A Diffusion-Based Pre-training Framework for Crystal Property Prediction

Zixing Song, Ziqiao Meng, Irwin King

The Chinese University of Hong Kong
zxsong@cse.cuhk.edu.hk, zqmeng@cse.cuhk.edu.hk, king@cse.cuhk.edu.hk

Abstract

Many significant problems involving crystal property prediction from 3D structures have limited labeled data due to expensive and time-consuming physical simulations or lab experiments. To overcome this challenge, we propose a pretrain-finetune framework for the crystal property prediction task named CrysDiff based on diffusion models. In the pre-training phase, CrysDiff learns the latent marginal distribution of crystal structures via the reconstruction task. Subsequently, CrysDiff can be fine-tuned under the guidance of the new sparse labeled data, fitting the conditional distribution of the target property given the crystal structures. To better model the crystal geometry, CrysDiff notably captures the full symmetric properties of the crystals, including the invariance of reflection, rotation, and periodic translation. Extensive experiments demonstrate that CrysDiff can significantly improve the performance of the downstream crystal property prediction task on multiple target properties, outperforming all the SOTA pre-training models for crystals with good margins on the popular JARVIS-DFT dataset.

Introduction

Predicting crystal properties remains a significant challenge in materials science (Schütt et al. 2014; Du et al. 2023). Unlike molecules, which are commonly represented as graphs (Wieder et al. 2020), crystals consist of a minimum unit cell repeating itself on a regular lattice in 3D space, which can be viewed as a point cloud of atoms with an orderly repeating pattern (Schütt et al. 2014). The properties of a crystal, such as its electronic, mechanical, and thermal characteristics, are fundamentally linked to its atomic arrangement or structure. Hence, the crystal property prediction task involves predicting these properties based on the crystal's 3D structure. Through the prediction of crystal properties, we may identify promising crystalline materials with tailored properties before actual synthesis or testing, speeding up the discovery process with reduced cost (Datta and Grant 2004; Oganov et al. 2019).

Recently, we have witnessed a surge of interest in developing deep-learning models for crystal property prediction (Chibani and Coudert 2020; Das et al. 2023a; Lin et al. 2023). These models often rival the accuracy of Density-functional theory (DFT) methods while being substantially faster (Orio, Pantazis, and Neese 2009). Recent state-of-the-art (SOTA) models (Louis et al. 2020; Xie and Grossman 2018; Choudhary and DeCost 2021; Yan et al. 2022) usually construct multi-edge graphs for a 3D crystal structure where edges are created between nearby atoms within a pre-specified distance threshold, and apply the Graph Neural Networks (GNNs) model to learn representations of crystal structures that are optimized for the downstream property prediction task. However, the success of such methods is largely tied to the large number of labeled crystal structure data with target property values, which demands extensive time and resources in physical simulations or laboratory experiments (Yu et al. 2023). Therefore, these crystal property predictors are often powerless for real scenarios when the annotations of the crystal properties are sparse, but the 3D crystal structure data are more readily available.

Consequently, a pertinent question is how to leverage vast unlabeled datasets of 3D crystal structures to enhance downstream crystal property prediction that also relies on 3D structures as input. Our answer is a form of self-supervised pre-training paradigm (Liu et al. 2023b) that constructs the surrogate task of exploring the intrinsic structure within the untagged crystal structures and generates useful hidden representations for the downstream prediction task. Recent studies (Jiao et al. 2023a; Zaidi et al. 2023; Liu et al. 2023a) have successfully applied similar strategies to molecular property prediction, and motivate this design of the pretrain-finetune framework in our method. For example, Zaidi et al. (2023) describe a pre-training technique based on denoising by utilizing large datasets of 3D molecular structures at equilibrium to learn meaningful representations for downstream prediction tasks. Yet, in the realm of crystal property prediction from 3D structures, such pre-training remains underexplored. To our knowledge, CrysGNN (Das et al. 2023b) is the only relevant published work for a pre-training framework that extracts distilled knowledge from unlabeled crystal structure data and injects it into different property predictors to enhance their prediction accuracy. Nonetheless, CrysGNN heavily relies on the existing train-from-scratch property predictors like Matformer (Yan et al. 2022) and fails to take the crystal symmetry properties into account.

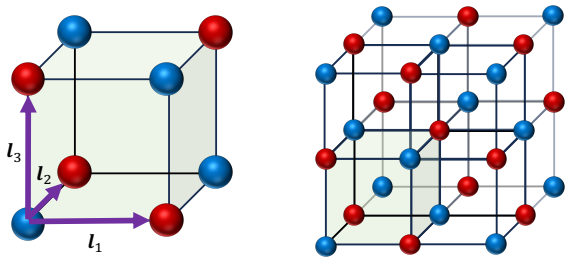
We aim to investigate the possibility of using crystal struc-

ture reconstruction as a pre-training task with abundant unlabeled crystal structure data, which is beneficial for downstream property prediction. The reconstruction of crystal structures, given the atom compositions only, enables the model to learn essential features and representations about the crystal structures. These characteristics, such as bond lengths, are often intricately correlated with the properties of the crystal. Therefore, a model pre-trained to understand and reconstruct these features will lay a foundation for predicting properties during the fine-tuning phase under the guidance of a small labeled dataset, leading to better data efficiency. Furthermore, reconstruction tasks require the model to be attentive to fine details in the crystal structure. This level of attention to detail can enhance the model's robustness, making it less prone to errors when predicting properties based on subtle features in the crystal structure.

However, it is challenging to design a backbone model for both the crystal structure reconstruction in the pre-training phase and the crystal property prediction in the fine-tuning phase. The main difficulty mainly stems from how to satisfy the specific symmetry property of the crystalline material world, where rotating, reflecting, or periodically translating the conformation of a 3D crystal structure does not change the law of its behavior. While molecular representations share symmetry properties like rotation and reflection invariance, the periodic translation invariance is distinct to crystal structures. This uniqueness necessitates the development of new definitions and innovative models to encode such periodic structures accurately and faithfully.

In this study, inspired by their profound impact in molecular biology (Zhang et al. 2023a) and material science (Jiao et al. 2023b), we resort to diffusion models for our backbone and propose CrysDiff, a diffusion-based pre-training framework via crystal structure reconstruction for the downstream property prediction task. In the pre-training phase, CrysDiff tries to reconstruct the original crystal structures via a joint denoising diffusion model given the atom compositions only, and thus is self-supervised without annotations of target properties. The denoising model leverages popular equivariant GNNs (EGNN) as building blocks so that all the unique symmetries of crystal structures can be maintained. In the fine-tuning phase, motivated by classification and regression diffusion (CARD) models (Han, Zheng, and Zhou 2022), CrysDiff retains the backbone diffusion model but diffuses on top of the new target property value with the associated crystal structure fixed. Hence, we can model the conditional distribution of the target property value after observing the corresponding crystal structure, effectively transferring the knowledge of crystal structure distribution gained from the pre-trained task.

To sum up, our contributions are three-fold. First, we pioneer the pre-training approach through crystal structure reconstruction, aiming to enhance downstream property predictions from 3D structures. Second, we adopt diffusion models in both the pre-training and fine-tuning phases, highlighting their versatility and efficacy across generative and predictive tasks. Third, we guarantee strict adherence to all the symmetric properties of crystals in both phases, thanks to the incorporation of an EGNN-based denoising component.



(a) Unit cell with the lattice $\mathbf{L} = [\boldsymbol{l}_1, \boldsymbol{l}_2, \boldsymbol{l}_3]$.

(b) Point cloud of atoms with the repeating pattern.

Figure 1: The illustration of periodic crystal structures.

