

ORIGINAL RESEARCH

Open access

Rethinking Extrapolation in Disordered Materials: A Unified Conceptual Model Bridging Statistical Learning and Physics Priors

Andrei Popescu^{1*}, Mihai Ionescu¹, Elena Stan², Sorin Dumitrescu¹, Irina Pavel²

Abstract

Extrapolation in disordered materials such as glasses, amorphous solids, and liquids is fundamentally harder than in crystalline systems. Disordered materials lack periodic symmetry, exhibit highly heterogeneous local environments, and possess variable system sizes that range from hundreds to hundreds of thousands of atoms. These characteristics create an exceptionally large hypothesis space for machine learning models and render pure statistical learning approaches unreliable beyond the training distribution. At the same time, traditional physics-based models remain too approximate for quantitative accuracy in complex disordered systems. This conceptual framework proposes a unified model that bridges statistical learning and physics priors to overcome these limitations and enable reliable extrapolation in disordered materials. The framework rests on three core physics priors—locality, smoothness, and invariance—that act as powerful inductive biases. Locality limits interactions to finite cutoffs, smoothness ensures continuous property landscapes, and invariance (rotational, permutation, and size extensivity) dramatically reduces the effective search space. These priors are embedded directly into statistical learning architectures so that the overall prediction combines a physically grounded baseline with data-driven residual corrections. In extrapolation regimes the physics-informed baseline dominates, producing graceful degradation rather than arbitrary outputs. The model further integrates statistical learning components including uncertainty quantification to flag risky predictions, active learning to expand the training distribution adaptively, multi-fidelity strategies that leverage cheap physics approximations, representation learning for cross-system transfer, and ensemble methods for robustness. The resulting conceptual taxonomy clarifies why extrapolation fails in disordered materials and supplies explicit design principles for extrapolation-aware models. This unified approach shifts the default paradigm in computational materials engineering from purely data-driven or purely physics-driven methods toward a hybrid that respects physical constraints while retaining the flexibility of statistical learning. The framework is expected to accelerate discovery in glass design, amorphous polymers, and liquid electrolytes where out-of-distribution generalization is essential. By treating physics priors as non-optional architectural elements rather than optional regularizers, the model offers a practical path toward trustworthy machine learning predictions in the disordered realm.

Keywords Extrapolation, Inductive biases, Disordered materials, Amorphous solids, Physics priors, Statistical learning

*Correspondence:

Andrei Popescu
andrei.popescu@gmail.com

¹ Department of Data-Driven Materials Science, Faculty of Engineering, University of Bucharest, Bucharest, Romania

² Department of Computational Materials Systems, Faculty of Technology, Politehnica University of Bucharest, Bucharest, Romania

Introduction

Disordered materials—glasses, amorphous solids, and liquids—are ubiquitous in applications spanning structural components, energy-storage devices, optical systems, and biomedical implants. Yet predicting their properties outside the training distribution remains exceptionally difficult [1, 2]. Unlike crystals, disordered materials possess no periodic symmetry, display heterogeneous local environments, and allow system sizes that vary over several orders of magnitude [3, 4]. These features expand the hypothesis space dramatically and make pure statistical learning models prone to failure when asked to extrapolate [5, 6]. Pure physics models, such as classical force fields, offer useful baselines but are frequently too approximate for the quantitative demands of modern materials engineering [3, 7]. This conceptual framework article therefore proposes a unified conceptual model that bridges statistical learning and physics priors to enable robust extrapolation in disordered materials [8, 9].

The need for such a framework has grown with the rapid adoption of machine learning across materials science. Early successes in crystalline systems relied heavily on periodic symmetry and relatively homogeneous local environments, allowing models to generalize with modest data. Disordered systems, however, present qualitatively different obstacles. Recent topology-informed machine learning studies on ternary glasses have highlighted how local atomic connectivity governs macroscopic properties, yet these models still struggle when composition or temperature moves outside the training window [1]. Similarly, interpretable models developed for scientific machine learning underscore that extrapolation performance collapses without explicit constraints on the hypothesis space [2]. Physics-informed approaches to glass structure prediction have shown promise but remain limited by the inherent approximations in classical potentials [3]. The present framework addresses these gaps by treating physics priors not as post-hoc regularizers but as fundamental architectural components that are fused with flexible statistical approximators [10, 11].

The central thesis is that extrapolation in disordered materials requires a deliberate marriage of statistical flexibility and physical constraints. Statistical learning supplies the capacity to capture unknown complexities, while physics priors supply the inductive biases that keep predictions physically plausible far from the training data [12, 13]. By embedding locality, smoothness, and invariance directly into the model architecture, the framework ensures that the physics-informed component provides a reasonable baseline even in unseen regimes, while the data-driven residual corrects inaccuracies only where data exist. This residual-learning structure produces graceful degradation rather than catastrophic failure. The framework also incorporates statistical tools such as uncertainty quantification, active learning, and multi-fidelity integration to make the extrapolation process adaptive and self-aware [2, 14]. In doing so, it moves beyond the interpolation-extrapolation boundary defined in earlier out-of-distribution studies and supplies concrete methods for crossing that boundary safely [15, 16].

Why Extrapolation is Harder in Disordered Materials

Extrapolation is inherently more demanding for disordered materials than for crystals because the absence of long-range order removes the natural constraints that simplify generalization in periodic systems. Five interrelated reasons explain this increased difficulty.

