arXiv:2501.18850v2 [cs.CE] 29 Aug 2025

# Equivariant Hypergraph Diffusion for Crystal Structure Prediction

**Yang Liu** [1] **Chuan Zhou** [1 2] **Shuai Zhang** [1] **Peng Zhang** [3] **Xixun Lin** [4] **Shirui Pan** [5]

## Abstract

Crystal Structure Prediction (CSP) remains a fundamental challenge with significant implications for the development of new materials and the advancement of various scientific disciplines. Recent developments have shown that generative models, particularly diffusion models, hold great promise for CSP. However, traditional graph-based representations, where atomic bonds are modeled as pairwise graph edges, fail to fully capture the intricate high-order interactions essential for accurately representing crystal structures. In this work, we propose a novel approach that utilizes hypergraphs to represent crystal structures, providing a more expressive abstraction for modeling multi-way atomic interactions. By adopting hypergraphs, we can effectively capture complex high-order relationships and symmetries, such as permutation and periodic translation invariance, which are crucial for characterizing crystal structures. In this work, we propose the **E**quivariant **H**ypergraph **Diff**usion Model (**EH-Diff**), a generative model designed to take advantage of the symmetry-preserving properties of hypergraphs. EH-Diff exploits these features to offer an efficient and accurate method for predicting crystal structures with a strong theoretical justification to preserve invariance properties. Empirically, we conduct extensive experiments on four benchmark datasets, and the results demonstrate that EH-Diff outperforms state-of-the-art CSP methods with only one sample.

## 1. Introduction

Crystal Structure Prediction (CSP), a long-standing challenge in the physical sciences since the early 2000s (Desiraju, 2002; Woodley & Catlow, 2008), aims to determine the stable 3D structure of a compound based solely on its chemical composition. Given the fundamental role of crystals in a wide range of materials (Steiger, 2005; Cartwright & Mackay, 2012), accurately predicting their spatial configuration is crucial (Wooster, 1953; Sato, 2016) for understanding the physical and chemical properties that govern their applications in various academic and industrial domains, including drug development (Zhang et al., 2023; Bannan et al., 2025), battery technologies (Xie & Grossman, 2018; Yin et al., 2024), and catalysis (Bao et al., 2023).

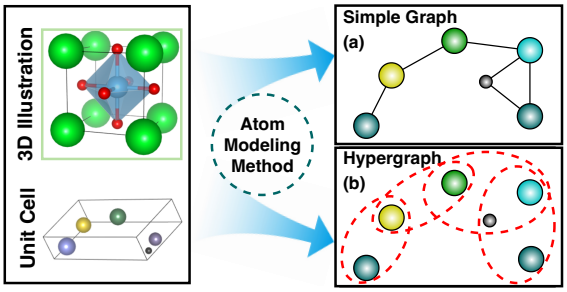



*Figure 1.* Illustration of Crystal Structure Modeling. (a) Simple Graph and (b) Hypergraph. Interference on the hypergraph demonstrates effects involving the atom, with pairwise (first-order) interference evident in (a), while higher-order (non-pairwise) interference is observed in (b).

Conventional CSP methods mainly use density functional theory (DFT) (Kohn & Sham, 1965) to iteratively evaluate the system energies. Techniques like random search (Pickard & Needs, 2011) and Bayesian optimization (Yamashita et al., 2018) explore the energy landscape to identify stable local or global minima (Oganov et al., 2019). Predicting crystal properties remains challenging because of the inherent complexity of crystalline materials. Crystalline structures, characterized by periodic repetition of a unit cell (Yan et al., 2022) in a 3D lattice, feature highly ordered atomic arrangements. The physical properties of these materials, including electronic, mechanical, and thermal characteristics, depend on their atomic configurations.

Recent advances in AI4Material research have utilized graph-based deep learning models (Xie & Grossman, 2018; Xie et al., 2021), with a particular focus on diffusion models (Lin et al., 2024; Jiao et al., 2024). These models employ graph-based diffusion techniques to generate multiple atoms

[1]Academy of Mathematics and Systems Science, Chinese Academy of Sciences [2]School of Cyber Security, University of Chinese Academy of Sciences [3]Guangzhou University [4]Institute of Information Engineering, Chinese Academy of Sciences [5]Griffith University. Correspondence to: Chuan Zhou <zhouchuan@amss.ac.cn>.

Preprint. Work in progress.

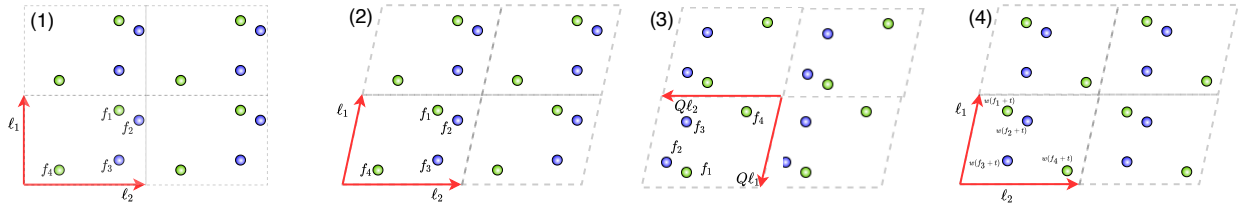



*Figure 2.* Crystal Structure Illustration. (1) crystal structure in a Cartesian coordinate system; (2) crystal structure in a fractional coordinate system. (2)→(3) orthogonal transformation of the lattice vectors; (2)→(4) periodic translation of the fractional coordinates.

while preserving the invariant properties within the models. However, traditional graph-based approaches, including graph neural networks and graph transformers, are restricted to pairwise relationships (Kim et al., 2022), where atoms are represented as nodes and atomic bonds as edges. This *limitation* restricts their ability to model higher-order interactions, which are crucial to capturing the structural complexity of crystal materials. Crystal structures often exhibit complex interactions, such as shared lattice points and multi-atomic coordination (Voet & Rich, 1970; Grimmer et al., 1974), which are difficult to represent comprehensively using only pairwise edges.

To overcome the limitation above, we employ *hypergraphs*, a robust graph data structure capable of capturing high-order relationships, to model the structure of crystal materials. Hypergraphs offer significant advantages over traditional graph representations by naturally encoding multi-atom interactions through hyperedges that link more than two atoms. In this framework, the model can offer a more expressive means of capturing the intricate coordination and structural dependencies within crystal structures. Hypergraphs represent atomic arrangements, enabling analysis of crystal stability, coordination geometry, and bond distributions. Their ability to incorporate symmetries and periodicity makes them ideal for modeling the fundamental properties and behaviors of crystalline materials.

However, several *challenges* exist in the introduction of hypergraphs into CSP with generative models. First, a major challenge is effectively constructing hypergraph-based material representations. The next challenge involves developing diffusion models that operate on hypergraph data. Unlike traditional graph-based methods, hypergraph representations must capture complex multi-atomic interactions while preserving key structural invariants, such as permutation and periodic translation symmetries inherent in crystalline materials. Moreover, designing a hypergraph-based diffusion model is critical, as it introduces additional complexities. The model must simultaneously ensure that these symmetries are maintained throughout both the forward and reverse diffusion processes. This necessitates the integration of equivariant architectures capable of handling the high-order interactions encoded in hypergraphs, thereby ensuring accurate and symmetry-consistent predictions of crystal structures.

In this work, we address the CSP problem by using a hypergraph-based modeling approach and we propose the Equivariant Hypergraph Diffusion Model, a novel framework designed to capture high-order interactions while preserving intrinsic symmetries and invariance properties of crystal structures (defined in Section 2.3). It simultaneously generates the lattice vectors and fractional coordinates of all atoms by employing a proposed denoising hypergraph model with periodic translation (defined in Section 3). Through an iterative process of expansion and refinement, our framework captures complex structural dependencies and trains a denoising diffusion model specifically for CSP, while preserving invariance properties (theoretical justification provided in Section 3.2). Empirically, we conduct comprehensive experiments across four real-world datasets and results position EH-Diff as a reliable approach for modeling complex crystalline materials, outperforming simple graph-based diffusion models.

Our contributions are summarized as follows:

- We formalize the Crystal Structure Prediction (CSP) problem by introducing a hypergraph modeling framework to describe high-order atomic interactions. To the best of our knowledge, this is the first work to represent crystal structures beyond traditional pairwise relationships, offering a novel perspective on capturing complex atomic coordination.
- We propose the **E**quivariant **H**ypergraph **Diff**usion Model (**EH-Diff**), a novel framework that integrates hypergraph-based representation learning with a diffusion process. EH-Diff effectively captures high-order atomic interactions while explicitly preserving key symmetries, including permutation and periodic translation invariance, which are fundamental properties of crystal structures.
- We evaluate the proposed framework through extensive experiments on four benchmark datasets, demonstrating its superior performance compared to state-of-the-art generative models with only one sample.

## 2. Problem Setup and Analysis

In this section, we formally define the problem and provide a concise theoretical analysis of the problem. Some basic

mathematical definitions are given in Appendix A.