Preliminaries

Representation of Crystal Structure

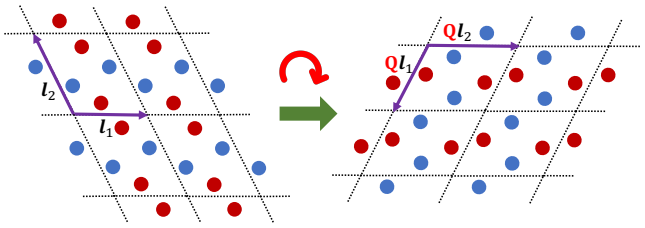
Cartesian Coordinate System A crystal can be viewed as the infinite point cloud data, with a periodic arrangement of atoms in 3D space (Figure 1). Formally, a crystal $\mathcal{C}$ can be completely specified by three matrices $\mathcal{C} = (\mathbf{A}, \mathbf{X}, \mathbf{L})$, and it is built up by repetitive translation of the unit cell $(\mathbf{A}, \mathbf{X})$, the smallest repeating unit, along the lattice matrix $\mathbf{L}$. For an unit cell with N atoms, $\mathbf{A} = [\boldsymbol{a}_1, \cdots, \boldsymbol{a}_N] \in \mathbb{R}^{d_a \times N}$ is the atom feature matrix and $\boldsymbol{a}_i \in \mathbb{R}^{d_a}$ denotes the atom type. $\mathbf{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N] \in \mathbb{R}^{3\times N}$ represents the 3D Cartesian coordinates of each atom position. $\mathbf{L} = [\boldsymbol{l}_1, \boldsymbol{l}_2, \boldsymbol{l}_3] \in \mathbb{R}^{3\times 3}$ is the lattice matrix containing three basic vectors to describe how the unit cell repeats itself along three directions. The infinite point cloud of atoms in the periodic crystal structure can be determined as: $\{(\boldsymbol{a}_i', \boldsymbol{x}_i') \mid \boldsymbol{a}_i' = \boldsymbol{a}_i, \boldsymbol{x}_i' = \boldsymbol{x}_i + \mathbf{L}\boldsymbol{k}, \quad \forall \boldsymbol{k} \in \mathbb{Z}^3\}$. The integer vector $\boldsymbol{k}$ serves as the translation vector, indicating how the unit cell can be replicated in three dimensions via the lattice matrix $\mathbf{L}$.

Fractional Coordinate System The position of an atom under the Cartesian coordinate system is defined by its distances along three standard orthogonal bases. However, in a crystal, the unit cell axes $(\boldsymbol{l}_1, \boldsymbol{l}_2, \boldsymbol{l}_3)$ are not necessarily perpendicular to each other. Instead, it is more natural to describe the position of an atom in terms of how far it is along the $\boldsymbol{l}_1$, $\boldsymbol{l}_2$, and $\boldsymbol{l}_3$ directions, utilizing $\mathbf{L}$ as bases. A point denoted by the fractional coordinate vector $\boldsymbol{f} = [f_1, f_2, f_3]^\intercal \in [0, 1)^3$ corresponds to the Cartesian vector $\boldsymbol{x} = \sum_{i=1}^{3} f_i \boldsymbol{l}_i$. We resort to the fractional coordinate system in this work since it can easily reflect the periodic nature of crystal structures. The crystal is now denoted by a tuple $\mathcal{C} = (\mathbf{A}, \mathbf{F}, \mathbf{L})$ with $\mathbf{F} = [\boldsymbol{f}_1, \cdots, \boldsymbol{f}_N] \in [0, 1)^{3\times N}$ containing the fractional coordinates of all atoms in the unit cell.

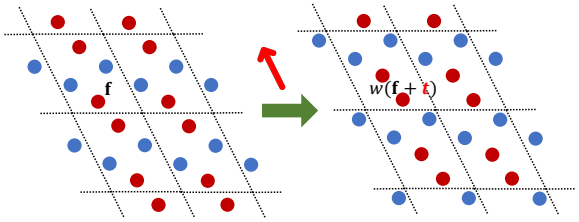
Symmetries of Crystal Structure

Unlike molecules, crystals are made up of a unit cell that repeats regularly along a three-dimensional lattice. Encoding such periodic structures poses unique properties of quantum symmetries, including O(3) invariance and periodic translation invariance under the fractional coordinate system.

Definition 1 (O(3) invariance)**.** A function $f\colon (\mathbf{A}, \mathbf{F}, \mathbf{L}) \mapsto \mathcal{X}$ is O(3) invariant such that for any orthogonal transformation $\mathbf{Q} \in \mathbb{R}^{3\times 3}$ with $\mathbf{Q}^\intercal\mathbf{Q} = \mathbf{I}$, we have, $f(\mathbf{A}, \mathbf{F}, \mathbf{L}) =$



(a) **Definition 1**: O(3) invariance of $\mathbf{L}$.



(b) **Definition 2**: Periodic translation invariance of $\mathbf{F}$.

Figure 2: Symmetry properties of the crystal. Both cases of transformations do not change the actual crystal structures. For simplicity, we use 2D crystals for illustration.

$f(\mathbf{A}, \mathbf{F}, \mathbf{QL})$. Namely, the structure of a crystal remains the same when applying rotations and reflections to lattice $\mathbf{L}$.

Definition 2 (Periodic translation invariance)**.** A function $f\colon (\mathbf{A}, \mathbf{F}, \mathbf{L}) \mapsto \mathcal{X}$ is periodic invariant such that for any translation vector $\boldsymbol{t} \in \mathbb{R}^3$, we have, $f(\mathbf{A}, \mathbf{F}, \mathbf{L}) = f(\mathbf{A}, w(\mathbf{F} + \boldsymbol{t}), \mathbf{L})$. Here, $\mathbf{F} + \boldsymbol{t}$ is a broadcasting addition and the function $w(\mathbf{F}) := \mathbf{F} - \lfloor \mathbf{F} \rfloor$ returns the fractional part of each entry in $\mathbf{F}$. Namely, the structure of a crystal remains the same when applying periodic translation to $\mathbf{F}$.

It is worth noting that the defined symmetries of crystals (Figure 2) are consistent with the ones given under the Cartesian coordinate system (Yan et al. 2022). The rotations or reflections of $\mathbf{L}$ have no effects on the fractional coordinates of atoms, and thus, the E(3) invariance under the Cartesian coordinate system can be reduced to O(3) invariance of $\mathbf{L}$ only. Besides, since the fractional coordinate system reflects the periodicity of the crystal structure intrinsically, we can now obtain a simpler formulation on the periodic translation invariance without the introduction of corner points under the Cartesian coordinate system (Jiao et al. 2023b).