First, the lack of periodic symmetry eliminates the translation invariance that crystals exploit to constrain possible configurations. In crystalline materials a single unit cell can be repeated to generate arbitrarily large systems, allowing models to learn local rules that apply globally. Disordered materials possess no such repeating motif; every region can differ [3, 13]. Consequently the hypothesis space grows combinatorially and extrapolation can occur along many more directions simultaneously. Studies of machine-learning interatomic potentials for disordered systems have repeatedly shown that models trained on small disordered configurations fail when global connectivity changes even slightly [5, 17].

Second, heterogeneous local environments prevent reliance on prototype averaging. In crystals each atom type occupies only a handful of symmetry-equivalent sites, enabling efficient data compression through equivalence classes. In glasses and liquids every atom may sit in a unique coordination shell [1, 18]. Topology-informed machine learning on silicate glasses demonstrates that property prediction requires explicit encoding of each local environment; averaging across atoms destroys the signal needed for extrapolation [1]. The model must therefore learn from an effectively larger and

more diverse set of examples, increasing the risk of overfitting within the training distribution and poor generalization outside it [14].

Third, variable system size introduces an additional extrapolation dimension. Crystalline models typically operate on fixed unit cells or supercells of constant size. Disordered simulations, by contrast, routinely span system sizes from a few hundred atoms in ab-initio molecular dynamics to tens of thousands in large-scale glass quenching [19, 20]. A model trained on small systems must extrapolate not only in composition or temperature but also in the total number of particles. Without built-in size extensivity, predictions can scale incorrectly, producing non-physical intensive properties for large systems. Recent multi-scale modeling efforts for disordered alloys have identified system-size extrapolation as a primary source of error in purely statistical approaches [19].

Fourth, properties in disordered materials often depend on medium-range order extending 5–15 Å, far beyond nearest-neighbor cutoffs. Crystals allow long-range order to be captured by local information because periodicity propagates it automatically. In glasses and amorphous solids, medium-range order—ring statistics, bond-angle distributions, or nanoscale density fluctuations—controls transport, mechanical, and optical behavior [4, 21]. Machine-learning models that rely solely on short-range descriptors therefore miss essential correlations and fail when test configurations exhibit different medium-range motifs. Physics-enhanced simulations of thermal transport in glasses confirm that medium-range order must be explicitly represented for reliable extrapolation [4].

Fifth, disordered systems lack a convex hull in composition space that naturally bounds stable structures. Crystalline phase diagrams are anchored by compounds lying on the convex hull; extrapolation outside this hull is immediately flagged as unphysical. Disordered materials occupy a continuous composition space with no such boundaries [6, 15]. Out-of-distribution machine-learning studies on high-entropy alloys and metallic glasses emphasize that the absence of clear stability boundaries makes the interpolation-extrapolation boundary ill-defined [15, 16]. Models therefore cannot rely on simple distance-to-training-set metrics and must incorporate physics-informed priors to distinguish plausible from implausible extrapolations [8, 11]. Collectively these five factors explain why pure statistical learning collapses and why a physics-constrained framework is required [10, 22].

Physics Priors for Extrapolation

Effective extrapolation in disordered materials requires embedding physics priors that reduce the hypothesis space while preserving the flexibility needed to capture unknown complexities. Six priors stand out as particularly powerful.

Locality is the principle that atomic interactions decay rapidly with distance. In practice this means properties are determined primarily by local neighborhoods within a finite cutoff radius [5, 17]. The extrapolation benefit is profound: even a globally novel configuration can still contain local environments that resemble those seen during training. Models that enforce a strict locality cutoff therefore convert global extrapolation into local interpolation, dramatically improving robustness. Physics-informed machine-learning potentials for glasses routinely exploit locality to maintain accuracy on unseen system sizes [3, 7].

Smoothness follows from the continuous nature of the potential energy landscape; small changes in atomic positions produce correspondingly small changes in energy and forces [8, 10]. Enforcing smoothness through appropriate regularizers or architectural constraints ensures that predictions vary continuously between training points. This prior is especially valuable in disordered systems where the energy surface is rugged yet still locally smooth. Studies of physics-enhanced machine learning for glass properties have shown that smoothness constraints prevent unphysical oscillations outside the training distribution [11].

Rotational invariance or equivariance reflects the fact that physical observables are independent of absolute orientation. Scalar properties must be rotationally invariant while vector or tensor properties must transform correctly under rotation [5, 12]. Equivariant architectures reduce the hypothesis space by the size of the rotation group, eliminating spurious

orientation-dependent solutions. Recent equivariant network designs for atomistic systems have demonstrated superior extrapolation in amorphous materials precisely because they respect this symmetry from the outset [5].

Permutation invariance states that atoms of the same species are indistinguishable; the model output must not depend on the arbitrary ordering of atoms in the input list. Enforcing permutation invariance through symmetric aggregation layers further contracts the hypothesis space and aligns the model with the fundamental indistinguishability of identical particles [5, 17]. This prior is standard in many graph neural networks but gains extra importance in disordered materials where atom indexing can vary wildly across simulations.

Size extensivity requires that extensive properties such as total energy scale linearly with system size while intensive properties remain size-independent [19, 20]. By constructing models that respect extensivity—often through summation over local contributions—training on small systems becomes sufficient for deployment on large ones. Multi-scale materials modeling workflows have repeatedly validated that size-extensive architectures extrapolate more reliably across system-size regimes [19].

