Conformal Crystal Graph Transformer with Robust Encoding of Periodic Invariance

Yingheng Wang[1], **Shufeng Kong**[1,3*], **John M. Gregoire**[2], **Carla P. Gomes**[1]
[1]Department of Computer Science, Cornell University, USA
[2]Liquid Sunlight Alliance, California Institute of Technology, USA
[3]School of Software Engineering, Sun Yat-sen University, China
{yw2349, sk2299}@cornell.edu, gregoire@caltech.edu, gomes@cs.cornell.edu

Abstract

Machine learning techniques, especially in the realm of materials design, hold immense promise in predicting the properties of crystal materials and aiding in the discovery of novel crystals with desirable traits. However, crystals possess unique geometric constraints—namely, E(3) invariance for primitive cell and periodic invariance—which need to be accurately reflected in crystal representations. Though past research has explored various construction techniques to preserve periodic invariance in crystal representations, their robustness remains inadequate. Furthermore, effectively capturing angular information within 3D crystal structures continues to pose a significant challenge for graph-based approaches. This study introduces novel solutions to these challenges. We first present a graph construction method that robustly encodes periodic invariance and a strategy to capture angular information in neural networks without compromising efficiency. We further introduce CrystalFormer, a pioneering graph transformer architecture that emphasizes angle preservation and enhances long-range information. Through comprehensive evaluation, we verify our model's superior performance in 5 crystal prediction tasks, reaffirming the efficiency of our proposed methods.

Introduction

Machine learning (ML) techniques offer the capacity to model intricate physical and chemical interactions, presenting a significant opportunity to expedite materials design vital for sustainable agriculture, low-carbon energy, and more. One notable application of ML in this domain is predicting properties of crystal materials, aiding in the discovery of new crystals with desired attributes. Unique to crystals are geometric priors that encompass two distinct physical constraints and symmetries: E(3) invariance within the primitive cell, and periodic invariance. These inherent constraints should be reflected in crystal representations like periodic graphs, often used as input for graph neural networks (GNNs) (Xie and Grossman 2018; Louis et al. 2020; Brody, Alon, and Yahav 2021; Schütt et al. 2017; Chen et al. 2019; Choudhary and DeCost 2021; Yan et al. 2022; Lin et al. 2023). While previous studies (Xie and Grossman 2018; Ying et al. 2021; Yan et al. 2022) have investigated construction techniques such as multi-edge and k-connected graph methodologies to establish periodic invariance in crystal structures, these methods often lack robustness when applied to a diverse range of crystal structures, leading to incomplete representations under perturbations. More recent research (Widdowson and Kurlin 2022; Widdowson et al. 2022) has introduced matrix-form crystal representations as a means to encode this completeness. However, these too are potentially susceptible to breaking the completeness in practical applications.

Another significant challenge for GNNs in processing 3D crystal structure is the difficulty of efficiently incorporating angular information. Choudhary and DeCost (2021) suggests an approach using a distinct line graph type for angular information. Despite its notable performance enhancements, this method is hampered by the considerable computational overhead associated with constructing the line graph. On the other hand, Yan et al. (2022) employs radial basis functions (RBF) and combines spherical Bessel functions (SBF) with spherical harmonics to manage angles. Nonetheless, this approach does not yield the anticipated performance boost.

This study addresses two key challenges: robustly encoding periodic invariance and effectively managing angular information. **Our contributions** include (1) introducing a graph construction method that preserves periodic invariance and completeness across diverse settings by effectively reducing stochasticity via physical properties; (2) developing a method to approximate conformal mapping in neural networks using harmonic mapping with minimized Dirichlet energy, preserving inherent angular information in crystal graphs; (3) unveiling CrystalFormer, a novel E(3)-invariant graph transformer architecture with a unique angular attention mechanism that enhances our angle preservation approach and fortifies long-range information; (4) conducting comprehensive experiments over 5 tasks on the Jarvis materials benchmark (Choudhary et al. 2020), highlighting the significant of our components and showing our model's superior performance, and subsequently verifying the effectiveness of our proposed construction and learning methods.

Preliminaries

Crystal Structures

A crystal structure can be represented as an infinite periodic arrangement of atoms in 3D space. Mathematically, this can

*Contact author

be captured using the lattice points defined by translation vectors and the basis (the set of atoms within the primitive cell). A lattice $\mathcal{L}$ is an infinite set of points in space that can be generated by translating a point using discrete combinations of three non-coplanar vectors, $\mathbf{l}_1$, $\mathbf{l}_2$, and $\mathbf{l}_3$. This set can be written as:

$$\mathcal{L} = \{\mathbf{T}|\mathbf{T} = n_1\mathbf{l}_1 + n_2\mathbf{l}_2 + n_3\mathbf{l}_3, \forall(n_1, n_2, n_3) \in \mathbb{Z}^3\}. \quad (1)$$

For a crystal structure, the basis $\mathcal{B}$ is a set of atoms associated with every lattice point. Each atom in the basis can be represented as:

$$\mathcal{B} = \{(\alpha, \mathbf{r}_\alpha^i)|\alpha \in \mathcal{A}, \mathbf{r}_\alpha^i \in \mathbb{R}^3\}, \quad (2)$$

where α represents the type of atom (e.g., Carbon, Oxygen, etc.), $\mathcal{A}$ is the set of all atom types in the basis, and $\mathbf{r}_\alpha^i$ represents the position of the i^{th} atom of type α within the primitive cell. Given the lattice and the basis, the crystal structure $\mathcal{C}$ is the set of all atoms in the 3D space, defined as:

$$\mathcal{C} = \{(\alpha, \mathbf{p}_\alpha^i(\mathbf{T}))|\mathbf{T} \in \mathcal{L}, (\alpha, \mathbf{r}_\alpha^i) \in \mathcal{B}\}, \quad (3)$$

where $\mathbf{p}_\alpha^i(\mathbf{T}) = \mathbf{r}_\alpha^i + \mathbf{T}$. This indicates that for every lattice point $\mathbf{T}$ and every atom in the basis $\mathcal{B}$, there exists an atom in the crystal structure at position $\mathbf{p}_\alpha^i(\mathbf{T})$.