### 2.1. Representation of Crystal Structure

We introduce two frameworks for representing crystal structures and define the CSP task, providing a foundation for modeling atomic configurations in crystal lattices and predicting material behaviors.

**Cartesian Coordinate System** A three-dimensional crystal is defined as an infinite periodic arrangement of atoms in a three-dimensional space, characterized by the smallest repeating unit called the unit cell. The unit cell can be represented by a triplet $\mathbf{M}_C = (\mathbf{A}, \mathbf{X}, \mathbf{L})$, where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N] \in \mathbb{R}^{h \times N}$ denotes the set of one-hot encoded representations of atom types, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N] \in \mathbb{R}^{3 \times N}$ represents the Cartesian coordinates of the atoms, and $\mathbf{L} = [\ell_1, \ell_2, \ell_3] \in \mathbb{R}^{3 \times 3}$ is the lattice matrix, which contains three basis vectors that describe the periodicity of the crystal. The infinite periodic structure of the crystal can be formally expressed as:

$$\{(\mathbf{a}_i', \mathbf{x}_i') | \mathbf{a}_i' = \mathbf{a}_i,\ \mathbf{x}_i' = \mathbf{x}_i + \mathbf{L}\mathbf{k},\ \forall \mathbf{k} \in \mathbb{Z}^3\}, \quad (1)$$

where the $j$-th element of the integer vector $\mathbf{k}$ represents the integer translation along the $j$-th basis vector $\ell_j$ in three-dimensional space for $j \in \{1, 2, 3\}$.

**Fractional Coordinate System** Analogous to the Cartesian coordinate system, a unit cell can also be defined as a triplet $\mathbf{M}_F = (\mathbf{A}, \mathbf{X}, \mathbf{L})$ in the fractional coordinate system. In this system, a point represented by the fractional coordinate vector $\mathbf{f} = [f_1, f_2, f_3]^\top \in [0, 1)^3$ corresponds to the Cartesian coordinate vector $\mathbf{x}$, given by $\mathbf{x} = \sum_{i=1}^{3} f_i \ell_i$, where $\ell_i$ denotes the $i$-th basis vector of the lattice.

**Remark 2.1.** *We adopt the fractional coordinate system, representing the crystal as $\mathbf{M} = \mathbf{M}_C = (\mathbf{A}, \mathbf{F}, \mathbf{L})$, where the fractional coordinates of atoms in the unit cell form the matrix $\mathbf{F} \in [0, 1)^{3 \times N}$. The two frameworks are shown in Figure 2.*

**Remark 2.2.** *The Cartesian coordinate system $\mathbf{X}$ uses three standard orthogonal basis vectors as coordinate axes. In crystallography, the fractional coordinate system is commonly employed to capture the periodicity of crystal structures (Nouira et al., 2018; Ren et al., 2022), using lattice vectors $(\ell_1, \ell_2, \ell_3)$ as a basis.*

**Crystal Structure Prediction (CSP)** CSP predicts the lattice matrix $\mathbf{L}$ and the fractional coordinate matrix $\mathbf{F}$ from the chemical composition $\mathbf{A}$. The goal is to learn the conditional probability distribution $p(\mathbf{L}, \mathbf{F} | \mathbf{A})$, where $\mathbf{L}$ represents the $3 \times 3$ lattice matrix that defines the unit cell's periodic structure, and $\mathbf{F}$ is the $3 \times N$ fractional coordinate matrix, where each column represents the fractional coordinates of an atom within the unit cell. Thus, CSP involves learning $p(\mathbf{L}, \mathbf{F} | \mathbf{A})$ such that: $p(\mathbf{L}, \mathbf{F} | \mathbf{A}) = \prod_{i=1}^{N} p(\mathbf{f}_i | \mathbf{A}, \mathbf{L})$, where $\mathbf{f}_i$ denotes the fractional coordinates of the $i$-th atom, and the prediction of $\mathbf{L}$ and $\mathbf{F}$ minimizes the error between generated and true atomic configurations.

### 2.2. Hypergraph Modeling of Crystal Structure

**Definition 2.3** (Hypergraph)**.** *Consider a set of individuals $\mathbf{V} = \{\mathbf{v}_i\}_{i=1}^{n}$ connected by hyperedges $\mathbf{E} = \{\mathbf{e}_k\}_{k=1}^{m}$, forming a hypergraph $\mathbf{H} = \{\mathbf{V}, \mathbf{E}\}$ consisting of $n$ nodes and $m$ hyperedges, where each hyperedge can connect any number of nodes.*

Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^T \in \mathbb{R}^n$ represent the node attributes, with $\mathbf{x}_v$ indicating the attribute of node $v$. The degree of node $v$, denoted as $d_v$, is defined by $d_v = |\{\mathbf{e} \in \mathbf{E} : \mathbf{v} \in \mathbf{e}\}|$. Moreover, let $\mathbf{D}$ and $\mathbf{D}_e$ represent the diagonal degree matrix for nodes in $\mathbf{V}$ and the sub-matrix corresponding to nodes within a hyperedge $\mathbf{e}$, respectively. More formal definition of hypergraph can be found in Appendix A.

Building on the previous definition of hypergraph (Kim et al., 2022), we extend this concept and its key attributes to the field of crystallography. This extension provides a more detailed framework for understanding the structural and dynamic properties of crystalline materials, thereby introducing new analytical tools and methodologies that enhance the study of crystallography.

**Definition 2.4** (Hypergraph-based Crystal)**.** *Formally, let the hypergraph $\mathbf{H}_{crystal} = (\mathbf{M}, \mathbf{V}, \mathbf{E})$ represent a crystal unit cell, where $\mathbf{M}$ denotes the unit cell, $\mathbf{V}$ represents the set of atoms, and $\mathbf{E}$ denotes the set of hyperedges capturing the bonds or interactions between these units, potentially involving more than two atoms simultaneously.*

**Remark 2.5.** *The hypergraph-based framework models crystal lattices by representing atoms and interactions as nodes and hyperedges, enabling the exploration of complex multi-atom relationships beyond pairwise bonding. This approach facilitates studies in material science, including topological properties, collective atomic behaviors, and novel crystalline structures, advancing our understanding of the structural and dynamic aspects of crystalline materials.*

### 2.3. Symmetries of Crystal Structure

Crystals differ fundamentally from molecules in that they consist of a repeating unit cell organized within a regular three-dimensional lattice. This periodic structure endows crystals with unique quantum symmetries, including invariance under the orthogonal group $O(3)$ and periodic translational invariance within the fractional coordinate system. These symmetries play a critical role in defining the physical and quantum mechanical properties of crystalline materials, distinguishing them from discrete molecular entities.

**Definition 2.6** (Atom Permutation Invariance)**.** *For any permutation $\mathbf{P} \in \mathbb{S}_N$ (where $\mathbb{S}_N$ represents the symmetric group of degree $N$), the distribution $p(\mathbf{L}, \mathbf{F}|\mathbf{A})$ remains invariant under the transformation $p(\mathbf{L}, \mathbf{FP}|\mathbf{AP})$. This indicates that reordering the atoms does not affect the underlying distribution, implying the inherent permutation symmetry of the atomic configuration.*

**Definition 2.7** ($O(3)$ Invariance))**.** *For any orthogonal transformation $\mathbf{Q} \in \mathbb{R}^{3\times3}$ satisfying $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}$, the distribution $p(\mathbf{QL}, \mathbf{F}|\mathbf{A})$ is equivalent to $p(\mathbf{L}, \mathbf{F}|\mathbf{A})$. This invariance indicates that any rotation or reflection of $\mathbf{L}$ does not alter the distribution.*

**Definition 2.8** (Periodic Translation Invariance)**.** *For any translation $\mathbf{T} \in \mathbb{R}^{3\times1}$, the distribution $p(\mathbf{L}, w(\mathbf{F} + \mathbf{T1}^\top)|\mathbf{A})$ is equivalent to $p(\mathbf{L}, \mathbf{F}|\mathbf{A})$. Here, the function $w(\mathbf{F}) = \mathbf{F} - \lfloor\mathbf{F}\rfloor \in [0, 1)^{3\times N}$ returns the fractional part of each element in $\mathbf{F}$, and $\mathbf{1} \in \mathbb{R}^{3\times1}$ is a vector with all elements set to one. This invariance indicates that any periodic translation of $\mathbf{F}$ does not alter the distribution $\overline{p}$.*

**Proposition 2.9.** $\mathbf{H}_{crystal}$ *exhibits permutation invariance with respect to the potentials, reflecting the inherent symmetry properties of crystalline structures.*

**Remark 2.10.** *It should be noted that the defined crystal symmetries are consistent with those described in the Cartesian coordinate system (Xie et al., 2021; Yan et al., 2022). Specifically, rotations or reflections of the lattice $\mathbf{L}$ have no impact on the fractional coordinates of the atoms. Consequently, the invariance $E(3)$ under the Cartesian coordinate system can be reduced to the invariance $O(3)$ of $\mathbf{L}$ alone. Furthermore, since the fractional coordinate system inherently reflects the periodicity of the crystal structure, we can derive a more straightforward formulation of periodic translation invariance without the need to introduce corner points, as previously required under the Cartesian coordinate system.*

