

ORIGINAL RESEARCH

Open access

# When Graph Convolutions Over-Smooth on Large Unit Cells: Failure Mode Analysis of Deep GNNs for Thermal Transport

Fatima Zahra Amrani<sup>1\*</sup>, Youssef Benali<sup>1</sup>

## Abstract

Deep graph neural networks (GNNs) have become a cornerstone of computational materials engineering because they can learn representations directly from crystal graphs. Yet a fundamental failure mode remains hidden when these models are applied to thermal transport: over-smoothing. After only a modest number of message-passing layers, node features converge to a single global vector, erasing the atom-resolved local environments that govern phonon scattering. The problem is dramatically amplified in large unit cells—structures containing 50 or more atoms, common in skutterudites, clathrates, and complex perovskites—because the graph diameter forces the model designer to choose between insufficient propagation depth (missing long-range interactions) and excessive depth (catastrophic smoothing). Lattice thermal conductivity  $\kappa$  emerges from local force constants, anharmonicity, and defect-sensitive scattering; over-smoothing replaces these variations with a featureless average, producing systematically erroneous predictions. This failure-mode analysis demonstrates that the central conflict is structural: thermal transport demands precisely the short- to medium-range distinctions that deep GNNs destroy. Four distinct failure modes are identified and mechanistically linked to the over-smoothing process. Detection principles based on feature similarity, depth sensitivity, and defect response are proposed. Mitigation strategies grounded in architectural modifications—residual connections, normalization, multi-scale pathways, and attention—are shown to preserve local information without sacrificing global coherence. The analysis reveals why standard crystal GNN architectures succeed for formation energies or band gaps yet fail systematically for thermal transport in realistic materials. By reframing over-smoothing as the dominant limitation rather than a minor training artifact, this work provides a conceptual roadmap for next-generation GNNs tailored to phonon physics.

**Keywords** Failure mode analysis, Over-smoothing, Deep GNNs, Large unit cells, Thermal transport, Lattice thermal conductivity

\*Correspondence:

Fatima Zahra Amrani  
fatima.amrani@gmail.com

<sup>1</sup> Department of Computational Materials Science, Faculty of Sciences and Engineering, Mohammed V University, Rabat, Morocco

## Introduction

Graph neural networks have rapidly emerged as the dominant machine-learning framework for crystal property prediction [1, 2]. By representing atoms as nodes and bonds as edges within a periodic graph, these models learn material properties directly from atomic connectivity [3]. Early successes demonstrated that even shallow crystal graph convolutional networks could achieve remarkable accuracy for formation energies, elastic moduli, and electronic band gaps [4, 5]. Subsequent architectures incorporating equivariant features and higher-order interactions further improved data efficiency and physical consistency [6, 7].

Yet the very mechanism that enables these successes—repeated message passing—harbours a critical failure mode when applied to thermal transport in complex crystals. As the number of layers increases, node embeddings converge toward a single global representation, a phenomenon known as over-smoothing [8, 9]. For small, high-symmetry unit cells the effect is often tolerable because local environments are redundant and global averages suffice. In large, low-symmetry unit cells—precisely the structures of greatest interest for thermoelectric and thermal-management applications—the graph diameter is large, forcing deeper networks. The result is systematic information loss.

Thermal transport, quantified by lattice thermal conductivity  $\kappa$ , is exquisitely sensitive to local atomic environments [10]. Phonon scattering rates depend on anharmonic force constants that vary from site to site within the unit cell [11]. Over-smoothing replaces these site-specific signatures with a uniform vector, destroying the very information required for accurate scattering predictions. The paradox is immediate: large unit cells require long-range information propagation and therefore many layers; many layers guarantee feature collapse; feature collapse renders  $\kappa$  predictions meaningless.

This failure-mode analysis dissects the interaction between network depth, unit-cell complexity, and phonon physics. It demonstrates that over-smoothing is not merely a training inconvenience but the fundamental reason deep GNNs underperform on thermal transport tasks. The central thesis is clear: deep graph convolutions over-smooth on large unit cells, and thermal transport is the property most vulnerable to this collapse. By mapping the mechanisms, typology, detection, and mitigation of this failure, the present work supplies a conceptual foundation for designing GNNs that respect the locality demands of phonon physics while retaining the representational power of graph learning [12–14].

## Over-Smoothing in Graph Neural Networks

Over-smoothing is the process by which node features in a graph neural network become increasingly similar with each additional message-passing layer until all nodes are represented by nearly identical vectors. The mathematical origin is simple: each graph convolution layer performs a weighted average of a node's own features with those of its immediate neighbors. After  $k$  layers a node's representation is a smoothed average over its entire  $k$ -hop neighborhood. In a connected graph, as  $k$  approaches the graph diameter, every node's receptive field covers the entire graph. The result is convergence to the dominant eigenvector of the (normalized) graph Laplacian—the lowest-frequency mode that carries only global information [8, 9].