A primitive cell is the smallest volume element of a crystal that, when translated through all the vectors of the Bravais lattice, can fill up the whole space without gaps or overlaps. Mathematically, a primitive cell $\mathcal{P}$ in terms of the basis is

$$\mathcal{P} = \{(\alpha, \mathbf{r}_\alpha^i)|(\alpha, \mathbf{r}_\alpha^i) \in \mathcal{B}\}, \quad (4)$$

In essence, the lattice represents the periodicity of the crystal, while the basis represents the arrangement of atoms within a single primitive cell. The combination of the lattice and the basis gives the complete infinite arrangement of atoms in the crystal structure. Moreover, we define $\mathbf{x}_\alpha^i \in \mathbb{R}^h$ as the feature vector of the i^{th} atom of type α in a primitive cell and naturally $\mathbf{x}_\alpha^i(\mathbf{T})$ for this atom and its periodic repeats.

Crystal Graph Construction

Given a crystal graph, $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of vertices representing atoms, and $\mathcal{E}$ is the set of edges representing spatial connections between atoms. $\mathcal{V}$ can be defined as

$$\mathcal{V} = \{v_\alpha^i(\mathbf{T})|(\alpha, \mathbf{p}_\alpha^i(\mathbf{T})) \in \mathcal{C}\}, \quad (5)$$

where $v_\alpha^i(\mathbf{T})$ is the vertex corresponding to the i^{th} atom of type α at position $\mathbf{p}_\alpha^i(\mathbf{T})$.

Multi-edge Graph In a multi-edge graph, an edge (or multiple edges) exists between two vertices if their spatial distance falls within a specified radius R:

$$\mathcal{E} = \{(v_\alpha^i(\mathbf{T}), v_\beta^j(\mathbf{T}'))|\text{dist}(\mathbf{p}_\alpha^i(\mathbf{T}), \mathbf{p}_\beta^j(\mathbf{T}')) \leq R)\} \quad (6)$$

where dist$(\cdot)$ is the Euclidean distance function $||\cdot||_2$, and $\mathbf{T}$ and $\mathbf{T}'$ are translation vectors corresponding to the positions of two atoms. Using this graph construction, the spatial relationships between atoms in a crystal structure within a specified radius can be captured. The multi-edge nature allows for more complex and nuanced relationships between atoms, capturing different types of interactions.

k-connected Graph k-connected graph shares the same concept with fully-connected graph presented in Ying et al. (2021) and Yan et al. (2022), in which each atom is connected to its k nearest neighbors. Since the term "fully-connected graph" has a different definition in traditional graph theory, here we switch to k-connected graph for clarity. In a k-connected graph, and the set of edges is determined by

$$\mathcal{E} = \{(v_\alpha^i(\mathbf{T}), v_\beta^j(\mathbf{T}'))|\mathbf{v}_\beta^j(\mathbf{T}') \in \mathcal{N}^k(\mathbf{v}_\alpha^i(\mathbf{T}))\} \quad (7)$$

where $\mathcal{N}^k(\mathbf{v}_\alpha^i(\mathbf{T}))$ represents the set of k closest neighbors of the atom located at $\mathbf{p}_\alpha^i(\mathbf{T})$.

Physical Constraints of Crystals

There are two essential properties of crystals in 3D space, i.e., E(3) invariance with respect to the primitive cell, and periodic invariance. The former refers to the invariance of the structure of a primitive cell under the 3D Euclidean transformations. Formally, it can be defined as follows:

Definition 1 (E(3) Invariance for Primitive Cell)**.** *Let $f: \mathcal{C} \to \mathcal{X}$ denote a function and $E(3)$ be the Euclidean group. The function f is E(3) invariant with respect to the primitive cell $\iff \forall\mathbf{A} \in E(3), \forall\mathbf{t} \in \mathbb{R}^3, f(\mathcal{C}') = f(\mathcal{C})$, where $\mathcal{C}' = \{(\alpha, \mathbf{A}\mathbf{p}_\alpha^i(\mathbf{T}) + \mathbf{t})|\mathbf{T} \in \mathcal{L}, (\alpha, \mathbf{r}_\alpha^i) \in \mathcal{B}\}$. $\mathbf{A}$ refers to a rotation or reflection transformation.*

The latter indicates the invariance of the crystal representation when the periodic boundaries of a primitive cell are shifted or scaled up. This can be formally defined as well.

Definition 2 (Periodic Invariance)**.** *Let $f: \mathcal{C} \to \mathcal{X}$ denote a function that is E(3) invariant with respect to the primitive cell and $\mathbf{\Lambda} \in \mathbb{R}^{3\times3}$ be a primitive-cell transformation (e.g., rotations, shearing, etc.). The function f is periodically invariant $\iff f(\mathcal{P}) = f(\mathcal{P}')$ where $\mathcal{P}' = \{(\alpha, \mathbf{\Lambda}\mathbf{r}_\alpha^i + \mathbf{o})|(\alpha, \mathbf{r}_\alpha^i) \in \mathcal{B}, \mathbf{o} \in \mathbb{R}^3\}$ where $\mathbf{o}$ denotes the translation of the reshaped primitive cell.*

The primitive-cell transformation, represented by $\mathbf{\Lambda}$, embodies a change of basis in the lattice. While the transformed primitive cell still reflects the periodic nature of the original crystal, it represents a different periodic unit by redefining the primitive vectors of the cell.

Robust Encoding for Periodic Invariance

Previously discussed graph construction methods (Xie and Grossman 2018; Ying et al. 2021; Yan et al. 2022), effectively encode periodic invariance in standard settings. However, they falter in specific scenarios where their methods fail to achieve this invariance. More recent works (Widdowson et al. 2022; Widdowson and Kurlin 2022) dedicate to solve incompleteness problem by introducing matrix-form crystal representations, which are still likely to break the completeness in practice.

In this section, we first examine the limitations of two dominant graph construction models—for multi-edge and k-connected graphs—and provides solutions to these challenges. Then we discuss how the matrix-form crystal representations potentially fail to preserve completeness practically. Lastly, we introduce a new graph construction method that ensures efficient and robust encoding of periodic invariance and completeness.

