending on the model architecture and the target properties. Therefore, exploring suitable conditional generation techniques is expected to be a valuable direction for future research.

6 Conclusion and Future Work

In this work, we explored a new diffusion model for generative inverse design of crystal structures, where the backbone is formulated based on a Transformer architecture. We explored two conditioning methods for generating crystal structures with target physical properties. Our models generally demonstrated comparable or superior performance compared to previous methods. Furthermore, it was found that the optimal conditioning method varies depending on the dataset, suggesting that the exploration of conditioning techniques depending on the dataset and property would be important for high-precision inverse design.

As future work, evaluation using DFT calculations will be necessary for more rigorous assessments. Additionally, Additionally, the novelty and diversity of the inversely designed crystal structures will also become important metrics in the discovery of new materials.

References

[AI4Science *et al.*, 2023] 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.

[Austin *et al.*, 2023] Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured denoising diffusion models in discrete state-spaces. *arXiv preprint arXiv:2107.03006*, 2023.

[Ba *et al.*, 2016] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

[Castelli *et al.*, 2012a] Ivano E Castelli, David D Landis, Kristian S Thygesen, Søren Dahl, Ib Chorkendorff, Thomas F Jaramillo, and Karsten W Jacobsen. New cubic perovskites for one-and two-photon water splitting using the computational materials repository. *Energy & Environmental Science*, 5(10):9034–9043, 2012.

[Castelli *et al.*, 2012b] Ivano E Castelli, Thomas Olsen, Soumendu Datta, David D Landis, Søren Dahl, Kristian S Thygesen, and Karsten W Jacobsen. Computational screening of perovskite metal oxides for optimal solar light capture. *Energy & Environmental Science*, 5(2):5814–5819, 2012.

[Chen *et al.*, 2020a] Chi Chen, Xunxing Zuo, Weike Ye, Xiangguo Li, Zhi Deng, and Shyue Ping Ong. A critical review of machine learning of energy materials. *Advanced Energy Materials*, 10:1903242, 2020.

[Chen *et al.*, 2020b] Nanxin Chen, Yu Zhang, Heiga Zen, Ron J. Weiss, Mohammad Norouzi, and William Chan. Wavegrad: Estimating gradients for waveform generation. *arXiv preprint arXiv:2009.00713*, 2020.

[Court *et al.*, 2020] 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.

[Dosovitskiy *et al.*, 2021] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2021.

[Gao *et al.*, 2023] Shanghua Gao, Pan Zhou, Ming-Ming Cheng, and Shuicheng Yan. Masked diffusion transformer is a strong image synthesizer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 23164–23173, October 2023.

[Guan *et al.*, 2023] Jiaqi Guan, Wesley Wei Qian, Xingang Peng, Yufeng Su, Jian Peng, and Jianzhu Ma. 3d equivariant diffusion for target-aware molecule generation and

affinity prediction. *arXiv preprint arXiv:2303.03543*, 2023.

[Hatamizadeh *et al.*, 2023] Ali Hatamizadeh, Jiaming Song, Guilin Liu, Jan Kautz, and Arash Vahdat. Diffit: Diffusion vision transformers for image generation. *arXiv preprint arXiv:2312.02139*, 2023.

[Ho *et al.*, 2020] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *arXiv preprint arXiv:2006.11239*, 2020.

[Hoffmann *et al.*, 2019] Jordan Hoffmann, Louis Maestrati, Yoshihide Sawada, Jian Tang, Jean Michel Sellier, and Yoshua Bengio. Data-driven approach to encoding and decoding 3-d crystal structures. *arXiv preprint arXiv:1909.00949*, 2019.

[Hoogeboom *et al.*, 2021] Emiel Hoogeboom, Didrik Nielsen, Priyank Jaini, Patrick Forré, and Max Welling. Argmax flows and multinomial diffusion: Learning categorical distributions. *arXiv preprint arXiv:2102.05379*, 2021.

[Hoogeboom *et al.*, 2022] Emiel Hoogeboom, Victor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. *arXiv preprint arXiv:2203.17003*, 2022.

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

[Jiao *et al.*, 2023] Rui Jiao, Wenbing Huang, Peijia Lin, Jiaqi Han, Pin Chen, Yutong Lu, and Yang Liu. Crystal structure prediction by joint equivariant diffusion. *arXiv preprint arXiv:2309.04475*, 2023.

[Kong *et al.*, 2021] Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. Diffwave: A versatile diffusion model for audio synthesis. *arXiv preprint arXiv:2009.09761*, 2021.

[Luo and Hu, 2021] Shitong Luo and Wei Hu. Diffusion probabilistic models for 3d point cloud generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2837–2845, June 2021.

[Luo *et al.*, 2023] Youzhi Luo, Chengkai Liu, and Shuiwang Ji. Towards symmetry-aware generation of periodic materials. *arXiv preprint arXiv:2307.02707*, 2023.

[Momma and Izumi, 2008] Koichi Momma and Fujio Izumi. *VESTA*: a three-dimensional visualization system for electronic and structural analysis. *Journal of Applied Crystallography*, 41(3):653–658, 2008.

[Noh *et al.*, 2019] Juhwan Noh, Jaehoon Kim, Helge S. Stein, Benjamin Sanchez-Lengeling, John M. Gregoire, Alan Aspuru-Guzik, and Yousung Jung. Inverse design of solid-state materials via a continuous representation. *Matter*, 1(5):1370–1384, 2019.

[Noh *et al.*, 2020] Juhwan Noh, Geun Ho Gu, Sungwon Kim, and Yousung Jung. Machine-enabled inverse design of inorganic solid materials: promises and challenges. *Chemical Science*, 11:4871–4881, 2020.