## 3. The Proposed Model: EH-Diff

### 3.1. Diffusion Model on Crystal Structure

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

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

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

**Diffusion on L.** We employ the Denoising Diffusion Probabilistic Model (DDPM) (Ho et al., 2020) to diffuse and generate $\mathbf{L}$. The forward diffusion process is defined as a Markov chain $q(\mathbf{L}_t|\mathbf{L}_{t-1}) = \mathcal{N}(\mathbf{L}_t; \sqrt{1-\beta_t}\mathbf{L}_{t-1}, \beta_t\mathbf{I})$, where the variance schedule $\{\beta_t\}_{t=1}^T$ is fixed with $\beta_t \in (0, 1)$. This process progressively transforms $L_0$ into the normal prior distribution $p(\mathbf{L}_T) = \mathcal{N}(0, \mathbf{I})$ for a sufficiently large $T$, since $q(\mathbf{L}_t|\mathbf{L}_0) = \mathcal{N}(\mathbf{L}_t; \sqrt{\alpha_t}\mathbf{L}_0, (1-\alpha_t)\mathbf{I})$. For the reverse generation process, we model the reverse dynamics as a conditional Markov chain with learnable transitions $p_\theta(\mathbf{L}_{t-1}|\mathbf{L}_t) = \mathcal{N}(\mathbf{L}_{t-1}; \mu_\theta(\mathbf{L}_t, t), \Sigma_\theta(\mathbf{L}_t, t))$. Here, the mean $\mu_\theta$ is given by $\mu_\theta(\mathbf{L}_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{L}_t - \frac{\beta_t}{\sqrt{1-\alpha_t}}\hat{\epsilon}_\mathbf{L}(\mathbf{L}_t, t)\right)$, and the denoising term $\hat{\epsilon}_\mathbf{L}$ is provided by the denoising model $\phi_\theta(\mathbf{A}, \mathbf{F}_t, \mathbf{L}_t, t)$, which will be detailed later. The training loss is formulated as the expected $\ell_2$-distance between the ground truth $\epsilon_\mathbf{L}$ and the estimated $\hat{\epsilon}_\mathbf{L}$:

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

### 3.2. Denoising Model

Here we introduce our well-designed denoising model $\phi(\mathbf{L}, \mathbf{F}, \mathbf{A}, t)$ that produces $\hat{\epsilon}_\mathbf{L}$ and $\hat{\epsilon}_\mathbf{F}$. Let $\mathbf{H}^{(\ell)} = \left[\mathbf{h}_1^{(\ell)}, \mathbf{h}_2^{(\ell)}, \cdots, \mathbf{h}_N^{(\ell)}\right]$ denote the node representations of the $\ell$-th layer. The input feature is given by $\mathbf{h}_\ell^{(0)} = \mathrm{MLP}(f_{\text{atom}}(\mathbf{a}_i), f_{\text{position}}(t))$, where $f_{\text{atom}}$ and $f_{\text{position}}$ are the atomic embedding and sinusoidal positional encoding (Ho et al., 2020), respectively.

Inspired by the Equivariant Hypergraph Neural Networks (EHNN) framework proposed by Kim et al. (2022), we construct our model using an unfolded layer architecture. The $\ell$-th layer of this framework (EHNN-MLP) can be formulated as follows:

$$\begin{cases} \mathbf{m}_\mathbf{i}^{(\ell)} = \mathrm{MLP}\left(\{\mathbf{h}_i^{(\ell-1)}\}_{i\in\mathbf{i}}, \mathbf{L}^\top\mathbf{L}, \psi(\sum_{i,j\in\mathbf{i}}(f_j - f_i))\right), \\ \mathbf{m}_i^{(\ell)} = \sum_{i\in\mathbf{i}} \frac{\mathbf{m}_\mathbf{i}^{(\ell)}}{|\mathbf{i}|}, \\ \mathbf{h}_i^{(\ell)} = \mathbf{h}_i^{(\ell-1)} + \mathrm{MLP}\left(\mathbf{h}_i^{(\ell-1)}, \{\mathbf{m}_\mathbf{i}^{(\ell)}\}_{i\in\mathbf{i}}\right), \end{cases} \tag{4}$$

where $\psi : (-1, 1)^3 \to [-1, 1]^{3\times K}$ is a Fourier Transformation function and $\psi$ is periodic translation invariant.

**Algorithm 1** EH-Diff Training Procedure

**Input:** atom features $\mathbf{A}$, fractional coordinates $\mathbf{F}$, lattice $\mathbf{L}$, number of diffusion steps $T_{\text{diffusion}}$.
Sample $\epsilon_{\mathbf{L}} \sim \mathcal{N}(0, \mathbf{I})$;
Sample $\epsilon_{\mathbf{F}} \sim \mathcal{N}(0, \mathbf{I})$;
Sample $t \sim \mathcal{U}(1, T_{\text{diffusion}})$;
Update $\mathbf{L}_t \leftarrow \sqrt{\bar{\alpha}_t}\mathbf{L}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_{\mathbf{L}}$;
Update $\mathbf{F}_t \leftarrow w\left(\mathbf{F}_0 + \sigma_t \epsilon_{\mathbf{F}}\right)$;
Update $\hat{\epsilon}_{\mathbf{L}}, \hat{\epsilon}_{\mathbf{F}} \leftarrow \phi_\theta(\mathbf{A}, \mathbf{F}_t, \mathbf{L}_t, t)$, see Eq. (5);
Minimize $\mathcal{L}_{\mathbf{L}} + \mathcal{L}_{\mathbf{F}}$, see Eq. (2) and (3);
**Return:** Model $\phi_\theta$.

After $L$ layers of message passing conducted on the fully connected graph, the lattice noise $\hat{\epsilon}_L$ is acquired by a linear combination of $\mathbf{L}$, with the weights given by the final layer, and the fractional coordinate score $\hat{\epsilon}_{\mathbf{F}}$ is output by $\hat{\epsilon}_{\mathbf{F}}[:, i] = \varphi_{\mathbf{F}}(\mathbf{h}_i^{(\mathbf{L})})$. We state them as follows:

$$\hat{\epsilon}_{\mathbf{L}} = \mathbf{L}\varphi\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{h}_i^{(L)}\right), \hat{\epsilon}_{\mathbf{F}} = \left[\varphi_F(h_i^L)\right]_{i=1}^{N}, \quad (5)$$

where $\varphi$ is an MLP with output shape as $3 \times 3$, $\hat{\epsilon}_{\mathbf{F}}[:, i]$ defines the $i$-th column of $\hat{\epsilon}_{\mathbf{F}}$, and $\varphi_F$ is an MLP on the final representation.

**Proposition 3.1.** *Propositions regarding $\hat{\epsilon}_{\mathbf{L}}$ and $\hat{\epsilon}_{\mathbf{F}}$:*

- *$\hat{\epsilon}_{\mathbf{L}}$ is $O(3)$ equivariant;*
- *$\hat{\epsilon}_{\mathbf{F}}$ is periodic translation invariant.*

*Hence, the generated distribution by EH-Diff is periodic $E(3)$ invariant.*

**Proposition 3.2.** *If $\hat{\epsilon}_{\mathbf{L}}$ exhibits $O(3)$ equivariance, then the corresponding learned marginal distribution $p_\theta(\mathbf{L}_0)$ will be invariant under $O(3)$ transformations. Similarly, if $\hat{\epsilon}_{\mathbf{F}}$ is invariant in periodic translations, the learned marginal distribution $p_\theta(\mathbf{F}_0)$ will also maintain periodic translation invariance.*

The full proofs of Proposition 3.1 and Proposition 3.2 can be found in Appendix D.

### 3.3. CSP Workflow

**Training via Hypergraph Diffusion.** Algorithm 1 outlines the training procedure for the EH-Diff model. The process begins by taking as input the atom features, fractional coordinates, and lattice of the crystal, along with the number of diffusion steps. Initially, noise is sampled for both the lattice and the fractional coordinates from a normal distribution. A random diffusion step is selected from a uniform distribution. The lattice and coordinates are then updated using their respective diffusion equations, incorporating noise to approximate their final states. Next, the model parameters are updated using a denoising function, and the objective is to minimize the combined loss functions for the lattice and fractional coordinates. This iterative training process continues until the model effectively captures the desired crystal properties, thus providing a learned representation capable of predicting complex crystal structures.