Limitations of Existing Graph Constructions

Two primary crystal graph construction methods are prevalent: the radius-based graph construction (Xie and Grossman 2018; Louis et al. 2020; Choudhary and DeCost 2021) and the k-connected graph construction (Ying et al. 2021; Rampášek et al. 2022; Lin et al. 2023). While both have been tailored to ensure periodic invariance under general conditions (Yan et al. 2022), they exhibit significant drawbacks in preserving this invariance in practical scenarios. These limitations include (1) the random selection of multiple neighbors based on the k^{th} smallest pairwise distance in k-connected graph construction; (2) the inability of the k-connected graph to leverage information beyond Euclidean distance, such as angular data; and (3) challenges posed by experimental crystal structures with minor atomic position deviations. We delve into these issues in the subsequent sections.

In k-connected graph construction, as illustrated in Equation (7), neighbors are selected based on the k nearest neighbors criterion. However, when several neighboring atoms of varying atomic types share the k^{th} smallest pairwise distance, the selection of a particular atom can be arbitrary, leading to inconsistencies across iterations. This introduces stochasticity, jeopardizing periodic invariance as the resulting graphs might differ each time. On the other hand, when neighboring atoms of the same atomic type share the k^{th} smallest distance, periodic invariance is typically preserved in standard scenarios where basic vertex and edge attributes, such as atomic embedding and interatomic distance, are considered—making these atoms indiscernible. Nonetheless, this periodic invariance is compromised when incorporating higher-order information, like unique atomic angles.

A straightforward solution would be to incorporate all atoms at the k^{th} smallest pairwise distance into the neighbor set, rather than solely including the exact k closest ones:

$$\mathcal{E} = (v_\alpha^i(\mathbf{T}), v_\beta^j(\mathbf{T}'))|\text{dist}(\mathbf{p}_\alpha^i(\mathbf{T}), \mathbf{p}_\beta^j(\mathbf{T}')) \leq d_k(\mathbf{p}_\alpha^i(\mathbf{T})) \tag{8}$$

where $d_k(\mathbf{p}_\alpha^i(\mathbf{T}))$ is the distance to the k^{th} nearest neighbor of the atom at $\mathbf{p}_\alpha^i(\mathbf{T})$. This method guarantees deterministic neighbor selection in every situation. However, empirical evidence suggests that its performance is suboptimal.

Another challenge lies within the robustness of periodic invariance in crystal structure determination. In materials science, crystal structures are predominantly determined through experimental methods like X-ray crystallography (Ackland and Jones 2006). Consequently, the derived crystal structure may exhibit minor, random perturbations. Such discrepancies, although subtle, can cause notable variations in the atomic positions across different measurements of the same crystal. Given that neighboring atoms are identified based on a fixed radius cutoff, these minor changes can significantly alter the resulting graph's topological structure. We provide illustrative figures and details in the Appendix.

To address this challenge, one could introduce a buffer or a stability margin during the neighbor selection, rather than relying on a strict cutoff. A straightforward method would be to add a tolerance to the cutoff, transforming it into a narrow range. However, setting a precise tolerance value requires a comprehensive understanding of the perturbations and their statistical patterns. This may be impractical given the extensive lab work involved.

An alternative approach is the parameter-free Voronoi tessellation (Voronoi 1908), which also offers a natural neighbor selection strategy. The crystal structure is initially partitioned via a Voronoi decomposition of its atomic sites. In this method, each atom's domain is characterized by a polyhedron. Its faces are demarcated by equidistant borders between the atom and its neighboring sites. Atoms sharing a face are considered neighbors, and edges are consequently drawn between them (see the Appendix for illustrative figures and details). Nevertheless, the Voronoi-based method still has drawbacks: (1) the number of edges and their respective pairwise distances can be highly variable and may exhibit significant fluctuations across atoms (Ruff et al. 2023); (2) computational efficiency diminishes noticeably in spaces beyond 2D (Vassiliades, Chatzilygeroudis, and Mouret 2017).

Addressing the challenge of incompleteness in crystal representation is another crucial issue. Recent advancements (Widdowson et al. 2022; Widdowson and Kurlin 2022), have developed matrix forms that could be complete. However, employing these representations to predict crystal properties while maintaining their completeness poses practical difficulties. (1) The proposed matrix-form representations for stable crystal structures neglect atom types, whose completeness relies on the premise that no two crystals with identical structures differ solely in atom type, which is only attainable in stable structures. (2) Ensuring completeness requires predefining a large number of neighbors that is impractical and costly (Balasingham, Zamaraev, and Kurlin 2022).

The Proposed Graph Construction Method

As previously discussed, a robust and complete graph construction technique should minimize variability and reduce inconsistency in neighbor selection. To address this, we present a new graph construction method that incorporates a hierarchical weighting approach. This method assigns normalized weights to represent the preference of each neighboring atom.

Reciprocal Weight We begin by considering the reciprocal of interatomic distances, normalized by the distance to the nearest neighbor. To limit the number of neighbors, we restrict our attention to atoms within the k^{th} smallest pairwise distance. For ease of notation, let d_{ij} represent $\text{dist}(\mathbf{p}_\alpha^i(\mathbf{T}), \mathbf{p}_\beta^j(\mathbf{T}'))$ and d_{ij}^k be $d_k(\mathbf{p}_\alpha^i(\mathbf{T}))$. Here, d_{ij}^0 stands for the smallest pairwise distance. With this notation in place, the reciprocal weight is formally given by:

$$w_{ij}^r = \frac{d_{ij}^0}{d_{ij}}, \text{ for } d_{ij} < (1+\delta)d_{ij}^k \tag{9}$$

where δ serves as a tolerance parameter[1], enhancing robustness against experimental discrepancies.

