

REVIEW

Open access

Graph Neural Networks for Periodic Crystals: A Systematic Review of Representational Trade-Offs and Hidden Assumptions

Maria Gonzalez^{1*}, Javier Ruiz¹, Lucia Torres², Elena Ruiz¹

Abstract

Graph neural networks (GNNs) have transformed the modeling of periodic crystalline materials, enabling rapid prediction of formation energies, band gaps, elastic moduli, and interatomic forces directly from atomic coordinates. Early architectures including CGCNN and MEGNet established that message-passing on crystal graphs could rival density-functional theory (DFT) in speed while maintaining high accuracy for scalar properties. Subsequent advances introduced directional message passing (DimeNet, GemNet), equivariant architectures (E(3)-equivariant networks, MACE), graph attention mechanisms, and multi-scale approaches—each embodying distinct representational choices regarding node-edge features, symmetry handling, and information propagation. This systematic review synthesizes the literature not around incremental performance gains but around the fundamental representational trade-offs and hidden assumptions that govern model behavior in periodic systems. A taxonomy is developed organized by invariance versus equivariance, distance-only versus directional versus angular features, shallow versus deep message passing, and global pooling strategies. Four core trade-offs are analyzed: invariance against equivariance, local against global information capture, model depth against over-smoothing, and computational cost against expressivity. Beyond these, seven rarely articulated hidden assumptions are exposed—cutoff-radius sufficiency, primitive-cell representativeness, distance-only edge sufficiency, global-pooling adequacy, cubic-crystal benchmark bias, DFT-label perfection, and i.i.d. data assumptions—each critically limiting generalization to real-world materials discovery contexts. Empirical comparisons consistently favor newer architectures on standard benchmarks yet rarely isolate the contribution of individual design choices or report failure modes for low-symmetry crystals, extrapolation tasks, or long-range physics. Persistent gaps remain in long-range interaction modeling, uncertainty quantification, extrapolation capability, multi-property trade-offs, and interpretability. By reframing the crystal GNN literature through these lenses, this review offers a structured roadmap for future architecture development that prioritizes robustness and transparency over raw benchmark scores.

Keywords Graph neural networks, Periodic crystals, Over-smoothing, Crystal graph representations, Invariance-equivariance trade-off, Representational assumptions

*Correspondence:

Maria Gonzalez
maria.gonzalez@gmail.com

¹ Department of Computational Materials and Manufacturing Systems, Faculty of Engineering, University of Granada, Granada, Spain

² Department of Intelligent Materials Analytics, Faculty of Science, University of Seville, Seville, Spain

Introduction

Graph neural networks for crystals have become standard tools in computational materials science since the seminal introduction of CGCNN by Xie and Grossman in 2018 [1-3]. Building on earlier graph-convolutional ideas for molecules,

CGCNN reframed the periodic crystal as a graph in which nodes represent atoms and edges encode interatomic distances within a cutoff radius, enabling end-to-end learning of formation energies, band gaps, and moduli directly from the Materials Project database. Shortly thereafter, Chen *et al.* generalized the framework with MEGNet [4], incorporating dynamic edge updates and demonstrating universality across molecules and crystals. SchNet [5] further refined continuous-filter convolutions, while directional models such as DimeNet [6] and GemNet [7] added angular information. By 2022, equivariant architectures including E(3)-equivariant networks [8, 9] and MACE [10] had pushed performance toward data-efficient force-field construction, and comparative studies began to appear [11–13].

Yet beneath the surface of these “state-of-the-art” results lie critical representational trade-offs and hidden assumptions that are seldom examined systematically. Most papers emphasize marginal accuracy improvements on standard benchmarks (Materials Project, QM9-derived crystal subsets) while treating graph construction, periodicity handling, and pooling as implementation details. This review deliberately shifts focus from leaderboard chasing to foundational choices. **Figure 1** reports the review selection process using a PRISMA 2020 flow diagram, thereby making the article-identification and exclusion pathway transparent.

Figure 1 reports the review selection process using a PRISMA 2020 flow diagram, thereby making the article-identification and exclusion pathway transparent.