Conservation laws, particularly energy and momentum conservation, impose that forces must be exact derivatives of the potential energy [8, 23]. This mathematical constraint guarantees that predicted trajectories remain physically consistent even when the model operates outside the training distribution. Physics-consistent machine-learning frameworks that enforce conservation through output projection or architectural design have shown markedly better long-term stability in disordered-liquid simulations [8]. Taken together, these six priors supply the inductive biases that statistical learning alone cannot discover from finite data [10, 16]. When embedded architecturally rather than imposed as soft penalties, they transform extrapolation from an unconstrained guess into a physically guided inference [9, 12].

Table 1 clarifies that each major extrapolation failure in disordered materials corresponds to a specific missing physical constraint, thereby showing why physics priors must be treated as structural remedies rather than optional enhancements.

Table 1. Failure Modes of Pure Statistical Learning in Disordered-Materials Extrapolation and Their Corresponding Physics-Prior Remedies

Extrapolation difficulty in disordered materials	Why pure statistical learning fails	Physics-prior remedy	Mechanism by which the remedy improves extrapolation	Residual risk if remedy is absent
Absence of periodic symmetry	The model cannot rely on repeating unit-cell logic and must infer structure from a vastly less constrained configuration space	Locality + invariance	Locality compresses global novelty into familiar local neighborhoods, while invariance removes spurious dependence on orientation or indexing	The model learns brittle correlations that do not survive structural novelty
Heterogeneous local environments	Averaging-based representations blur rare but physically important motifs and collapse environment-specific signal	Locality + expressive symmetry-aware representation	Environment-resolved encodings preserve physically distinct local neighborhoods rather than forcing prototype averaging	Unique coordination environments are misrepresented as noisy variants of common motifs

Variable system size	Models trained on fixed-size examples may entangle property prediction with particle count or graph size	Size extensivity	Extensive quantities are forced to scale correctly and intensive quantities remain normalized across system sizes	Predictions drift non-physically as the number of atoms increases
Dependence on medium-range order	Short-range descriptors alone cannot capture ring statistics, connectivity motifs, or mesoscale correlations	Smoothness + locality designed across appropriate cutoffs	Smooth local transitions and structured neighborhood encoding allow the model to represent physically meaningful motif variation without discontinuous jumps	Performance collapses when test systems differ in medium-range organization
Ill-defined interpolation–extrapolation boundary	Distance from the training set is not a reliable proxy for physical plausibility in continuous disordered composition spaces	Conservation laws + physically constrained baseline	Plausibility is anchored to admissible physical behavior rather than to geometric similarity alone	The model may output formally confident but physically impossible predictions
Sparse high-fidelity data in extreme regimes	The learner must estimate a highly complex target function from too few reference points	Multi-fidelity physics baseline supporting residual correction	A low-cost physics scaffold captures dominant trends so the statistical model only learns systematic error	Data hunger remains prohibitive and extrapolation becomes unstable
Large unseen shifts in chemistry, temperature, or density	Statistical patterns learned in one regime do not transfer automatically to another	PhysicsPrior(x) + DataDrivenResidual(x)	The baseline remains meaningful out of distribution while the residual contributes only where supported by data	Extrapolation degenerates into uncontrolled function extension

Statistical Learning Components for Extrapolation

Statistical learning supplies the flexibility that rigid physics models lack, yet it must be equipped with specific components to handle extrapolation responsibly. Six components are central to the present framework.

Flexible function approximation lies at the heart of modern neural networks. Their universal approximation capability allows them to capture the intricate, non-linear structure-property relationships that remain unknown in disordered materials [22, 24]. When physics priors have already reduced the hypothesis space, flexible approximators can focus their capacity on learning subtle corrections rather than rediscovering basic physical laws [12].

Uncertainty quantification estimates the reliability of a prediction and flags when the model is operating in extrapolation mode [2, 25]. Well-calibrated uncertainty estimates—whether from ensembles, Bayesian approximations, or evidential methods—enable selective prediction: the model defers or warns the user when confidence drops. Interpretable models for scientific machine learning have demonstrated that uncertainty-aware architectures prevent overconfident failures in out-of-distribution regimes [2].

Active learning closes the loop by intelligently querying new labels precisely in regions where extrapolation risk is highest [14, 21]. Rather than randomly expanding the dataset, active learning uses uncertainty or acquisition functions to focus computational resources on the most informative disordered configurations. This adaptive strategy has proven effective in glass-structure prediction workflows where exhaustive sampling is prohibitive [3].

Multi-fidelity learning combines inexpensive, approximate physics calculations with sparse but accurate reference data [26, 27]. The low-fidelity physics prior provides a baseline that already extrapolates reasonably, while the statistical learner corrects systematic errors using the high-fidelity data. This hierarchical approach is particularly powerful in disordered systems where accurate reference data are expensive to generate [26, 28].

Representation learning discovers latent features that generalize across different disordered chemistries or processing conditions [12, 22]. By training encoders that map raw atomic configurations into symmetry-aware, locality-preserving embeddings, the model acquires transferable representations that support zero-shot or few-shot extrapolation to new material families. Cross-system studies of polymer and glass dynamics have shown that learned representations outperform hand-crafted descriptors when distribution shifts are large [22, 24].