Distance-penalizing Weight While the reciprocal weight accounts for penalizing distances from close to distant neighbors, it overlooks a crucial factor: the radius r_α^i of the

[1] δ is distinct from the tolerance value discussed earlier. Rather than an absolute distance, it's a relatively minor parameter, conveniently set at 0.1 by default for all tasks.

species at $\mathbf{p}_\alpha^i(\mathbf{T})$. To integrate this, we introduce a distance-penalizing weight that incorporates atomic radii:

$$w_{ij}^d = \sqrt{\cos\frac{\pi(d_{ij} - d_{lc})}{2(d_{hc} - d_{lc})}} \tag{10}$$

where $d_{lc} = r_\alpha^i + r_\beta^j + d_{ij}^0$ and $d_{hc} = r_\alpha^i + r_\beta^j + d_{ij}^k$. Note that the choice of radius type often hinges on the available structural information and is typically ranked by descending preference. We expound upon this in the Appendix.

In the realm of materials science, electronegativity plays a pivotal role in determining inter-atomic interactions. Following Pan et al. (2021), we represent the difference in electronegativity as: $w_{ij}^e = 1 + \lambda\sqrt{|\chi_\alpha^i - \chi_\beta^j|/3.3}$ where χ_α^i denotes the Pauling electronegativity of the atom at $\mathbf{p}_\alpha^i(\mathbf{T})$ and λ is a parameter that adjusts the preference for neighboring atoms with pronounced electronegativity differences. The normalization factor is set to 3.3 in the denominator, as it corresponds to the maximal electronegativity difference observed between any two atoms. Through this formulation, atoms exhibiting larger electronegativity disparities with the central atom are assigned greater weights.

Lastly, we perform neighbor selection following Pan et al. (2021). We first normalize the weights to the range $[0, 1]$ using $w_{ij} = w_{ij}^r \cdot w_{ij}^d \cdot w_{ij}^e / \max(w_{ij})$. For the final neighbor selection, a straightforward approach entails calculating the largest weight gap between two consecutive weights, once they are sorted in descending order. Yet, given the dominant interactions in a proximal coordination environment, an alternative strategy might be preferred. Pan et al. (2021) proposes to project the normalized weights onto a quarter-circle placed in the first quadrant. Then, neighbors are chosen based on the highest AUROC observed between successive weights (see the Appendix for details). In this way, we eliminate the stochasticity to the maximal extent. In practice, both techniques yield similar performance. Additionally, to encapsulate periodic patterns, we add self-connected edges for each vertex (Yan et al. 2022).

Preserving Implicit Angular Information

Modeling angular information within GNNs has previously demonstrated notable performance improvements. However, these gains often come at the cost of increased computational demands, largely due to the line graph construction (Choudhary and DeCost 2021). More recent methodologies (Yan et al. 2022) introduce efficient techniques for processing angles in graph transformers, but intriguingly, these methods don't harness the expected performance gains from the additional information.

Here, we demonstrate that any constructed graph intrinsically encapsulates angular information within its topological structure and associated edge attributes. Subsequently, we put forth a novel method designed to implicitly integrate and preserve this angular information throughout the model.

Implicit Angular Information

From the outset, the angular information has been incorporated implicitly within the constructed crystal graphs' edge attributes. Let $\vec{e}_{ij}$ denote the edge attribute, signifying the relative position vector between two neighboring atoms.

Definition 3. *An **angle**, $\angle ikj$, arises from two adjacent edges: $\vec{e}_{ki}$ and $\vec{e}_{kj}$. Here, $v_\gamma^k(\mathbf{T}'')$ is the mutual neighbor of vertices $v_\alpha^i(\mathbf{T})$ and $v_\beta^j(\mathbf{T}')$.*

Given that the edge attributes in $\vec{e}_{ki}$ and $\vec{e}_{kj}$ denote relative positions, the magnitude of $\angle ikj$ can be quantified by its cosine value: $\cos\angle ikj = \vec{e}_{kj} \cdot \vec{e}_{kj} / \|\vec{e}_{ki}\| \, \|\vec{e}_{kj}\|$. Leveraging the capacity of neural networks as universal approximators (Hornik, Stinchcombe, and White 1989; Scarselli and Tsoi 1998), one can derive a nonlinear mapping from edge attributes to the cosine value of angles. This is made feasible through GNNs, in which a multi-layer perceptron is employed to interpret messages from attributes of neighboring vertices and edges.

Nevertheless, a persistent challenge lies in ensuring the preservation of the angular information throughout multiple GNN layers. In practical scenarios, not only might the information not be consistently preserved, but expressive GNN architectures could also inadvertently neglect this angular information amidst a sea of input features.

The Preservation Approach

Conformal mapping (Schinzinger and Laura 2012) serves as a potent tool for preserving angular information and has found extensive applications in areas such as image rectification (Zhang, Li, and Wang 2023) and texture generation (Lévy et al. 2023). However, the rigorous nature of conformal mapping poses challenges for its direct integration into neural networks (Nehari 2012).

Lemma 1. *Let U and V denote real normed spaces. A map $f : U \to V$ is conformal at a point $u_0 \in U$ if it is differentiable at u_0 and its derivative at that point is an isomorphism that preserves angles and orientations between vectors.*

Rather than imposing rigid constraints to achieve conformal mapping, which would compromise the differentiability of neural networks, we can consider approximating it concurrently with the optimization of training objectives. One potential approximation is harmonic mapping, attributed to the harmonicity of conformal mapping, characterized by functions that adhere to the Laplace equation across the domain (Nehari 2012). Nonetheless, this unconstrained approximation may introduce inherent distortion, potentially undermining the preservation of angles.