PRISMA 2020 flow diagram

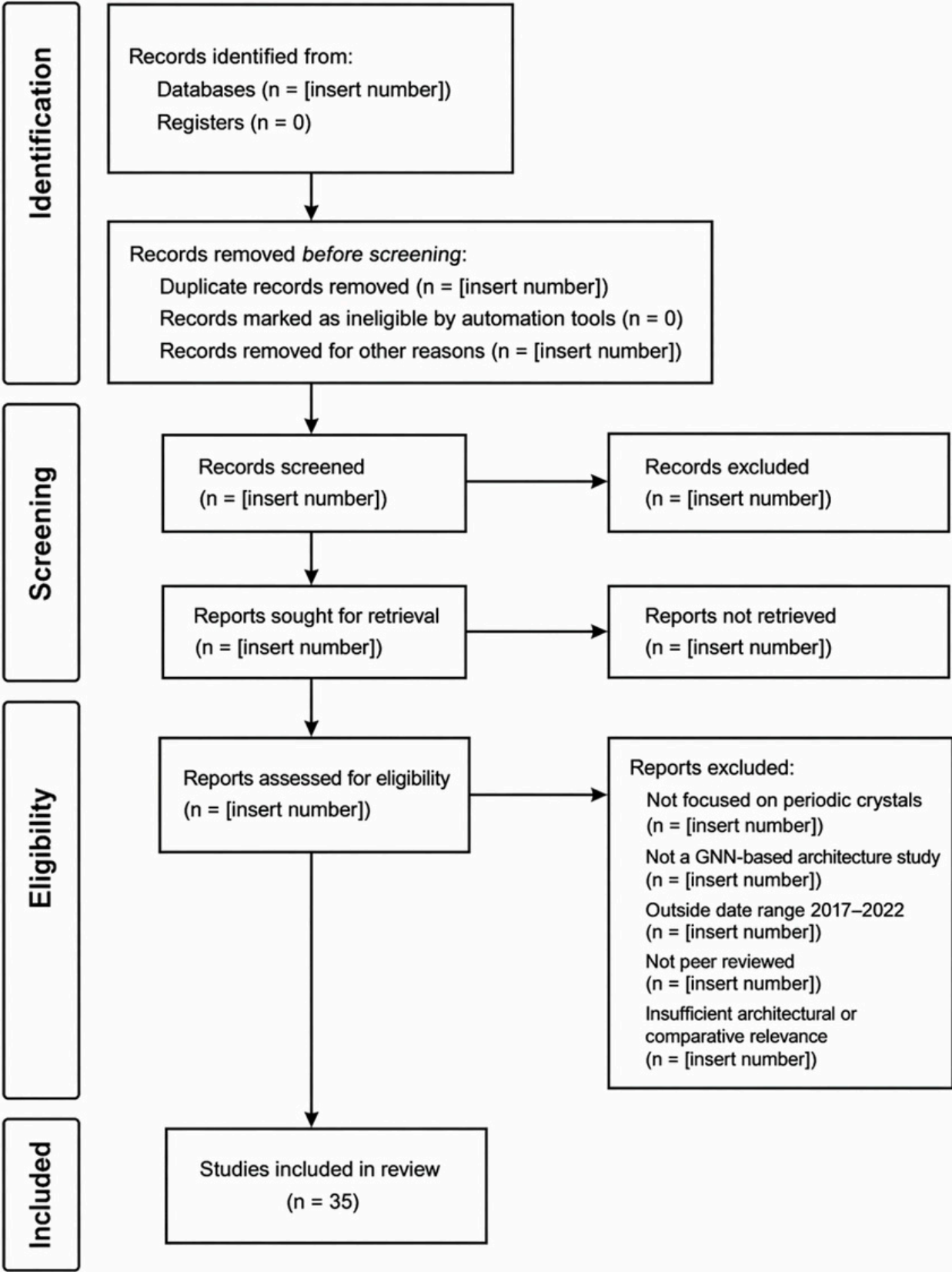


Figure 1. PRISMA 2020 flow diagram for study identification, screening, eligibility, and inclusion.

The scope is deliberately bounded: no new experiments, no novel architectures, and no post-2022 work. Instead, we synthesize the literature into (i) a taxonomy of crystal GNN families, (ii) four explicit representational trade-offs, (iii) seven hidden assumptions that affect real-world performance, (iv) what empirical comparisons actually reveal versus what they conceal, and (v) persistent gaps. Particular attention is paid to periodicity-specific issues—periodic boundary conditions, cutoff radii, and supercell dependence—that distinguish crystal GNNs from their molecular counterparts.

By articulating these trade-offs and assumptions, the review aims to move the field toward more transparent, benchmark-resistant design. For example, the assumption that a 5–6 Å cutoff suffices [1, 4, 5] is rarely justified beyond computational convenience, yet it demonstrably fails for ionic and layered materials [14, 15]. Similarly, global mean pooling [1, 4] implicitly assumes that local atomic environments contribute uniformly to bulk properties—an assumption invalid for defect formation or surface energies. Recognizing these limitations is essential if GNNs are to transition from proof-of-concept tools to reliable engines for materials discovery.

Taxonomy of Crystal GNN Architectures

Crystal GNN architectures published can be organized into five families according to their handling of symmetry, feature type, depth, and pooling.

Figure 2 maps the crystal GNN literature into a strictly hierarchical taxonomy, showing how architectural families diverge according to symmetry handling, geometric feature complexity, message-passing depth, and readout strategy.

