

ORIGINAL RESEARCH

Open access

Revisiting Graph Attention Mechanisms for Crystal Property Prediction: Beyond Mean Aggregation

Jose Martinez^{1*}, Carmen Lopez¹

Abstract

Most crystal graph neural networks rely on mean aggregation to combine information from neighboring atoms when predicting material properties. This simple averaging of neighbor features works well for many bulk properties that depend on average coordination environments. Yet it carries fundamental limitations: it erases information about local variance, extreme values from individual neighbors, and complex interaction patterns that are essential in real crystals. Graph attention mechanisms, which learn to weight neighbors dynamically according to relevance, have emerged as a promising way to overcome these shortcomings. However, even standard attention often reduces to a form of weighted averaging and therefore inherits many of the same restrictions. This conceptual framework revisits graph attention specifically for crystal property prediction. It first clarifies precisely what mean aggregation captures and, more importantly, what it systematically misses in crystalline systems. It then introduces a structured set of attention variants that move beyond simple averaging while remaining fully compatible with the symmetries and periodicity of crystals. Five attention variants are defined—variance-aware attention, max-attention for extreme values, multi-head attention with mixed aggregations, directional attention, and set-based attention—each tailored to different physical demands of crystal properties. A core contribution is a set of six design principles that guide the selection and implementation of these mechanisms. The principles emphasize matching the aggregation strategy to the target property, preserving permutation invariance, accounting for directional bond information, balancing expressivity against computational cost, and systematically testing aggregation sensitivity. By shifting attention from default mean aggregation to property-aware, richer aggregation strategies, the framework aims to unlock higher predictive fidelity for diverse crystal properties ranging from formation energies in ordered compounds to defect sensitivities and anisotropic tensor properties in disordered or low-symmetry materials. The proposed conceptual lens provides a blueprint for next-generation crystal graph neural networks that better respect the rich local physics of atomic arrangements.

Keywords Graph neural networks, Crystal property prediction, Graph attention mechanisms, Mean aggregation limitations, Variance-aware attention, Directional attention

*Correspondence:

Jose Martinez

jose.martinez@gmail.com

¹ Department of Computational Materials Analytics, Faculty of Engineering, University of Salamanca, Salamanca, Spain

Introduction

Graph neural networks have established themselves as a central tool in computational materials science for predicting the properties of crystals directly from atomic structure [1–6]. In nearly all early and many current implementations, the core operation at each node is to aggregate messages from neighboring atoms by simply taking their mean or sum. This approach is computationally light, numerically stable, and sufficient for properties that depend primarily on average local environments, such as formation energies of ordered compounds or band gaps in semiconductors [1–3]. Studies applying

these models to large crystal databases have demonstrated practical success and scalability across thousands of structures [7, 8].

Yet as the community pushes toward more challenging predictions—defect formation energies, properties of high-entropy alloys, anisotropic transport coefficients, or responses in low-symmetry crystals—the limitations of mean aggregation become increasingly apparent [9–11]. Mean aggregation treats every neighbor identically and collapses all local diversity into a single scalar summary. It cannot distinguish a situation in which one neighbor is dramatically different from the rest, nor can it preserve information about the spread or directional character of the local environment. These omissions are not minor; they reflect fundamental mismatches between the aggregation operation and the physics of crystals.

Graph attention mechanisms, originally developed for general graph learning and later adapted to materials, offer a natural path forward [12, 13]. By computing learned weights that depend on both the central atom and each neighbor, attention allows the model to emphasize more relevant neighbors and suppress less informative ones. Several recent works have incorporated attention layers into crystal graphs and reported improved performance on selected tasks [14–18]. At the same time, closer inspection reveals that standard attention still normalizes weights to sum to one and therefore remains a weighted average. The expressivity gain is real but limited [11, 19, 20].

This paper revisits graph attention mechanisms through a materials-specific lens. Rather than treating attention as an automatic upgrade, we analyze exactly where and why mean aggregation falls short in crystalline systems [9, 10, 21]. We then develop a conceptual framework that extends attention beyond weighted averaging. The framework rests on five core attention variants plus a set-based approach, each designed to capture distinct aspects of local atomic environments that crystals demand. Accompanying design principles translate these variants into practical guidelines for architecture choice [11, 22, 23].

The motivation is both theoretical and applied. Theoretically, richer aggregation directly increases the expressive power of crystal graph neural networks without sacrificing permutation invariance or periodicity [11, 19]. Applied, it promises better generalization to disordered, defective, or anisotropic materials that dominate modern materials discovery pipelines. By grounding the discussion in recent literature on attention-adapted crystal models and on the expressive limits of aggregation functions [12–16, 18], the framework provides a clear roadmap for moving beyond the mean-aggregation default that has dominated the field. The ultimate goal is not merely incremental accuracy gains but a deeper alignment between network operations and the physical realities of crystal chemistry.

Figure 1 conceptualizes crystal graph attention as an expanded aggregation design space in which moving beyond mean-like averaging becomes necessary whenever the target property depends on variance, extremal neighbors, bond directionality, or higher-order coordination structure.