Empirical studies confirm the rapidity of this collapse [15]. Even in moderately sized molecular graphs, feature cosine similarity across nodes exceeds 0.95 after 8–10 layers in standard GCN architectures [13, 16]. For crystal graphs the effect is exacerbated by periodicity: the graph is effectively infinite, yet the unit cell is treated as a finite but large graph. Message passing therefore propagates information both within and across periodic images, accelerating the averaging process.

The consequence for materials modelling is profound [17, 18]. Once node features are indistinguishable, the network can no longer distinguish between chemically or vibrationally distinct atomic sites. Any downstream prediction—whether energy, force, or property—must rely solely on the global graph-level readout. Node-level or edge-level distinctions required for properties that depend on local heterogeneity are lost [14].

Conceptually, the progression can be visualised as follows: at layer 1, each node retains a strong local signature coloured by its immediate coordination environment; by layer 5, colours begin to blend across neighboring atoms; by layer 10, the entire unit cell is rendered in a single uniform shade, indicating complete loss of local contrast. This diagram (imagined as a sequence of four heat-map panels overlaid on the same crystal structure) illustrates the irreversible transition from information-rich embeddings to a featureless global average.

Over-smoothing is therefore not an occasional pathology but the default asymptotic behaviour of repeated graph convolutions. Architectures that rely on depth to capture long-range interactions inevitably trade local fidelity for global

uniformity [8, 9, 13, 15]. In the context of materials science, this trade-off has been tolerated for properties that are largely global (e.g., total energy). For thermal transport, however, the trade-off is fatal.

## Why Large Unit Cells Amplify Over-Smoothing

Large unit cells introduce a structural mismatch between the information-propagation requirements of the material and the smoothing dynamics of the GNN. A unit cell with  $N$  atoms has a graph diameter that scales roughly with the cube root of  $N$  in three dimensions. For a typical skutterudite or clathrate containing 50–100 atoms, the diameter exceeds 10–15 hops. Capturing interactions across the cell therefore demands at least that many message-passing layers. Yet standard GNNs already exhibit severe over-smoothing after 8–12 layers [8, 9, 16]. The required depth for long-range propagation therefore lies deep inside the over-smoothing regime.

Low-symmetry cells (triclinic, monoclinic) exacerbate the problem further. Each atom occupies a unique local environment with no symmetry-equivalent copies. The loss of node-specific features is therefore total rather than partially masked by redundancy. In contrast, high-symmetry cells (cubic, hexagonal) retain some invariance even after moderate smoothing, delaying the observable failure.

The amplification is multiplicative. Larger cells  $\rightarrow$  larger diameter  $\rightarrow$  more layers required  $\rightarrow$  earlier onset of feature collapse. Once collapse occurs, additional layers contribute nothing but numerical instability and gradient vanishing. Shallow networks avoid smoothing but cannot propagate information across the cell; deep networks propagate but erase distinctions. No intermediate depth satisfies both constraints simultaneously [19, 20].

Empirical evidence from crystal GNN literature supports this analysis [21]. Models trained on small-unit-cell datasets (e.g., <20 atoms) achieve reasonable accuracy for global properties, yet the same architectures applied to supercells or complex crystals show precipitous performance drops once depth exceeds the smoothing threshold [4, 5]. The failure is not dataset-specific but architecture-intrinsic and size-dependent.

Thus large unit cells do not merely challenge GNNs—they expose the fundamental incompatibility between message-passing depth and local-information preservation. Thermal transport prediction inherits this incompatibility at full strength.

## Thermal Transport: Why It Demands Local Information

Lattice thermal conductivity  $\kappa$  quantifies heat flow carried by lattice vibrations (phonons). Within the simplified kinetic theory picture,

$$\kappa = \left(\frac{1}{3}\right) C_v v l^{(1)}$$

where  $C_v$  is the volumetric heat capacity,  $v$  is the phonon group velocity, and  $l$  is the phonon mean free path. Although the formula is approximate, it highlights the three essential ingredients.  $C_v$  is relatively insensitive to local detail, but both  $v$  and  $l$  are governed by local atomic environments.

Phonon group velocity  $v$  derives from the curvature of the dispersion relation, which in turn depends on local force constants (bond stiffnesses) [22]. In a large unit cell these stiffnesses vary significantly between framework atoms, rattler sites, and defect regions. The mean free path  $l$  is even more local: it is inversely proportional to the scattering rate, which arises from anharmonic three-phonon processes and defect scattering—both determined by atom-resolved anharmonicity and coordination [23].