Methodology

Problem Setup and Framework Overview

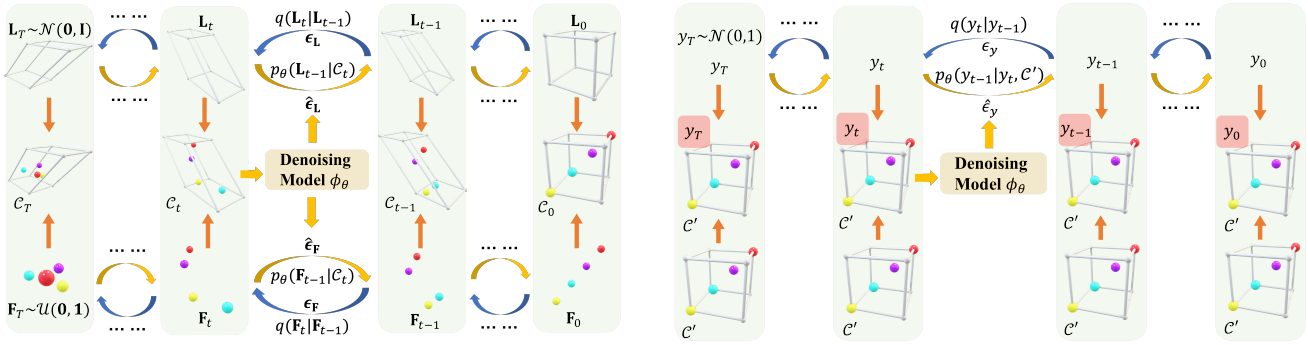
The crystal property prediction aims to predict a target property value $y \in \mathbb{R}$ based on its structure $(\mathbf{A}, \mathbf{F}, \mathbf{L})$. The proposed method follows the popular pre-training and fine-tuning framework and utilizes the heated diffusion model as the backbone model in both stages (Figure 3). It is common knowledge in crystallography that the structure of a crystal determines its properties. Following this fundamental fact, the model first learns the latent distribution of the crystal structure based on a large amount of *unlabeled* data $(\mathbf{A}, \mathbf{F}, \mathbf{L})$ during the pre-training stage. Next, the model is fine-tuned under the guidance of new limited *labeled* data $(\mathbf{A}', \mathbf{F}', \mathbf{L}', y)$, adjusting the pre-learned representations to perform well on the specific property prediction task.

More specifically, in the pre-training phase of the proposed CrysDiff model, we reconstruct the original crystal structure via a joint denoising diffusion model on both the lattice $\mathbf{L}$ and the fractional coordinates of all atoms $\mathbf{F}$ given the composition atom type $\mathbf{A}$. We define the forward diffusion process to transform $(\mathbf{L}, \mathbf{F})$ into random noise simultaneously and then a reverse generation process to reconstruct the clean $(\mathbf{L}, \mathbf{F})$ via an EGNN-based denoising model. To transfer the knowledge regarding crystal structures learned in the pre-trained task, we stick to the same backbone diffusion model in the fine-tuning phase. But the diffusion process is now performed on top of target value y so that the fine-tuned model can recover the predictive conditional distribution $p(y \mid \mathbf{A}', \mathbf{F}', \mathbf{L}')$ after observing the labeled data.

Pre-training via Crystal Structure Reconstruction

In the pre-training phase, we perform the task of crystal structure reconstruction to model the distribution $p(\mathbf{F}, \mathbf{L} \mid \mathbf{A})$ based on the diffusion model from a large amount of unlabeled data. The motivations are as follows. First, the crystal structure reconstruction task enables the pre-trained model to learn meaningful representations of crystal structures, which are beneficial for the crystal property prediction task. The model has already learned to recognize important patterns in crystal structures, which expedites the learning process in the fine-tuning phase. Second, some previous studies (Song and Ermon 2020) have shown that the usage of the diffusion models is equivalent to learning a particular force field directly from equilibrium structures with some desirable properties in quantum chemistry. This force field plays a decisive factor in crystal properties (Nagy 1998). Third, a preliminary study (Zaidi et al. 2023) also employs a score-based diffusion model to denoise the molecular structures and significantly improves the downstream molecular property prediction. Following the previous work (Jiao et al. 2023b), the pre-trained model is built upon the diffusion model where we simultaneously diffuse the lattice $\mathbf{L}$ and the fractional coordinates $\mathbf{F}$. $\mathcal{C}_t = (\mathbf{A}, \mathbf{F}_t, \mathbf{L}_t)$ represents the intermediate state at time step t. In the forward diffusion process, we gradually add noise to $\mathcal{C}_0$ until we obtain the random noise. In the reverse generation process, we randomly sample from the prior distribution $\mathcal{C}_T$ and gradually learn to eliminate the noise through a parameterized denoising model until the original structure $\mathcal{C}_0$ is recovered.

Diffusion on L We follow the denoising diffusion probabilistic model (DDPM) (Ho, Jain, and Abbeel 2020) to diffuse and generate $\mathbf{L}$. We define the forward diffusion process as a Markov chain $q(\mathbf{L}_t \mid \mathbf{L}_{t-1}) = \mathcal{N}(\mathbf{L}_t; \sqrt{1-\beta_t}\mathbf{L}_{t-1}, \beta_t \mathbf{I})$ according to a fixed variance schedule $\{\beta_t\}_{t=1}^T$ with $\beta_t \in (0, 1)$. It progressively diffuses $\mathbf{L}_0$ to the normal prior distribution $p(\mathbf{L}_T) = \mathcal{N}(\mathbf{0}, \mathbf{I})$ with sufficiently large T since $q(\mathbf{L}_t \mid \mathbf{L}_0) = \mathcal{N}(\mathbf{L}_t; \sqrt{\bar{\alpha}_t}\mathbf{L}_0, (1 - \bar{\alpha}_t)\mathbf{I})$ with $\bar{\alpha}_t = \Pi_{s=1}^t \alpha_s = \Pi_{s=1}^t (1 - \beta_s)$. For backward generation process, we formulate this reverse dynamics as a conditional Markov chain with learnable transitions $p_\theta(\mathbf{L}_{t-1} \mid \mathcal{C}_t) = \mathcal{N}(\mathbf{L}_{t-1}; \mu_\theta(\mathcal{C}_t), \sigma_t^2 \mathbf{I})$ to reconstruct $\mathbf{L}_0$



(a) Pre-training Phase of the CrysDiff model.

(b) Fine-tuning Phase of the CrysDiff model.

Figure 3: In the pre-training phase, we perform the crystal structure reconstruction task based on the diffusion model. The forward process is carried out on lattice $\mathbf{L}$ and fractional coordinates $\mathbf{F}$ simultaneously and gradually transforms the given $\mathbf{L}_0$ and $\mathbf{F}_0$ to some random noises $\mathbf{L}_T$ and $\mathbf{F}_T$. The backward process predicts the noise terms $\hat{\boldsymbol{\epsilon}}_{\mathbf{L}}$ and $\hat{\boldsymbol{\epsilon}}_{\mathbf{F}}$ to reconstruct from the noisy structure $\mathcal{C}_t$ via an EGNN-based denoising model ϕ_θ. In the fine-tuning phase, we conduct the crystal property prediction task. We perform the diffusion on the target y to model the conditional $p(y \mid \mathcal{C}')$ by keeping the new labeled data $\mathcal{C}'$ fixed.

from noisy $\mathcal{C}_T$. Here, $\mu_\theta(\mathcal{C}_t) = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{L}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\hat{\boldsymbol{\epsilon}}_{\mathbf{L}}(\mathcal{C}_t)\right)$ and the denoising term $\hat{\boldsymbol{\epsilon}}_{\mathbf{L}}$ is given by the denoising model $\phi_\theta(\mathbf{A}, \mathbf{F}_t, \mathbf{L}_t, t)$, which will be introduced later. The training loss is defined as the expected ℓ_2 distance between ground-truth $\boldsymbol{\epsilon}_{\mathbf{L}}$ and the estimated $\hat{\boldsymbol{\epsilon}}_{\mathbf{L}}$:

