

ORIGINAL RESEARCH

Open access

Why Diffusion Models Fail for Inverse Materials Design: Mode Collapse, Compositional Violation, and Stability Loss

Hassan Ali¹, Mariam Farooq^{1*}, Usman Shah²

Abstract

Diffusion-based generative models have become a leading approach in artificial intelligence for sampling complex data distributions and have been widely adopted in inverse materials design to generate crystal structures with targeted compositions and properties. Although early implementations suggest scalability beyond traditional combinatorial methods, these models exhibit systematic limitations when applied to crystalline materials. This work identifies three interconnected failure modes that undermine their reliability. Mode collapse leads to overrepresentation of a narrow set of high-symmetry structures, neglecting structurally diverse candidates. Compositional violation results in chemically invalid outputs, including non-integer stoichiometries and charge imbalance. Stability loss arises when generated structures are thermodynamically or dynamically unstable, such as those lying above the convex hull or exhibiting imaginary phonon modes. These issues originate from a fundamental mismatch between diffusion models—designed for continuous, unconstrained data—and the discrete, periodic, and physically constrained nature of crystal systems. Based on peer-reviewed literature, this study provides a conceptual analysis of these failure modes and highlights that robust inverse design requires embedding physical constraints directly within the generative process rather than relying on post hoc filtering.

Keywords Diffusion models, Inverse materials design, Crystal generation, Mode collapse, Compositional violation, Stability loss

*Correspondence:

Mariam Farooq

mariam.farooq@outlook.com

¹ Department of Computational Materials Engineering, Faculty of Engineering, University of Lahore, Lahore, Pakistan

² Department of Data-Driven Materials Analytics, Faculty of Science and Technology, National University of Sciences and Technology, Islamabad, Pakistan

Introduction

Diffusion models have achieved remarkable success in image generation, where their iterative denoising process produces samples of unprecedented fidelity and diversity [1, 2]. Buoyed by these results, researchers in computational materials science have adapted the same framework for inverse materials design—the task of generating novel crystal structures that meet specified target properties such as bandgap, conductivity, or catalytic activity [3–6]. Recent works have reported impressive demonstrations: diffusion-based generators proposing thousands of candidate inorganic crystals, many of which appear chemically plausible at first glance [7–9]. The appeal is obvious. Traditional high-throughput screening is limited by the discrete nature of chemical space, whereas generative models promise to explore that space continuously and intelligently.

Yet the hype obscures systematic failures [10]. Diffusion models, as currently designed, are not well-suited for crystalline materials [11]. They suffer from mode collapse (generating only a few crystal types), compositional violation (invalid stoichiometry, charge imbalance), and stability loss (thermodynamically or dynamically unstable structures). These are not occasional bugs but predictable consequences of applying a continuous, unconstrained generative paradigm to discrete, constrained, periodic data that must obey strict physical laws [1, 12-14].

The literature already hints at these issues. Reviews of generative models for inorganic materials acknowledge that while diffusion approaches outperform earlier VAEs and GANs in raw sample quality, they still produce large fractions of invalid or unstable outputs [4, 5]. Benchmark studies of stability reveal that many diffusion-generated crystals relax to entirely different structures under DFT optimization [15]. Papers focused on compositional constraints document persistent charge-imbalance and fractional-stoichiometry problems even in state-of-the-art crystal diffusion models [13]. Mode-collapse analyses show that diversity metrics for generated crystal space-group distributions are markedly narrower than those of training databases [13, 16].

Table 1 clarifies that mode collapse, compositional violation, and stability loss are mechanistically distinct yet structurally linked consequences of the same model–domain mismatch.

Table 1. Failure-mode architecture for diffusion-based inverse crystal design: mechanism, violation signature, and design consequence

Failure mode	Immediate mechanistic origin	What physical/material rule is being violated	Primary observable signature	What the failure destroys in inverse design	Why this failure is structurally predictable
Mode collapse	The model concentrates sampling probability on dominant training-set patterns during iterative denoising	The practical requirement that inverse design explore the breadth of crystal prototype space rather than only high-frequency modes	Over-representation of common cubic or otherwise high-symmetry prototypes; under-representation of low-symmetry and rare structure families; narrowed prototype coverage	Novelty, exploratory breadth, and access to technologically valuable tail-region structures	Public crystal datasets are distributionally skewed, and diffusion learning reproduces high-density modes more easily than sparse structural tails
Compositional violation	Atom-type generation is learned statistically without embedded conservation constraints	Integer stoichiometry, charge neutrality, chemically plausible oxidation-state balance, and required site occupancy	Fractional occupancies, charge imbalance, chemically impossible valence requirements, impossible pairings, or missing framework sites	Chemical validity and synthesizability	Continuous denoising does not inherently respect discrete atom counts or exact compositional rules
Stability loss	Sampling is guided by denoising likelihood rather	Thermodynamic viability and dynamical stability	Distance above the convex hull, imaginary phonon modes, or strong	Engineering usefulness, downstream screening	Stability is only implicitly represented in the training corpus,