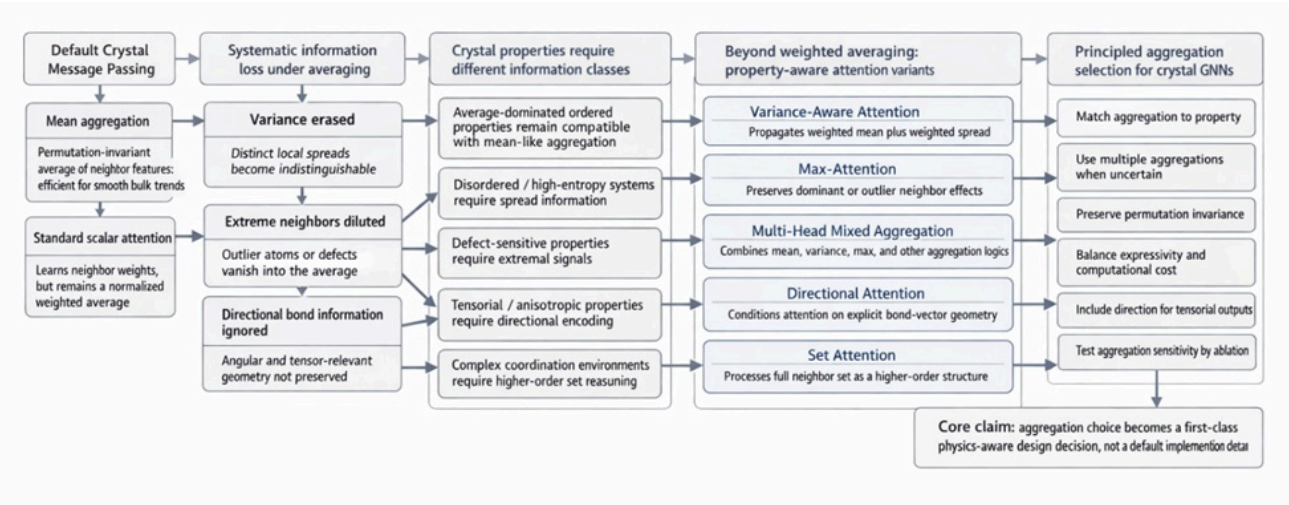


Figure 1. Hierarchical Design Space for Crystal Graph Attention beyond Mean Aggregation

What Mean Aggregation Does (And Doesn't Do)

Mean aggregation, in its simplest form, collects the feature vectors of all neighboring atoms around a central atom and replaces them with their arithmetic average [1-3, 21]. This operation is permutation-invariant, computationally cheap, and produces a single summary vector that reflects the average local chemical environment. For properties governed by collective averages—such as cohesive energy or average coordination trends—it captures the essential signal efficiently. Early crystal graph models relied heavily on this operation precisely because it balances simplicity with sufficient descriptive power for many bulk thermodynamic quantities [1-3].

Nevertheless, mean aggregation discards critical higher-order information. It cannot represent the variance among neighbors: two environments with identical averages can differ wildly in how spread out the neighbor features are. It also erases extreme values; a single outlier neighbor whose feature vector is far from the rest is diluted into the average and effectively disappears. Higher statistical moments such as skewness or the presence of rare but decisive local configurations are lost entirely. In addition, the operation assumes every neighbor contributes equally, ignoring any notion of relative importance or directional preference [11, 19].

These omissions become decisive in several classes of crystal problems. In defect detection, a single impurity or vacancy can dominate the local electronic structure or mechanical response; mean aggregation averages the impurity signal away, making the defect invisible to downstream layers. In high-entropy alloys or compositionally disordered systems, the diversity of local environments—not merely their average—determines configurational entropy and phase stability; mean aggregation collapses this diversity into a uniform summary. For anisotropic properties such as elastic constants or thermal conductivity in low-symmetry crystals, the directional arrangement of neighbors carries essential information that mean aggregation cannot encode because it is rotationally insensitive at the aggregation step [9, 10].

The hidden assumption underlying mean aggregation is that all neighbors surrounding a given atom are equally informative and that their collective effect is well represented by a linear average. For many perfect crystals with high symmetry this assumption holds reasonably well. For the vast majority of technologically relevant materials—those containing defects, interfaces, disorder, or directional bonding—it does not. Recent analyses of aggregation functions in graph neural networks have formalized these limitations, showing that networks relying exclusively on mean or sum operations are provably unable to distinguish certain non-isomorphic local neighborhoods that differ in variance or extremal structure [11, 19, 21, 22].

In short, mean aggregation excels at capturing smooth, average trends but systematically fails to preserve the fine-grained, local heterogeneity that governs many crystal properties. Recognizing these blind spots is the necessary first step toward designing aggregation strategies that respect the full richness of atomic environments in crystals.