Over-smoothing replaces these site-specific force constants and anharmonicities with a single average vector. The network therefore predicts identical velocities and scattering rates everywhere inside the unit cell. The resulting  $\kappa$  is an

unphysical global average that cannot reproduce measured anisotropy, temperature dependence, or defect sensitivity.

Consider a conceptual diagram of a large unit cell containing two distinct atomic sites: atom A in a stiff, light coordination environment (high  $v$ , long  $l$ ) and atom B in a soft, heavy environment (low  $v$ , short  $l$ ). Before smoothing, the GNN embeddings reflect these differences (visualised as contrasting colours). After 10 layers, both sites share the same colour and therefore the same predicted contributions to  $\kappa$ . The diagram makes the information loss visually immediate: local contrast vanishes, and the predicted conductivity becomes spatially uniform and quantitatively wrong.

Thermal transport is therefore not a global emergent property in the same way as cohesive energy. It is a local-to-global property whose accuracy hinges on preserving atom-resolved variations. Deep GNNs, by design, erase precisely those variations [4, 11, 24].

**Table 1** clarifies why thermal transport is not merely another crystal-property task for GNNs, but the property class most structurally vulnerable to over-smoothing because it depends on preserving local heterogeneity across a globally connected graph.

**Table 1.** Property-Specific Vulnerability of Crystal GNNs to Over-Smoothing

| Property class                            | Dominant information scale               | Reliance on atom-resolved heterogeneity | Tolerance to deep message passing | Expected effect of over-smoothing                                                                         | Relevance to manuscript argument                                             |
|-------------------------------------------|------------------------------------------|-----------------------------------------|-----------------------------------|-----------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| Formation energy                          | Global / cell-averaged                   | Low to moderate                         | Relatively high                   | Often modest degradation; global averages remain informative                                              | Explains why deep crystal GNNs can still perform well on standard benchmarks |
| Electronic band gap                       | Mixed local + global                     | Moderate                                | Moderate                          | Performance may degrade, but not always catastrophically                                                  | Shows that some properties partially tolerate smoothing                      |
| Elastic modulus                           | Global with local bonding contributions  | Moderate                                | Moderate                          | Loss of fine bonding contrast can reduce fidelity, but some global structure survives                     | Useful comparison case                                                       |
| Lattice thermal conductivity ( $\kappa$ ) | Local-to-global                          | Very high                               | Low                               | Severe failure once atom-level force constants, anharmonicity, and scattering signatures are averaged out | Core target property of the manuscript                                       |
| Defect-sensitive thermal response         | Highly local                             | Extremely high                          | Very low                          | Defects become masked by neighborhood averaging; response becomes artificially weak                       | Explains defect insensitivity failure mode                                   |
| Size-dependent phonon                     | Multi-scale with strong local dependence | Very high                               | Low                               | Global embedding collapse suppresses meaningful size scaling                                              | Explains supercell and large-unit-cell breakdown                             |

|                  |  |  |  |  |
|------------------|--|--|--|--|
| transport trends |  |  |  |  |
|------------------|--|--|--|--|

## Mechanisms of information loss

The degradation of predictive fidelity in graph-based models of thermal transport is not an incidental artifact but emerges from a set of tightly coupled mechanisms that progressively erode physically meaningful distinctions. At the core of this process lies the recursive averaging inherent to graph convolution, through which each node representation is continuously reconstituted as a convex mixture of its local neighborhood. With increasing depth, this iterative mixing drives the system toward the stationary distribution of a random walk, effectively collapsing the embedding space onto a global mean. Under these conditions, subtle but consequential variations in interatomic force constants and anharmonic interactions are not merely attenuated but suppressed at an exponential rate, rendering the representation increasingly insensitive to local perturbations that govern phonon dynamics [8, 9].

This homogenization acquires greater significance when considered alongside the spatial scales required for accurate thermal transport modeling. Phonon scattering processes depend critically on medium-range interactions spanning approximately 10–20 Å, a regime that necessitates relatively deep architectures to capture sufficient structural context. Yet, the same depth amplifies smoothing effects, creating a structural incompatibility within the model: the layers required to encode physically relevant interaction ranges simultaneously induce the very feature collapse that erases them. In practice, representations begin to lose discriminative capacity before the necessary spatial extent is reached, leaving the model unable to reconcile locality with range in a physically coherent manner [14, 20].