	than explicit free-energy or lattice-dynamical feasibility		structural rearrangement after relaxation	efficiency, and confidence in candidate viability	while the generative process itself has no built-in stability objective
Interaction effect: collapse–stability trade-off	Restricting generation toward safer, common low-energy prototypes can suppress instability but intensify structural concentration	The field's practical need to balance diversity with viability	High validity rates accompanied by poor structural variety	Multi-objective performance across novelty and feasibility	Correcting one failure mode without a joint objective can worsen another
Interaction effect: validity–diversity tension	Encouraging exploration into under-sampled regions increases exposure to chemically or structurally fragile candidates	Simultaneous satisfaction of discovery breadth and physical admissibility	More rare prototypes but lower compositional or stability pass rates	Robust inverse design prioritization	Tail exploration pushes the generator into poorly constrained regions of crystal space
System-level interpretation	The three failures are not independent output defects but parallel manifestations of model-domain mismatch	The requirement that crystal generation be physically constrained at generation time, not merely filtered afterward	High aggregate rejection rates across post-generation diagnostic checks	End-to-end reliability of the generative pipeline	The diffusion architecture was inherited from unconstrained continuous domains rather than designed around crystalline rules

Figure 1 shows that mode collapse, compositional violation, and stability loss do not arise as isolated output defects, but emerge as parallel consequences of the deeper mismatch between continuous diffusion-based generation and the discrete, constrained, periodic, and physically governed nature of crystal structures.

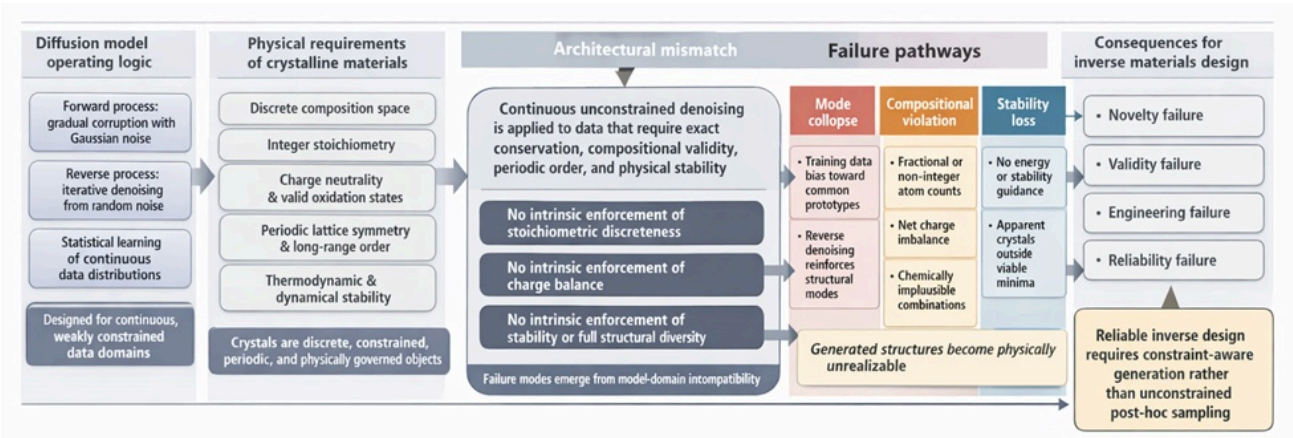


Figure 1. Architectural mismatch and three-path failure emergence in diffusion-based inverse crystal design

How Diffusion Models Work (And Why Crystals are Different)

Diffusion models operate through a two-stage Markov process. In the forward process, Gaussian noise is gradually added to a clean data sample over many time steps until the sample becomes indistinguishable from pure random noise. In the reverse process, a neural network is trained to predict and remove the noise at each step, effectively learning to generate new samples by starting from noise and iteratively denoising [1]. This framework offers stable training dynamics and high-quality sample generation, advantages that have driven its rapid adoption beyond images into scientific domains [5, 10, 11, 16].