[Nouira *et al.*, 2018] Asma Nouira, Nataliya Sokolovska, and Jean-Claude Crivello. Crystalgan: Learning to discover crystallographic structures with generative adversarial networks. *arXiv preprint arXiv:1810.11203*, 2018.

[OpenAI, 2023] OpenAI. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[Peebles and Xie, 2022] William Peebles and Saining Xie. Scalable diffusion models with transformers. *arXiv preprint arXiv:2212.09748*, 2022.

[Peng *et al.*, 2023] Xingang Peng, Jiaqi Guan, Qiang Liu, and Jianzhu Ma. Moldiff: Addressing the atom-bond inconsistency problem in 3d molecule diffusion generation. *arXiv preprint arXiv:2305.07508*, 2023.

[Pickard, 2020] Chris J. Pickard. Airss data for carbon at 10gpa and the c+n+h+o system at 1gpa. 2020.

[Ren *et al.*, 2022] Zekun Ren, Siyu Isaac Parker Tian, Juhwan Noh, Felipe Oviedo, Guangzong Xing, Jiali Li, Qiaohao Liang, Ruiming Zhu, Armin G. Aberle, Shijing Sun, Xiaonan Wang, Yi Liu, Qianxiao Li, Senthilnath Jayavelu, Kedar Hippalgaonkar, Yousung Jung, and Tonio Buonassisi. An invertible crystallographic representation for general inverse design of inorganic crystals with targeted properties. *Matter*, 5(1):314–335, 2022.

[Rombach *et al.*, 2022] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. *arXiv preprint arXiv:2112.10752*, 2022.

[Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[Shaw *et al.*, 2018] Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. Self-attention with relative position representations. *1803.02155*, 2018.

[Song and Ermon, 2019] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *arXiv preprint arXiv:21907.05600*, 2019.

[Song and Ermon, 2020] Yang Song and Stefano Ermon. Improved techniques for training score-based generative models. *arXiv preprint arXiv:2006.09011*, 2020.

[Song *et al.*, 2019] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2019.

[Taniai *et al.*, 2024] Tatsunori Taniai, Ryo Igarashi, Yuta Suzuki, Naoya Chiba, Kotaro Saito, Yoshitaka Ushiku,

and Kanta Ono. Crystalformer: Infinitely connected attention for periodic structure encoding. *arXiv preprint arXiv:2403.11686*, 2024.

[Toyao *et al.*, 2020] Takashi Toyao, Zen Maeno, Satoru Takakusagi, Takashi Kamachi, Ichigaku Takigawa, and Ken ichi Shimizu. Machine learning for catalysis informatics: Recent applications and prospects. *ACS Catalysis*, 10(3):2260–2297, 2020.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017.

[Vignac *et al.*, 2023] Clement Vignac, Igor Krawczuk, Antoine Siraudin, Bohan Wang, Volkan Cevher, and Pascal Frossard. Digress: Discrete denoising diffusion for graph generation. *arXiv preprint arXiv:2209.14734*, 2023.

[Watson *et al.*, 2023] Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles Basile I. M., Wicky, Nikita Hanikel, Samuel J. Pellock, Alexis Courbet, William Sheffler, Jue Wang, Preetham Venkatesh, Isaac Sappington, Susana Vázquez Torres, Anna Lauko, Valentin De Bortoli, Emile Mathieu, Sergey Ovchinnikov, Regina Barzilay, Tommi S. Jaakkola, Frank DiMaio, Minkyung Baek, and David Baker. De novo design of protein structure and function with rfdiffusion. *Nature*, 620:1089–1100, 2023.

[Xie *et al.*, 2021] Tian Xie, Xiang Fu, Octavian-Eugen Ganea, Regina Barzilay, and Tommi Jaakkola. Crystal diffusion variational autoencoder for periodic material generation. *arXiv preprint arXiv:2110.06197*, 2021.

[Yan *et al.*, 2022] Keqiang Yan, Yi Liu, Yuchao Lin, and Shuiwang Ji. Periodic graph transformers for crystal material property prediction. *arXiv preprint arXiv:2209.11807*, 2022.

[Yan *et al.*, 2024] Keqiang Yan, Cong Fu, Xiaofeng Qian, Xiaoning Qian, and Shuiwang Ji. Complete and efficient graph transformers for crystal material property prediction. *arXiv preprint arXiv:2403.11857*, 2024.

[Yang *et al.*, 2023] Mengjiao Yang, KwangHwan Cho, Amil Merchant, Pieter Abbeel, Dale Schuurmans, Igor Mordatch, and Ekin Dogus Cubuk. Scalable diffusion for materials generation. *arXiv preprint arXiv:2311.09235*, 2023.

[Ying *et al.*, 2021] Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. Do transformers really perform bad for graph representation? *arXiv preprint arXiv:2106.05234*, 2021.

[Zeni *et al.*, 2023] Claudio Zeni, Robert Pinsler, Daniel Zügner, Andrew Fowler, Matthew Horton, Xiang Fu, Sasha Shysheya, Jonathan Crabbé, Lixin Sun, Jake Smith, Bichlien Nguyen, Hannes Schulz, Sarah Lewis, Chin-Wei Huang, Ziheng Lu, Yichi Zhou, Han Yang, Hongxia Hao, Jielan Li, Ryota Tomioka, and Tian Xie. Mattergen: a generative model for inorganic materials design. *arXiv preprint arXiv:2312.03687*, 2023.

[Zhao *et al.*, 2021] Yong Zhao, Mohammed Al-Fahdi, Ming Hu, dirisuriya M. D. Siriwardane, Yuqi Song, Alireza Nasiri, and Jianjun Hu. High-throughput discovery of novel cubic crystal materials using deep generative neural networks. *Advanced Science*, 8:2100566, 2021.