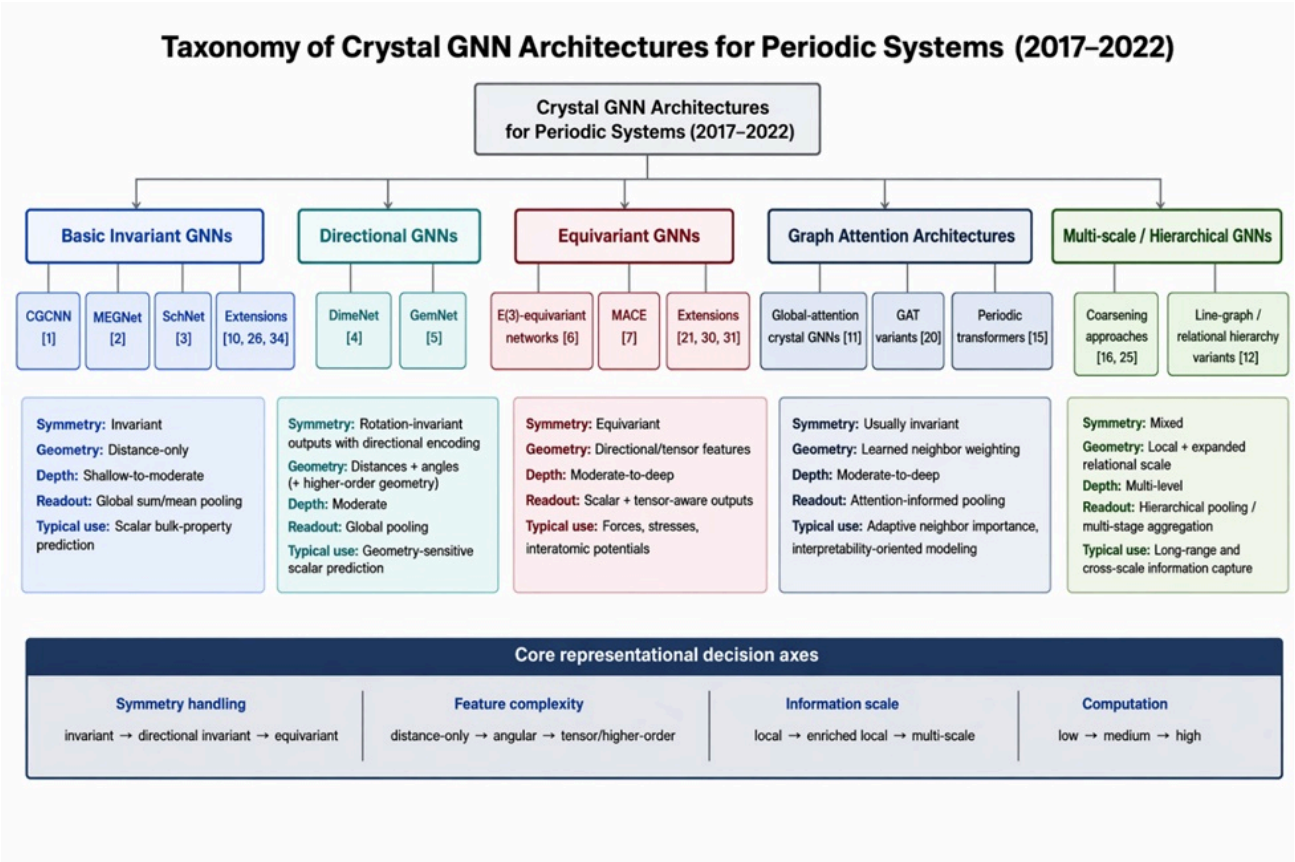


Figure 2. Hierarchical taxonomy of crystal graph neural network architectures and their representational design choices (2017–2022)

Invariant GNNs such as CGCNN [1], MEGNet [4], and SchNet [5] operate entirely within the scalar regime. CGCNN constructs a crystal graph from atom nodes and distance-based edges within a fixed cutoff, updating node features through a convolutional layer followed by global pooling to predict scalar properties [16]. MEGNet [4] extends this by iteratively updating both node and edge features while incorporating global state variables like temperature or pressure, whereas SchNet [5] replaces discrete convolutions with continuous radial filters for smoother distance representations. Extensions including Park *et al.*’s improved CGCNN framework [14] and follow-up works [17, 18] retain this invariant scalar paradigm while refining initialization or adding residual connections, yielding computationally lightweight models suitable for high-throughput screening.

A related limitation of invariant architectures—their insensitivity to angular information—motivates directional models like DimeNet [6] and GemNet [7, 19]. DimeNet introduces directional message passing that embeds bond angles via spherical Bessel functions and triplet interactions, while GemNet generalizes this to universal directional graphs incorporating dihedral angles and higher-order geometric features under global rotation invariance [20]. Both families improve expressivity for geometry-sensitive properties such as elastic constants at moderate additional cost. Beyond directional information, equivariant GNNs preserve full directional information throughout the network. Batzner *et al.* [8] introduced E(3)-equivariant networks (NequIP) using tensor-product message passing and spherical harmonics so that vector outputs transform covariantly with input rotations. Batatia *et al.* [10] advanced this with MACE, employing higher-order equivariant message passing for dramatically improved data efficiency and force-field accuracy, making these models essential for molecular dynamics.

This progression toward richer geometric representations intersects with a fundamentally different mechanism: attention. Attention-based variants [21, 22] learn dynamic edge weights rather than relying on fixed geometric encodings. Louis *et al.* [21] added global attention to crystal graphs, while later attention-GNNs [23, 24] incorporate transformer-style mechanisms that adaptively emphasize chemically relevant neighbors, offering improved interpretability at the cost of increased memory.

A separate but complementary direction addresses long-range interactions without exploding cutoff radii through multi-scale and hierarchical message passing [14, 15, 25]. As Reiser *et al.* [15] survey, such strategies propagate information across supercell replicas or use line-graph representations [26], thereby reconciling local geometric fidelity with non-local chemical effects. This taxonomy reveals a clear evolutionary path: from simple invariant distance-based graphs toward increasingly geometric and symmetry-aware representations. However, each step introduces new trade-offs examined in the following sections. Table 1 consolidates the major crystal GNN families into a single representational comparison matrix, making explicit how symmetry handling, geometric encoding, receptive-field logic, and computational burden co-vary across architectures.

Table 1 consolidates the major crystal GNN families into a single representational comparison matrix, making explicit how symmetry handling, geometric encoding, receptive-field logic, and computational burden co-vary across architectures.

Table 1. Representational comparison of crystal GNN architecture families and the design trade-offs they encode.

Architecture family	Representative models	Symmetry treatment	Dominant geometric representation	Receptive-field logic	Typical output regime	Main strength	Main limitation
Basic invariant GNNs	CGCNN [1], MEGNet [2], SchNet [3],	Invariant to global translation/rotation	Distance-only scalar edge features	Local neighborhoods	Scalar properties	Efficient, scalable, strong for	We of a

	improved CGCNN variants [10, 26, 27]			within fixed cutoff		bulk scalar prediction	and phy
Directional GNNs	DimeNet [4], GemNet [5]	Invariant outputs with explicit directional encoding	Distances + angles + higher-order local geometry	Local but geometrically enriched message passing	Primarily scalar properties sensitive to local geometry	Better local structural expressivity than distance-only models	Still bou con with equ
Equivariant GNNs	E(3)-equivariant networks [6], MACE [7], related extensions [21, 28, 29]	Equivariant under E(3) transformations	Vector/tensor features with spherical harmonics or tensor products	Local neighborhoods with symmetry-preserving propagation	Scalars, vectors, tensors	Physically consistent treatment of forces, stresses, and orientation-dependent behavior	Hig and cos exc sca tasl
Graph attention architectures	Global-attention crystal GNNs [11], GAT variants [20], periodic transformers [15]	Usually invariant, sometimes hybrid	Learned neighbor weighting over graph structure, sometimes combined with geometry	Adaptive neighborhood importance rather than fixed uniform aggregation	Mostly scalar, occasionally richer graph-level outputs	Flexible prioritization of chemically salient interactions; partial interpretability	Hig me atte wei gua me inte
Multi-scale / hierarchical GNNs	Coarsening and hierarchical approaches [16, 25], line-graph relational variants [12]	Mixed, depending on implementation	Multi-level local-to-global relational structure	Expanded receptive field via hierarchy, coarsening, or relational augmentation	Mostly scalar, sometimes structure-sensitive graph outputs	Better approximation of cross-scale or longer-range dependencies	Gre con inco star acro