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

Diffusion on $\mathbf{F}$ To obey the bounded domain of $\mathbf{F}$, we exploit the score-based generative model (Song and Ermon 2020) with the wrapped normal distribution (Jiao et al. 2023b), which has proven effective in the molecular conformer generation task (Jing et al. 2022). The forward diffusion process is now defined as $q(\mathbf{F}_t \mid \mathbf{F}_0) \propto \sum_{\mathbf{Z}\in\mathbb{Z}^{3\times N}} \exp\left(-\|\mathbf{F}_t - \mathbf{F}_0 + \mathbf{Z}\|_F^2/(2\sigma_t^2)\right)$. Here, σ_t is an exponential annealing schedule and $\mathbf{F}_t = w(\mathbf{F}_0 + \sigma_t\boldsymbol{\epsilon}_{\mathbf{F}})$ with $\boldsymbol{\epsilon}_{\mathbf{F}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. $q(\mathbf{F}_t \mid \mathbf{F}_0)$ approaches the uniform distribution $\mathcal{U}(\mathbf{0}, \mathbf{1})$ when T is sufficiently large. For the backward process, we apply the predictor-corrector method (Song et al. 2021a) to reconstruct $\mathbf{F}_0$ from $\mathbf{F}_T \sim \mathcal{U}(\mathbf{0}, \mathbf{1})$. The estimated score $\hat{\boldsymbol{\epsilon}}_{\mathbf{F}}$ is still given by the denoising model $\phi_\theta(\mathbf{A}, \mathbf{F}_t, \mathbf{L}_t, t)$. The training loss for score matching is given as,

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

Here, $\lambda_t = 1/\mathbb{E}_{\mathbf{F}_t}\left[\left\|\nabla_{\mathbf{F}_t} \log q\left(\mathbf{F}_t \mid \mathbf{F}_0\right)\right\|_2^2\right]$ can be approximated via Monte Carlo sampling methods.

EGNN-based Denoising Model The denoising model $\phi_\theta(\mathbf{A}, \mathbf{F}, \mathbf{L}, t)$ that outputs $\hat{\boldsymbol{\epsilon}}_{\mathbf{L}}$ and $\hat{\boldsymbol{\epsilon}}_{\mathbf{F}}$ is built upon EGNN (Satorras, Hoogeboom, and Welling 2021) to satisfy the symmetry properties. Considering the representations $[\boldsymbol{h}_1^{(s)}, \cdots, \boldsymbol{h}_N^{(s)}]$ for each atom in the unit cell at the k-th layer. The message-passing paradigm is presented as follows.

$$\boldsymbol{m}_{ij}^{(k)} = \varphi_\theta^m\left(\boldsymbol{h}_i^{(k-1)}, \boldsymbol{h}_j^{(k-1)}, \mathbf{L}^\top\mathbf{L}, \psi_{\mathrm{FT}}\left(\boldsymbol{f}_j - \boldsymbol{f}_i\right)\right),$$

$$\boldsymbol{h}_i^{(k)} = \boldsymbol{h}_i^{(k-1)} + \varphi_\theta^h\left(\boldsymbol{h}_i^{(k-1)}, \sum_{j=1}^N \boldsymbol{m}_{ij}^{(k)}\right).$$

Here, $\boldsymbol{h}_0$ is initialized as the $\boldsymbol{a}_i$ with the Transformer sinusoidal position embedding (Vaswani et al. 2017). ψ_{FT} is set as the Fourier transform to ensure periodic translation invariance. φ_θ^m and φ_θ^h are multilayer perceptions(MLPs). The final output denoising terms $\hat{\boldsymbol{\epsilon}}_{\mathbf{L}}$ and $\hat{\boldsymbol{\epsilon}}_{\mathbf{F}}$ are decoded from the atom representations by K EGNN layers.

$$\hat{\boldsymbol{\epsilon}}_{\mathbf{L}} = \mathbf{L}\varphi_\theta^L\left(\frac{1}{N}\sum_{i=1}^N \boldsymbol{h}_i^{(K)}\right), \quad \hat{\boldsymbol{\epsilon}}_{\mathbf{F}} = [\varphi_\theta^F(\boldsymbol{h}_i^{(K)})]_{i=1}^N. \quad (3)$$

Here, φ_θ^L and φ_θ^F are two MLPs with proper output sizes. The final training algorithm in the pre-training phase of the proposed CrysDiff model is summarized in Algorithm 1.

Fine-tuning for Crystal Property Prediction

To fully leverage the knowledge about the crystal structure during the pre-training phase, we stick to the backbone diffusion model when fine-tuning. However, the downstream task becomes a deterministic task. Hence, we aim to model $p(y \mid \mathbf{A}', \mathbf{F}', \mathbf{L}')$ with the newly observed labeled data, following the previous work (Han, Zheng, and Zhou 2022).

Algorithm 1: Pre-training Phase of CrysDiff

Input: Atom features $\mathbf{A}$, fractional coordinates $\mathbf{F}_0$, lattice $\mathbf{L}_0$, number of diffusion steps T.

1: Sample $\boldsymbol{\epsilon}_{\mathbf{L}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\boldsymbol{\epsilon}_{\mathbf{F}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $t \sim \mathcal{U}(1, T)$.
2: $\mathbf{L}_t \leftarrow \sqrt{\bar{\alpha}_t}\mathbf{L}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_{\mathbf{L}}$, $\mathbf{F}_t \leftarrow w\left(\mathbf{F}_0 + \sigma_t\boldsymbol{\epsilon}_{\mathbf{F}}\right)$.
3: $\hat{\boldsymbol{\epsilon}}_{\mathbf{L}}, \hat{\boldsymbol{\epsilon}}_{\mathbf{F}} \leftarrow \phi_\theta\left(\mathbf{A}, \mathbf{F}_t, \mathbf{L}_t, t\right)$ via Eq. (3).
4: Minimize $\mathcal{L}_{\mathbf{L}} + \mathcal{L}_{\mathbf{F}}$ via Eq. (1) and Eq. (2).
5: **return** Pre-trained model ϕ_θ.

Algorithm 2: Fine-tuning Phase of CrysDiff (Training)

Input: Atom features $\mathbf{A}'$, fractional coordinates $\mathbf{F}'$, lattice $\mathbf{L}'$, corresponding target crystal property y, number of diffusion steps T.

1: Sample $\epsilon_y \sim \mathcal{N}(0,1)$, $t \sim \mathcal{U}(1,T)$.
2: $y_t \leftarrow \sqrt{\bar{\alpha}_t} y_0 + (1-\sqrt{\bar{\alpha}_t})\bar{y} + \sqrt{1-\bar{\alpha}_t}\epsilon_y$.
3: $\hat{\epsilon}_y \leftarrow \phi_\theta(\mathbf{A}', \mathbf{F}', \mathbf{L}', y_t, t)$ via Eq. (4).
4: Minimize $\mathcal{L}_y$ via Eq. (5).
5: **return** Fine-tuned model ϕ_θ.