When applied to crystal generation, the same denoising mechanism is used, but the data being generated are fundamentally different. Crystal structures are not continuous pixel arrays. They inhabit a discrete composition space defined by specific chemical elements. They must satisfy strict stoichiometry constraints so that atom counts in the unit cell are integers [14]. They must obey charge neutrality, meaning the sum of oxidation states across all atoms equals zero. They are periodic, repeating identically in three dimensions according to space-group symmetry. Most critically, they must be thermodynamically and dynamically stable; otherwise they cannot exist in nature or be synthesized [3, 13, 15, 17, 18].

These differences create an inherent architectural mismatch. Diffusion models were engineered for continuous, unconstrained, non-periodic data where small perturbations in pixel values remain semantically meaningful. Crystals, by contrast, live in a highly constrained, discrete, periodic space where even tiny violations of stoichiometry or charge render the entire structure invalid. The denoising network therefore learns statistical correlations present in the training data (typically drawn from databases such as the Materials Project) but has no intrinsic mechanism to enforce the conservation laws that crystals demand [4, 8, 19].

Training data bias compounds the problem. Most publicly available crystal databases are dominated by high-symmetry, stable structures with common stoichiometries. Diffusion models, which tend to capture the mode rather than the full diversity of the underlying distribution, internalize these biases [12, 20, 21]. The periodicity requirement is usually handled by representing the crystal in a supercell or using equivariant networks, yet the core denoising process remains agnostic to long-range translational symmetry [19]. Stability, meanwhile, is only implicitly present through the training set; the generative process itself contains no explicit energy or phonon guidance [15].

The result is a generative pipeline that appears powerful on paper but systematically violates the physical realities of materials. Subsequent sections dissect the three dominant failure modes that emerge from this mismatch.

Mode Collapse

Mode collapse occurs when a generative model produces only a subset of the possible crystal types, ignoring the vast majority of the structural diversity that exists in chemical space [22, 23]. In the context of crystals, this means the model repeatedly generates only a handful of high-symmetry prototypes—most often cubic perovskites or rock-salt structures—while almost never proposing orthorhombic, tetragonal, monoclinic, or triclinic variants, even when those prototypes are well-represented in training data [12, 16].

The mechanism is rooted in both data bias and the inductive bias of diffusion models. Training datasets such as the Materials Project are heavily skewed toward high-symmetry, stable compounds. Diffusion models learn to reproduce the dominant modes of this distribution with high fidelity but under-sample the tails. Unlike GANs, which can suffer catastrophic mode collapse through adversarial dynamics, diffusion models avoid complete collapse yet still exhibit partial collapse because their reverse process is guided by likelihood rather than explicit diversity enforcement [1, 12, 22, 23].

The iterative denoising steps reinforce common patterns while the probability of sampling rare space groups diminishes with each step.

Detection is straightforward through diversity metrics. When the distribution of space groups or structural prototypes in the generated set deviates significantly from the training distribution, mode collapse is present. Pairwise descriptor distances between generated structures shrink, and prototype coverage drops [5, 24]. For example, if the training set contains approximately 30 % cubic, 20 % hexagonal, 15 % tetragonal, and the remainder distributed across lower-symmetry groups, a collapsed diffusion model might output 80 % cubic, 10 % hexagonal, 5 % tetragonal, with orthorhombic and triclinic structures virtually absent [12].

The practical consequence is severe. Inverse design is supposed to expand the known materials space, yet mode collapse confines discovery to already-explored regions. Rare crystal types that might possess unique electronic, magnetic, or mechanical properties are never proposed [6, 9]. The model therefore fails at its core purpose: proposing truly novel candidates [4, 20].

Compositional Violation

Compositional violation arises when the generated crystal structure contains invalid stoichiometry, charge imbalance, or chemically impossible element combinations [13, 14]. Four subtypes are common. Type A involves non-integer atom counts (e.g., $\text{Li}_{0.3}\text{CoO}_2$), which are physically impossible in a perfect crystal. Type B features net charge imbalance (e.g., LiCoO_3 requiring cobalt in an unrealistic +5 oxidation state). Type C combines elements that cannot coexist stably within a single phase. Type D omits required sites, such as an A-site cation in a perovskite framework [4, 13].

These violations occur because diffusion models treat atom types as categorical variables without built-in conservation laws [13, 19]. The denoising network learns statistical co-occurrence patterns from training data but has no explicit mechanism to enforce integer stoichiometry or charge neutrality. During the reverse process, small noise perturbations can push atom counts or oxidation states across physically forbidden boundaries [3, 8].