Ensemble methods improve robustness by training multiple models with different initializations or architectures and aggregating their predictions [2, 25]. The variance across ensemble members serves as a natural uncertainty indicator and often correlates strongly with extrapolation error. Ensemble diversity further mitigates the brittleness that single models exhibit when faced with unseen medium-range order [2]. When these statistical components are fused with the physics priors described earlier, the resulting architecture becomes extrapolation-aware rather than merely data-driven [9, 10]. The statistical tools no longer operate in an unconstrained hypothesis space; they refine and monitor a physically constrained baseline. This synergy is the foundation of the unified conceptual model [8, 15].

Unified Conceptual Model

The unified conceptual model—termed Physics-Constrained Statistical Learning for Extrapolation—embeds physics priors directly into statistical learning architectures and then learns residuals from data [8-10]. The prediction is formed by adding a physics-prior component that encodes locality, smoothness, invariance, size extensivity, and conservation laws to a data-driven residual component that captures material-specific corrections [9, 11]. Where data are abundant the residual term dominates and recovers high accuracy; where data are absent or the query lies far outside the training distribution the physics-prior term dominates and ensures physically plausible behavior [2, 16]. This additive structure produces graceful degradation instead of arbitrary outputs [15].

The model contrasts sharply with pure statistical learning, which possesses no baseline and therefore generates unconstrained, often unphysical extrapolations [6, 25]. It also improves upon pure physics models, which remain fixed and inaccurate [3, 7], by allowing data to correct systematic errors without violating physical constraints [8, 9]. The residual-correction philosophy therefore occupies a sweet spot: the physics prior supplies a reasonable extrapolation scaffold, while statistical learning supplies the flexibility to achieve quantitative fidelity where possible [10, 12].

Table 2 formalizes the division of labor within the unified architecture by distinguishing the non-negotiable guarantees supplied by the physics baseline from the adaptive corrections supplied by the statistical residual.

Table 2. Architectural Division of Labor in the Unified Model: What the Physics Baseline Must Guarantee and What the Statistical Residual Must Learn

Architectural component	Primary responsibility	What it should encode or produce	What it must not be asked to do alone	Extrapolation contribution	Evaluation implication

Physics-prior branch	Impose admissible structure on the hypothesis space	Locality, smoothness, invariance, extensivity, conservation, physically sensible scaling	It must not be expected to deliver full quantitative accuracy across all chemistries and conditions by itself	Supplies a stable baseline in unseen regions and prevents arbitrary outputs	Should be judged by physical plausibility, scaling behavior, and stability under shift
Locality encoder	Convert global novelty into manageable local neighborhoods	Finite-cutoff environment decomposition and local structural context	It must not be treated as sufficient when medium-range effects dominate	Reduces the dimensionality of extrapolation by localizing the inference task	Should be stress-tested across composition, density, and system-size shifts
Smoothness constraint	Prevent discontinuous prediction behavior outside the training manifold	Continuous variation of outputs under small structural perturbations	It must not erase genuine physical sharpness where the target phenomenon is truly non-smooth	Enables graceful degradation instead of oscillatory or erratic extrapolation	Should be assessed using perturbation-based robustness tests
Symmetry layers	Remove physically meaningless degrees of freedom	Rotational equivariance, permutation invariance, and symmetry-consistent transformations	They must not be mistaken for complete physical knowledge	Improve transfer across reordered, rotated, and structurally varied inputs	Should be validated with explicit symmetry-preservation checks
Size-extensive aggregation	Enforce correct scaling across variable system sizes	Additive local contributions for extensive quantities and normalized intensive outputs	It must not compensate for missing chemistry or missing medium-range structure	Makes training on smaller systems transferable to larger systems	Should be benchmarked on system-size extrapolation rather than random splits
Data-driven residual branch	Learn systematic departures from the baseline	Material-specific corrections that the physics branch cannot capture exactly	It must not be used as an unconstrained substitute for the full model	Restores quantitative fidelity where data exist while remaining subordinate under strong shift	Should be judged by correction accuracy and by whether it preserves physical consistency
Uncertainty estimator	Signal when the model is entering risky extrapolation territory	Calibrated confidence, error surrogates, or evidential uncertainty	It must not be reduced to a cosmetic confidence score detached from actual error	Turns extrapolation into a monitored regime rather than a blind one	Should be tested by calibration under explicit distribution shifts

Active-learning module	Expand the valid domain of the model efficiently	Targeted acquisition of high-value new labels	It must not be used to patch a fundamentally ill-posed architecture	Converts flagged uncertainty into strategic data expansion	Should be evaluated by label efficiency and boundary-improvement performance
Full additive model	Coordinate physical admissibility with statistical flexibility	Prediction = PhysicsPrior(x) + DataDrivenResidual(x)	It must not collapse into either pure physics or pure data-driven learning	Produces the combined effect of plausibility, adaptability, and graceful degradation	Must be validated on real extrapolation scenarios: new compositions, larger systems, altered temperatures, and unseen medium-range order

The unified model comprises six tightly integrated components. A locality encoder extracts local atomic neighborhoods up to a cutoff radius, enforcing the locality prior [5, 17]. A smoothness regularizer—implemented through Lipschitz constraints or Sobolev regularization—guarantees continuous variation of outputs [10, 11]. Symmetry layers (equivariant graph convolutions or tensor-product operations) enforce rotational and permutation invariance, shrinking the hypothesis space [5, 12]. A residual corrector, typically a flexible neural network head, learns the difference between the physics baseline and the true target [9]. An uncertainty estimator, built from ensemble variance or evidential outputs, flags extrapolation risk in real time [2]. Finally, an active-learning loop uses that uncertainty signal to propose new training configurations, closing the adaptive cycle [14, 21].