**Algorithm 2** EH-Diff Inference Procedure

**Input:** atom features $\mathbf{A}'$, fractional coordinates $\mathbf{F}'$, lattice $\mathbf{L}'$, empirical mean $\bar{y}$, number of diffusion steps, $T_{\text{diffusion}}$, trained model $\phi_\theta$, hyperparameters $\bar{\alpha}_t, \tilde{\beta}_t, \gamma_0, \gamma_1, \gamma_2$.
Sample $y_0 \sim \mathcal{N}(\bar{y}, 1)$.
**for** $t = T_{\text{diffusion}}$ to 1 **do**
  **if** $t > 1$ **then**
    Sample $\epsilon_y \sim \mathcal{N}(0, 1)$.
  **end if**
  $\hat{\epsilon}_y \leftarrow \phi_\theta(\mathbf{A}', \mathbf{F}', \mathbf{L}', y_t, t)$ via trained model $\phi_\theta$.
  $\hat{y}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(y_t - (1-\sqrt{\bar{\alpha}_t})\bar{y} - \sqrt{1-\bar{\alpha}_t}\hat{\epsilon}_y\right)$.
  $y_{t-1} = \gamma_0\hat{y}_0 + \gamma_1 y_t + \gamma_2\bar{y} + \sqrt{\tilde{\beta}_t}\epsilon_y$.
**end for**
**Return:** Prediction $\hat{y}_0$.

**Inference.** Algorithm 2 presents the inference procedure for the EH-Diff model, which is used to predict the crystal properties. The process starts with the input of atom features, fractional coordinates, lattice information, an empirical mean, hyperparameters, and the number of diffusion steps. Initially, a sample is drawn from a normal distribution centered on the empirical mean. The diffusion steps are then performed in reverse order, starting from the maximum step to one. For each step, the noise is sampled if the current step is greater than one, and the predicted value is updated using the denoising function derived from the trained model. At each step, the prediction is refined by adjusting the combination of the previous state, mean, and noise components, modulated by hyperparameters. The final output, after all of the diffusion steps have been completed, is the predicted value representing the crystal property of interest.

## 4. Experiments

### 4.1. Experimental Setup

**Dataset.** We conducted experiments on four datasets, each with distinct levels of complexity. The first dataset, Perov-5, comprises 18,928 perovskite materials with structurally similar configurations, containing 5 atoms per unit cell (Castelli et al., 2012a;b). The second dataset, Carbon-24, includes 10,153 carbon-based materials, with unit cells ranging from 6 to 24 atoms. The third dataset, MP-20, is derived from the Materials Project and contains 45,231 stable inorganic mate-

*Table 1.* Results on Stable Structure Prediction Task. *MR* denotes the match rate. We use **boldface** to denote the best result and underline to denote the second best result. "-" indicates missing data in original papers.

| Method | *num of Samples* | Perov-5 | | Carbon-24 | | MP-20 | | MPTS-52 | |
|---|---|---|---|---|---|---|---|---|---|
| | | *MR* ↑ | *RMSE* ↓ | *MR* ↑ | *RMSE* ↓ | *MR* ↑ | *RMSE* ↓ | *MR* ↑ | *RMSE* ↓ |
| **RS** | 20 | 29.22 | 0.2924 | 14.63 | 0.4041 | 8.73 | 0.2501 | 2.05 | 0.3329 |
| | 5,000 | 36.56 | 0.0886 | 14.63 | 0.4041 | 11.49 | 0.2822 | 2.68 | 0.3444 |
| **BO** | 20 | 21.03 | 0.2830 | 0.44 | 0.3653 | 8.11 | 0.2402 | 2.05 | 0.3024 |
| | 5,000 | 55.09 | 0.2037 | 12.17 | 0.4089 | 12.68 | 0.2816 | 6.69 | 0.3444 |
| **PSO** | 20 | 20.90 | 0.0836 | 6.40 | 0.4204 | 4.05 | 0.1567 | 1.06 | 0.2339 |
| | 5,000 | 21.88 | 0.0844 | 6.50 | 0.4211 | 4.35 | 0.1670 | 1.09 | 0.2390 |
| **P-cG-SchNet** | 1 | 48.22 | 0.4179 | 17.29 | 0.3846 | 15.39 | 0.3762 | 3.67 | 0.4115 |
| **CDVAE** | 1 | 45.31 | 0.1138 | 17.09 | 0.2969 | 33.90 | 0.1045 | 5.34 | 0.2106 |
| **EquiCSP** | 1 | 52.02 | 0.0707 | - | - | 57.59 | 0.0510 | 14.85 | **0.1169** |
| **DiffCSP** | 1 | 52.02 | 0.0760 | 17.54 | 0.2759 | 51.49 | 0.0631 | 12.19 | 0.1786 |
| **EH-Diff** (Sphere) | 1 | **54.59** | 0.0728 | 17.28 | **0.2623** | 56.81 | 0.0559 | 13.97 | 0.1285 |
| **EH-Diff** (Cube) | 1 | 53.43 | **0.0706** | **18.11** | 0.2757 | **58.12** | **0.0460** | **16.99** | 0.1340 |

rials, which include most of the experimentally synthesized compounds with up to 20 atoms per unit cell (Jain et al., 2013). Lastly, the MPTS-52 dataset represents a more challenging extension of MP-20, consisting of 40,476 structures with up to 52 atoms per unit cell, ordered chronologically by their earliest publication year. For the Perov-5, Carbon-24 and MP-20 data sets, we adhered to the 0.3/0.1/0.1 split as suggested by Xie et al. (2021). For the MPTS-52 dataset, a chronological split of 1/2/3 was used for training, validation, and testing, respectively.

**Baseline.** This study distinguishes itself from previous research by focusing on a comparison of two main methodologies. The first methodology follows a sequential predict-and-optimize strategy. In this approach, a model is initially trained to predict material properties, and then optimization algorithms are employed to determine the most favorable structures. We adapt the approach described by Cheng et al. (2022), using MEGNet (Chen et al., 2019) to predict formation energies. To convert these predictions into optimal structures, we apply three optimization techniques: random search (RS), Bayesian optimization (BO), and particle swarm optimization (PSO), each of which is run for 5,000 iterations to thoroughly explore the solution space. The second methodology utilizes generative models. Building on the enhancements proposed by Xie & Grossman (2018), we leverage the cG-SchNet framework (Gebauer et al., 2022), which incorporates SchNet (Schütt et al., 2018) as its fundamental model, along with a ground-truth lattice initialization to capture periodic information, resulting in the P-cG-SchNet model. We also explore conditional diffusion-based generative models such as CDVAE (Xie et al., 2021), EquiCSP (Lin et al., 2024) and DiffCSP (Jiao et al., 2024).

**Evaluation Metrics.** We evaluate the performance of different methods by reporting the Match Rate (*MR*) and Root Mean Square Error (*RMSE*), as shown in Table 1. Following established protocols (Xie et al., 2021; Jiao et al., 2024), the assessment involves comparing predicted candidates with ground truth structures. Specifically, for each structure in the test set, we generate $k$ samples of the same chemical composition and determine a match if at least one sample aligns with the ground-truth structure. This matching is carried out using the StructureMatcher class from pymatgen (Ong et al., 2013), with the thresholds set at stol $= 0.5$, $\text{angle}_{\text{tol}} = 10$, and ltol $= 0.3$. The *MR* is defined as the ratio of successfully matched structures to the total number of structures in the test set. The *RMSE* is calculated between the ground-truth structure and the best-matching candidate, normalized by $\sqrt{\frac{V}{n}}$, where $V$ denotes the lattice volume, and averaged across all matched structures. For optimization methods, we select the 20 lowest-energy structures from the 5,000 generated during testing as the candidates. In contrast, for generative baselines and our EH-Diff approach, we evaluate performance using $k = 1$.

**Hypergraphs as Crystals** To describe a crystal structure as a hypergraph, we can model the crystal as a hypergraph where each node represents an atom or a unit cell, and each hyperedge represents a bond or a cluster of atoms that interact with each other[1] which captures the complex interactions and relationships within the crystal structure, which may involve multiple atoms or unit cells interacting simultaneously. In addition, we present a few examples from each data set in Figure 3.

[1]In graph domain, it can be abstract as clique or star.

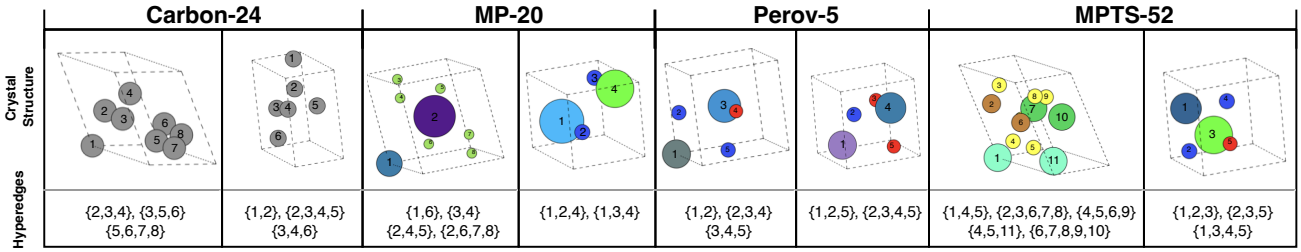



*Figure 3.* Visualization of the hypergraph-based crystal structure for each dataset and we list the corresponding hyperedges whose order are greater than one.

(1) The nodes represent atoms or unit cells in the crystal.

(2) Hyperedges represent interacting sets of atoms or unit cells, including chemical bonds, van der Waals forces, hydrogen bonds, and others.

(3) Hyperedges are assigned weights, and stronger interactions (e.g. covalent bonds) have higher weights than weaker ones (e.g. van der Waals).