Training via Diffusion on y Different from the vanilla diffusion model, we set the endpoint of the forward process or the standpoint of the backward process as $p_\theta(y_T \mid \mathcal{C}_t) = \mathcal{N}(\bar{y}, 1)$. Here, $\bar{y}$ is the empirical mean of target value y (or y_0) in the fine-tuning data. More complex choices to approximate $\mathbb{E}(y \mid \mathcal{C})$ could be used to incorporate the prior knowledge. With a diffusion schedule $\{\beta_t\}_{t=1}^T \in (0,1)^T$, the forward process of the conditional distribution (Pandey et al. 2022) is specified as $q(y_t \mid y_{t-1}, \bar{y}) = \mathcal{N}(y_t; \sqrt{1-\beta_t}y_{t-1} + (1-\sqrt{1-\beta_t})\bar{y}, \beta_t)$. The closed-form sampling distribution is $q(y_t \mid y_0, \bar{y}) = \mathcal{N}(y_t; \sqrt{\bar{\alpha}_t}y_0 + (1-\sqrt{\bar{\alpha}_t})\bar{y}, (1-\bar{\alpha}_t))$ with $\bar{\alpha}_t = \Pi_{s=1}^t \alpha_s = \Pi_{s=1}^t (1-\beta_s)$. This mean term can be regarded as an interpolation between true target y_0 and the prior assumption $\bar{y}$, which gradually changes from the former to the latter throughout the forward process. The goal of the reverse process is to gradually recover the distribution of the noise term $\epsilon_y \sim \mathcal{N}(0,1)$. The denoising model $\phi_\theta(\mathbf{A}', \mathbf{F}', \mathbf{L}', y_t, t)$ follows the same EGNN model as the one in the pre-training phase with fixed clean input crystal structure $\mathcal{C}'$, and the predicted denoising term is again decoded from the hidden representations of atoms via another MLP φ_θ^y.

$$\hat{\epsilon}_y = y_t \varphi_\theta^y \left(\frac{1}{N} \sum_{i=1}^N \boldsymbol{h}_i^{(K)} \right). \tag{4}$$

The final training loss in the fine-tuning phase is:

$$\mathcal{L}_y = \mathbb{E}_{\epsilon_y \sim \mathcal{N}(0,1), t \sim \mathcal{U}(1,T)} \left[\|\epsilon_y - \hat{\epsilon}_y(\mathcal{C}', y_t, t)\|_2^2 \right]. \tag{5}$$

Algorithm 2 presents how we fine-tune the proposed CrysDiff model with new labeled data.

Inference via Generation of y To make predictions on the test data $\mathcal{C}'$, we generate the samples that match the true conditional $p(y \mid \mathcal{C}')$ learned by the denoising model in the fine-tuning stage. The formulation of the forward diffusion process on y leads to a tractable posterior:

$$q(y_{t-1} \mid y_t, y_0, \bar{y}) = \mathcal{N}(y_{t-1}; \gamma_0 y_0 + \gamma_1 y_t + \gamma_2 \bar{y}, \tilde{\beta}_t). \tag{6}$$

Here, we have $\gamma_0 = \frac{\beta_t \sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_t}$, $\gamma_1 = \frac{(1-\bar{\alpha}_{t-1})\sqrt{\alpha_t}}{1-\bar{\alpha}_t}$, $\gamma_2 = \left(1 + \frac{(\sqrt{\bar{\alpha}_t}-1)(\sqrt{\alpha_t}+\sqrt{\bar{\alpha}_{t-1}})}{1-\bar{\alpha}_t}\right)$, and $\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$. The final inference procedure in the CrysDiff is carried out in a standard DDPM manner via sampling as Algorithm 3.

Theoretical Guarantees on Crystal Symmetries

We prove that the learned marginal distribution $p_\theta(\mathcal{C}_0)$ ($p_\theta(\mathbf{L}_0)$ and $p_\theta(\mathbf{F}_0)$) satisfies the proposed symmetries of crystal structures in the pre-training phase, and the learned conditional distribution $p_\theta(y_0 \mid \mathcal{C}')$ also satisfy these two symmetry properties in the fine-tuning phase. Hence, our method can explicitly acknowledge these symmetries guaranteed by underlying physical principles, leading to more accurate and interpretable predictions.

Our proposed method, CrysDiff, utilizes the same EGNN model (Satorras, Hoogeboom, and Welling 2021) as the backbone of the parameterized denoising model ϕ_θ in both phases. Thanks to the symmetry properties of EGNN, we have the following proposition for the estimated noise terms by ϕ_θ, following the previous work (Jiao et al. 2023b).

Proposition 1. $\hat{\boldsymbol{\epsilon}}_{\mathbf{L}}$ by Eq. (3) is O(3) equivariant. $\hat{\boldsymbol{\epsilon}}_{\mathbf{F}}$ by Eq. (3) is periodic translation invariant. $\hat{\epsilon}_y$ by Eq. (4) is both O(3) equivariant and periodic translation invariant.

Meanwhile, due to the symmetry of the Markov process (Xu et al. 2022), we obtain the following proposition.

Proposition 2. If $\hat{\boldsymbol{\epsilon}}_{\mathbf{L}}$ is O(3) equivariant, the learned marginal distribution $p_\theta(\mathbf{L}_0)$ is O(3) invariant. If $\hat{\boldsymbol{\epsilon}}_{\mathbf{F}}$ is periodic translation invariant, the learned marginal distribution $p_\theta(\mathbf{F}_0)$ is periodic translation invariant. If $\hat{\epsilon}_y$ is both O(3) equivariant and periodic translation invariant, the learned conditional distribution $p_\theta(y_0 \mid \mathcal{C}')$ is both O(3) invariant and periodic translation invariant.

From Proposition 1 and 2, we immediately prove that the learned marginal distribution $p_\theta(\mathcal{C}_0)$ in the pre-training phase is both O(3) invariant and periodic translation invariant assuming the independence of $p_\theta(\mathbf{L}_0)$ and $p_\theta(\mathbf{F}_0)$. Similarly, we obtain that the conditional distribution $p_\theta(y_0 \mid \mathcal{C}')$ in the fine-tuning phase also maintains both symmetries.

Algorithm 3: Fine-tuning Phase of CrysDiff (Inference)

q **Input**: Atom features $\mathbf{A}'$, fractional coordinates $\mathbf{F}'$, lattice $\mathbf{L}'$, empirical mean $\bar{y}$, number of diffusion steps T, fine-tuned model ϕ_θ.

1: Sample $y_t \sim \mathcal{N}(\bar{y}, 1)$.
2: **for** $t = T$ to 1 **do**
3: Sample $\epsilon_y \sim \mathcal{N}(0,1)$ if $t > 1$.
4: $\hat{\epsilon}_y \leftarrow \phi_\theta(\mathbf{A}', \mathbf{F}', \mathbf{L}', y_t, t)$ via fine-tuned model ϕ_θ.
5: $\hat{y}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(y_t - (1-\sqrt{\bar{\alpha}_t})\bar{y} - \sqrt{1-\bar{\alpha}_t}\hat{\epsilon}_y\right)$.
6: $y_{t-1} = \gamma_0 \hat{y}_0 + \gamma_1 y_t + \gamma_2 \bar{y} + \sqrt{\tilde{\beta}_t}\epsilon_y$
7: **end for**
8: **return** Prediction $\hat{y}_0$.

Experiments

We evaluate how CrysDiff performs on diverse crystal properties from two popular benchmark materials datasets. We first briefly discuss the datasets used both in pre-training and downstream property prediction tasks. Then we compare CrysDiff with various state-of-the-art (SOTA) models with or without the pretrain-finetune framework on the downstream property prediction task. Next, we illustrate the effectiveness of CrysDiff even with sparse fine-tuning data and little experimental data. We finally conduct some abla-

	Train-from-scratch Models				Pretrain-finetune Models			
	CGCNN (2018)	GATGNN (2020)	ALIGNN (2021)	Matformer (2022)	Pretrain-GNN (2020)	CrysXPP (2022)	CrysGNN (2023b)	CrysDiff (Ours)
Formation Energy	0.063	0.047	0.033	0.033	0.764	0.062	0.056	**0.029**
Bandgap (OPT)	0.200	0.170	0.142	0.137	0.688	0.190	0.183	**0.131**
Total Energy	0.078	0.056	0.037	0.035	1.451	0.072	0.069	**0.034**
Ehull	0.179	0.129	0.076	0.064	1.112	0.139	0.130	**0.062**
Bandgap (MBJ)	0.413	0.513	0.310	0.302	1.493	0.378	0.371	**0.287**
Bulk Modulus (Kv)	14.47	14.32	10.40	11.21	20.34	13.61	13.42	**9.875**
Shear Modulus (Gv)	11.75	12.48	9.481	10.76	16.51	11.20	11.07	**9.193**
SLME (%)	8.022	7.504	5.145	5.260	9.853	5.110	5.452	**5.030**
Spillage	0.454	0.431	0.389	0.398	0.481	0.363	0.374	**0.358**

Table 1: The prediction performance (MAE) of nine properties on the JARVIS-DFT dataset for the proposed CrysDiff model against existing train-from-scratch and pretrain-finetune models.

tion studies to show the influence of different design choices in predicting the crystal properties.