Definition 4. *Let $u : \Omega \to \mathbb{R}^n$ be a differentiable function defined over a domain Ω in $\mathbb{R}^n$. The Dirichlet energy of u is given by $E_{dir}(u) = \frac{1}{2}\int_\Omega \|\nabla u\|^2 \, dV$, where ∇u represents the gradient of u and dV denotes the area element in Ω.*

Proposition 1. *Given a domain Ω in $\mathbb{R}^n$, a conformal mapping $f : \Omega \to \mathbb{R}^n$ can be approximated by a harmonic mapping that minimizes the Dirichlet energy. If the Dirichlet energy has been sufficiently minimized and the mapping is bijective, then the mapping is approximately conformal within Ω (Wu and Yau 2020).*

Although conformal mapping can be efficiently and accurately approximated by harmonic mapping using the minimizer of Dirichlet energy, obtaining a closed form for this

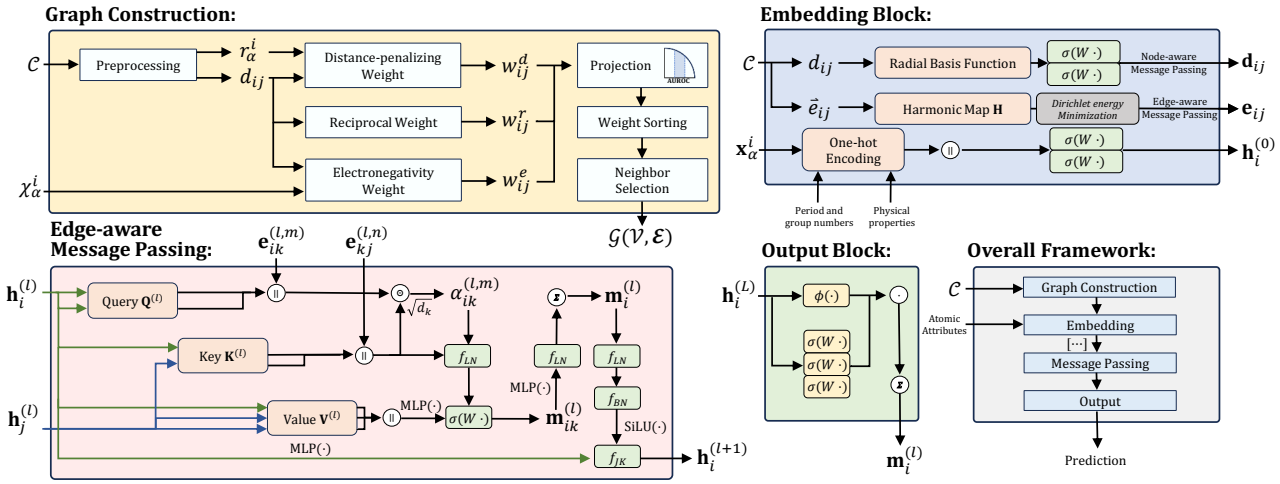

Figure 1: Architecture Overview. CrystalFormer accepts an input crystal structure $\mathcal{C}$. It preserves both periodic invariance and E(3) primitive cell invariance in its predictions. This is achieved by successively processing the input through stages: a graph construction step, an embedding block, multiple (node, edge)-aware message-passing layers, and finally, an output block.

optimization objective on graph-structured data, especially for multi-edge crystal graphs, remains challenging. Here, we introduce a formulation for the Dirichlet energy on graph data: given a harmonic mapping $\mathbf{H} \in \mathbb{R}^{d\times d'}$ that is shared across edges, the Dirichlet energy can be written as (we denote $v_\alpha^j(\mathbf{T}), v_\beta^j(\mathbf{T}'), v_\gamma^j(\mathbf{T}'')$ as v_i, v_j, v_k below for brevity):

$$\begin{aligned} E_{dir}(\mathbf{H}) &= \frac{1}{2}\sum_{(v_i,v_j)\in\mathcal{U}} \kappa_{ij}\|\mathbf{H}\vec{e}_{ij}\|^2 \\ &= \frac{1}{2}\sum_{(v_i,v_j)\in\mathcal{U}} (\cot\angle ikj)\,\|\mathbf{H}\vec{e}_{ij}\|^2 \end{aligned} \tag{11}$$

where $\mathcal{U} = \{(v_i, v_j)\,|\,\forall v_i, v_j \in \mathcal{V}, \exists v_k \in \mathcal{V}, \text{s.t. } (v_i, v_j) \notin \mathcal{E} \text{ and } (v_i, v_k), (v_j, v_k) \in \mathcal{E}\}$, κ_{ij} represents the elastic coefficient between the unconnected vertex pair (v_i, v_j), and cot denotes the cotangent function.

In the context of multi-edge crystal graphs, a key issue is the presence of *multiple potential relay vertices* v_k and *multiple corresponding topological connections* $\vec{e}_{ik}$ *and* $\vec{e}_{kj}$, which will introduce much complexity. For the latter, we consider all existing connections, which ensures that angular information associated with distant atoms can be sufficiently retained, beneficial for encoding long-range interactions and periodicity inherent to crystal structures. For the former, i.e. when selecting relay vertices, it's essential to follow the underlying physical principle: atoms in closer proximity exhibit stronger interactions. Denote the neighborhood of vertices as $\mathcal{N}_i := \mathcal{N}(\mathbf{v}_\alpha^i(\mathbf{T}))$ and $\mathcal{N}_j := \mathcal{N}(\mathbf{v}_\beta^j(\mathbf{T}))$. The selection for k can be articulated as:

$$k = \underset{o\in\mathcal{N}_i\cap\mathcal{N}_j}{\arg\max}\left[\frac{\vec{e}_{io}\cdot\vec{e}_{jo}}{\|\vec{e}_{io}\|\,\|\vec{e}_{jo}\|}\right]. \tag{12}$$

Above we only consider the situation in which we utilize interatomic distances as edge attributes. For those seeking to initialize atomic positions as vertex attributes, we present a more computationally efficient method to handle matrix operations (with the proof available in the Appendix):

$$\begin{aligned} E_{dir}(\mathbf{H}) &= \frac{1}{2}\sum_{(v_i,v_j)\in\mathcal{U}} \kappa_{ij}\|\mathbf{H}\vec{p}_i - \mathbf{H}\vec{p}_j\|^2 \\ &= \mathrm{tr}\left(\mathbf{F}^T(\mathbf{D}-\mathbf{K})\mathbf{F}\right). \end{aligned} \tag{13}$$

Here, $\vec{p}_i$ signifies the atomic position. The function tr(·) returns the trace of a matrix. Additionally, $\mathbf{K}$ is defined as $(\kappa_{ij})_{n\times n}$; $\mathbf{F}$ can be expressed as $(\mathbf{H}\vec{p}_1; \mathbf{H}\vec{p}_2; \cdots; \mathbf{H}\vec{p}_n)$; and $\mathbf{D}$ is a diagonal matrix denoted as $\mathrm{diag}(d_1, d_2, ..., d_N)$, where each d_i is given by $\sum_{j=1}^{n}\kappa_{ij}$.