Attention Mechanisms for Crystals

Graph attention introduces a learnable weighting step before aggregation [12, 13, 20]. For each central atom and each of its neighbors, a scalar attention coefficient is computed from the features of both atoms. These coefficients are normalized across the neighbor set so they sum to one, and the weighted combination of neighbor features then replaces the simple average. The result is an adaptive summary that emphasizes neighbors deemed more relevant by the learned scoring function. This mechanism was first popularized in general graph learning and has since been integrated into several crystal-specific architectures [14-18].

What attention adds is the capacity for context-dependent weighting. Instead of treating every neighbor identically, the model can, for example, up-weight neighbors that share similar electronic character with the central atom or down-weight

those that are geometrically distant in feature space. This adaptive behavior allows the network to focus on chemically or structurally decisive interactions. Several studies applying attention layers to crystal graphs have shown that these learned weights often align with intuitive chemical concepts such as bond strength or coordination type [14, 15, 24, 25].

Standard attention, however, still operates within the constraints of a weighted average. Because the coefficients are forced to sum to one, the final aggregated vector remains a convex combination of neighbor features. Consequently, variance and extreme values are still diluted; the mechanism cannot propagate the raw maximum or the spread of neighbor features unless additional modifications are introduced. Moreover, the attention coefficient is typically a scalar that ignores the three-dimensional direction of each bond. In crystals, where bonds possess well-defined vectors and where tensorial properties depend on those directions, scalar attention discards geometrically critical information [9, 10, 17].

Crystal structures also introduce unique complications. Periodic boundary conditions mean that some neighbors may be periodic images rather than distinct atoms, requiring careful handling of symmetry. Coordination shells often contain multiple bond types—nearest neighbors, next-nearest neighbors, or angular relationships—that standard scalar attention does not distinguish [17, 24, 26]. These factors suggest that attention mechanisms must be customized for the periodic, directional, and multi-scale nature of crystals rather than imported directly from generic graph learning.

Recent crystal models that incorporate attention have begun to explore these adaptations, yet most still rely on the core weighted-average template [14–18, 24, 25, 27, 28]. The present framework argues that the real opportunity lies in extending attention beyond this template. By re-thinking the aggregation step itself—incorporating variance, extrema, multiple heads, directional encoding, or set-level operations—attention can become a truly expressive tool tailored to the physics of crystals [11, 22, 23].

Beyond Mean Aggregation: Five Attention Variants

To overcome the limitations of both mean aggregation and standard weighted-average attention, this framework introduces a set of attention mechanisms tailored to crystalline systems, each modifying the aggregation step to preserve richer statistical and geometric information [11, 20, 22, 23].

Variance-aware attention extends the conventional weighted mean by jointly estimating a weighted variance of neighboring features, yielding a representation that couples central tendency with local dispersion. In disordered crystals and high-entropy alloys, this additional spread becomes physically meaningful, as it encodes configurational diversity governing entropy and phase stability, and can be propagated as an auxiliary signal for downstream inference on local heterogeneity. In a related direction, max-attention departs from averaging altogether by enabling a soft selection mechanism in which attention weights emphasize the most extreme neighbor feature, thereby preserving dominant outliers that are often decisive in defect-driven properties where single atomic environments control the macroscopic response.

Beyond these formulations, multi-head attention introduces parallel aggregation pathways that apply distinct operators such as mean, variance, max, or directional encoding, allowing the model to adaptively reconcile heterogeneous physical dependencies when the governing mechanisms are not known a priori. Directional attention further conditions the weighting process on explicit three-dimensional bond vectors, for instance through spherical-harmonic representations, ensuring that anisotropic interactions relevant to tensorial properties like elastic response or phonon transport are explicitly encoded. Set attention, in contrast, treats each coordination environment as an unordered collection and applies permutation-invariant set operations inspired by deep-sets and set-transformer paradigms, capturing higher-order collective effects that arise within finite coordination shells in crystalline structures [22, 23]. This progression moves beyond the weighted-average constraint that has limited earlier graph-based formulations [13–16, 18], offering a broader and more physically expressive aggregation space for crystal learning systems.

Table 1 distinguishes aggregation mechanisms by the specific classes of crystal-relevant information they preserve, revealing that the real design issue is not attention versus no attention, but which local statistics or geometric relations survive aggregation.

Table 1. Information Classes Preserved or Lost Across Crystal Aggregation Mechanisms