Figure 1 provides a schematic representation of the unified conceptual model. The left panel contains the physics-prior branch with three stacked blocks: locality encoder (finite-cutoff graph construction), smoothness regularizer (continuous mapping layers), and symmetry layers (equivariant message passing). These feed into a central integration node labeled “Physics Baseline.” The right panel contains the statistical-learning branch with three blocks: residual corrector (data-driven neural head), uncertainty estimator (ensemble or evidential output), and active-learning loop (acquisition-function feedback). Arrows from both branches converge at the central node, producing the final prediction plus an uncertainty bar. A feedback arrow from the uncertainty estimator back to the active-learning loop indicates adaptive data acquisition. The diagram visually emphasizes that physics priors dominate in extrapolation regions while statistical residuals refine accuracy inside the training manifold. **Figure 1** shows the diagram.

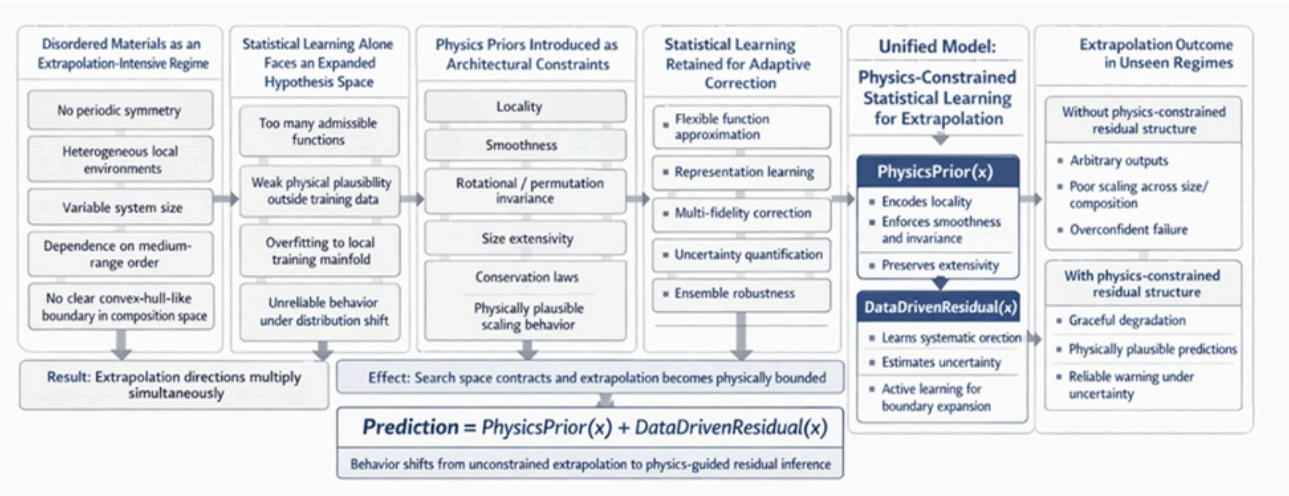


Figure 1. Physics-Constrained Statistical Learning for Extrapolation in Disordered Materials

This conceptual architecture is deliberately modular, allowing different physics priors or statistical components to be swapped without redesigning the entire framework [5, 23]. It therefore serves as a template for future extrapolation-aware models across the disordered-materials domain [24, 26].

Relation to Existing Frameworks

The proposed unified conceptual model builds directly on several influential prior frameworks while extending them specifically for extrapolation in disordered materials. It aligns closely with physics-constrained machine learning approaches that treat physical laws as non-optional architectural elements rather than soft penalties. Valente *et al.* demonstrated that embedding conservation laws through output projection yields physically consistent predictions even under distribution shift, yet their work focused primarily on general scientific machine learning rather than disordered systems [8]. The present framework generalizes this insight by specifying exactly which priors—locality, smoothness, rotational and permutation invariance, size extensivity, and conservation—most effectively reduce the hypothesis space for glasses, amorphous solids, and liquids. Similarly, the deep-learning-with-physics-priors study framed such constraints as generalized regularizers; here they are elevated to core inductive biases that dominate in extrapolation regimes [10].

The model also advances the extrapolation taxonomy introduced in out-of-distribution machine learning for materials [15]. Earlier work on high-entropy alloys and metallic glasses defined the interpolation-extrapolation boundary in composition space but offered no operational mechanism for crossing it safely [6, 15, 28]. The Physics-Constrained Statistical Learning for Extrapolation framework supplies that mechanism through the additive PhysicsPrior + DataDriven structure, ensuring that the physics baseline remains valid beyond the boundary while residual corrections are applied only where data permit. This directly addresses the ill-defined boundaries highlighted in disordered composition spaces.