CrystalFormer

In this section, we unveil our proposed crystal graph transformer architecture, CrystalFormer, as illustrated in Figure 1. Drawing from the graph construction method and the angle-preserving framework described earlier, CrystalFormer seamlessly integrates 3D geometric data based on atomic relative positions while preserving periodic invariance. This is achieved without the need to apply restrictive architectural constraints. Notably, CrystalFormer inherits the theoretical guarantees of the preservation approach, ensuring it upholds the intrinsic invariant physical constraints within crystal graphs. The combination of these features provides the architecture with better modeling flexibility. Furthermore, the inclusion of a plug-and-play angle-preserving layer amplifies the model's performance in evaluating crystal properties.

Embedding Block Expanding on the foundational work of Xie and Grossman (2018), we integrate additional domain-specific knowledge to initialize the vertex attribute of each vertex v_i. To this end, we adopt atomic embedding, achieved by concatenating the one-hot encoding of both group and period numbers together with a suite of physical properties. Further details are available in the Appendix.

In our pursuit to aptly encapsulate the 3D topology inherent in atomic systems, we employ a blend of atom relative positions, denoted as $\vec{e}_{ij}$, and the interatomic distances denoted by d_{ij}. Our approach entails the application of harmonic mapping directly on $\vec{e}_{ij}$ to obtain $\mathbf{e}_{ij}$. Concurrently, we harness a 2-layer MLP on RBF derived from d_{ij} to get $\mathbf{d}_{ij}$. $\mathbf{e}_{ij}$ will be used as the edge embedding for our edge-aware message-passing layer while $\mathbf{d}_{ij}$ will be the edge embedding for the node-aware message-passing layer (Yan et al. 2022). Note that, in practice, replacing $\mathbf{d}_{ij}$ with $\mathbf{e}_{ij}$ achieves better prediction performance.

Edge-aware Message Passing Within each interaction block, messages are propagated from neighboring vertices $v_j \in \mathcal{N}_i$ to the center vertice v_i via a unique graph transformer block. To establish an invariant architecture, we implement an edge-aware attention mechanism that integrates embeddings from vertices and their adjacent edges $(\vec{e}_{ik}, \vec{e}_{kj})$ in both the query and key. This is followed by an inner product operation to compute the attention coefficient.

For a given vertex embedding $\mathbf{h}_i^{(l)}$ and the m^{th} edge embedding at the l^{th} layer $\mathbf{e}_{ij}^{(l,m)}$, we compute query $\mathbf{q}_{ik}^{(l,m)}$, key $\mathbf{k}_{ik}^{(l,m)}$, and value $\mathbf{v}_{ik}^{(l,m)}$ in line with Vaswani et al. (2017):

$$\begin{aligned}\mathbf{q}_{ik}^{(l,m)} &= \left(\mathbf{Q}^{(l)}\mathbf{h}_i^{(l)} || \mathbf{Q}^{(l)}\mathbf{h}_i^{(l)} || \mathbf{e}_{ik}^{(l,m)}\right) \\ \mathbf{k}_{ik}^{(l,m)} &= \left(\mathbf{K}^{(l)}\mathbf{h}_i^{(l)} || \mathbf{K}^{(l)}\mathbf{h}_j^{(l)} || \mathbf{e}_{kj}^{(l,n)}\right) \\ \mathbf{v}_{ik}^{(l,m)} &= \left(\mathbf{V}^{(l)}\mathbf{h}_i^{(l)} || \mathbf{V}^{(l)}\mathbf{h}_j^{(l)} || \mathbf{V}^{(l)}\mathbf{h}_j^{(l)}\right)\end{aligned} \tag{14}$$

where $\mathbf{Q}^{(l)}$, $\mathbf{K}^{(l)}$, and $\mathbf{V}^{(l)}$ are linear transformations. This formulation allows us to integrate self-attention for the central vertex, cross-attention between adjacent vertices, and edge-aware attention, all within a single transformer block.

Subsequently, the attention coefficients $\alpha_{ik}^{(l,m)}$ and messages $\mathbf{m}_{ik}^{(l,m)}$ are derived as:

$$\begin{aligned}\alpha_{ik}^{(l,m)} &= \frac{\mathbf{q}_{ik}^{(l,m)} \odot \mathbf{k}_{ik}^{(l,m)}}{\sqrt{d_{\mathbf{k}}}} \\ \mathbf{m}_{ik}^{(l)} &= \sigma\left(f_{LN}\left(\alpha_{ik}^{(l,m)}\right) \text{MLP}\left(\mathbf{v}_{ik}^{(l,m)}\right)\right)\end{aligned} \tag{15}$$

where f_{LN} denotes layer normalization, while $d_{\mathbf{k}}$ represents the embedding dimension of $\mathbf{k}_{ik}^{(l,m)}$. Then we impose jumping knowledge (Xu et al. 2018) into feature update after aggregating the messages:

$$\begin{aligned}\mathbf{m}_i^{(l)} &= \sum_{k \in \mathcal{N}_i} \sum_m f_{LN}\left(\text{MLP}\left(\mathbf{m}_{ik}^{(l)}\right)\right) \\ \mathbf{h}_i^{(l+1)} &= f_{JK}\left(\text{MLP}\left(\mathbf{h}_i^{(l)}\right), \text{SiLU}\left(f_{BN}\left(\mathbf{m}_i^{(l)}\right)\right)\right)\end{aligned} \tag{16}$$

where f_{JK} and f_{BN} are the jumping knowledge layer and batch normalization. Note that we stack node/edge-aware message passing layer alternately.

Output Block We derive the final atom predictions (y) by channeling the obtained atomic representations through two dense layers. This process effectively maps the embeddings to the appropriate dimensionality. Given that the prediction tasks for crystal properties are graph-level, we employ a weighted average strategy for the atomic representations.