Aggregation mechanism	Central operation	Information preserved well	Information systematically lost or weakened	Crystal-property regimes where it is most appropriate	Main conceptual limitation
Mean aggregation	Arithmetic average over neighbor features	Average coordination trends; smooth collective environments; stable bulk signals	Variance among neighbors; outlier neighbors; directional dependence; higher-order coordination structure	Ordered compounds, average-dominated scalar properties, broad thermodynamic trends	Collapses heterogeneous local environments into one central tendency
Standard scalar attention	Learned neighbor weighting followed by normalized weighted sum	Relative neighbor relevance; context-dependent emphasis; chemically informative weighting patterns	Raw extrema remain diluted; spread is not explicitly represented; bond direction typically absent; higher-order set effects remain compressed	Tasks where some neighbors matter more than others, but averaging still approximates the physics	More selective than mean, but still remains an averaging-based operator
Variance-aware attention	Weighted mean plus weighted variance	Central tendency and local spread; heterogeneity of coordination shells; configurational diversity	Exact outlier identity may still be softened; geometry still limited unless explicitly encoded	High-entropy alloys, disordered crystals, entropy-sensitive stability problems	Adds spread information, but not full directional or set-level structure
Max-attention	Attention-guided preservation of dominant or extremal neighbor features	Rare decisive local events; defect-like signals; impurity-dominated responses	Global spread and full neighborhood balance may be underrepresented	Defect formation energies, impurity-sensitive band changes, localized failure modes	Can over-prioritize extremes when properties depend on distributed neighborhood effects
Multi-head mixed aggregation	Parallel heads with different aggregation logics combined	Simultaneous capture of averages, spread, extremes, and other complementary signals	Efficiency and interpretability may decline as heads proliferate	Heterogeneous benchmarks; uncertain structure–property mappings; broad screening tasks	Gains flexibility at the cost of greater architectural and computational complexity
Directional attention	Attention weights	Angular dependence;	Non-directional global set statistics may still	Elastic constants, phonons, transport	Requires careful symmetry-aware

	conditioned on bond-vector geometry	anisotropy; tensor-relevant neighbor structure	need additional modules	tensors, low-symmetry crystals	geometric encoding
Set attention	Permutation-invariant processing of the whole neighbor set	Higher-order interactions; collective coordination patterns; complex neighbor-set structure	Computational simplicity; scalability for larger or denser neighborhoods	Moderate-coordination shells where collective configuration governs response	Highest expressivity often comes with the largest computational burden

Design Principles for Crystal Attention

Effective use of attention variants requires deliberate design choices grounded in crystal chemistry, where a set of six interlocking principles provides the methodological guidance [9-11, 21, 23]. Within this framework, the central idea of aligning aggregation strategy with the target property becomes particularly consequential, since distinct crystal properties encode sensitivity to different aspects of the local environment. Formation energies in ordered compounds are typically governed by averaged coordination statistics and are therefore well captured through mean aggregation [1-3, 7], whereas defect formation energies or impurity-induced bandgap shifts depend on rare but influential local configurations and thus benefit from max-attention. In contrast, configurational entropy and the stability of disordered systems are more faithfully represented when variance-aware attention is introduced to preserve distributional structure. This perspective, in practice, requires identifying the dominant statistical or geometric character of the target property before any aggregation mechanism is selected. **Table 2** operationalizes this view by translating it into a property-aware architecture selection matrix, reinforcing that aggregation design should emerge from the underlying physical character of the prediction task rather than being inherited as a default modeling convention.

Table 2. Property-to-Aggregation Design Matrix for Physics-Aware Crystal Graph Architecture Selection

Target property or prediction context	Dominant physical signal that must survive aggregation	Recommended primary aggregation strategy	Recommended secondary strategy when uncertainty remains	Why mean aggregation alone is insufficient	Expected empirical implication
Formation energy in ordered compounds	Average local coordination and bulk compositional regularity	Mean aggregation	Multi-head mixed aggregation for robustness testing	Mean often already captures the main signal; richer schemes may add little	Performance gap between mean and richer variants should be small
Defect formation energy	Single impurity, vacancy, or localized abnormal neighbor effect	Max-attention	Multi-head mixed aggregation with max branch	Averaging dilutes the decisive outlier neighbor	Max-sensitive models should outperform mean in dilute defect settings

High-entropy alloy stability or entropy-sensitive behavior	Local environmental spread and configurational diversity	Variance-aware attention	Multi-head mixed aggregation with variance branch	Mean cannot represent neighborhood heterogeneity	Variance-preserving models should improve predictions in disordered systems
Elastic constants, piezoelectricity, anisotropic transport, phonon-related tensor properties	Bond direction, angular dependence, low-symmetry geometry	Directional attention	Directional plus multi-head mixed aggregation	Scalar averaging discards tensor-relevant geometric information	Direction-aware models should outperform isotropic aggregators on tensor targets
Broad benchmark with mixed property families and uncertain mechanisms	Unknown combination of averages, extremes, spread, and direction	Multi-head mixed aggregation	Set attention on moderate coordination shells	A single fixed aggregator risks underfitting one or more physical regimes	Mixed-head models should deliver the most robust cross-task performance
Coordination-shell problems where collective neighborhood configuration matters	Higher-order structure of the full unordered neighbor set	Set attention	Set attention plus directional encoding when geometry matters	Mean and scalar attention compress the shell too aggressively	Set-level models should help where pairwise summaries miss cooperative effects
High-throughput screening under strict computational budgets	Need for acceptable fidelity with scalable inference	Mean aggregation or lightweight variance-aware attention, depending on property	Small multi-head design only if justified by ablation	Richer operators may cost more than their benefit on simple tasks	Cost-sensitive benchmarks should reveal when added expressivity is worth deploying
New property with unclear mechanism	Mechanism not known in advance; safe exploration needed	Multi-head mixed aggregation	Systematic ablation back to mean, max, variance, and directional forms	Choosing mean by default bakes in an untested physical assumption	Aggregation-sensitivity analysis becomes necessary model-selection evidence