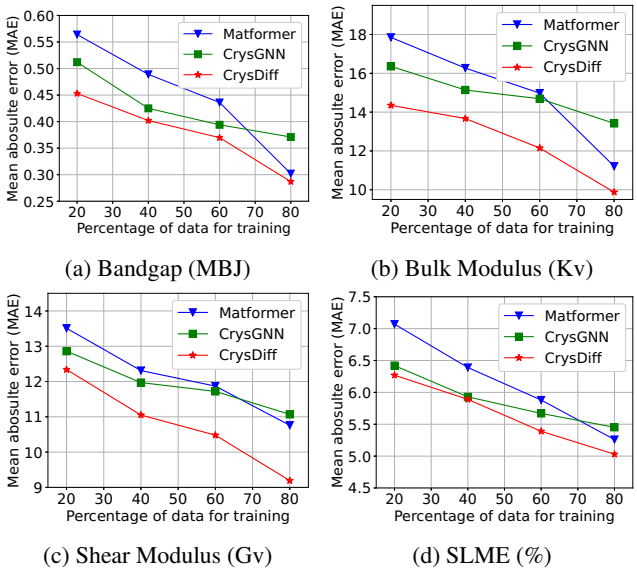
Experimental Settings

Datasets We collect 800K untagged crystal graph data from two popular materials databases, Materials Project (MP) (Jain et al. 2013) and OQMD (Saal et al. 2013), to pre-train the CrysDiff model. Further, to evaluate the fine-tuning performance of CrysDiff compared with other crystal property predictors, we select the 2021.8.18 version of JARVIS-DFT (Choudhary et al. 2020), another popular materials database, for the downstream property prediction task. JARVIS-DFT consists of 55,722 materials with 19 properties, like formation energy, bandgap, total energy, bulk modulus, etc., which depend significantly on crystal structures and atom features. Moreover, all these properties in the above datasets are based on DFT calculations of chemicals. Therefore, to investigate how our model helps to mitigate the DFT error, we also take a small dataset OQMD-EXP (Kirklin et al. 2015), containing 1,500 available experimental data of formation energy.

Baselines To evaluate the effectiveness of CrysDiff, we choose four SOTA train-from-scratch models for crystal property prediction, CGCNN (Xie and Grossman 2018), GATGNN (Louis et al. 2020), ALIGNN (Choudhary and DeCost 2021), Matformer (Yan et al. 2022) and three SOTA pretrain-finetune models, Pretrain-GNN (Hu et al. 2020) CrysXPP (Das et al. 2022), CrysGNN (Das et al. 2023b). We report mean absolute error (MAE) of the predicted and actual value of a particular property to compare the performance of different participating methods.

Downstream Task Evaluation

Results Using DFT Data For train-from-scratch models, we directly train all the baseline models on the tagged JARVIS-DFT dataset. For pretrain-finetune models, we pretrain all the baseline models on the curated 800K untagged crystal data and fine-tune them on nine properties in the JARVIS dataset. For each property, all the models are trained on 80% data, validated on 10%, and tested on 10% of the data. In Table 1, we report the MAE of different predicted



(a) Bandgap (MBJ) (b) Bulk Modulus (Kv)

(c) Shear Modulus (Gv) (d) SLME (%)

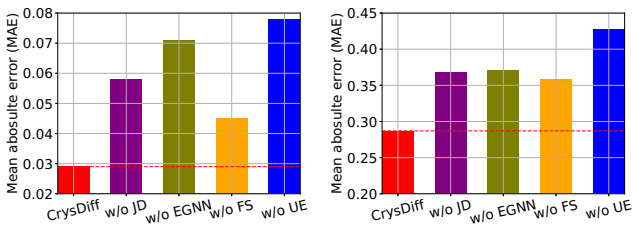
Figure 4: The prediction performance (MAE) of four properties on the JARVIS-DFT dataset with sparse data.

crystal properties. We have several insightful observations from these results in Table 1. First, our proposed CrysDiff model achieves the SOTA performance over all nine properties, outperforming other pretrain-finetune baselines with large margins. Second, many existing SOTA pretrain-finetune models cannot beat some strong train-from-scratch models like Matformer. We speculate that there exists a mismatch between the pre-trained and fine-tuned dataset. Third, CrysDiff significantly improves the prediction performance on some particular properties like Bulk Modulus and Shear Modulus. The potential reason is that these properties are heavily dependent on the corresponding crystal structures, and CrysDiff successfully learns the distribution of crystal structures in the pre-training phase via the diffusion process.

Results Using Sparse Data To demonstrate the effectiveness of the pre-training in limited data settings, we conduct an additional set of experiments under different train-

Experiment Settings	CrysXPP (2022)	CrysGNN (2023b)	CrysDiff (Ours)
Train on DFT Test on Expt	0.307	0.253	**0.211**
Train on DFT + 20% Expt Test on 80% Expt	0.158	0.135	**0.102**
Train on DFT + 80% Expt Test on 20% Expt	0.110	0.096	**0.087**

Table 2: The prediction performance (MAE) of different pretrain-finetune models with full DFT data and different percentages of experimental data for formation energy on the OQMD-EXP dataset. Expt stands for experimental data.



(a) Formation Energy (b) Bandgap (MBJ)

Figure 5: The prediction performance (MAE) of two properties for CrysDiff variants on the JARVIS-DFT dataset.

ing data splits during the fine-tuning phase. To be more precise, we vary available labeled training data from 20% to 80%, train different models, and check their performance on the test dataset. We choose two strong baselines, Matformer from the train-from-scratch models and CrysGNN from the pretrain-finetine models. From Figure 4, we observe that both CrysDiff and CrysGNN can significantly outperform Matformer when the given labeled training data are quite limited, like with only 20% or 40% tagged training data, even though CrysGNN shows inferior performance than Matformer when the ratio of training data is high. This demonstrates that the benefits of the pretrain-finetine framework can be revealed under the sparse training data setting because the knowledge learned with the abundant unlabeled data during the pre-training phase can be essentially transferred and adapted during the fine-tuning phase, even with limited labeled data. Besides, the proposed CrysDiff model consistently performs the best among the three, with all training data ratio cases, illustrating our pre-training framework's extraordinary robustness to spare data.

Results Using Experimental Data One of the fundamental issues in material science is that experimental data for crystal properties are scarce. Hence existing SOTA models rely on DFT-calculated data to train their parameters. However, mathematical approximations in DFT calculation lead to erroneous predictions compared to the actual experimental values. CrysXPP has shown that pre-training helps to remove DFT error bias when fine-tuned with experimental data. Hence, we investigate whether the proposed CrysDiff can remove the DFT error using a small amount of available experimental data. We consider the OQMD-EXP dataset to conduct an experiment, where we train each pretrain-finetune model with available DFT data and different percentages of experimental data for formation energy. We report the MAE of different methods in Table 2. We observe, with more amount of experimental training data, all the models minimize the error consistently. Moreover, we note a larger degree of improvement with more amount of experimental training data in almost all the models, and our proposed model still achieves the lowest MAE in all cases.