**Hyperedge Details (Data Processing).** The modeling of hyperedges in crystal structures is detailed in Figure 4, where we utilize either a sphere or a cube to scan a localized region of space. A hyperedge is established when certain atoms are encompassed within the same sphere or cube. In the left panel, the sphere-based hyperedge is depicted. Here, a spherical volume is employed to scan the space. If multiple atoms are found within this spherical region, they are considered to be part of the same hyperedge. This method leverages the spherical geometry to define the connectivity between atoms, offering a more flexible and potentially more accurate representation of atomic interactions, particularly in cases where the atomic distribution is not strictly uniform or regular. The right panel shares similar characteristics with the cube-based hyperedge. Both methods—sphere-based and cube-based hyperedges—provide different ways to model the connectivity within crystal structures, each with its own advantages depending on the specific characteristics of the atomic arrangement being studied.

**Remark 4.1** (Symmetry and Periodicity)**.** *Crystals have inherent symmetry and periodicity, which can be modeled in a hypergraph. The repeating unit cell of a crystal can be represented by a subgraph, with nodes for atoms and hyperedges for atomic bond. By applying symmetry operations such as rotations and reflections, the hypergraph can be transformed to generate the entire crystal structure from this subgraph, effectively capturing both the periodicity and symmetry of the crystal in a compact form.*

### 4.2. Main Results

Table 1 demonstrates the superior performance of EH-Diff (Sphere and Cube) in the Stable Structure Prediction Task compared to other methods in four datasets. EH-Diff consistently achieves the best or second-best Match Rate and RMSE with only one sample, outperforming both optimization-based methods (RS, BO, PSO) that require thousands of samples and generative models like CDVAE and EquiCSP. Specifically, EH-Diff (Cube) excels in complex datasets such as MP-20 and MPTS-52, achieving the highest MR (58.12 and 16.99, respectively) and the lowest RMSE (0.0460 and 0.1340).

*Table 2.* Overall results over the 15 selected compounds in MP-20.

| **Method** | *MR* ↑ | *Avg.RMSE* ↓ | *Avg.Time* ↓ |
|---|---|---|---|
| USPEX | 53.33 | 0.0159 | 12.5h |
| DiffCSP | 73.33 | 0.0172 | **10s** |
| EH-Diff (Sphere) | **82.69** | **0.0092** | 13 ∼ 14s |
| EH-Diff (Cube) | 79.77 | 0.0109 | 13 ∼ 14s |

Table 2 shows that we achieve superior accuracy with competitive runtimes. EH-Diff (Sphere) has the highest MR (82.69) and lowest RMSE (0.0092), while maintaining an efficient runtime of 13∼14 seconds, comparable to DiffCSP's 10 seconds. This shows that EH-Diff provides significantly better prediction accuracy without sacrificing efficiency.

### 4.3. Results with Ab Initio Crystal Generation

Although our method is initially designed to tackle CSP with fixed composition $\mathbf{A}$, it can be extended to address the ab initio generation task by also generating $\mathbf{A}$. This extension is achieved by optimizing the one-hot representation of $\mathbf{A}$ using a DDPM-based approach. More details can be found in Appendix B[2]. In addition, we define a ranking score for each model as "Rank Score" $= \frac{\sum_{m\in\mathcal{M}} \text{rank}_m}{|\mathcal{M}|}$, where $\mathcal{M}$ represents the set of metrics and $\text{rank}_m$ denotes the rank of the model according to the metric $m$.

[2]Composition-based metrics are irrelevant for Carbon-24, as its structures consist solely of carbon. Thus, the corresponding results in Table 3 are represented by a "-" symbol.

*Table 3. Results on ab initio generation task.* We use **boldface** to denote the best result and underline to denote the second best result. We also use the **red** to denote the best rank score for the best model.

| Data | Method | Validity(%) ↑ | | Coverage(%) ↑ | | Property ↓ | | | Rank ↓ |
|---|---|---|---|---|---|---|---|---|---|
| | | Structure | Composition | COV-Recall | COV-Precision | $d_\rho$ | $d_E$ | $d_{\text{elem}}$ | Score |
| Perov-5 | FTCP | 0.24 | 54.24 | 0.00 | 0.00 | 10.27 | 156.0 | 0.6297 | 9.9 |
| | Cond-DFC-VAE | 73.60 | 82.95 | 73.92 | 10.13 | 2.268 | 4.111 | 0.8373 | 8.4 |
| | G-SchNet | 99.92 | 98.79 | 0.18 | 0.23 | 1.625 | 4.746 | 0.0368 | 7.0 |
| | P-G-SchNet | 79.63 | **99.13** | 0.37 | 0.25 | 0.2755 | 1.388 | 0.4552 | 6.7 |
| | CDVAE | **100.0** | 98.59 | 99.45 | 98.46 | 0.1258 | 0.0264 | 0.0628 | 5.0 |
| | SyMat | **100.0** | 97.40 | 99.68 | 98.64 | 0.1893 | 0.2364 | 0.0177 | 4.4 |
| | EquiCSP | **100.0** | 98.60 | 99.60 | 98.76 | 0.1110 | 0.0257 | 0.0503 | 3.3 |
| | DiffCSP | **100.0** | 98.85 | 99.74 | 98.27 | 0.1110 | 0.0263 | 0.0128 | 2.6 |
| | EH-Diff (Sphere) | **100.0** | 98.60 | 99.68 | **98.82** | 0.1116 | 0.0879 | **0.0118** | 2.9 |
| | EH-Diff (Cube) | **100.0** | 98.75 | **99.77** | **98.82** | **0.1103** | **0.0214** | 0.0620 | **2.1** |
| Carbon-24 | FTCP | 0.08 | - | 0.00 | 0.00 | 5.206 | 19.05 | - | 8.0 |
| | G-SchNet | 99.94 | - | 0.00 | 0.00 | 0.9427 | 1.320 | - | 7.0 |
| | P-G-SchNet | 48.39 | - | 0.00 | 0.00 | 1.533 | 134.7 | - | 7.8 |
| | CDVAE | **100.0** | - | 99.80 | 83.08 | 0.1407 | 0.2850 | - | 4.6 |
| | SyMat | **100.0** | - | **100.0** | 97.59 | 0.1195 | 3.9576 | - | 3.2 |
| | EquiCSP | **100.0** | - | 99.75 | 97.12 | 0.0734 | **0.0508** | - | 3.2 |
| | DiffCSP | **100.0** | - | 99.90 | 97.27 | 0.0805 | 0.0820 | - | 3.0 |
| | EH-Diff (Sphere) | **100.0** | - | 99.90 | **99.13** | 0.0655 | 0.0561 | - | 1.8 |
| | EH-Diff (Cube) | **100.0** | - | 99.90 | 98.03 | **0.0601** | 0.0533 | - | **1.6** |
| MP-20 | FTCP | 1.55 | 48.37 | 4.72 | 0.09 | 23.71 | 160.9 | 0.7363 | 8.9 |
| | G-SchNet | 99.65 | 75.96 | 38.33 | 99.57 | 3.034 | 42.09 | 0.6411 | 7.4 |
| | P-G-SchNet | 77.51 | 76.40 | 41.93 | 99.74 | 4.04 | 2.448 | 0.6234 | 6.9 |
| | CDVAE | **100.0** | 86.70 | 99.15 | 99.49 | 0.6875 | 0.2778 | 1.432 | 5.3 |
| | SyMat | **100.0** | 88.26 | 98.97 | **99.97** | 0.3805 | 0.3506 | 0.5067 | 3.6 |
| | EquiCSP | 99.97 | 82.20 | 99.65 | 99.68 | **0.1300** | 0.0848 | 0.3978 | 4.1 |
| | DiffCSP | **100.0** | 83.25 | 99.71 | 99.76 | 0.3502 | 0.1247 | **0.3398** | 2.7 |
| | EH-Diff (Sphere) | 99.90 | **89.50** | 99.69 | 99.89 | 0.3996 | **0.0590** | 0.4119 | 3.0 |
| | EH-Diff (Cube) | **100.0** | 82.50 | **99.81** | 99.86 | 0.2880 | 0.0804 | 0.3474 | **2.3** |

In Table 3, we present a comprehensive evaluation of various methods on the ab initio crystal generation task, emphasizing the strong performance of EH-Diff (Sphere and Cube) in this domain. Both variants of EH-Diff—Sphere and Cube—demonstrate exceptional consistency, achieving 100% validity in terms of both structure and composition across all datasets. These results consistently outperform or at least match the performance of other state-of-the-art methods, including DiffCSP, CDVAE, and EquiCSP. Regarding coverage, both EH-Diff (Sphere) and EH-Diff (Cube) exhibit outstanding COV-Recall and COV-Precision, with particularly impressive results on datasets such as Perov-5 and Carbon-24, where they achieve near-perfect scores. In addition, EH-Diff (Cube) outperforms the other methods in terms of key property accuracy metrics, such as $d_\rho$, $d_E$, and $d_{\text{elem}}$, further highlighting the physical and chemical plausibility of the generated structures. Most notably, EH-Diff (Cube) achieves the highest overall rank score across all three datasets, showcasing its superiority in this task. Compared with traditional and advanced methods, EH-Diff stands out for its ability to capture high-order interactions and effectively preserve crystal symmetries, solidifying its position as a highly robust and effective approach for crystal generation tasks. Based on our evaluation, we recommend the Cube variant of EH-Diff, as it consistently outperforms the Sphere variant across all metrics, offering a more reliable and precise solution.