Where the mapping from structure to property is not fully resolved, it becomes necessary to move beyond a single aggregation choice and instead rely on multiple aggregation pathways that can be jointly optimized during training. In such settings, multi-head architectures combining heterogeneous aggregation types provide a mechanism for the model to autonomously infer the most appropriate compositional strategy, a capability that proves particularly valuable in large-scale materials screening where governing mechanisms differ across chemical subspaces.

A further constraint arises from the requirement to preserve permutation invariance, which is fundamental to the representation of crystal graphs defined over unordered neighbor sets. All considered aggregation variants maintain this

property either through symmetric statistical operators such as mean, variance, and max, or through explicit set-based processing formulations [19, 20, 22], thereby preventing any artificial dependence on arbitrary ordering of neighbors that would otherwise compromise physical consistency. Alongside this invariance requirement, computational considerations impose a practical boundary on architectural complexity: while mean aggregation remains linear in coordination number, variance and max-based mechanisms introduce only moderate overhead, whereas directional encoding and set-attention mechanisms exhibit quadratic scaling and are therefore more appropriately applied in regimes with moderate coordination or when clear accuracy gains justify the additional cost. This trade-off becomes especially salient in high-throughput screening contexts, where efficiency constraints are as critical as predictive fidelity.

When tensorial properties are involved, scalar aggregation alone becomes insufficient, since quantities such as elastic moduli, piezoelectric responses, and thermal conductivity inherently encode directional dependence; in these cases, directional attention explicitly integrates bond vectors to retain geometric information that would otherwise be lost [9, 10, 17, 24]. Even for scalar targets such as total energy, the inclusion of directional information can remain beneficial without introducing adverse effects. Complementing these architectural considerations, robustness demands that aggregation choices be subjected to systematic sensitivity analysis prior to deployment, wherein each variant is selectively removed or replaced by mean aggregation to evaluate its contribution to predictive performance. The resulting degradation patterns provide direct insight into the statistical or geometric features that dominate the learning process, establishing a form of empirical interpretability that should be routinely incorporated into crystal graph neural network evaluation. Through this lens, attention mechanisms are no longer generic architectural components but become tightly coupled to the underlying physics of the material system, ensuring that aggregation design is dictated by structural and property-level realism rather than inherited modeling convention.

Relation to Existing Work

Several recent studies have incorporated graph attention mechanisms into crystal graph neural networks [14–16, 18]. Several recent studies have incorporated graph attention mechanisms into crystal graph neural networks [14–18]. Yet nearly all of them retain the core weighted-average template inherited from standard graph attention networks. For instance, Banik *et al.* [14] developed the crystal edge graph attention neural network and demonstrated gains on multiscale classification tasks, while Zhang *et al.* [15] introduced an adaptive scalar-moment crystal graph attention network that improved material property prediction through learned neighbor weighting. Cheng *et al.* [18] and Jia *et al.* [16] similarly applied attention layers to oxidation potentials and topology-guided crystal graphs, respectively, showing that attention can help the model focus on chemically relevant neighbors compared with plain mean aggregation. These contributions confirm that attention provides a measurable benefit over uniform averaging in selected crystal tasks [12, 24, 25, 27, 28].

Nevertheless, in every case the final aggregation step remains a normalized weighted sum. Consequently, the same fundamental limitations discussed earlier—dilution of variance, loss of extreme values, and absence of directional sensitivity—persist [9, 10, 21]. This framework builds directly on those efforts but argues for moving beyond the weighted-mean paradigm. Where earlier crystal attention models treat attention primarily as a soft selection mechanism within averaging, the variants proposed here explicitly augment or replace the averaging operation itself.

The relation to foundational graph attention work by Veličković *et al.* [13] is therefore one of principled extension: the original GAT formulation supplies the scoring function, yet the aggregation logic is generalized to variance-aware, max, multi-head, directional, and set-based forms. Connections to set-based approaches are equally direct. Kortvelesy *et al.* [22] explored generalised f-mean aggregation in graph neural networks and showed that richer symmetric functions over neighbor sets increase expressivity without violating permutation invariance. Set attention (Variant 5) aligns closely with these ideas and with set-transformer concepts, offering a practical way to capture higher-order statistics over the moderate coordination shells typical of crystals.

On the theoretical side, the framework resonates with formal analyses of aggregation expressivity [11, 19–21]. Joshi et al. [11] and Xu et al. [19] established clear bounds on what mean- or sum-based message passing can and cannot distinguish; the five variants address those gaps by preserving additional moments and geometric features. Recent crystal-specific examinations [9, 10] further underscore that standard aggregation choices limit performance on diverse property classes. Equivariant attention mechanisms explored in some crystal models implicitly incorporate directional information and can be viewed as a special case of the directional attention variant introduced here.