Ablation Study

We perform an ablation study to investigate the influence of some key designed components in the proposed CrysDiff model. First, in the pre-training phase, we remove the joint diffusion of $\mathbf{L}$ and $\mathbf{F}$ and build two separate denoising models instead. The final representations for atoms in the unit cell are calculated as the average of the representations from the two models. We name this variant as *CrysDiff w/o joint diffusion (JD)*. Second, we replace the EGNN model in the denoising model with ts non-equivariant GNN cousin, and we name it *CrysDiff w/o EGNN*. Third, we try to jointly diffuse y, $\mathbf{F}'$ and $\mathbf{L}'$ in the fine-tuning phase without fixing the clean crystal structure $\mathbf{F}'$ and $\mathbf{L}'$ as $\mathcal{C}'$. We refer to it as *CrysDiff w/o fixing structures (FS)*. Fourth, we keep the pre-training phase unchanged but freeze the encoder EGNN embeddings $\boldsymbol{h}_i$ for atoms and only update the top-level MLP decoder (Eq. (4)) during the fine-tuning. We call this variant *CrysDiff w/o updating EGNN (UE)*. We report the MAE of all these CrysDiff variants on the JARVIS-DFT dataset in Figure 5. It clearly shows that all the key design choices in CrysDiff play significant roles in predicting the target property. Among them, freezing the atom embeddings learned from the pre-training phase during the fine-tuning phase (CrysDiff w/o UE) leads to the most substantial performance drop. This proves the pretrain-finetune architecture matters most in the success of CrysDiff because the abundant unlabeled crystal structure data only lay the foundation for learning hidden representations of the crystal, but the limited labeled crystal property data can guide the adjustments of these representations, enabling the decoder to predict the target value more easily and efficiently. We also find that the design of joint diffusion and the EGNN encoder also play crucial roles because they make the diffusion model fit the distribution of crystals with the scientific guidelines.

Conclusion

To conclude, we introduce CrysDiff, a novel pretrain-finetune framework based on diffusion models for predicting crystal properties from 3D structures. Future endeavors may explore the fusion of CrysDiff with efficient pretraining techniques. In a broader sense, we could investigate diffusion models for other graph-related applications (Song, Zhang, and King 2023c,a,b; Zhang et al. 2023d; Liang et al. 2023a; Song, Ma, and King 2022; Zhang et al. 2023b; Ma et al. 2023; Song, Zhang, and King 2022; Song et al. 2021b; Song and King 2022; Yang et al. 2023b,a; Liu et al. 2022; Chen et al. 2023a,b; Liang et al. 2023b; Zhang et al. 2023c, 2022) as well.

Acknowledgments

The work described here was partially supported by grants from the National Key Research and Development Program of China (No. 2018AAA0100204) and from the Research Grants Council of the Hong Kong Special Administrative Region, China (CUHK 14222922, RGC GRF No. 2151185).

References

Chen, Y.; Fang, Y.; Zhang, Y.; and King, I. 2023a. Bipartite Graph Convolutional Hashing for Effective and Efficient Top-N Search in Hamming Space. In *WWW*, 3164–3172. ACM.

Chen, Y.; Zhang, Y.; Yang, M.; Song, Z.; Ma, C.; and King, I. 2023b. WSFE: Wasserstein Sub-graph Feature Encoder for Effective User Segmentation in Collaborative Filtering. In *SIGIR*, 2521–2525. ACM.

Chibani, S.; and Coudert, F.-X. 2020. Machine learning approaches for the prediction of materials properties. *APL Materials*, 8(8): 080701.

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

Choudhary, K.; Garrity, K. F.; Reid, A. C.; DeCost, B.; Biacchi, A. J.; Hight Walker, A. R.; Trautt, Z.; Hattrick-Simpers, J.; Kusne, A. G.; Centrone, A.; et al. 2020. The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design. *npj computational materials*, 6(1): 173.

Das, K.; Goyal, P.; Lee, S.-C.; Bhattacharjee, S.; and Ganguly, N. 2023a. CrysMMNet: Multimodal Representation for Crystal Property Prediction. In *Uncertainty in Artificial Intelligence*, 507–517. PMLR.

Das, K.; Samanta, B.; Goyal, P.; Lee, S.; Bhattacharjee, S.; and Ganguly, N. 2023b. CrysGNN: Distilling Pre-trained Knowledge to Enhance Property Prediction for Crystalline Materials. In *AAAI*, 7323–7331. AAAI Press.

Das, K.; Samanta, B.; Goyal, P.; Lee, S.-C.; Bhattacharjee, S.; and Ganguly, N. 2022. CrysXPP: An explainable property predictor for crystalline materials. *npj Computational Materials*, 8(1): 43.

Datta, S.; and Grant, D. J. 2004. Crystal structures of drugs: advances in determination, prediction and engineering. *Nature Reviews Drug Discovery*, 3(1): 42–57.

Du, Y.; Wang, Y.; Huang, Y.; Li, J. C.; Zhu, Y.; Xie, T.; Duan, C.; Gregoire, J. M.; and Gomes, C. P. 2023. M^2Hub: Unlocking the Potential of Machine Learning for Materials Discovery. *CoRR*, abs/2307.05378.

Han, X.; Zheng, H.; and Zhou, M. 2022. CARD: Classification and Regression Diffusion Models. In *NeurIPS*.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising Diffusion Probabilistic Models. In *NeurIPS*.

Hu, W.; Liu, B.; Gomes, J.; Zitnik, M.; Liang, P.; Pande, V. S.; and Leskovec, J. 2020. Strategies for Pre-training Graph Neural Networks. In *ICLR*. OpenReview.net.

Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; et al. 2013. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL materials*, 1(1).

Jiao, R.; Han, J.; Huang, W.; Rong, Y.; and Liu, Y. 2023a. Energy-Motivated Equivariant Pretraining for 3D Molecular Graphs. In *AAAI*, 8096–8104. AAAI Press.

Jiao, R.; Huang, W.; Lin, P.; Han, J.; Chen, P.; Lu, Y.; and Liu, Y. 2023b. Crystal Structure Prediction by Joint Equivariant Diffusion on Lattices and Fractional Coordinates. In *Workshop on "Machine Learning for Materials" ICLR 2023*.

Jing, B.; Corso, G.; Chang, J.; Barzilay, R.; and Jaakkola, T. S. 2022. Torsional Diffusion for Molecular Conformer Generation. In *NeurIPS*.

Kirklin, S.; Saal, J. E.; Meredig, B.; Thompson, A.; Doak, J. W.; Aykol, M.; Rühl, S.; and Wolverton, C. 2015. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Computational Materials*, 1(1): 1–15.

Liang, L.; Hu, X.; Xu, Z.; Song, Z.; and King, I. 2023a. Predicting Global Label Relationship Matrix for Graph Neural Networks under Heterophily. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Liang, L.; Xu, Z.; Song, Z.; King, I.; Qi, Y.; and Ye, J. 2023b. Tackling Long-tailed Distribution Issue in Graph Neural Networks via Normalization. *IEEE Transactions on Knowledge and Data Engineering*.

Lin, Y.; Yan, K.; Luo, Y.; Liu, Y.; Qian, X.; and Ji, S. 2023. Efficient Approximations of Complete Interatomic Potentials for Crystal Property Prediction. In Krause, A.; Brunskill, E.; Cho, K.; Engelhardt, B.; Sabato, S.; and Scarlett, J., eds., *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, 21260–21287. PMLR.

Liu, J.; Zhou, M.; Fournier-Viger, P.; Yang, M.; Pan, L.; and Nouioua, M. 2022. Discovering Representative Attribute-stars via Minimum Description Length. In *ICDE*, 68–80. IEEE.