A further consequence emerges in the treatment of anharmonicity, which is encoded not in dominant structural features but in fine-scale deviations from harmonic baselines [23]. These deviations manifest as delicate variations in node embeddings, carrying the signatures of temperature-dependent scattering processes. When subjected to repeated averaging, such distinctions are systematically flattened, leaving behind representations that approximate harmonic behavior even in regimes where anharmonic effects are dominant. The resulting model, while structurally consistent, becomes fundamentally misaligned with the physics governing  $\kappa(T)$ , particularly under conditions where scattering processes dictate transport behavior [11].

This erosion of detail extends to the representation of phonon modes themselves [25]. Distinctions between acoustic and optical branches, as well as between transverse and longitudinal modes, are encoded in localized displacement patterns that require high-resolution feature spaces to remain separable. As smoothing progresses, these mode-specific signatures converge toward a shared representation, effectively obscuring their individual contributions. The model, deprived of this differentiation, loses the capacity to assign appropriate weights to distinct transport channels, leading to systematic distortions in mode-resolved predictions [24]. What emerges from this sequence is not merely a loss of precision but a structural insensitivity to the atomic-scale mechanisms that underpin thermal conductivity.

## A typology of failure modes

The theoretical erosion of information described above becomes empirically visible through a set of recurring failure patterns that reveal how representational collapse translates into predictive bias. One of the most immediate manifestations appears in the systematic underestimation of thermal conductivity. In heterogeneous or anisotropic materials, high-conductivity pathways—often associated with light-atom chains or directionally aligned bonding networks—are averaged together with more resistive regions. This blending suppresses the contribution of dominant transport channels, yielding predictions that consistently fall below both experimental measurements and first-principles calculations, particularly in layered or structurally complex systems [26].

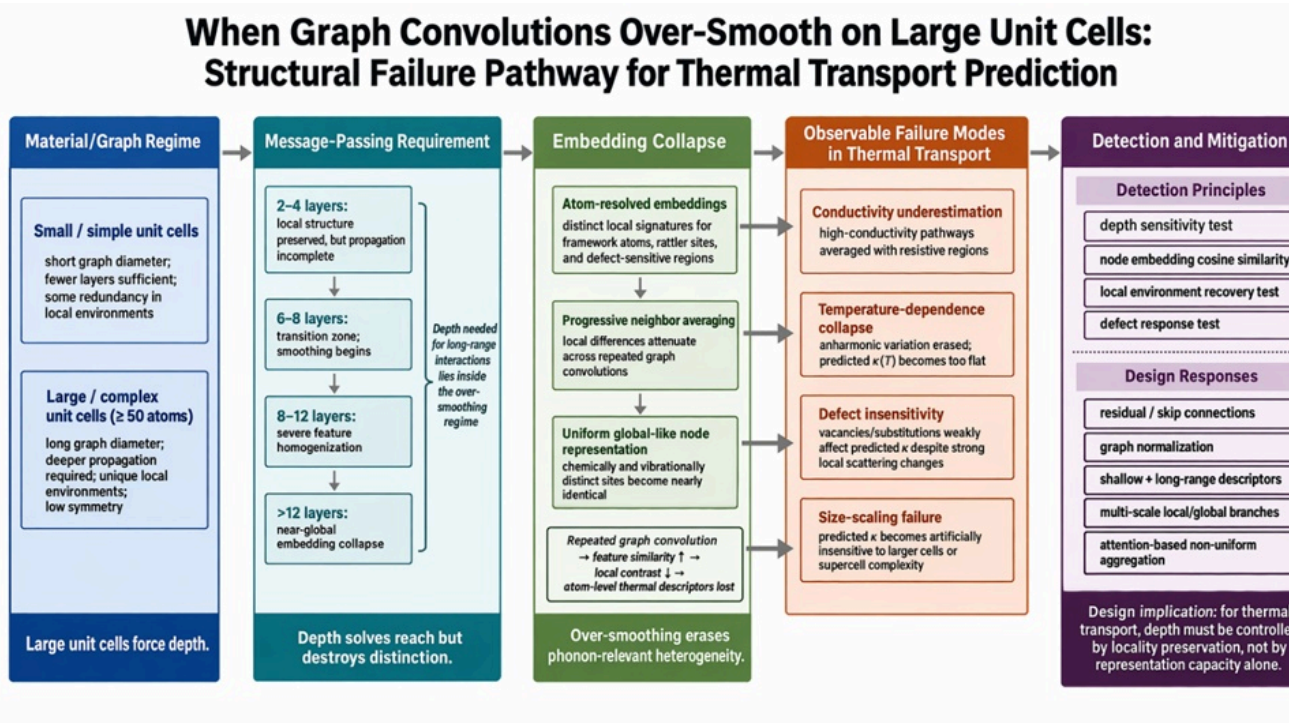
The implications extend further when temperature dependence is considered. In real materials, thermal conductivity typically decreases with increasing temperature due to enhanced phonon–phonon scattering. However, when anharmonic signatures are diminished within the learned representation, the model no longer captures the variability required to

reproduce this trend. Instead, it produces a flattened response in which  $\kappa$  appears weakly dependent on temperature, reflecting the loss of localized scattering information rather than any underlying physical invariance [11]. This divergence signals a deeper misalignment between learned features and thermodynamic behavior.