Collectively, the existing literature validates the need for attention in crystal property prediction while leaving open the richer design space of non-average aggregation. This conceptual framework synthesizes those insights [11, 13–16, 18, 19, 22] into a coherent set of variants and principles, positioning attention not as an incremental upgrade but as a modular toolkit matched to the physical demands of crystalline systems.

Empirical Predictions

The proposed framework induces a set of empirically testable expectations regarding the relative behavior of the introduced attention variants compared with conventional mean aggregation across distinct crystal property classes, with each expectation emerging from the intrinsic informational constraints of mean-based representations and the complementary strengths of the alternative mechanisms [9–11, 19]. In the case of defect formation energies, the presence of localized perturbations such as vacancies or impurities suggests a systematic advantage for max-attention, since retaining extreme feature responses avoids the averaging-out effect that can obscure energetically dominant local environments, a behavior that becomes particularly pronounced in dilute defect regimes. A different regime emerges for entropy-driven phenomena in high-entropy alloys and disordered solid solutions, where variance-aware attention is expected to provide superior fidelity by explicitly preserving the spread of local environments that governs configurational entropy and phase stability, in contrast to mean aggregation which suppresses precisely this distributional information. Tensorial response functions, including elastic constants and thermal conductivity in low-symmetry systems, further shift the requirement toward directional attention, as conditioning on bond vectors introduces angular sensitivity that scalar aggregation cannot capture, thereby aligning the representation with anisotropic physical laws governing such properties. In broader chemical spaces where the governing structure–property relationships are heterogeneous or not a priori identifiable, multi-head configurations that integrate mean, variance, max, and directional components are expected to exhibit enhanced robustness by enabling adaptive selection of the most informative aggregation pathways during training, thereby mitigating the risk of representational under-specification. Against these enriched mechanisms, a residual baseline emerges for scalar properties primarily governed by averaged coordination environments, such as formation energies in ordered intermetallics, where mean aggregation remains sufficient and more complex variants may yield marginal gains at best, reinforcing the principle that representational complexity must be matched to property-specific structure [1–3]. Rather than speculative assertions, these outcomes follow directly from the fundamental mismatch between the information compression inherent in mean aggregation and the distinct statistical or geometric requirements of each property class, suggesting that systematic ablation studies replacing mean aggregation with the proposed variants should reveal consistent performance hierarchies and thereby empirically substantiate the theoretical framework.

Implications for GNN Architecture Design

The conceptual shift away from default mean aggregation carries concrete implications for three stakeholder groups: model developers, practitioners, and benchmark designers [9–11].

For model developers, the primary recommendation is to abandon the reflexive choice of mean aggregation. Instead, architectures should include a modular aggregation layer that can instantiate any of the five variants (or combinations thereof) with minimal code overhead. This modularity allows rapid experimentation and makes the aggregation strategy itself a hyperparameter. Developers should also embed aggregation-sensitivity reporting into standard model cards, documenting how performance changes when each variant is ablated back to mean aggregation [13, 19, 29].

Practitioners tasked with predicting a new crystal property now have explicit guidance. The first step is to classify the property according to its dominant statistical or geometric character: average, extreme, diverse, or directional. For disordered systems, variance-aware attention becomes the default; for defect-sensitive tasks, max-attention is preferred. When the mapping is uncertain, multi-head mixed aggregation offers a safe starting point. Practitioners should also consider coordination number: high-degree nodes may favor lightweight variants to preserve training efficiency.

Benchmark designers can accelerate progress by curating tasks that deliberately stress different aggregation needs. Current crystal benchmarks often emphasize average-coordination properties where mean aggregation already performs well [1-3]. Future suites should include dedicated subsets for defect formation, high-entropy alloy entropy, low-symmetry tensor properties, and extreme-value scenarios. Performance tables should report not only overall metrics but also per-aggregation ablation results, enabling direct comparison of expressive power. Such benchmarks would make the hidden assumptions of mean aggregation visible and reward architectures that implement the richer variants proposed here.

Taken together, these implications transform aggregation choice from an afterthought into a first-class design decision. Crystal graph neural network research can thereby evolve from “attention as a drop-in replacement” to “property-aware aggregation engineering,” ultimately producing models whose internal operations more closely mirror the physics they aim to predict.

Open Questions

Despite the conceptual clarity of the proposed attention variants and their governing principles, a number of deeper methodological and theoretical issues remain unresolved within the broader research landscape [11, 19, 22]. A central concern is the identification of aggregation functions that achieve an optimal balance between expressive power and computational tractability under realistic coordination regimes in crystal graphs. Although set attention is, in principle, capable of representing any permutation-invariant statistic, its quadratic scaling introduces significant limitations in large unit cells and highly coordinated or high-pressure phases, making the search for more efficient yet expressive alternatives an open problem. Closely related is the question of whether aggregation itself can be inferred directly from data, where meta-learning or reinforcement learning strategies could enable adaptive selection—or even generation—of aggregation schemes tailored to specific properties or chemical families, thereby reducing dependence on manually engineered choices and moving toward more autonomous materials graph learning systems.