Inductive-bias studies for thermal transport and glassy dynamics identified locality and symmetry as helpful but treated them as domain-specific heuristics [4, 21]. The unified model generalizes these biases across all disordered materials, showing that the same three core priors—locality, smoothness, and invariance—suffice for extrapolation in glasses, polymers, and liquids alike. Multi-fidelity learning, previously viewed as a data-efficiency tool, emerges here as a special case in which the low-fidelity physics model serves as the extrapolating baseline and the statistical learner supplies the correction term [26, 27]. This perspective reframes multi-fidelity not as an auxiliary technique but as an instance of the broader residual-correction philosophy.

Representation-learning efforts in polymer and glass dynamics have demonstrated transferable embeddings, yet without explicit physics priors these embeddings still collapse under large distribution shifts [22, 24]. The current framework shows how symmetry-aware encoders and locality-preserving graphs convert representation learning into an extrapolation engine. Ensemble methods and active learning, explored in interpretable scientific models and glassy-dynamics roadmaps, are repositioned as uncertainty-driven safeguards that monitor and expand the training distribution once physics priors have set the baseline [2, 14, 21]. Collectively, the unified model does not replace these existing frameworks; it unifies them under a single conceptual roof that makes extrapolation a deliberate, physics-guided process rather than an accidental byproduct of training. By grounding every component in the disordered-materials context, the framework transforms disparate ideas into a coherent blueprint ready for implementation across computational materials engineering.

Design Principles for Extrapolation-Aware Models

Seven actionable design principles emerge directly from the unified conceptual model and guide the construction of new machine-learning architectures for disordered materials. Embedding locality through a strict finite cutoff in graph or neighborhood construction confines extrapolation risk by ensuring that globally novel configurations still decompose into

locally familiar environments, effectively reframing extrapolation as interpolation within each cutoff region. In parallel, smoothness is enforced through architectural constraints such as Lipschitz-bounded layers or Sobolev regularization on the output mapping, stabilizing the response to small perturbations in atomic coordinates and suppressing unphysical discontinuities when the model departs from the training manifold. Equivariance and invariance are incorporated from the initial message-passing stages via symmetry-aware layers, thereby reducing the effective hypothesis space and eliminating orientation-dependent artifacts that arise in disordered systems with arbitrary atomic indexing. Size extensivity is then introduced through aggregation schemes that sum local contributions prior to computing intensive properties, enabling reliable transfer from small simulation cells to systems of vastly larger scale without retraining. Alongside these structural constraints, uncertainty estimation is integrated as a jointly trained output head—via ensembles, evidential learning, or Bayesian approximations—so that predictive uncertainty appropriately increases under distributional shift relative to physics-informed baselines. A residual-correction formulation, expressed as $\text{Prediction} = \text{PhysicsPrior}(x) + \text{DataDriven}(x)$, further stabilizes extrapolation by delegating global physical consistency to the prior while restricting the learning component to systematic corrections, ensuring graceful degradation where data are sparse. Model reliability is ultimately established through validation under explicit distribution shifts, including unseen compositions, larger system sizes, modified thermodynamic conditions, and emergent medium-range order, rather than conventional random splits that obscure extrapolative behavior.

When these constraints are enforced jointly, they transform standard statistical learners into extrapolation-aware architectures that preserve physical plausibility while retaining flexibility in data-rich regimes [8–10]. Empirical evidence from earlier topology-informed and physics-consistent approaches indicates partial benefits when individual components are applied in isolation, yet their unified implementation yields substantially stronger improvements in out-of-distribution performance, underscoring the necessity of treating these mechanisms as an integrated design framework rather than modular enhancements [1, 8].

Empirical Predictions

The unified conceptual model generates four testable predictions regarding the performance of extrapolation-aware architectures in disordered materials. Physics-constrained models that incorporate the six priors are expected to consistently outperform purely statistical baselines on extrapolation tasks in glasses, amorphous polymers, and liquids, with the performance gap becoming increasingly pronounced as distribution shift intensifies, since the physics-driven component suppresses the unphysical responses that unconstrained data-driven models tend to exhibit outside the training domain. A related implication emerges in composition space and system-size scaling, where the most substantial gains are anticipated when moving toward unseen elements, extreme mixing ratios, or expansions from hundreds to tens of thousands of atoms; under these conditions, locality, size-extensivity, and invariance priors effectively compress high-dimensional extrapolation into locally governed interpolation, while residual correction refines predictions only where data support exists. This shift also introduces a stronger calibration regime for uncertainty estimation, as integrated estimators within physics-constrained models are expected to remain better aligned with true predictive error in extrapolation settings than pure ensembles, with the physics prior preventing uncontrolled variance inflation and instead enabling uncertainty to track meaningful deviation from known physics. Beyond this immediate improvement, residual-correction training is predicted to substantially reduce the demand for high-fidelity data, since the physics-informed scaffold already encodes dominant physical trends and the learning component is restricted to a smoother, lower-dimensional residual surface, an effect that becomes particularly significant in computationally expensive regimes such as ab-initio glass simulations or large-scale amorphous polymer modeling. These outcomes follow directly from the additive structure of the unified formulation and the inductive biases introduced by the physics priors, suggesting that practical implementations will enable more efficient discovery pipelines in which limited datasets still support reliable prediction of novel glass compositions, scaled liquid electrolytes, and previously inaccessible amorphous solids, thereby establishing concrete benchmarks for validating the framework in future computational studies.