A similar pattern arises in the presence of defects, where local disruptions to the lattice should introduce significant scattering and thereby reduce thermal conductivity. Within an over-smoothed representation, the defect region is effectively diluted into the surrounding pristine matrix, preventing the model from registering its localized impact. As a result, even substantial experimental reductions in  $\kappa$ —often exceeding fifty percent—are translated into only marginal changes in prediction, indicating a failure to encode the sensitivity required for defect-aware modeling [27].

The limitations become particularly pronounced when examining system size. In first-principles approaches, thermal conductivity converges with increasing supercell size once the relevant scattering phase space is adequately sampled. By contrast, graph-based models subject to strong smoothing effects exhibit a striking invariance: predictions remain effectively unchanged regardless of system scale. This behavior reflects the dominance of global averaging, which fixes the representation independently of structural extent and eliminates any meaningful notion of size-dependent convergence [28]. Taken together, these failure modes do not arise in isolation but trace back to a common origin in representational collapse, underscoring the extent to which architectural constraints can obscure the very physics such models are intended to capture.

**Figure 1** summarizes the manuscript's central causal argument: large unit cells force deeper message passing, deeper message passing accelerates over-smoothing, and over-smoothing systematically destroys the local information required for reliable thermal-transport prediction.



**Figure 1.** Large unit cells require deeper message passing, but deeper message passing drives over smoothing collapses atom resolved representations; collapsed representations generate four predictable thermal-transport failure modes; detection and mitigation must therefore be built into GNN design

These four modes are not occasional bugs but the predictable consequence of feature collapse on large graphs. They appear consistently across different GNN families once depth exceeds the smoothing threshold [9, 11].

**Table 2** consolidates the manuscript’s analytical core by showing that each observed prediction failure can be traced to a specific mechanism of embedding collapse and linked to a concrete diagnostic procedure.

**Table 2.** Mechanism-to-Failure Crosswalk for Over-Smoothing in Thermal-Transport GNNs

| Mechanistic source of information loss         | What is erased in the embedding space                                  | Thermal-transport quantity distorted           | Observable failure mode                                             | Practical detection test                                           |
|------------------------------------------------|------------------------------------------------------------------------|------------------------------------------------|---------------------------------------------------------------------|--------------------------------------------------------------------|
| Repeated feature averaging across neighbors    | Site-specific force-constant contrast                                  | Phonon group velocity estimates                | Conductivity underestimation in heterogeneous pathways              | Depth sensitivity test plus embedding similarity analysis          |
| Depth required to span large graph diameter    | Locality before meaningful long-range context can be integrated safely | Medium-range scattering pathway representation | Abrupt performance drop beyond intermediate depth                   | Accuracy-versus-depth sweep across 2, 4, 8, 16 layers              |
| Collapse of fine-grained anharmonic signatures | Atom-resolved deviations from harmonic behavior                        | Temperature-dependent scattering rates         | Temperature-dependence collapse in $\kappa(T)$                      | Compare predicted $\kappa(T)$ slope across shallow and deep models |
| Homogenization of defect neighborhoods         | Vacancy/substitution-induced local perturbation                        | Defect scattering contribution                 | Defect insensitivity                                                | Vacancy/substitution perturbation test                             |
| Mixing of mode-relevant local patterns         | Acoustic/optical and branch-specific local signatures                  | Mode-resolved transport weighting              | Misallocated transport contribution and flattened physical contrast | Recovery of local environment descriptors from final embeddings    |
| Saturation to global graph-level average       | Cell-specific complexity signals                                       | Size and supercell dependence                  | Size-scaling failure                                                | Compare predicted $\kappa$ across increasing cell/supercell size   |

## Detection Principles

A set of diagnostic procedures has emerged as particularly effective in revealing the onset of representational collapse before a model is exposed to downstream prediction tasks. One of the most direct signals appears when model depth is systematically varied while holding all other factors constant. Under stable conditions, increasing depth should enhance the capacity to encode extended interactions; however, when predictive accuracy for thermal transport exhibits a pronounced decline beyond intermediate depths, the inflection point marks the transition at which smoothing begins to dominate the learning dynamics. This behaviour reflects not a limitation in expressive power per se, but a shift in the balance between information propagation and information dilution, with prior studies consistently locating this transition within a relatively narrow architectural window [8, 9].