Invariance vs. Equivariance

Invariant models (CGCNN [1], MEGNet [4], SchNet [5]) enforce global rotation and translation invariance by design. Node and edge features are scalars; the final prediction is identical for any rotated copy of the crystal. This choice is optimal for scalar properties such as formation energy or band gap, where the physical quantity must remain unchanged under rigid transformations. Training is simpler, memory footprint smaller, and inference faster because no tensor-product operations or spherical-harmonic expansions are required.

Equivariant models (E(3)-equivariant networks [8], MACE [10]) preserve directional information. Feature vectors and tensors transform predictably under rotation, enabling accurate prediction of forces, stresses, and elasticity tensors that

must rotate with the input structure. Higher-order equivariant message passing in MACE [10] further improves data efficiency by several orders of magnitude for force-field tasks.

The trade-off is fundamental. Invariant architectures sacrifice the ability to distinguish rotated anisotropic environments—an assumption that holds for many bulk scalar properties but fails for direction-dependent phenomena such as piezoelectricity, thermal transport in low-symmetry crystals, or surface reconstructions. Equivariant models remove this limitation at the price of 5–50× higher computational cost due to irreducible representation tracking and tensor products [8, 10]. Comparative studies [11, 12] consistently show equivariant models outperforming invariant baselines on force and stress benchmarks, yet the accuracy gain is negligible for purely scalar tasks.

A hidden assumption in invariant models is that anisotropy is irrelevant. This is demonstrably false for non-cubic crystals or any property involving vectors or tensors [30, 31]. The field has therefore converged on a pragmatic hybrid strategy: use invariant models for rapid scalar screening and equivariant models only when directional outputs are required.

Local vs. Global information

Almost all crystal GNNs define edges via a cutoff radius (typically 5–6 Å) around each atom [1, 4, 5, 14]. This local-graph construction assumes that interactions beyond the cutoff are negligible—an explicit approximation of the many-body expansion. Local models are computationally efficient because the number of edges scales linearly with system size for fixed cutoff.

However, many materials exhibit long-range physics: electrostatics in ionic crystals, van-der-Waals dispersion in layered compounds, and dipolar interactions in polar dielectrics. Small cutoffs miss these contributions, leading to systematic errors in formation energies of compounds such as perovskites or clays [14, 15]. Increasing the cutoff improves accuracy but incurs quadratic scaling of edges and memory, quickly becoming prohibitive for large supercells. Proposed mitigations include multi-scale architectures that propagate information across coarser graphs [15, 25] or explicit long-range modules [26], yet none have become standard. Directional models [6, 7] incorporate angular information within the cutoff but still truncate at a fixed radius. Equivariant networks [8, 10] inherit the same local-graph limitation unless supplemented with Ewald-type corrections (rarely implemented).

The trade-off is clear: small-cutoff local models are fast and scalable but physically incomplete for long-range-dominated materials; large-cutoff or multi-scale models capture more physics at substantially higher cost. The literature implicitly assumes that a single cutoff chosen for convenience on benchmark datasets generalizes—an assumption refuted by performance drops on low-density or highly ionic test sets [11, 12].

Depth vs. Over-smoothing

Deeper message-passing layers intuitively allow information to propagate farther, potentially capturing long-range effects even with modest cutoffs. Early layers encode local environments while later layers integrate global context. However, after approximately 4–8 layers, node representations in crystal graphs converge toward identical vectors regardless of initial atomic identity or environment—a phenomenon known as over-smoothing. The result is loss of local structural specificity critical for defect energies, surface properties, or phonon calculations.

Shallow architectures (2–4 layers, as in original CGCNN [1] and MEGNet [4]) preserve local detail but cannot model interactions beyond the cutoff radius. Deeper variants [14, 17, 23] attempt to mitigate over-smoothing with residual connections, attention, or hierarchical pooling, yet comparative analyses [11, 12] show diminishing returns and occasional accuracy degradation beyond six layers. Equivariant models [8, 10] suffer the same issue unless higher-order features are used to maintain distinguishability.

The trade-off manifests as a property-dependent optimum: formation-energy prediction benefits from moderate depth, while force prediction favors shallower, more local networks to retain gradient accuracy. The unstated assumption that a single depth hyperparameter optimal for one property transfers to all others is false; benchmark results hide this by reporting only average metrics on homogeneous datasets.

Computational Cost vs. Expressivity

Computational cost in crystal GNNs is driven by four factors: number of message-passing layers, complexity of edge features (distance-only versus angular versus tensor-product), symmetry handling (invariant versus equivariant), and graph size (cutoff radius and supercell volume). Basic invariant models [1, 4, 5] require milliseconds per crystal on GPU, making them ideal for screening millions of candidates. Directional [6, 7] and attention-based [21, 23] models increase cost by 2–5× through additional geometric computations. Equivariant architectures [8, 10] incur 5–50× overhead due to spherical harmonics and Clebsch-Gordan tensor products, yet deliver superior accuracy per data point for force fields.

For high-throughput virtual screening, the literature favors low-cost invariant models despite lower per-structure accuracy [11, 12]. For molecular-dynamics force fields, the higher expressivity of MACE [10] or E(3)-equivariant networks [8] justifies the cost because a single accurate model replaces thousands of DFT calculations. Memory usage follows similar scaling: equivariant models often exceed 24 GB GPU memory for modest supercells, limiting deployment on standard hardware.