Experiments

Experimental setup

Baselines Baseline methods encompass CGCNN (Xie and Grossman 2018), SchNet (Schütt et al. 2017), MEGNET (Chen et al. 2019), and others up to PotNet (Lin et al. 2023). For these methods, results are sourced from their respective publications or directly from the authors unless noted otherwise. All CrystalFormer models employ the Adam optimizer (Kingma and Ba 2015), weight decay (Loshchilov and Hutter 2018), and a one-cycle learning rate scheduler (Smith and Topin 2018). Learning rates and training epochs are mildly adjusted, starting from 0.0005 and 1000 respectively, depending on the task. Specific configurations for each task can be found in the Appendix.

Datasets We test on five crystal property prediction tasks using the JARVIS (Choudhary et al. 2020) benchmark, specifically its DFT-2021.8.18 3D version, which features 55,722 crystals. Our evaluations focus on essential crystal properties, including two types of bandgaps (OPT functional and MBJ potential), total energy, formation energy, and Ehull. We adopt data splits as per (Lin et al. 2023; Yan et al. 2022) to ensure a fair comparison.

Metrics We evaluate performance and scalability. The primary performance metric is mean absolute error (MAE). Scalability assesses training time per epoch and throughput during inference[2], labeled *Train* and *Infer* respectively.

Model Evaluation

Experimental Results The results for Jarvis are presented in Table. 1. CrystalFormer surpasses baseline methods in 4 out of 5 tasks: bandgap (OPT), bandgap (MBJ), total energy, and Ehull. Its performance on formation energy is slightly below the state-of-the-art but remains competitive. CrystalFormer's robust graph construction technique accurately encodes unique crystal structures and effectively differentiates between crystals. Unlike PotNet, which focuses solely on interatomic geometric information, CrystalFormer also retains angular data, leading to an average improvement of 26.9% across the 5 tasks. In comparison to Matformer, CrystalFormer not only boasts the aforementioned advantages but also processes long-range information through 2-hop adjacent edges, akin to PotNet's approach with complete interatomic potentials. This capability significantly elevates CrystalFormer and PotNet above Matformer. Contrasting ALIGNN, CrystalFormer sidesteps the need for large-scale line graphs, resulting in lower computational costs. Notably, CrystalFormer achieved the best performance for Bandgap(MBJ) in JARVIS with just 14,537 training samples, underscoring its adaptability to varied data scales. In essence, these exceptional results underscore the efficacy of our Matformer's periodic pattern encoding in message passing.

[2]Throughput refers to the model's average processing speed in samples per second during its forward pass.

	Bandgap (OPT)	Bandgap (MBJ)	Total Energy	Formation Energy	Ehull
Method	eV	eV	eV/atom	eV/atom	eV
CGCNN (2018)	0.20	0.41	0.078	0.063	0.17
SchNet (2017)	0.19	0.43	0.047	0.045	0.14
MEGNET (2019)	0.145	0.34	0.058	0.047	0.084
GATGNN (2020)	0.17	0.51	0.056	0.047	0.12
ALIGNN (2021)	0.142	0.31	0.037	<u>0.033</u>	0.076
Matformer (2022)	0.137	0.30	0.035	<u>0.033</u>	0.064
PotNet (2023)	<u>0.127</u>	<u>0.27</u>	<u>0.032</u>	**0.029**	<u>0.055</u>
CrystalFormer	**0.101** (-20.5%)	**0.13** (-51.9%)	**0.017** (-46.9%)	<u>0.033</u> (+12.1%)	**0.040** (-27.3%)

Table 1: Comparison between CrystalFormer and baselines regarding MAE (↓) on JARVIS. The best results are shown in bold and the second best results are shown with underlines. We also provide the performance gain in percentage improvement (%) between CrystalFormer and the state-of-the-art. All scores have $\pm 0.5\%$ confidence intervals.

	Time Efficiency (↓)	Graph Size (↑)
Methods	ms/structure	# edges
Voronoi	79.24	126.1
Ours	**8.67**	**207.9**

Table 2: Comparison of the two robust graph construction methods (Voronoi and ours).

	MAE (↓)	Train (↓)	Infer (↑)
Methods	eV	s/epoch	samples/s
ALIGNN	0.31	107	35.7
Matformer (SBF)	0.31	56	55.9
Matformer (RBF)	0.30	53	56.7
CrystalFormer	**0.13**	**43**	**69.6**

Table 3: Comparison of CrystalFormer and baselines that use angular information.

Graph Construction Efficiency We examined the efficiency of two graph construction techniques: the Voronoi-based method and our proposal, comparing them in terms of time efficiency and neighbor count, as detailed in Table 2. Our method constructs a crystal structure in approximately 8.67 milliseconds, which is about 8.14 times faster than the Voronoi-based approach that takes 79.24 milliseconds. Moreover, our technique extracts an average of 207.9 edges per graph, approximately 0.65 times more than the Voronoi counterpart. Previous studies (Xie and Grossman 2018) have illustrated that an increased neighbor count can enhance performance. Consequently, our method proves superior in both time efficiency and neighbor extraction.

Model Efficiency We assessed CrystalFormer's efficiency against other models integrating angular information, using bandgap (MBJ) as the consistent dataset for all models. Results presented in Table 3 show that, CrystalFormer considerably surpasses ALIGNN in training speed (43 vs. 107 s/epoch) and inference efficiency. It also marginally exceeds Matformer, which incorporates preprocessed angular data using SBF with Spherical Harmonics (Gasteiger, Groß, and Günnemann 2020) and RBF. Notably, CrystalFormer avoids the complexities of line graph construction and angular data preprocessing, enhancing its efficiency relatively.

Graph Construction	Robustness	MAE (↓)
Multi-edge	×	0.28
k-connected	×	0.30
Modified k-connected	✓	0.29
Ours	✓	**0.13**

Table 4: Comparison of different graph construction methods. We provide the availability of robustness and model performance (MAE) on bandgap (MBJ).