A complementary perspective is obtained by examining the internal geometry of the learned representation. When node embeddings within a unit cell converge toward near-identical vectors, as evidenced by pairwise cosine similarities approaching unity, the model effectively loses its ability to distinguish between distinct atomic environments. Such uniformity is not an indicator of successful abstraction but rather of degeneracy in the feature space, signalling that the representational capacity required to encode local physics has been exhausted [13, 16]. This collapse becomes even more evident when one attempts to reconstruct basic structural descriptors from the final-layer embeddings. Quantities such as coordination numbers or bond lengths, which are trivially accessible from the input graph, should remain recoverable if the representation retains meaningful locality. Their systematic irrecoverability therefore provides a stringent confirmation that essential structural information has been irreversibly lost during message passing [19].

The implications of this loss become particularly tangible when the model is probed with controlled perturbations. Introducing a single vacancy into an otherwise pristine structure should induce a localized increase in phonon scattering, leading to a measurable reduction in thermal conductivity. When the predicted response remains negligible, the conclusion is unambiguous: the representation has averaged away the perturbation to such an extent that it no longer registers its physical consequences. In this sense, defect insensitivity functions as a practical stress test, exposing the extent to which local variations have been subsumed into a homogenized embedding space [27]. Incorporating such diagnostic checks into routine model development reframes validation as a probe of physical fidelity rather than purely statistical performance, reducing the risk that architectures fundamentally misaligned with transport physics are advanced to deployment.

Efforts to mitigate these limitations have increasingly focused on architectural strategies that preserve informational heterogeneity while still enabling the capture of extended spatial correlations. One influential approach introduces residual pathways that allow features from earlier layers to bypass successive transformations, thereby maintaining access to high-frequency, locality-sensitive information even as deeper layers aggregate broader context [15]. This mechanism not only stabilizes gradient flow but also counteracts the progressive homogenization characteristic of deep message passing [16]. Closely related interventions operate at the level of feature scaling, where normalization applied across layers moderates the contraction of the embedding space and slows convergence toward the global mean, effectively extending the depth range over which meaningful distinctions can be retained [13].

A different line of reasoning seeks to decouple spatial reach from architectural depth altogether. By enriching shallow graph models with explicit geometric descriptors—pairwise distances and angular relationships extending to medium-range cutoffs—it becomes possible to encode interaction scales relevant to phonon scattering without incurring the penalties associated with deep stacking. In practice, such hybridization preserves local resolution while injecting the non-local information required for accurate transport modelling [20]. This logic extends naturally to multi-branch architectures, where representations evolve along parallel pathways characterized by different receptive fields. Local features can be processed through shallow channels that preserve fine-grained detail, while deeper branches capture more global dependencies, with the two streams ultimately reconciled at the readout stage. The resulting structure reflects a deliberate alignment between architectural hierarchy and the multi-scale nature of lattice dynamics [29].

Further refinement arises through the introduction of attention mechanisms, which replace uniform aggregation with learned, context-dependent weighting [7]. By allowing the model to selectively amplify or suppress contributions from neighbouring nodes, attention preserves salient local variations that would otherwise be diluted under isotropic averaging. This selective propagation extends the lifetime of physically meaningful distinctions within the representation, particularly in heterogeneous environments where not all interactions contribute equally to transport processes [6]. Even with such modifications, however, the dynamics of smoothing cannot be entirely eliminated, making the timing of training termination a critical consideration. Monitoring performance directly on thermal-transport metrics, rather than surrogate objectives such as energy or force prediction, provides a more faithful indicator of when additional depth ceases to yield physically relevant gains. Halting training at this point prevents the model from entering regimes where representational collapse outweighs any incremental benefit in expressivity [14].

Taken together, these interventions do not function as isolated remedies but as components of a broader design philosophy in which architectural choices are explicitly informed by the competing demands of locality and range inherent to phonon-mediated transport. Their combined application enables the construction of models that retain sensitivity to atomic-scale perturbations while still capturing the extended interactions necessary for accurate prediction, thereby restoring alignment between learned representations and the underlying physics.

**Table 3** converts the mitigation discussion into a design decision framework by specifying what each intervention protects, when it should be used, and why hybrid architectures are most plausible for large-unit-cell thermal transport.