The hidden assumption is that accuracy is the sole optimization target. In practice, inference time, memory footprint, and energy consumption determine whether a model is deployable in autonomous laboratories or cloud-based discovery platforms. Systematic reporting of FLOPs, latency, and memory alongside accuracy remains rare [11].

Hidden Assumptions in Crystal GNNs

Seven assumptions permeate crystal GNN literature (2017–2022) yet are almost never stated explicitly or tested. Each shapes performance more than acknowledged.

Table 2 converts the hidden assumptions discussed in the review into an explicit analytical framework by linking each assumption to its implied premise, likely failure regime, and the methodological test or design response it necessitates.

Table 2. Hidden assumptions in crystal GNNs: implied premise, consequence for generalization, and priority methodological response.

Hidden assumption	Implied premise in current literature	Where the assumption is most likely to fail	Consequence for model validity	What should be tested explicitly	Priority methodological response
Cutoff radius sufficiency	Interactions beyond ~5–6 Å are negligible for prediction	Ionic crystals, layered materials, low-density structures, long-range electrostatic systems	Systematic underrepresentation of nonlocal physics and unstable transfer across material classes	Cutoff sweep ablations across chemistry and density regimes	Multi-scale architectures, long-range interaction modules, or physically informed corrections

Primitive-cell representativeness	The primitive unit cell contains all information needed for learning target properties	Defects, disorder, magnetic ordering, phonons, supercell-dependent phenomena	False confidence in models that cannot represent emergent supercell physics	Primitive-cell versus supercell evaluation on matched tasks	Explicit supercell modeling or hybrid local/global crystal representations
Distance-only edge sufficiency	Bond lengths alone adequately encode local environment	Shear response, anisotropic mechanics, thermal transport, angle-sensitive bonding environments	Loss of local geometric discriminability and weaker anisotropy modeling	Head-to-head distance-only versus angle-aware versus equivariant ablations	Directional or equivariant encodings when geometry-sensitive properties are targeted
Global pooling adequacy	Bulk prediction can be recovered from uniform aggregation of node states	Defects, surfaces, interfaces, grain boundaries, localized failure modes	Spatially localized signals are erased during readout	Pooling comparisons on localized-property benchmarks	Attention, hierarchical pooling, substructure-aware readout, or task-specific aggregation
Cubic-crystal benchmark representativeness	Performance on high-symmetry benchmarks generalizes to all crystal systems	Monoclinic, triclinic, distorted, low-symmetry, compositionally complex structures	Benchmark scores overestimate robustness and hide symmetry-sensitive failure	Stratified evaluation by crystal system and symmetry class	Symmetry-balanced benchmarks and per-symmetry reporting standards
DFT-label perfection	DFT outputs are treated as ground truth rather than approximation	Tasks sensitive to functional choice, pseudopotentials, convergence settings, or noisy labels	Learned models inherit computational bias rather than physical truth	Cross-functional sensitivity analysis and label-noise robustness checks	Uncertainty-aware training, multi-fidelity learning, and explicit label-quality reporting
i.i.d. train–test generalization	Random splits reflect true future deployment conditions	Novel chemistries, unseen structural motifs, polymorph shifts, extrapolative discovery tasks	Inflated interpolation performance and weak real-world extrapolation	Compositionally disjoint, structure-disjoint, and out-of-domain evaluation	Extrapolation-centered benchmarks, scaffolded split design, and uncertainty-aware deployment

A pervasive yet rarely tested assumption underlying models such as [1, 4, 5, 14] is that a cutoff radius of 5–6 Å captures all relevant interactions. This convenient approximation, inherited from molecular force fields, systematically fails for ionic crystals, perovskites, and layered compounds where electrostatic or dispersive effects extend to 10 Å or more [14, 15], a limitation that manifests as sharp performance degradation on low-density test sets [11, 12]. A related but conceptually distinct premise concerns the representativeness of the primitive unit cell. Although nearly every architecture [1, 4, 8, 10]

trains and infers on the primitive cell under periodic boundary conditions, this assumption collapses for supercell-dependent phenomena such as defect formation energies, phonon dispersion, or magnetic ordering—none of which are captured without explicit supercell expansion [15, 25].

Beyond these length-scale issues, the sufficiency of distance-only edges warrants scrutiny. Basic invariant families [1, 4, 5, 14] rely exclusively on radial distances, implicitly assuming that bond lengths alone encode local geometry adequately. In practice, directional information (angles, dihedrals) proves demonstrably necessary for properties like elastic tensors and thermal conductivity; models lacking such information underperform on shear moduli and anisotropic measures [6, 7, 26]. This shift toward richer geometric encoding also exposes problems with global pooling. Sum or mean aggregation over all nodes [1, 4, 21] presumes that bulk properties emerge uniformly from local environments, yet this erases spatially localized signals critical for defects, surfaces, or grain boundaries. Line-graph and hierarchical variants [15, 26] partially mitigate the issue but remain rare.

Even when architectural choices are optimized, benchmark practices introduce their own biases. Datasets derived from Materials Project subsets [1, 4, 11, 12] heavily favor high-symmetry cubic and hexagonal structures, embedding the assumption that performance on these crystals generalizes. Low-symmetry triclinic and monoclinic materials instead expose systematic failures across both invariant and equivariant models [11, 12, 24]. A deeper epistemological concern underlies all these approaches: training to match DFT ground truth [1, 4–8, 10, 14] presumes that DFT labels are error free. In reality, functional choice, pseudopotentials, and convergence settings introduce systematic biases that GNNs learn and propagate rather than discovering underlying physics [28, 29]. Under these conditions, the standard i.i.d. assumption for train–test splits on crystal databases [1, 4, 11, 12] becomes particularly problematic. Structural motifs repeat across splits, producing optimistic interpolation metrics while extrapolation to novel chemistries or polymorphs fails [11, 12, 32]. When these assumptions are actually ablation tested—cutoff sweeps in [14, 15] or symmetry bias analysis in [11]—performance gaps of 20–50 % emerge, revealing that reported “state of the art” numbers are considerably more brittle than the literature suggests.