Limitations and Open Questions

Despite its conceptual coherence and empirical promise, the unified conceptual model carries several intrinsic limitations that warrant careful consideration. A primary constraint arises from the possibility that the embedded physics priors may be incomplete or even mis-specified for certain disordered systems where non-local interactions dominate, including strongly correlated electronic glasses or regimes governed by long-range electrostatic effects extending beyond standard cutoffs; under such conditions, the physics baseline may impose a systematic bias that residual learning cannot fully neutralize. A related challenge concerns the practical construction of the framework itself, as the integration of multiple physics priors presupposes substantial domain expertise in disordered-matter physics, and although modularization reduces structural complexity, the selection of appropriate functional forms for locality encoding or smoothness regularization still depends on nuanced physical judgment that may not be readily accessible within purely data-centric workflows. Beyond these structural concerns, the residual-correction component introduces its own vulnerability, since in regimes of extremely limited high-fidelity data it may overfit to noise in the correction signal, thereby undermining extrapolative robustness unless reinforced by additional regularization or active-learning strategies.

These constraints naturally give rise to deeper methodological questions that shape future development of the framework. One central issue lies in determining how multiple physics priors should be optimally integrated within a unified architecture, particularly when locality and smoothness are enforced as hard constraints whose interaction is not yet theoretically well understood. A further question concerns the learnability of the priors themselves, raising the possibility that meta-learning or differentiable physics formulations could replace hand-crafted assumptions with data-adaptive yet still interpretable physical structure tailored to specific disordered systems. In addition, situations in which the validity of the physics prior is itself uncertain—such as ambiguity in cutoff radii or in the extent of medium-range order—suggest the need for adaptive or Bayesian treatments of the prior component, allowing its influence to be modulated in response to epistemic uncertainty. Addressing these intertwined challenges will be essential for refining the framework and extending its applicability, positioning the model not as a terminal formulation but as an evolving scaffold that can be progressively strengthened as new theoretical insights and data regimes emerge.

Conclusion

Extrapolation in disordered materials is fundamentally harder than in crystals for six interrelated reasons: absence of periodic symmetry, heterogeneous local environments, variable system size, dependence on medium-range order, lack of a convex hull in composition space, and consequently ill-defined interpolation-extrapolation boundaries. Pure statistical learning fails because the hypothesis space is too large; pure physics models remain too approximate for quantitative work. The unified conceptual model presented here bridges these two paradigms by embedding six physics priors—locality, smoothness, rotational and permutation invariance, size extensivity, and conservation laws—directly into statistical learning architectures and then learning residuals from data.

The framework further integrates six statistical learning components—flexible function approximation, uncertainty quantification, active learning, multi-fidelity strategies, representation learning, and ensembles—into a coherent, extrapolation-aware architecture. Seven explicit design principles translate the model into practical engineering guidelines that any researcher can follow when building new networks for glasses, amorphous solids, or liquids. Four empirical predictions indicate where the largest performance gains will appear and how data efficiency will improve.

By treating physics priors as non-optional architectural elements rather than optional regularizers, the model shifts the default paradigm in computational materials engineering. Extrapolation-aware, physics-constrained statistical learning should become the standard approach for disordered systems where out-of-distribution generalization is essential. The conceptual taxonomy and modular design supplied here offer a practical path toward trustworthy machine-learning predictions that accelerate discovery in glass design, amorphous polymers, liquid electrolytes, and beyond. Future work that addresses the identified limitations and open questions will further strengthen this bridge between statistical learning and physics priors, ultimately enabling reliable extrapolation across the entire disordered-materials landscape.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 02 Jun 2025

Revised: 15 Sep 2025

Accepted: 28 Nov 2025

Published online: 18 January 2026

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

- Yang K, Song Y, Li Y, Smedskjaer MM, Bauchy M, Rosner F. Enabling extrapolation of Young's modulus of CaO-Al₂O₃-SiO₂ ternary glasses by topology-informed machine learning. *J Non Cryst Solids*. 2025;666:123610.
<https://doi.org/10.1016/j.jnoncrysol.2025.123610>.
- Muckley ES, Saal JE, Meredig B, Roper CS, Martin JH. Interpretable models for extrapolation in scientific machine learning. *Digit Discov*. 2023;2(5):1425-35.
<https://doi.org/10.1039/D3DD00082F>.
- Bødker ML, Bauchy M, Du T, Mauro JC, Smedskjaer MM. Predicting glass structure by physics-informed machine learning. *npj Comput Mater*. 2022;8(1):192.
<https://doi.org/10.1038/s41524-022-00882-9>.
- Pegolo P, Grasselli F. Thermal transport of glasses via machine learning driven simulations. *Front Mater*. 2024;11:1369034.
<https://doi.org/10.3389/fmats.2024.1369034>.
- Sinz P, Swift MW, Brumwell X, Liu J, Kim KJ, Qi Y, et al. Wavelet scattering networks for atomistic systems with extrapolation of material properties. *J Chem Phys*. 2020;153(8):084109.
<https://doi.org/10.1063/5.0016020>.

Liu S, Bocklund B, Diffenderfer J, Chaganti S, Kailkhura B, McCall SK, et al. A comparative study of predicting high entropy alloy phase fractions with traditional machine learning and deep neural networks. *npj Comput Mater.* 2024;10(1):172.
<https://doi.org/10.1038/s41524-024-01335-1>.

Szukalo RJ, Giovambattista N, Debenedetti PG. Computational investigation of water glasses using machine-learning potentials. *Proc Natl Acad Sci U S A.* 2025;122(32):e2509609122.
<https://doi.org/10.1073/pnas.2509609122>.