**Table 3.** Architecture Design Matrix for Mitigating Over-Smoothing in Large-Unit-Cell Thermal-Transport GNNs

| Design intervention                                                           | Primary purpose                                            | What it preserves                                    | Best-use scenario                                               | Key limitation if used alone                        | Strategic value for thermal transport                          |
|-------------------------------------------------------------------------------|------------------------------------------------------------|------------------------------------------------------|-----------------------------------------------------------------|-----------------------------------------------------|----------------------------------------------------------------|
| Residual / skip connections                                                   | Bypass excessive smoothing across depth                    | Early-layer local features and gradient flow         | Moderately deep models that still need some propagation         | Does not by itself guarantee multi-scale fidelity   | Strong baseline safeguard against feature collapse             |
| Graph normalization                                                           | Slow convergence to homogeneous embeddings                 | Feature diversity across layers                      | Architectures already prone to magnitude collapse               | Can stabilize without fully restoring lost locality | Useful companion to residual design                            |
| Shallow GNN + explicit long-range descriptors                                 | Avoid excessive depth while supplying medium-range context | Local atomic distinctions plus geometric reach       | Large unit cells where 8–15 layers would otherwise be required  | Requires careful descriptor engineering             | Highly aligned with phonon-physics constraints                 |
| Multi-scale dual-path architecture                                            | Separate local and global information streams              | Simultaneous short-range and long-range structure    | Complex crystals with multiple transport-relevant length scales | Higher architectural complexity                     | Best conceptual match to local-to-global thermal transport     |
| Attention-based aggregation                                                   | Replace uniform averaging with selective weighting         | Salient local environments and asymmetric relevance  | Low-symmetry cells with strongly non-equivalent sites           | Attention may still degrade if stacked too deeply   | Extends usable depth without purely uniform smoothing          |
| Early stopping based on $\kappa$ -specific validation                         | Halt training before collapse dominates                    | Task-relevant representational fidelity              | Any supervised $\kappa$ prediction workflow                     | Depends on having robust validation criteria        | Essential operational safeguard                                |
| Hybrid strategy: residual + normalization + shallow-long-range or multi-scale | Jointly balance locality, stability, and range             | Local heterogeneity with controlled global coherence | Recommended default for large-unit-cell thermal transport       | Greater implementation complexity                   | Most credible path toward physics-aware next-generation models |

## Implications for Thermal Transport Prediction

For model developers the central recommendation is unambiguous: avoid deep GNNs ( $>8$  layers) for thermal transport in large unit cells unless skip connections and normalization are explicitly included. Depth sensitivity must be reported as a standard diagnostic.

Benchmark designers should incorporate large-unit-cell thermal-transport tasks ( $\geq 50$  atoms) and defect-sensitivity tests. Performance metrics must be stratified by GNN depth.

Practitioners evaluating existing models should apply the four detection principles before trusting predictions on new complex crystals. When in doubt, prefer shallow GNNs augmented with explicit long-range features over deeper but smoother architectures.

Collectively these implications shift the field from “deeper is better” to “depth-aware design” for phonon-related properties.

## Conclusion

Deep graph neural networks capture crystal structure effectively, but in thermal transport they face a fundamental limitation. Increasing depth to model medium-range interactions in large unit cells drives embeddings toward uniformity, erasing the local variations in force constants and anharmonicity that control phonon behaviour. This creates an inherent conflict between spatial reach and physical fidelity.

The resulting distortions are systematic:  $\kappa$  is underestimated as pathways are averaged, temperature dependence weakens with lost anharmonicity, defect effects are muted, and size dependence disappears. These failures stem directly from feature collapse beyond a critical depth.

This limitation is both diagnosable and manageable. Monitoring depth sensitivity, embedding similarity, structural recoverability, and defect response reveals when representations lose meaning, while residual connections, normalization, shallow architectures with geometric encoding, and multi-scale or attention mechanisms help preserve locality.

For thermal transport, performance depends less on depth than on maintaining local resolution alongside sufficient range. Over-smoothing is therefore a central constraint, and addressing it is essential for physically reliable prediction.