Further complexity arises when considering the joint use of directional and variance-sensitive mechanisms, since while their conceptual integration is straightforward, naive parameter-level fusion risks redundancy and inefficiency, motivating the need for compact representations that preserve symmetry while remaining computationally viable at scale. This challenge becomes even more pronounced in high-coordination environments such as complex bulk crystals or surface reconstructions, where large neighbor sets may dilute the discriminative value of individual contributions, potentially reducing the marginal benefit of sophisticated attention relative to simpler mean-based aggregation and necessitating systematic analysis of coordination-dependent performance regimes [29]. In parallel, questions of generalization remain unsettled, particularly whether enriched attention primarily enhances interpolation within known compositional domains or whether it can meaningfully improve extrapolation to unseen structures, given that the inherent simplicity of mean aggregation may sometimes confer greater robustness under distribution shift.

Addressing these intertwined issues will require a combination of theoretical formalization, architectural innovation, and rigorously controlled benchmarking protocols, as the current framework is best understood not as a finalized solution but as a structured design space in which attention mechanisms for crystal property prediction continue to evolve.

Conclusion

Mean aggregation has served the crystal graph neural network community well for properties governed by average local environments, yet it systematically discards variance, extreme values, directional cues, and higher-order interactions that are essential across much of materials space. Graph attention mechanisms offer a pathway beyond this limitation, but only when the aggregation step itself is re-designed rather than merely re-weighted.

This conceptual framework revisits attention for crystal property prediction through five targeted variants: variance-aware attention that preserves local diversity, max-attention that propagates extreme features, multi-head attention that mixes aggregation types, directional attention that respects bond vectors, and set attention that captures arbitrary set statistics. Accompanying design principles—match aggregation to property, use multiple aggregations when uncertain, preserve permutation invariance, consider computational cost, respect direction for tensorial outputs, and test sensitivity—translate these variants into actionable architecture choices.

By adopting property-aware aggregation strategies, next-generation crystal graph models can align their internal computations more closely with the physics of atomic environments. The result will be higher fidelity across diverse tasks, from defect-sensitive energies to anisotropic tensor properties and entropy-driven phase stability. The community is therefore encouraged to move beyond the mean-aggregation default and to treat aggregation design as a central, principled element of crystal graph neural network development. This shift promises not only incremental accuracy gains but a deeper conceptual alignment between network operations and the rich local structure of real crystals.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 03 Feb 2025

Revised: 18 Apr 2025

Accepted: 07 Jun 2025

Published online: 18 July 2025

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

- Schmidt J, Pettersson L, Verdozzi C, Botti S, Marques MA. Crystal graph attention networks for the prediction of stable materials. *Sci Adv.* 2021;7(49):eabi7948.
<https://doi.org/10.1126/sciadv.abi7948>.
- 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>.
- Omeel SS, Louis SY, Fu N, Wei L, Dey S, Dong R, et al. Scalable deeper graph neural networks for high-performance materials property prediction. *Patterns (N Y).* 2022;3(5):100491.
<https://doi.org/10.1016/j.patter.2022.100491>.
- Zhou Y, Li Q, Zhou W, Zang H, Xu L, Ren Y, et al. Reinforce crystal material property prediction with comprehensive message passing via deep graph networks. *Comput Mater Sci.* 2024;239:112958.
<https://doi.org/10.1016/j.commatsci.2024.112958>.
- Feng H, Tian H. Improving crystal property prediction from a multiplex graph perspective. *J Chem Inf Model.* 2024;64(19):7376-85.
<https://doi.org/10.1021/acs.jcim.4c01200>.
- Gurney K. An introduction to neural networks. London: CRC Press; 2018.
- Teng P, Fu C, Shen S, Zang H, Fang J, Ren Y, et al. MatGNet: A graph neural network for crystal property prediction as an alternative to first-principles calculations. *Mater Today Commun.* 2025;44:112021.
<https://doi.org/10.1016/j.mtcomm.2025.112021>.
- Das K, Goyal P, Lee SC, Bhattacharjee S, Ganguly N. CrysMMNet: Multimodal representation for crystal property prediction. In: *Proceedings of the 39th Conference on Uncertainty in Artificial Intelligence*; 2023 Jul 31-Aug 4; Pittsburgh, PA. PMLR. 2023;216:507-17.
- Gong S, Yan K, Xie T, Shao-Horn Y, Gomez-Bombarelli R, Ji S, et al. Examining graph neural networks for crystal structures: Limitations and opportunities for capturing periodicity. *Sci Adv.* 2023;9(45):eadi3245.
<https://doi.org/10.1126/sciadv.adi3245>.
- Du H, Wang J, Hui J, Zhang L, Wang H. DenseGNN: Universal and scalable deeper graph neural networks for high-performance property prediction in crystals and molecules. *NPJ Comput Mater.* 2024;10(1):292.
<https://doi.org/10.1038/s41524-024-01444-x>.
- Joshi CK, Bodnar C, Mathis SV, Cohen T, Lio P. On the expressive power of geometric graph neural networks. In: *Proceedings of the 40th International Conference on Machine Learning*; 2023 Jul 23-29; Honolulu, HI. PMLR. 2023;202:15330-55.
- Cong G, Ben-Naim T, Fung V, Gupta A, Neumann R, Steiner M. Extensive attention mechanisms in graph neural networks for materials discovery. In: *2022 IEEE International Conference on Data Mining Workshops*; 2022 Nov 28-Dec 1; Orlando, FL. IEEE; 2022. p. 658-65.
<https://doi.org/10.1109/ICDMW58026.2022.00090>.
- Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. In: *International Conference on Learning Representations*; 2018 Apr 30-May 3; Vancouver, Canada.
- Banik S, Dhabal D, Chan H, Manna S, Cherukara M, Molinero V, et al. CEGANN: Crystal Edge Graph Attention Neural Network for multiscale classification of materials environment. *NPJ Comput Mater.* 2023;9(1):23.
<https://doi.org/10.1038/s41524-023-00975-z>.