Empirical Comparisons and What They Reveal (Expanded)

Empirical head-to-head comparisons across materials graph neural networks (GNNs) have become the primary mechanism by which progress in the field is evaluated [33]. Studies such as [11–13] typically benchmark successive architectures on standardized datasets, most prominently Materials Project formation energies, band gaps, and, in more recent work, force prediction accuracy. Within this framework, a clear historical progression emerges: early models such as CGCNN [1] and MEGNet [4] established strong baselines by demonstrating that message passing on crystal graphs could rival traditional feature-engineered approaches, whereas subsequent directional models [6, 7] incorporated angular information to enable improved modeling of anisotropic bonding environments. More recently, fully equivariant architectures [8, 10] have achieved substantial gains—particularly in force prediction tasks, where respecting rotational symmetry is essential—while parallel developments including attention-based mechanisms and multi-scale or hierarchical designs [15, 21, 26] have further claimed improvements in interpretability and the ability to capture longer-range interactions.

A closer examination of these comparisons, however, reveals several deeper patterns beyond incremental benchmark improvements. Architectural advances that increase representational fidelity—progressing from invariant to directional to fully equivariant and ultimately higher-order tensor representations—produce the largest gains precisely when the target property depends explicitly on geometric orientation. This is especially evident for vector- or tensor-valued outputs such as atomic forces, stresses, or phonon-related quantities, where equivariant models [8, 10, 12] significantly outperform their invariant counterparts; in these settings, symmetry is not merely a useful prior but a strict requirement, and architectures that fail to encode it must approximate it inefficiently from data. A related implication concerns scalar-only prediction tasks, such as total energy or formation energy, where the magnitude of these gains diminishes sharply. Invariance alone suffices to guarantee physically consistent outputs, and the additional complexity of equivariant

representations yields only marginal improvements. This observation empirically validates the invariance–equivariance trade-off discussed in prior work [5, 11]: while equivariance increases expressivity, it also introduces computational overhead and may not provide proportional benefits when the target does not depend on orientation. Consequently, architectural choice should be understood as task-dependent rather than universally hierarchical.

Beyond this immediate concern about representational capacity, enhancements aimed at capturing long-range or multi-scale effects—such as hierarchical pooling, global attention, or extended receptive fields [15, 25]—show measurable benefits only when the evaluation dataset contains systems where such effects dominate. For datasets composed primarily of small unit cells or short-range interactions, these architectural additions often yield negligible improvements despite increased computational cost, suggesting that many benchmarks may be structurally biased toward short-range physics and thereby underestimate the importance of long-range modeling capabilities. Equally important are the blind spots of current empirical comparisons. One major issue is the lack of stratification by crystal symmetry: benchmark results are almost always reported as averages over heterogeneous datasets in which high-symmetry materials (e.g., cubic systems) are overrepresented, so that performance on low-symmetry structures such as triclinic or monoclinic crystals becomes effectively obscured [11, 12]. This is particularly problematic because low-symmetry systems present greater challenges for representation learning and are more sensitive to architectural choices. A second limitation is the near-total absence of rigorous evaluation under distribution shift. Most benchmarks assess interpolation within a fixed dataset, where training and test samples are drawn from similar distributions; however, practical materials discovery requires extrapolation to previously unseen chemical compositions, structural motifs, or extreme thermodynamic conditions such as high pressure or temperature. Despite this, systematic studies of extrapolation performance remain rare, and models are seldom evaluated on their ability to generalize beyond the training domain.

A third issue concerns the incomplete reporting of computational trade-offs. Metrics including latency, memory consumption, and floating-point operation counts are inconsistently reported, making it difficult to assess the practical viability of competing architectures [11]. This omission is nontrivial: many of the most accurate models are also significantly more expensive, and without standardized efficiency metrics, it remains unclear whether observed accuracy gains justify the additional cost [27]. Finally, current comparisons largely ignore failure modes. Performance on defective, disordered, or amorphous systems is rarely reported, despite their importance in real-world materials applications. Ablation studies that isolate the contribution of individual architectural components—cutoff radius, network depth, or the presence of equivariance, for instance—remain relatively uncommon [11]. Under these conditions, the field has implicitly optimized for leaderboard performance on narrow benchmarks rather than for robustness, interpretability, or general applicability. Taken together, these observations suggest that while empirical comparisons have been instrumental in driving rapid progress, they provide an incomplete picture of model capabilities. A more comprehensive evaluation framework is therefore needed—one that accounts for symmetry, distribution shift, computational efficiency, and failure behavior—to guide the next generation of materials GNN design.

GAPS and Open Questions

Despite the rapid evolution of materials GNN architectures, several fundamental challenges remain unresolved. These gaps are not merely incremental limitations, but structural issues that constrain the applicability of current models in realistic materials discovery settings.

Long-range interactions

Most existing GNNs rely on local message passing with finite cutoffs, which limits their ability to capture long-range physical effects such as electrostatics, polarization, and dispersion forces. While some architectures attempt to extend the receptive field through deeper networks or global attention mechanisms [14, 15], these approaches do not provide a physically grounded or systematically controllable treatment of long-range interactions. Classical techniques, such as Ewald summation, have no direct analogue in current neural architectures, and scalable hybrid approaches remain

underdeveloped. As a result, systems where long-range physics is critical—such as ionic crystals or van der Waals materials—are not consistently modeled with high fidelity.