## Acknowledgements

None

## Conflict of interest

None

## Financial support

None

## Ethics statement

None

Received: 02 May 2022

Revised: 19 Aug 2022

Accepted: 11 Nov 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

- Liang F, Qian C, Yu W, Griffith D, Golmie N. Survey of graph neural networks and applications. *Wirel Commun Mob Comput*. 2022;2022:9261537.  
<https://doi.org/10.1155/2022/9261537>.
- Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, et al. Graph neural networks: A review of methods and applications. *AI Open*. 2020;1:57-81.  
<https://doi.org/10.1016/j.aiopen.2021.01.001>.
- Gu GH, Jang J, Noh J, Walsh A, Jung Y. Perovskite synthesizability using graph neural networks. *NPJ Comput Mater*. 2022;8(1):71.  
<https://doi.org/10.1038/s41524-022-00757-z>.
- 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>.
- 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>.
- 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>.
- Satorras VG, Hoogeboom E, Welling M. E(n) equivariant graph neural networks. In: Proceedings of the 38th international conference on machine learning; 2021 Jul 18-24; Virtual. *Proc Mach Learn Res*. 2021;139:9323-32.
- Li Q, Han Z, Wu XM. Deeper insights into graph convolutional networks for semi-supervised learning. *Proc AAAI Conf Artif Intell*. 2018;32(1).  
<https://doi.org/10.1609/aaai.v32i1.11604>.
- 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>.
- Qian X, Zhou J, Chen G. Phonon-engineered extreme thermal conductivity materials. *Nat Mater*. 2021;20(9):1188-202.  
<https://doi.org/10.1038/s41563-021-00918-3>.
- Shin S, Wang Q, Luo J, Chen R. Advanced materials for high-temperature thermal transport. *Adv Funct Mater*. 2020;30(8):1904815.  
<https://doi.org/10.1002/adfm.201904815>.

Xu K, Li J, Zhang M, Du SS, Kawarabayashi KI, Jegelka S. What can neural networks reason about? arXiv preprint arXiv:1905.13211. 2019.  
<https://doi.org/10.48550/arXiv.1905.13211>.

Chen D, Lin Y, Li W, Li P, Zhou J, Sun X. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. *Proc AAAI Conf Artif Intell.* 2020;34(4):3438-45.  
<https://doi.org/10.1609/aaai.v34i04.5747>.

Oono K, Suzuki T. Graph neural networks exponentially lose expressive power for node classification. arXiv preprint arXiv:1905.10947. 2019.  
<https://doi.org/10.48550/arXiv.1905.10947>.

Liu M, Gao H, Ji S. Towards deeper graph neural networks. In: *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*; 2020. p. 338-48.  
<https://doi.org/10.1145/3394486.3403076>.

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. *Proc Mach Learn Res.* 2020;119:1725-35.

Liao M, Duan J, Zhang R, Zhou X, Wu X, Wang X, et al. A hypergraph-embedded convolutional neural network for ice crystal particle habit classification. *Intell Automat Soft Comput.* 2021;29(3):787-801.  
<https://doi.org/10.32604/iasc.2021.018190>.

Han K, Wang Y, Guo J, Tang Y, Wu E. Vision GNN: An image is worth graph of nodes. *Adv Neural Inf Process Syst.* 2022;35:8291-303.  
<https://doi.org/10.52202/068431-0603>.

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>.

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>.

Peng J, Gunda NSH, Bridges CA, Lee S, Haynes JA, Shin D. A machine learning approach to predict thermal expansion of complex oxides. *Comput Mater Sci.* 2022;210:111034.  
<https://doi.org/10.1016/j.commatsci.2021.111034>.

Zhang L, Han J, Wang H, Car R, E W. Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics. *Phys Rev Lett.* 2018;120(14):143001.  
<https://doi.org/10.1103/PhysRevLett.120.143001>.

Liu Z, Yang X, Zhang B, Li W. High thermal conductivity of wurtzite boron arsenide predicted by including four-phonon scattering with machine learning potential. *ACS Appl Mater Interfaces.* 2021;13(45):53409-15.  
<https://doi.org/10.1021/acsami.1c11595>.

Korotaev P, Shapeev A. Lattice dynamics of YbxCo4Sb12 skutterudite by machine-learning interatomic potentials: Effect of filler concentration and disorder. *Phys Rev B.* 2020;102(18):184305.  
<https://doi.org/10.1103/PhysRevB.102.184305>.

Tong Z, Li S, Ruan X, Bao H. Comprehensive first-principles analysis of phonon thermal conductivity and electron-phonon coupling in different metals. *Phys Rev B.* 2019;100(14):144306.  
<https://doi.org/10.1103/PhysRevB.100.144306>.

Chaabene WB, Flah M, Nehdi ML. Machine learning prediction of mechanical properties of concrete: Critical review. *Constr Build Mater.* 2020;260:119889.

<https://doi.org/10.1016/j.conbuildmat.2020.119889>.

Zhang M, Chen Y. Link prediction based on graph neural networks. *Adv Neural Inf Process Syst.* 2018;31:5165-75.

<https://doi.org/10.5555/3327345.3327423>.

Zhang M, Li P. Nested graph neural networks. *Adv Neural Inf Process Syst.* 2021;34:15734-47.

<https://doi.org/10.5555/3540261.3541465>.

Wu L, Cui P, Pei J, Zhao L, Guo X. Graph neural networks: Foundation, frontiers and applications. In: *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*; 2022. p. 4840-1.

<https://doi.org/10.1145/3534678.3542609>.