## 5. Conclusion

We introduce a hypergraph-based framework to model crystal structures and propose the novel Equivariant Hypergraph Diffusion Model (EH-Diff) for Crystal Structure Prediction (CSP). This approach effectively captures the intricate, higher-order interactions inherent in crystalline materials, offering a more comprehensive representation of their structural properties. Using hypergraphs, EH-Diff provides a more intuitive representation of crystal structures, effectively capturing essential symmetries and periodicities. Our experimental results demonstrate that EH-Diff out- performs existing state-of-the-art CSP methods.

## Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

## References


Bannan, C. C., Ovanesyan, G., Darden, T. A., Graves, A. P., Edge, C. M., Russo, L., Copley, R. C., Manas, E., Skillman, A. G., Nicholls, A., et al. Crystal structure prediction of drug molecules in the cloud: A collaborative blind challenge study. *Crystal Growth & Design*, 2025.

Bao, C., Serrano-Lotina, A., Niu, M., Portela, R., Li, Y., Lim, K. H., Liu, P., Wang, W.-j., Bañares, M. A., and Wang, Q. Microwave-associated chemistry in environmental catalysis for air pollution remediation: A review. *Chemical Engineering Journal*, 466:142902, 2023.

Cartwright, J. H. and Mackay, A. L. Beyond crystals: the dialectic of materials and information, 2012.

Castelli, I. E., Landis, D. D., Thygesen, K. S., Dahl, S., Chorkendorff, I., Jaramillo, T. F., and Jacobsen, K. W. New cubic perovskites for one-and two-photon water splitting using the computational materials repository. *Energy & Environmental Science*, 5(10):9034–9043, 2012a.

Castelli, I. E., Olsen, T., Datta, S., Landis, D. D., Dahl, S., Thygesen, K. S., and Jacobsen, K. W. Computational screening of perovskite metal oxides for optimal solar light capture. *Energy & Environmental Science*, 5(2): 5814–5819, 2012b.

Chen, C., Ye, W., Zuo, Y., Zheng, C., and Ong, S. P. Graph networks as a universal machine learning framework for molecules and crystals. *Chemistry of Materials*, 31(9): 3564–3572, 2019.

Cheng, G., Gong, X.-G., and Yin, W.-J. Crystal structure prediction by combining graph network and optimization algorithm. *Nature communications*, 13(1):1492, 2022.

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

De Bortoli, V., Mathieu, E., Hutchinson, M., Thornton, J., Teh, Y. W., and Doucet, A. Riemannian score-based generative modelling. *Advances in Neural Information Processing Systems*, 35:2406–2422, 2022.

Desiraju, G. R. Cryptic crystallography. *Nature materials*, 1(2):77–79, 2002.

Gebauer, N. W., Gastegger, M., Hessmann, S. S., Müller, K.-R., and Schütt, K. T. Inverse design of 3d molecular structures with conditional generative neural networks. *Nature communications*, 13(1):973, 2022.

Grimmer, H., Bollmann, W., and Warrington, D. Coincidence-site lattices and complete pattern-shift in cubic crystals. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 30(2):197–207, 1974.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Hoogeboom, E., Satorras, V. G., Vignac, C., and Welling, M. Equivariant diffusion for molecule generation in 3d. In *International conference on machine learning*, pp. 8867–8887. PMLR, 2022.

Hu, J., Yang, W., and Dilanga Siriwardane, E. M. Distance matrix-based crystal structure prediction using evolutionary algorithms. *The Journal of Physical Chemistry A*, 124 (51):10909–10919, 2020.

Hu, J., Yang, W., Dong, R., Li, Y., Li, X., Li, S., and Siriwardane, E. M. Contact map based crystal structure prediction using global optimization. *CrystEngComm*, 23 (8):1765–1776, 2021.

Jacobsen, T., Jørgensen, M., and Hammer, B. On-the-fly machine learning of atomic potential in density functional theory structure optimization. *Physical review letters*, 120 (2):026102, 2018.

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

Jiao, R., Huang, W., Lin, P., Han, J., Chen, P., Lu, Y., and Liu, Y. Crystal structure prediction by joint equivariant diffusion. *Advances in Neural Information Processing Systems*, 36, 2024.

Jing, B., Corso, G., Chang, J., Barzilay, R., and Jaakkola, T. Torsional diffusion for molecular conformer generation. *Advances in Neural Information Processing Systems*, 35: 24240–24253, 2022.

Kim, J., Oh, S., Cho, S., and Hong, S. Equivariant hypergraph neural networks. In *European Conference on Computer Vision*, pp. 86–103. Springer, 2022.

Kim, S., Noh, J., Gu, G. H., Aspuru-Guzik, A., and Jung, Y. Generative adversarial networks for crystal structure prediction. *ACS central science*, 6(8):1412–1420, 2020.

Kohn, W. and Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133, 1965.

Lin, P., Chen, P., Jiao, R., Mo, Q., Jianhuan, C., Huang, W., Liu, Y., Huang, D., and Lu, Y. Equivariant diffusion for crystal structure prediction. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 29890–29913, 2024.

Nouira, A., Sokolovska, N., and Crivello, J.-C. Crystalgan: learning to discover crystallographic structures with generative adversarial networks. *arXiv preprint arXiv:1810.11203*, 2018.

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

Ong, S. P., Richards, W. D., Jain, A., Hautier, G., Kocher, M., Cholia, S., Gunter, D., Chevrier, V. L., Persson, K. A., and Ceder, G. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013.

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

Podryabinkin, E. V., Tikhonov, E. V., Shapeev, A. V., and Oganov, A. R. Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning. *Physical Review B*, 99(6):064114, 2019.

Ren, Z., Tian, S. I. P., Noh, J., Oviedo, F., Xing, G., Li, J., Liang, Q., Zhu, R., Aberle, A. G., Sun, S., et al. An invertible crystallographic representation for general inverse design of inorganic crystals with targeted properties. *Matter*, 5(1):314–335, 2022.

Sato, O. Dynamic molecular crystals with switchable physical properties. *Nature chemistry*, 8(7):644–656, 2016.

Satorras, V. G., Hoogeboom, E., and Welling, M. E (n) equivariant graph neural networks. In *International conference on machine learning*, pp. 9323–9332. PMLR, 2021.

Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A., and Müller, K.-R. Schnet–a deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24), 2018.

Song, Y. and Ermon, S. Improved techniques for training score-based generative models. *Advances in neural information processing systems*, 33:12438–12448, 2020.

Song, Z., Meng, Z., and King, I. A diffusion-based pretraining framework for crystal property prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 8993–9001, 2024.

Steiger, M. Crystal growth in porous materials—i: The crystallization pressure of large crystals. *Journal of crystal growth*, 282(3-4):455–469, 2005.

Thomas, N., Smidt, T., Kearnes, S., Yang, L., Li, L., Kohlhoff, K., and Riley, P. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018.

Tran, R., Lan, J., Shuaibi, M., Wood, B. M., Goyal, S., Das, A., Heras-Domingo, J., Kolluru, A., Rizvi, A., Shoghi, N., et al. The open catalyst 2022 (oc22) dataset and challenges for oxide electrocatalysts. *ACS Catalysis*, 13 (5):3066–3084, 2023.

Voet, D. and Rich, A. The crystal structures of purines, pyrimidines and their intermolecular complexes. *Progress in nucleic acid research and molecular biology*, 10:183–265, 1970.

Woodley, S. M. and Catlow, R. Crystal structure prediction from first principles. *Nature materials*, 7(12):937–946, 2008.

Wooster, W. Physical properties and atomic arrangements in crystals. *Reports on Progress in Physics*, 16(1):62, 1953.

Xie, T. and Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters*, 120 (14):145301, 2018.

Xie, T., Fu, X., Ganea, O.-E., Barzilay, R., and Jaakkola, T. Crystal diffusion variational autoencoder for periodic material generation. *arXiv preprint arXiv:2110.06197*, 2021.

Yamashita, T., Sato, N., Kino, H., Miyake, T., Tsuda, K., and Oguchi, T. Crystal structure prediction accelerated by bayesian optimization. *Physical Review Materials*, 2 (1):013803, 2018.

Yan, K., Liu, Y., Lin, Y., and Ji, S. Periodic graph transformers for crystal material property prediction. *Advances in Neural Information Processing Systems*, 35:15066–15080, 2022.

Yin, Y., Xu, G., Xie, Y., Luo, Y., Wei, Z., and Li, Z. Utilizing deep learning for crystal system classification in lithium-ion batteries. *Journal of Theory and Practice of Engineering Science*, 4(03):199–206, 2024.

Zhang, Z., Yang, X., Zhao, Y., Ye, F., and Shang, L. Liquid crystal materials for biomedical applications. *Advanced Materials*, 35(36):2300220, 2023.