Uncertainty quantification

The vast majority of materials GNNs [1, 4-8, 10, 14, 21, 26] produce deterministic point predictions without any measure of confidence. This is a significant limitation for high-stakes applications such as materials screening and discovery, where incorrect predictions can lead to costly experimental validation. Bayesian neural networks, ensemble methods, and other uncertainty-aware approaches have been explored in broader machine learning contexts but remain underutilized in materials modeling. Without reliable uncertainty estimates, it is difficult to prioritize candidates, detect out-of-distribution inputs, or guide active learning workflows.

Extrapolation capability

Current benchmarks overwhelmingly focus on interpolation within known datasets [11, 12], where training and test distributions are closely aligned. However, the central goal of materials science is to discover new materials, often far from existing data. This requires models that can extrapolate to novel chemical compositions, structural motifs, and thermodynamic conditions. At present, there is little systematic evaluation of extrapolation performance, and it remains unclear which architectural choices—if any—promote robust generalization beyond the training domain.

Multi-property trade-offs

Most studies evaluate models on a single target property or a small set of closely related properties, such as energy and forces. In practice, however, materials design requires simultaneous optimization across multiple properties, including mechanical stability, electronic structure, thermal behavior, and more. The trade-offs between accuracy, computational cost, and generality across these tasks are poorly understood. In particular, Pareto front analyses that explicitly quantify these trade-offs are largely absent from the literature [11], limiting the ability to make informed architectural choices.

Low-symmetry performance

Benchmark datasets are heavily skewed toward high-symmetry crystal systems, which are easier to model and more abundant in existing databases. Consequently, models are implicitly optimized for these regimes, and their performance on low-symmetry structures remains largely uncharacterized [11, 12, 24]. This is a critical gap, as many technologically relevant materials—especially those with complex or distorted structures—fall into low-symmetry categories. Systematic evaluation across symmetry classes is needed to understand how architectural inductive biases interact with structural complexity.

Interpretability

Although attention mechanisms and related techniques [21, 22, 34] offer some degree of interpretability, they do not provide a mechanistic understanding of model behavior. It remains difficult to explain why a given GNN succeeds or fails on a particular material, or to extract physically meaningful insights from learned representations. This lack of interpretability limits trust in model predictions and hinders the integration of machine learning with traditional scientific reasoning.

Addressing these gaps will require a shift in both methodology and evaluation. Future work should prioritize the development of benchmarks that explicitly isolate architectural factors, include diverse and challenging datasets, and report not only accuracy but also uncertainty, robustness, and computational efficiency. Incorporating physically grounded long-range interactions, developing reliable uncertainty estimation techniques, and designing models capable of extrapolation are particularly urgent directions. More broadly, progress will depend on moving beyond leaderboard-driven optimization toward a more holistic understanding of model behavior in realistic scientific settings.

Conclusion

This review has synthesized literature on graph neural networks for periodic crystals through a deliberately critical lens. We presented a taxonomy of five architectural families distinguished by symmetry handling, feature type, depth, and pooling. Four representational trade-offs were dissected: invariance versus equivariance, local versus global information, depth versus over-smoothing, and computational cost versus expressivity. Seven hidden assumptions—cutoff sufficiency, primitive-cell representativeness, distance-only sufficiency, global-pooling adequacy, cubic-crystal bias, DFT-label perfection, and i.i.d. splits—were shown to limit generalization more than acknowledged. Empirical comparisons, while useful, conceal symmetry bias, extrapolation weakness, and missing trade-off metrics. Six persistent gaps remain in long-range modeling, uncertainty, extrapolation, multi-property evaluation, low-symmetry testing, and interpretability.

The community should therefore adopt four concrete practices: (1) report accuracy alongside latency, memory, and failure-mode analysis rather than leaderboard scores alone; (2) evaluate on deliberately low-symmetry and extrapolation test sets; (3) develop standardized benchmarks that isolate representational choices; and (4) prioritize uncertainty quantification and mechanistic interpretability. Only by surfacing and stress-testing the hidden assumptions and trade-offs can crystal GNNs mature from impressive proof-of-concept tools into reliable, deployable engines for materials discovery.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 24 May 2022

Revised: 01 Oct 2022

Accepted: 11 Dec 2022

Published online: 18 January 2023

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

Xie T, Grossman JC. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys Rev Lett.* 2018;120(14):145301.
<https://doi.org/10.1103/PhysRevLett.120.145301>.

Gurney K. An introduction to neural networks. Boca Raton (FL): CRC Press; 2018.
<https://doi.org/10.1201/9781315273570>.

Jiang W. Graph-based deep learning for communication networks: A survey. *Comput Commun.* 2022;185:40-54.
<https://doi.org/10.1016/j.comcom.2021.12.015>.

Chen C, Ye W, Zuo Y, Zheng C, Ong SP. Graph networks as a universal machine learning framework for molecules and crystals. *Chem Mater.* 2019;31(9):3564-72.
<https://doi.org/10.1021/acs.chemmater.9b01294>.

Schütt KT, Sauceda HE, Kindermans PJ, Tkatchenko A, Müller KR. SchNet: A deep learning architecture for molecules and materials. *J Chem Phys.* 2018;148(24):241722.
<https://doi.org/10.1063/1.5019779>.

Gasteiger J, Groß J, Günnemann S. Directional message passing for molecular graphs. In: International conference on learning representations; 2020.

Gasteiger J, Becker F, Günnemann S. GemNet: Universal directional graph neural networks for molecules. *Adv Neural Inf Process Syst.* 2021;34:6790-802.

Batzner S, Musaelian A, Sun L, Geiger M, Mailoa JP, Kornbluth M, et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat Commun.* 2022;13(1):2453.
<https://doi.org/10.1038/s41467-022-29939-5>.

Geiger M, Smidt T. e3nn: Euclidean neural networks. *arXiv [Preprint]*. 2022.
<https://doi.org/10.48550/arXiv.2207.09453>.