Liu, S.; Du, W.; Ma, Z.-M.; Guo, H.; and Tang, J. 2023a. A Group Symmetric Stochastic Differential Equation Model for Molecule Multi-modal Pretraining. In Krause, A.; Brunskill, E.; Cho, K.; Engelhardt, B.; Sabato, S.; and Scarlett, J., eds., *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, 21497–21526. PMLR.

Liu, X.; Zhang, F.; Hou, Z.; Mian, L.; Wang, Z.; Zhang, J.; and Tang, J. 2023b. Self-Supervised Learning: Generative or Contrastive. *IEEE Trans. Knowl. Data Eng.*, 35(1): 857–876.

Louis, S.-Y.; Zhao, Y.; Nasiri, A.; Wang, X.; Song, Y.; Liu, F.; and Hu, J. 2020. Graph convolutional neural networks with global attention for improved materials property prediction. *Physical Chemistry Chemical Physics*, 22(32): 18141–18148.

Ma, Y.; Song, Z.; Hu, X.; Li, J.; Zhang, Y.; and King, I. 2023. Graph Component Contrastive Learning for Concept Relatedness Estimation. In *AAAI*, 13362–13370. AAAI Press.

Nagy, Á. 1998. Density functional. Theory and application to atoms and molecules. *Physics Reports*, 298(1): 1–79.

Oganov, A. R.; Pickard, C. J.; Zhu, Q.; and Needs, R. J. 2019. Structure prediction drives materials discovery. *Nature Reviews Materials*, 4(5): 331–348.

Orio, M.; Pantazis, D. A.; and Neese, F. 2009. Density functional theory. *Photosynthesis research*, 102: 443–453.

Pandey, K.; Mukherjee, A.; Rai, P.; and Kumar, A. 2022. DiffuseVAE: Efficient, Controllable and High-Fidelity Generation from Low-Dimensional Latents. *Trans. Mach. Learn. Res.*, 2022.

Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; and Wolverton, C. 2013. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). *Jom*, 65: 1501–1509.

Satorras, V. G.; Hoogeboom, E.; and Welling, M. 2021. E(n) Equivariant Graph Neural Networks. In *ICML*, volume 139 of *Proceedings of Machine Learning Research*, 9323–9332. PMLR.

Schütt, K. T.; Glawe, H.; Brockherde, F.; Sanna, A.; Müller, K.-R.; and Gross, E. K. 2014. How to represent crystal structures for machine learning: Towards fast prediction of electronic properties. *Physical Review B*, 89(20): 205118.

Schütt, K. T.; Glawe, H.; Brockherde, F.; Sanna, A.; Müller, K. R.; and Gross, E. K. U. 2014. How to represent crystal structures for machine learning: Towards fast prediction of electronic properties. *Phys. Rev. B*, 89: 205118.

Song, Y.; and Ermon, S. 2020. Improved Techniques for Training Score-Based Generative Models. In *NeurIPS*.

Song, Y.; Sohl-Dickstein, J.; Kingma, D. P.; Kumar, A.; Ermon, S.; and Poole, B. 2021a. Score-Based Generative Modeling through Stochastic Differential Equations. In *ICLR*. OpenReview.net.

Song, Z.; and King, I. 2022. Hierarchical Heterogeneous Graph Attention Network for Syntax-Aware Summarization. In *AAAI*, 11340–11348. AAAI Press.

Song, Z.; Ma, Y.; and King, I. 2022. Individual Fairness in Dynamic Financial Networks. In *NeurIPS 2022 Workshop: New Frontiers in Graph Learning*.

Song, Z.; Meng, Z.; Zhang, Y.; and King, I. 2021b. Semi-supervised Multi-label Learning for Graph-structured Data. In *CIKM*, 1723–1733. ACM.

Song, Z.; Zhang, Y.; and King, I. 2022. Towards an Optimal Asymmetric Graph Structure for Robust Semi-supervised Node Classification. In *KDD*, 1656–1665. ACM.

Song, Z.; Zhang, Y.; and King, I. 2023a. No Change, No Gain: Empowering Graph Neural Networks with Expected Model Change Maximization for Active Learning. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Song, Z.; Zhang, Y.; and King, I. 2023b. Optimal Blockwise Asymmetric Graph Construction for Graph-based Semi-supervised Learning. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Song, Z.; Zhang, Y.; and King, I. 2023c. Towards Fair Financial Services for All: A Temporal GNN Approach for Individual Fairness on Transaction Networks. In *CIKM*, 2331–2341. ACM.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is All you Need. In *NIPS*, 5998–6008.

Wieder, O.; Kohlbacher, S.; Kuenemann, M.; Garon, A.; Ducrot, P.; Seidel, T.; and Langer, T. 2020. A compact review of molecular property prediction with graph neural networks. *Drug Discovery Today: Technologies*, 37: 1–12.

Xie, T.; and Grossman, J. C. 2018. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.*, 120: 145301.

Xu, M.; Yu, L.; Song, Y.; Shi, C.; Ermon, S.; and Tang, J. 2022. GeoDiff: A Geometric Diffusion Model for Molecular Conformation Generation. In *ICLR*. OpenReview.net.

Yan, K.; Liu, Y.; Lin, Y.; and Ji, S. 2022. Periodic Graph Transformers for Crystal Material Property Prediction. In *NeurIPS*.

Yang, M.; Zhou, M.; Pan, L.; and King, I. 2023a. κHGCN: Tree-likeness Modeling via Continuous and Discrete Curvature Learning. In *KDD*, 2965–2977. ACM.

Yang, M.; Zhou, M.; Ying, R.; Chen, Y.; and King, I. 2023b. Hyperbolic Representation Learning: Revisiting and Advancing. In *ICML*, volume 202 of *Proceedings of Machine Learning Research*, 39639–39659. PMLR.

Yu, H.; Song, Y.; Hu, J.; Guo, C.; and Yang, B. 2023. A Crystal-Specific Pre-Training Framework for Crystal Material Property Prediction. *CoRR*, abs/2306.05344.

Zaidi, S.; Schaarschmidt, M.; Martens, J.; Kim, H.; Teh, Y. W.; Sanchez-Gonzalez, A.; Battaglia, P. W.; Pascanu, R.; and Godwin, J. 2023. Pre-training via Denoising for Molecular Property Prediction. In *ICLR*. OpenReview.net.

Zhang, M.; Qamar, M.; Kang, T.; Jung, Y.; Zhang, C.; Bae, S.; and Zhang, C. 2023a. A Survey on Graph Diffusion Models: Generative AI in Science for Molecule, Protein and Material. *CoRR*, abs/2304.01565.

Zhang, Y.; Chen, Y.; Song, Z.; and King, I. 2023b. Contrastive Cross-scale Graph Knowledge Synergy. In *KDD*, 3422–3433. ACM.

Zhang, Y.; Zhu, H.; Song, Z.; Koniusz, P.; and King, I. 2022. COSTA: Covariance-Preserving Feature Augmentation for Graph Contrastive Learning. In *KDD*, 2524–2534. ACM.

Zhang, Y.; Zhu, H.; Song, Z.; Koniusz, P.; and King, I. 2023c. Spectral Feature Augmentation for Graph Contrastive Learning and Beyond. In *AAAI*, 11289–11297. AAAI Press.

Zhang, Y.; Zhu, H.; Song, Z.; Koniusz, P.; King, I.; et al. 2023d. Mitigating the Popularity Bias of Graph Collaborative Filtering: A Dimensional Collapse Perspective. In *Thirty-seventh Conference on Neural Information Processing Systems*.