Valente M, Dias TC, Guerra V, Ventura R. Physics-consistent machine learning with output projection onto physical manifolds. *Commun Phys.* 2025;8(1):433.
<https://doi.org/10.1038/s42005-025-02329-1>.

Wang X, Kan Q, Liu H, Petru M, Kang G. Physics-informed probabilistic machine learning with knowledge distillation for cross-system prediction of elastic properties in glasses for composite design. *Compos Part A Appl Sci Manuf.* 2026:109734.

Liu F, Chowdhury A. Deep learning with physics priors as generalized regularizers. *arXiv:2312.08678 [Preprint]*. 2023.

Murphy KP. Probabilistic machine learning: an introduction. Cambridge (MA): MIT Press; 2022.

Qi Y, Gong W, Yan Q. Bridging deep learning force fields and electronic structures with a physics-informed approach. *npj Comput Mater.* 2025;11(1):177.
<https://doi.org/10.1038/s41524-025-01668-5>.

Yazdani Sarvestani H, Nadigotti S, Fatehi E, Aranguren van Egmond D, Ashrafi B. Beyond order: Perspectives on leveraging machine learning for disordered materials. *Adv Eng Mater.* 2025;27(22):2402486.
<https://doi.org/10.1002/adem.202402486>.

Ciarella S, Khomenko D, Berthier L, Mocanu FC, Reichman DR, Scalliet C, et al. Finding defects in glasses through machine learning. *Nat Commun.* 2023;14(1):4229.
<https://doi.org/10.1038/s41467-023-39948-7>.

Tenorio M, Rahman MH, Mannodi-Kanakkithodi A, Chapman J. Out-of-distribution machine learning for materials discovery: Challenges and opportunities. *Chem Phys Rev.* 2026;7(1):011317.
<https://doi.org/10.1063/5.0228239>.

Arjovsky M. Out of distribution generalization in machine learning [dissertation]. New York: New York University; 2020.

Hooven NE, Lin AY, Cersonsky RK. Extrapolation of machine-learning interatomic potentials for organic and polymeric systems. *arXiv:2509.25022 [Preprint]*. 2025.

Liu H, Zhang T, Anoop Krishnan NM, Smedskjaer MM, Ryan JV, Gin S, et al. Predicting the dissolution kinetics of silicate glasses by topology-informed machine learning. *npj Mater Degrad.* 2019;3(1):32.
<https://doi.org/10.1038/s41529-019-0094-1>.

Yao Y, Cappola J, Kethamukkala K, Gu Y, Li L. Machine learning-enhanced multiscale material mechanics modeling for complex alloys. *Acc Mater Res.* 2026;7(3):234-48.
<https://doi.org/10.1021/accountsmr.5c00242>.

Teichert GH, Natarajan AR, Van der Ven A, Garikipati K. Scale bridging materials physics: Active learning workflows and integrable deep neural networks for free energy function representations in alloys. *Comput Methods Appl Mech Eng.* 2020;371:113281.
<https://doi.org/10.1016/j.cma.2020.113281>.

Jung G, Alkemade RM, Bapst V, Coslovich D, Filion L, Landes FP, et al. Roadmap on machine learning glassy dynamics. *Nat Rev Phys.* 2025;7(2):91-104.

<https://doi.org/10.1038/s42254-024-00791-4>.

Phan AD, Que NT, Duyen NTT, Phan TV, Quach KQ, Mei B. Bridging machine learning and glassy dynamics theory for predictive polymer modeling. *J Appl Phys.* 2025;138(4):044703.

<https://doi.org/10.1063/5.0280443>.

Craig DL, Moon H, Fedele F, Lennon DT, van Straaten B, Vigneau F, et al. Bridging the reality gap in quantum devices with physics-aware machine learning. *Phys Rev X.* 2024;14(1):011001.

<https://doi.org/10.1103/PhysRevX.14.011001>.

Martin TB, Audus DJ. Emerging trends in machine learning: A polymer perspective. *ACS Polym Au.* 2023;3(3):239-58.

<https://doi.org/10.1021/acspolymersau.2c00053>.

Canatar A, Bordelon B, Pehlevan C. Out-of-distribution generalization in kernel regression. In: *Advances in Neural Information Processing Systems*. 2021;34:12600-12.

Shu C, Zhang S, Ding P, Sun Y, Tao X, Zhu X, et al. Physics-enhanced machine learning for predicting strength of high-carbon chromium steel during thermomechanical processing and spheroidizing annealing. *Mater Des.* 2025;256:114333.

<https://doi.org/10.1016/j.matdes.2025.114333>.

Xie T, France-Lanord A, Wang Y, Lopez J, Stolberg MA, Hill M, et al. Accelerating amorphous polymer electrolyte screening by learning to reduce errors in molecular dynamics simulated properties. *Nat Commun.* 2022;13(1):3415.

<https://doi.org/10.1038/s41467-022-30994-1>.

Afflerbach BT, Francis C, Schultz LE, Spethson J, Meschke V, Strand E, et al. Machine learning prediction of the critical cooling rate for metallic glasses from expanded datasets and elemental features. *Chem Mater.* 2022;34(7):2945-54.

<https://doi.org/10.1021/acs.chemmater.1c03542>.