# Appendix

## A. Preliminaries

Let $\{x_1, x_2, \dots\}$ represent a set, and $(x_1, x_2, \dots)$ represent an ordered tuple. The notation $[n]$ is used to denote the set $\{1, 2, \dots, n\}$. The space of order-$k$ tensors is denoted by $\mathbb{R}^{n^k \times d}$, where $d$ represents the feature dimensionality. For an order-$k$ tensor $\mathbf{A} \in \mathbb{R}^{n^k \times d}$, a multi-index $\mathbf{i} = (i_1, \dots, i_k) \in [n]^k$ is used to index the element $\mathbf{A}_\mathbf{i} = \mathbf{A}_{i_1,\dots,i_k} \in \mathbb{R}^d$. Let $\mathbb{S}_n$ denote the set of all permutations of the set $[n]$. A permutation $\pi \in \mathbb{S}_n$ acts on a multi-index $\mathbf{i}$ by permuting its components, i.e., $\pi(\mathbf{i}) = (\pi(i_1), \dots, \pi(i_k))$, and it acts on the tensor $\mathbf{A}$ by permuting the corresponding tensor indices, such that the transformed tensor $(\pi \cdot \mathbf{A})_\mathbf{i} = \mathbf{A}_{\pi^{-1}(\mathbf{i})}$.

Consider hypergraph data represented as $\mathbf{G} = (\mathbf{V}, \mathbf{A})$, where $\mathbf{V}$ denotes a set of $n$ nodes, and $\mathbf{A} \in \mathbb{R}^{n^k \times d}$ is a tensor that encodes the features of the hyperedges (Kim et al., 2022). The order $k$ of the tensor $\mathbf{A}$ corresponds to the type of the hypergraph. Specifically, a first-order tensor encodes a set of node-specific features (e.g., point clouds), where $\mathbf{A}_i$ represents the feature associated with node $i$. A second-order tensor encodes pairwise edge features (e.g., adjacency), where $\mathbf{A}_{i_1,i_2}$ denotes the feature of the edge $(i_1, i_2)$. More generally, an order-$k$ tensor encodes the features of hyperedges (e.g., mesh normals), where $\mathbf{A}_{i_1,\dots,i_k}$ represents the feature of the hyperedge $(i_1, \dots, i_k)$. Ultimately, this framework leads to the well-established concept of hypergraphs with undirected hyperedges of arbitrary order.

**Hypergraph as a Sequence of Higher-order Tensors.** The sequence representation of a hypergraph $(\mathbf{V}, \mathbf{E}, \mathbf{X})$ with a maximum hyperedge order $K$ consists of a sequence of $k$-uniform hypergraphs for all $k \leq K$, denoted as

$$(\mathbf{V}, \mathbf{E}^{(k)}, \mathbf{X}^{(k)})_{k \leq K} = (\mathbf{V}, \mathbf{E}^{(:K)}, \mathbf{X}^{(:K)}),$$

where $\mathbf{E}^{(k)}$ represents the set of all order-$k$ hyperedges in $E$, and $\mathbf{X}^{(k)}$ is a matrix formed by stacking the features $\{\mathbf{X}_\mathbf{e} \mid \mathbf{e} \in \mathbf{E}^{(k)}\}$ for each hyperedge $\mathbf{e}$ in $\mathbf{E}^{(k)}$.

In practical scenarios involving undirected hypergraphs, a hypergraph $\mathbf{G} = (\mathbf{V}, \mathbf{E}, \mathbf{X})$ is typically defined by a set of $n$ nodes $\mathbf{V}$, a set of $m$ hyperedges $\mathbf{E}$, and a feature matrix $\mathbf{X} \in \mathbb{R}^{m \times d}$ that encodes the attributes of the hyperedges. Each hyperedge $\mathbf{e} \in \mathbf{E}$ is a subset of the node set $\mathbf{V}$, and the order of the hyperedge, $|\mathbf{e}|$, characterizes its type. For example, a first-order edge $\{i\}$ corresponds to the $i$-th node, a second-order edge $\{i, j\}$ represents a pairwise link between the $i$-th and $j$-th nodes, and in general, an order-$k$ edge $\{i_1, \dots, i_k\}$ refers to a hyperedge that connects $k$ nodes. The feature associated with each hyperedge $\mathbf{e}$ is denoted by $\mathbf{X}_\mathbf{e} \in \mathbb{R}^d$. We assume that both node and hyperedge features are $d$-dimensional. To accommodate possible variations in dimensionalities, we define $d = d_\mathbf{v} + d_\mathbf{e}$, where $d_\mathbf{v}$ and $d_\mathbf{e}$ represent the dimensionalities of the node and hyperedge features, respectively. The node features are placed in the first $d_\mathbf{v}$ channels, while the hyperedge features occupy the remaining $d_\mathbf{e}$ channels. It is important to note that the concept of hypergraphs $(\mathbf{V}, \mathbf{E}, \mathbf{X})$ presented here does not directly correspond to the higher-order tensors $\mathbf{A} \in \mathbb{R}^{n^k \times d}$ discussed earlier. In contrast to these tensors, the hypergraphs we focus on are sparse, have undirected hyperedges, and each hyperedge involves a unique set of node indices.

**Permutation Equivariance.** In the context of (hyper)graph learning, the goal is to design a function $f$ that takes a higher-order tensor $A$ as input and produces an output. Since the tensor representation of a graph undergoes substantial changes when the node labeling is permuted, it is essential that the function $f$ exhibit invariance or equivariance with respect to node permutations. Specifically, if the output is a scalar or vector, $f$ must be permutation-invariant, meaning that

$$f(\pi \cdot A) = f(A)$$

holds for any permutation $\pi \in \mathbb{S}_n$. Conversely, if the output is a tensor, $f$ must be permutation-equivariant, ensuring that

$$f(\pi \cdot A) = \pi \cdot f(A)$$

for all $\pi \in \mathbb{S}_n$ (Kim et al., 2022). When $f$ is implemented as a neural network, typically consisting of a sequence of linear transformations and non-linear activation functions, the task simplifies to the design of linear layers that are both invariant and equivariant under permutations.

## B. Ab Initio Generation

We extend EH-Diff to the domain of ab initio crystal generation by incorporating discrete diffusion on atom types, denoted as $\mathbf{A}$. To evaluate its performance, we compare EH-Diff against five state-of-the-art generative methods: FTCP (Ren et al.,

2022), Cond-DFC-VAE (Court et al., 2020), G-SchNet (Gebauer et al., 2022), its periodic extension P-G-SchNet, and the original version of the generative framework (Xie et al., 2021).

For EH-Diff, we estimate the statistical distribution of atom counts from the training dataset and sample the number of atoms according to this pre-computed distribution, inspired by the approach of Hoogeboom et al. (2022). This enables EH-Diff to generate crystal structures with variable sizes. Consistent with the evaluation protocol in (Xie et al., 2021), we assess generation performance using three key metrics: Validity, Coverage, and Property Statistics. Validity measures the percentage of generated crystals that are physically plausible. Coverage quantifies the similarity between the generated samples and the test dataset in terms of structural diversity. Property statistics evaluate the physical and chemical properties of the generated crystals, including density, formation energy, and the distribution of elemental compositions.

## C. Hypergraph Modeling Variants

In Figure 4, a two-dimensional illustration is provided to elucidate the process of hyperedge modeling within crystal structures. To effectively identify these hyperedges, a scanning volume, either spherical or cubic, is utilized to encompass specific sets of atoms. A hyperedge is defined when a group of designated atoms, depicted as green and blue nodes, are enclosed within the same scanning volume, regardless of whether it assumes a spherical or cubic shape. The left panel of the figure demonstrates the formation of hyperedges using a spherical scanning approach, where the scanning region is a sphere centered on the atoms of interest. Conversely, the right panel depicts the hyperedge formation using a cubic scanning method, where the scanning region is a cube that encapsulates the relevant atoms. This comparative representation highlights the differences in hyperedge identification based on the geometry of the scanning volume.

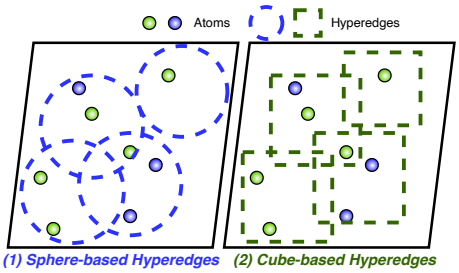



*Figure 4.* Hypergraph Modeling Variants.

## D. Proofs of Proposition 3.1 and Proposition 3.2

In this section, we provide the whole proofs for the main proposition of EH-Diff — Proposition 3.1 and Proposition 3.2.

### D.1. Proof of Proposition 3.1

*Proof.* We begin by establishing the orthogonal invariance of the inner product $\mathbf{L}^\top\mathbf{L}$. Let $\mathbf{Q} \in \mathbb{R}^{3\times3}$ be any orthogonal transformation, satisfying $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}$. Under these conditions, it follows that: $(\mathbf{QL})^\top(\mathbf{QL}) = \mathbf{L}^\top\mathbf{L}$.