Zhang W, Tang L, Xu M. Adaptive scalar-moment crystal graph attention neural network for material property prediction. *IEEE Trans Emerg Top Comput Intell.* 2025;9(4):2728-42.
<https://doi.org/10.1109/TETCI.2024.3436869>.

Jia B, Li Q, Zhou W, Niu Z, Zang H, Xu J, et al. Local and global attention mechanisms synergy for material property prediction. *Mol Phys.* 2025;123(17):e2458645.
<https://doi.org/10.1080/00268976.2025.2458645>.

Lin X, Jiang H, Wang L, Ren Y, Ma W, Zhan S. 3D-structure-attention graph neural network for crystals and materials. *Mol Phys.* 2022;120(11):e2077258.
<https://doi.org/10.1080/00268976.2022.2077258>.

Cheng D, Sha W, Han Q, Tang S, Zhong J, Du J, et al. ACGNet: An interpretable attention crystal graph neural network for accurate oxidation potential prediction. *Electrochim Acta.* 2024;473:143459.
<https://doi.org/10.1016/j.electacta.2023.143459>.

Xu K, Hu W, Leskovec J, Jegelka S. How powerful are graph neural networks? *arXiv [Preprint]*. 2018.
<https://doi.org/10.48550/arXiv.1810.00826>.

Huang J, Zhu X. Adaptive node-level weighted learning for directed graph neural network. *Neural Netw.* 2025;187:107393.
<https://doi.org/10.1016/j.neunet.2025.107393>.

Morris M, Horrocks I. Sound logical explanations for mean aggregation graph neural networks. *arXiv [Preprint]*. 2025.
<https://doi.org/10.48550/arXiv.2511.11593>.

Kortvelesy R, Morad S, Prorok A. Generalised f-mean aggregation for graph neural networks. *Adv Neural Inf Process Syst.* 2023;36:34439-50.
<https://doi.org/10.52202/075280-1494>.

Keren Taraday M, David A, Baskin C. Sequential signal mixing aggregation for message passing graph neural networks. *Adv Neural Inf Process Syst.* 2024;37:93985-4021.
<https://doi.org/10.52202/079017-2981>.

Wang T, Chen J, Teng J, Shi J, Zeng X, Snoussi H. Crysformer: An attention-based graph neural network for properties prediction of crystals. *Chin Phys B.* 2023;32(9):090703.
<https://doi.org/10.1088/1674-1056/ace247>.

Xu Y, Deng W, Zheng J. Study of crystal property prediction based on dual attention mechanism and transfer learning. *AIP Adv.* 2024;14(11):115001.
<https://doi.org/10.1063/5.0232308>.

Cai L, Li J, Wang J, Ji S. Line graph neural networks for link prediction. *IEEE Trans Pattern Anal Mach Intell.* 2022;44(9):5103-13.
<https://doi.org/10.1109/TPAMI.2021.3080635>.

Liu J, Tang Y, Tretiak S, Duan W, Zhou L. SA-GAT-SR: Self-adaptable graph attention networks with symbolic regression for high-fidelity material property prediction. *NPJ Comput Mater.* 2025;11(1):377.
<https://doi.org/10.1038/s41524-025-01854-5>.

Kang J, Yang L, Zeng J, Sun Z, Lin H, Li J. Graph isomorphism attention network combined with pre-trained language models: A novel approach for crystal material property prediction. *Neural Comput Appl.* 2025;37(30):25061-76.
<https://doi.org/10.1007/s00521-025-11532-8>.

Chen M, Wei Z, Huang Z, Ding B, Li Y. Simple and deep graph convolutional networks. In: *Proceedings of the 37th International Conference on Machine Learning*; 2020 Jul 13-18; Virtual. PMLR. 2020;119:1725-35.