Ablation Studies We conducted an ablation study to assess the significance of various components in our method. This analysis confirms the importance of both the robust encoding of periodic invariance (Table 4) and our angle-preserving graph transformer block to the model's performance. Notably, we contrasted CrystalFormer, built using our graph construction method, with one leveraging previous multi-edge and k-connected graph construction (Yan et al. 2022). We also considered a modified version of the k-connected graph for a thorough assessment. Additionally, we conduct another ablation study by excluding the angular-preserving graph transformer blocks while retaining our graph construction. We found CrystalFormer surpassed its counterpart by approximately 53% (0.13 vs. 0.28 in terms of MAE).

Conclusion

In our study on predicting crystal materials properties, we introduced a graph construction method ensuring robust periodic invariance and a technique for preserving angular information in GNNs. The developed CrystalFormer highlights an innovative angular attention mechanism, reinforcing long-range data. Rigorous testing on the Jarvis materials benchmark confirmed our method's superior performance, heralding a potential to accelerate materials design.

Acknowledgements

This project is partially supported by the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Futures program; the National Science Foundation, the Air Force Office of Scientific Research; the Department of Energy (DOE); and the Toyota Research Institute (TRI).

References

Ackland, G.; and Jones, A. 2006. Applications of local crystal structure measures in experiment and simulation. *Physical Review B*, 73(5): 054104.

Balasingham, J.; Zamaraev, V.; and Kurlin, V. 2022. Compact Graph Representation of molecular crystals using Point-wise Distance Distributions. *arXiv preprint arXiv:2212.11246*.

Brody, S.; Alon, U.; and Yahav, E. 2021. How Attentive are Graph Attention Networks? In *International Conference on Learning Representations*.

Chen, C.; Ye, W.; Zuo, Y.; Zheng, C.; and Ong, S. P. 2019. Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals. *Chemistry of Materials*, 31(9): 3564–3572.

Choudhary, K.; and DeCost, B. 2021. Atomistic Line Graph Neural Network for improved materials property predictions. *npj Computational Materials*, 7(1): 1–8.

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): 1–13.

Gasteiger, J.; Groß, J.; and Günnemann, S. 2020. Directional message passing for molecular graphs. *arXiv preprint arXiv:2003.03123*.

Hornik, K.; Stinchcombe, M.; and White, H. 1989. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5): 359–366.

Kingma, D. P.; and Ba, J. 2015. Adam: A Method for Stochastic Optimization. In *Proceedings of the 3rd International Conference on Learning Representations*.

Lévy, B.; Petitjean, S.; Ray, N.; and Maillot, J. 2023. Least squares conformal maps for automatic texture atlas generation. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, 193–202.

Lin, Y.; Yan, K.; Luo, Y.; Liu, Y.; Qian, X.; and Ji, S. 2023. Efficient Approximations of Complete Interatomic Potentials for Crystal Property Prediction. *arXiv preprint arXiv:2306.10045*.

Loshchilov, I.; and Hutter, F. 2018. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*.

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.

Nehari, Z. 2012. *Conformal mapping*. Courier Corporation.

Pan, H.; Ganose, A. M.; Horton, M.; Aykol, M.; Persson, K. A.; Zimmermann, N. E.; and Jain, A. 2021. Benchmarking coordination number prediction algorithms on inorganic crystal structures. *Inorganic chemistry*, 60(3): 1590–1603.

Rampášek, L.; Galkin, M.; Dwivedi, V. P.; Luu, A. T.; Wolf, G.; and Beaini, D. 2022. Recipe for a general, powerful, scalable graph transformer. *Advances in Neural Information Processing Systems*, 35: 14501–14515.

Ruff, R.; Reiser, P.; Stühmer, J.; and Friederich, P. 2023. Connectivity Optimized Nested Graph Networks for Crystal Structures. *arXiv preprint arXiv:2302.14102*.

Scarselli, F.; and Tsoi, A. C. 1998. Universal approximation using feedforward neural networks: A survey of some existing methods, and some new results. *Neural networks*, 11(1): 15–37.

Schinzinger, R.; and Laura, P. A. 2012. *Conformal mapping: methods and applications*. Courier Corporation.

Schütt, K.; Kindermans, P.-J.; Sauceda Felix, H. E.; Chmiela, S.; Tkatchenko, A.; and Müller, K.-R. 2017. SchNet: A Continuous-filter Convolutional Neural Network for Modeling Quantum Interactions. *Advances in Neural Information Processing Systems*, 30.

Smith, L. N.; and Topin, N. 2018. Super-convergence: Very fast training of residual networks using large learning rates.

Vassiliades, V.; Chatzilygeroudis, K.; and Mouret, J.-B. 2017. Using centroidal voronoi tessellations to scale up the multidimensional archive of phenotypic elites algorithm. *IEEE Transactions on Evolutionary Computation*, 22(4): 623–630.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Voronoi, G. 1908. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les parallélloèdres primitifs. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 1908(134): 198–287.

Widdowson, D.; and Kurlin, V. 2022. Resolving the data ambiguity for periodic crystals. *Advances in Neural Information Processing Systems*, 35: 24625–24638.

Widdowson, D.; Mosca, M. M.; Pulido, A.; Cooper, A. I.; and Kurlin, V. 2022. Average Minimum Distances of periodic point sets-foundational invariants for mapping periodic crystals. *MATCH-COMMUNICATIONS IN MATHEMATICAL AND IN COMPUTER CHEMISTRY*, 87(3): 529–559.

Wu, T.; and Yau, S.-T. 2020. Computing harmonic maps and conformal maps on point clouds. *arXiv preprint arXiv:2009.09383*.

Xie, T.; and Grossman, J. C. 2018. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Physical review letters*, 120(14): 145301.

Xu, K.; Li, C.; Tian, Y.; Sonobe, T.; Kawarabayashi, K.-i.; and Jegelka, S. 2018. Representation learning on graphs with jumping knowledge networks. In *International conference on machine learning*, 5453–5462. PMLR.

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

Ying, C.; Cai, T.; Luo, S.; Zheng, S.; Ke, G.; He, D.; Shen, Y.; and Liu, T.-Y. 2021. Do Transformers Really Perform Badly for Graph Representation? *Advances in Neural Information Processing Systems*, 34.

Zhang, Q.; Li, H.; and Wang, Q. 2023. Wide-Angle Rectification via Content-Aware Conformal Mapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17357–17365.