For the Fourier Transformation $\psi$, we assume $\psi(x)[c,k] = \sin(2\pi mx)$ and consider $k$ is odd, we have:

$$\psi(w(f_j+t)-w(f_i+t))[c,k] = \cos\big(2\pi m(w(f_{j,c}+t_c)-w(f_{i,c}+t_c))\big) = \cos(2\pi m(f_{j,c}+f_{i,c})) = \psi(f_j-f_i)[c,k].$$

When $k$ is even, we assume $\psi(x)[c,k] = \cos(2\pi mx)$, and we have:

$$\psi(w(f_j+t)-w(f_i+t))[c,k] = \sin\big(2\pi m(w(f_{j,c}+t_c)-w(f_{i,c}+t_c))\big) = \sin(2\pi m(f_{j,c}-f_{i,c})) = \psi(f_j-f_i)[c,k].$$

Therefore, $\psi$ is periodic translation invariant. Also, the message passing layers in EHGNN-MLP defined in Eq.4 is periodic $E(3)$ invariant. Hence, we can directly prove that the coordinate denoising term $\hat{\epsilon}_{\mathbf{F}}$ is periodic translation invariant.

For the lattice denoising term $\hat{\epsilon}_{\mathbf{L}}$, we have $\hat{\epsilon}_{\mathbf{L}}(\mathbf{QL},\mathbf{F},\mathbf{A}) = \mathbf{Q}\hat{\epsilon}_{\mathbf{L}}(\mathbf{L},\mathbf{F},\mathbf{A})$.

Above all, we observe that $\hat{\epsilon}_{\mathbf{L}}$ is equivariant with respect to $O(3)$, and $\hat{\epsilon}_{\mathbf{F}}$ is invariant under periodic translations. As a result, the distribution generated by EH-Diff is invariant under the periodic $E(3)$ symmetry group. □

### D.2. Proof of Proposition 3.2

**Lemma D.1** ((Jiao et al., 2024)). *Consider the generative Markov process defined as $p(x_0) = p(x_T)\int p(x_{0:T-1} \mid x_t)\, dx_{1:T}$. Suppose that the prior distribution $p(x_T)$ is G-invariant and that the Markov transitions $p(x_{t-1} \mid x_t)$, for $0 < t \le T$, are G-equivariant. Then, the marginal distribution $p(x_0)$ is guaranteed to be G-invariant as well.*

*Proof.* In the context of the transition probability defined in Eq.3, we can express it as follows:

$$p(\mathbf{L}_{t-1} \mid \mathbf{L}_t, \mathbf{F}_t, \mathbf{A}) = \mathcal{N}\Big(\mathbf{L}_{t-1} \mid a_t\big(\mathbf{L}_t - b_t \hat{\epsilon}_{\mathbf{L}}(\mathbf{L}_t, \mathbf{F}_t, \mathbf{A}, t)\big), \sigma_t^2 \mathbf{I}\Big),$$

where the parameters are defined as: $a_t = \frac{1}{\sqrt{\alpha_t}}, \quad b_t = \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}, \quad \sigma_t^2 = \beta_t \cdot \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}$.

Here:

- $\alpha_t$ and $\beta_t$ represent the time-dependent coefficients associated with the variance schedule.
- $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$ accumulates the product of $\alpha_s$ values up to time $t$, capturing the cumulative effect of the noise process.
- $\hat{\epsilon}_{\mathbf{L}}(\mathbf{L}_t, \mathbf{F}_t, \mathbf{A}, t)$ is the predicted noise function conditioned on the current state $\mathbf{L}_t$, auxiliary features $\mathbf{F}_t$, adjacency matrix $\mathbf{A}$, and time step $t$.
- $\sigma_t^2$ represents the variance of the Gaussian distribution, ensuring consistency with the underlying stochastic process.

As the denoising term $\hat{\epsilon}_{\mathbf{L}}$ is $O(3)$-equivariant, we have $\hat{\epsilon}_{\mathbf{L}}(\mathbf{Q}\mathbf{L}_t, \mathbf{F}_t, \mathbf{A}, t) = \mathbf{Q}\hat{\epsilon}_{\mathbf{L}}(\mathbf{L}_t, \mathbf{F}_t, \mathbf{A}, t)$, for any orthogonal transformation $\mathbf{Q} \in \mathbb{R}^{3\times 3}$, $\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}$.

For the variable $\mathbf{L} \sim \mathcal{N}(\bar{\mathbf{L}}, \sigma^2 \mathbf{I})$, we have $\mathbf{Q}\mathbf{L} \sim \mathcal{N}(Q\bar{\mathbf{L}}, \mathbf{Q}(\sigma^2 \mathbf{I})\mathbf{Q}^\top) = \mathcal{N}(\mathbf{Q}\bar{\mathbf{L}}, \sigma^2 \mathbf{I})$. That is, $\mathcal{N}(\mathbf{L} \mid \bar{\mathbf{L}}, \sigma^2 \mathbf{I}) = \mathcal{N}(\mathbf{Q}\mathbf{L} \mid \mathbf{Q}\bar{\mathbf{L}}, \sigma^2 \mathbf{I})$.

For the transition probability $p(\mathbf{L}_{t-1} \mid \mathbf{L}_t, \mathbf{F}_t, \mathbf{A})$, we have

$$p(\mathbf{Q}\mathbf{L}_{t-1} \mid \mathbf{Q}\mathbf{L}_t, \mathbf{F}_t, \mathbf{A}) = \mathcal{N}(\mathbf{L}_{t-1} \mid a_t(\mathbf{L}_t - b_t \hat{\epsilon}_{\mathbf{L}}(\mathbf{L}_t, \mathbf{F}_t, \mathbf{A}, t)), \sigma_t^2 \mathbf{I}) = p(\mathbf{L}_{t-1} \mid \mathbf{L}_t, \mathbf{F}_t, \mathbf{A}).$$

Given that the transition function exhibits $O(3)$-equivariance and the prior distribution $\mathcal{N}(0, \mathbf{I})$ is $O(3)$-invariant, it follows from Lemma D.1 that the marginal distribution $p(\mathbf{L}_0)$ is also $O(3)$-invariant. This conclusion leverages the interplay between the symmetry of the transition function and the invariance of the prior. □

## E. Related Works

**Crystal Structure Prediction** Traditional computational methodologies, as exemplified by Pickard & Needs (2011); Yamashita et al. (2018), have combined Density Functional Theory (DFT) (Kohn & Sham, 1965) with optimization algorithms to locate local minima on the potential energy surface. While these methods achieve high accuracy, the computational demands of DFT calculations present a significant challenge, often necessitating a trade-off between computational efficiency and predictive precision. The proliferation of extensive crystal structure databases has enabled the deployment of machine learning techniques as alternative predictors of material energies, serving as a complement or substitute to DFT calculations, typically followed by optimization procedures (Jacobsen et al., 2018; Podryabinkin et al., 2019; Cheng et al., 2022). These data-driven approaches have shown promise in enhancing the speed and scalability of materials discovery and design. Moving beyond the conventional predict-optimize framework, a new class of methods has emerged that utilizes deep generative models to infer stable crystal structures directly from data. These models employ various representations to encapsulate the structural information of crystals, including three-dimensional voxel grids (Gebauer et al., 2022), distance matrices (Hu et al., 2020; 2021), and direct 3D coordinate systems (Nouira et al., 2018; Kim et al., 2020). By leveraging these advanced data representations (Jiao et al., 2024; Song et al., 2024), deep generative models contribute to the development of more efficient and accurate methods for predicting and designing novel crystalline materials.

**Equivariant Models** Geometrically Equivariant Graph Neural Networks (EGNNs), which maintain $E(3)$ symmetry, have emerged as powerful frameworks for the representation of physical entities (Schütt et al., 2018; Thomas et al., 2018; Satorras et al., 2021). These networks have exhibited exceptional performance in the modeling of three-dimensional structures (Tran et al., 2023). To adapt these capabilities to the context of periodic materials, Xie & Grossman (2018) introduced a multi-graph edge construction method designed to capture periodicity through the connection of edges between neighboring lattices. Expanding on this foundation, Yan et al. (2022); Jiao et al. (2024); Song et al. (2024) incorporated periodic pattern encoding into a Transformer-based architecture, thereby further enhancing the ability to model the complex periodic properties of materials.

**Remark E.1.** *CDVAE (Xie et al., 2021) and DiffCSP (Jiao et al., 2024) are similar to our work in their use of equivariant generative models to capture the underlying data distribution and generate crystal structures through a score-matching diffusion process. However, their primary focus is on ab initio crystal generation, relying on random sampling of chemical compositions, which differs from the CSP problem we address. The major distinction in our approach lies in how we model the crystal structures. Unlike CDVAE, DiffCSP, and EquiCSP, which use traditional molecular graph representations, we utilize hypergraph models to represent crystal structures. This allows us to capture high-order interactions effectively, providing a richer and more expressive framework for modeling complex crystalline materials, thus enhancing the generation process and achieving better predictive performance.*