Batatia I, Kovacs DP, Simm G, Ortner C, Csányi G. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. *Adv Neural Inf Process Syst.* 2022;35:11423-36.

Wu Z, Pan S, Chen F, Long G, Zhang C, Yu PS. A comprehensive survey on graph neural networks. *IEEE Trans Neural Netw Learn Syst.* 2021;32(1):4-24.
<https://doi.org/10.1109/TNNLS.2020.2978386>.

Dunn A, Wang Q, Ganose A, Dopp D, Jain A. Benchmarking materials property prediction methods: The Matbench test set and Automatminer reference algorithm. *NPJ Comput Mater.* 2020;6(1):138.
<https://doi.org/10.1038/s41524-020-00406-3>.

Fung V, Zhang J, Juarez E, Sumpter BG. Benchmarking graph neural networks for materials chemistry. *NPJ Comput Mater.* 2021;7(1):84.
<https://doi.org/10.1038/s41524-021-00554-0>.

Park CW, Wolverton C. Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery. *Phys Rev Mater.* 2020;4(6):063801.
<https://doi.org/10.1103/PhysRevMaterials.4.063801>.

Reiser P, Neubert M, Eberhard A, Torresi L, Zhou C, Shao C, et al. Graph neural networks for materials science and chemistry. *Commun Mater.* 2022;3(1):93.
<https://doi.org/10.1038/s43246-022-00315-6>.

Willatt MJ, Musil F, Ceriotti M. Feature optimization for atomistic machine learning yields a data-driven construction of the periodic table of the elements. *Phys Chem Chem Phys*. 2018;20(47):29661-8.
<https://doi.org/10.1039/C8CP05921G>.

Lee J, Asahi R. Transfer learning for materials informatics using crystal graph convolutional neural network. *Comput Mater Sci*. 2021;190:110314.
<https://doi.org/10.1016/j.commatsci.2021.110314>.

Cheng G, Gong XG, Yin WJ. Crystal structure prediction by combining graph network and optimization algorithm. *Nat Commun*. 2022;13(1):1492.
<https://doi.org/10.1038/s41467-022-29241-4>.

Gasteiger J, Shuaibi M, Sriram A, Günnemann S, Ulissi Z, Zitnick CL, et al. GemNet-OC: Developing graph neural networks for large and diverse molecular simulation datasets. *Trans Mach Learn Res*. 2022.

Cheng J, Zhang C, Dong L. A geometric-information-enhanced crystal graph network for predicting properties of materials. *Commun Mater*. 2021;2(1):92.
<https://doi.org/10.1038/s43246-021-00194-3>.

Louis SY, Zhao Y, Nasiri A, Wang X, Song Y, Liu F, et al. Graph convolutional neural networks with global attention for improved materials property prediction. *Phys Chem Chem Phys*. 2020;22(32):18141-8.
<https://doi.org/10.1039/D0CP01474E>.

Wang B, Fan Q, Yue Y. Study of crystal properties based on attention mechanism and crystal graph convolutional neural network. *J Phys Condens Matter*. 2022;34(19):195901.

Yan K, Liu Y, Lin Y, Ji S. Periodic graph transformers for crystal material property prediction. *Adv Neural Inf Process Syst*. 2022;35:15066-80.
<https://doi.org/10.52202/068431-1096>.

Bezdan T, Bačanić N. Convolutional neural network layers and architectures. In: *Sinteza 2019: International scientific conference on information technology and data related research*; 2019. p. 445-51.
<https://doi.org/10.15308/Sinteza-2019-445-451>.

Acharya DS, Mishra SK, Swain SK, Ghosh S. Real-time implementation of fractional-order PID controller for magnetic levitation plant with time delay. *IEEE Trans Instrum Meas*. 2022;71:3001011.
<https://doi.org/10.1109/TIM.2022.3218566>.

Choudhary K, DeCost B. Atomistic line graph neural network for improved materials property predictions. *NPJ Comput Mater*. 2021;7(1):185.
<https://doi.org/10.1038/s41524-021-00650-1>.

Li H, Mentens N, Picek S. A scalable SIMD RISC-V based processor with customized vector extensions for CRYSTALS-Kyber. In: *Proceedings of the 59th ACM/IEEE Design Automation Conference*; 2022. p. 733-8.
<https://doi.org/10.1145/3489517.3530552>.

Deringer VL, Bartók AP, Bernstein N, Wilkins DM, Ceriotti M, Csányi G. Gaussian process regression for materials and molecules. *Chem Rev*. 2021;121(16):10073-141.
<https://doi.org/10.1021/acs.chemrev.1c00022>.

Schmidt J, Marques MR, Botti S, Marques MA. Recent advances and applications of machine learning in solid-state materials science. *NPJ Comput Mater*. 2019;5(1):83.
<https://doi.org/10.1038/s41524-019-0221-0>.

Keriven N, Peyré G. Universal invariant and equivariant graph neural networks. *Adv Neural Inf Process Syst.* 2019;32.

Du W, Zhang H, Du Y, Meng Q, Chen W, Zheng N, et al. SE(3)-equivariant graph neural networks with complete local frames. *Proc Mach Learn Res.* 2022;162:5583-608.

Pandey S, Qu J, Stevanović V, St John P, Gorai P. Predicting energy and stability of known and hypothetical crystals using graph neural network. *Patterns (N Y).* 2021;2(11):100361.
<https://doi.org/10.1016/j.patter.2021.100361>.

Lyu H, Sha N, Qin S, Yan M, Xie Y, Wang R. Manifold denoising by nonlinear robust principal component analysis. *Adv Neural Inf Process Syst.* 2019;32.

Islam MK, Aridhi S, Smail-Tabbone M. A comparative study of similarity-based and GNN-based link prediction approaches. *arXiv [Preprint].* 2020.