Detection relies on simple post-generation checks: verifying that atom counts are integers within a tolerance, computing the net charge using predicted oxidation states, and screening for chemically implausible pairings. A generated “ LiCoO_3 ” structure, for instance, would fail charge-balance verification because cobalt oxides do not support the +5 state required for neutrality [13].

The downstream consequence is immediate: such structures are unsynthesizable [17, 18]. Experimental validation is impossible, and even computational relaxation often reveals the model has produced an artifact rather than a viable material. Compositional violation therefore wastes downstream resources and erodes trust in generative pipelines [5, 15].

Stability Loss

Stability loss occurs when the generated crystal is thermodynamically unstable (formation energy lies significantly above the convex hull) or dynamically unstable (imaginary phonon modes exist) [15, 18]. Even when the structure appears chemically valid, it may represent a high-energy configuration that spontaneously relaxes to a different phase or decomposes [15].

The root cause is the absence of explicit stability constraints in the diffusion process. Training data consist of stable or metastable crystals, but the generative mechanism does not optimize for formation energy or phonon dispersion. Consequently, the model can produce structures that lie outside the convex hull or exhibit soft modes that signal dynamical instability [3, 14, 15].

Detection signatures are clear: formation energy more than 0.1 eV/atom above the hull, presence of imaginary phonon frequencies, or significant structural change (>0.1 Å RMSD) upon DFT relaxation. A generated “cubic” structure that relaxes to tetragonal under optimization is a textbook case of latent instability [15].

Stability loss interacts with mode collapse. Models that collapse onto only the most stable (low-energy) prototypes may avoid stability loss but at the cost of diversity. Conversely, models encouraged to explore diverse structures often generate unstable candidates. This trade-off highlights the need for multi-objective approaches [12, 16].

The practical outcome is that large fractions of generated structures are not viable for synthesis or application. Without addressing stability loss, inverse design remains an academic exercise rather than an engineering tool [4, 5].

Detection Principles

Detection of these failure modes rests on systematic post-generation diagnostics that operate without retraining or full-scale simulation [18, 24]. Diversity assessment begins by comparing the distribution of space groups and structural prototypes in the generated ensemble against the training distribution; marked narrowing—such as over-representation of cubic forms and near-absence of lower-symmetry groups—reveals mode collapse, while pairwise descriptor distances and prototype coverage further expose the model's convergence onto dominant training modes.

A complementary compositional filter assigns oxidation states according to established chemical rules and verifies net charge neutrality across the unit cell, immediately flagging violations that arise when diffusion pathways escape the constraints of charge balance [14, 19]. Closely related, stoichiometry validation confirms that atom counts for each element remain integers within numerical tolerance, exposing fractional occupancies that breach the discrete requirements of crystal chemistry.

Thermodynamic viability is assessed through convex hull distance estimation, whether via fast surrogates or selective DFT, where structures exceeding 0.1 eV/atom above the hull indicate stability loss despite apparent chemical plausibility. Dynamical integrity follows from phonon calculations, whose imaginary frequencies directly signal mechanical instability even in chemically valid candidates.

Finally, geometry relaxation quantifies latent instability by measuring root-mean-square atomic displacements; deviations beyond 0.1 Å demonstrate that the generated configuration was never a true local minimum. Applied in concert, these diagnostics form a coherent pipeline that integrates seamlessly into existing diffusion-based generators and highlights, consistent with benchmark studies, that 70–90 % of outputs fail at least one check [5, 15, 24], thereby enabling early rejection and targeted model refinement.

Mitigation Principles

Mitigation must address each failure mode at its root while preserving the core strengths of diffusion models [10, 11, 25]. Ten interlocking principles, drawn from analyses across the generative materials literature, offer a practical roadmap.

For mode collapse, Principle 1 introduces class-conditional generation: the diffusion process is conditioned explicitly on target space group or prototype label during both training and sampling. This forces the model to cover the full diversity of crystal types rather than defaulting to dominant modes. Principle 2 applies diversity-promoting sampling through classifier-free guidance with an explicit diversity weight that penalizes over-representation of common prototypes [26]. Principle 3 restores balance to the training data by oversampling rare crystal types or using synthetic augmentation of underrepresented space groups, ensuring the model learns the tails of the distribution [12, 16].

For compositional violation, Principle 4 embeds a charge neutrality constraint directly into the denoising objective by adding a differentiable penalty for non-zero net charge at each reverse step [19]. Principle 5 incorporates oxidation-state

prediction as an auxiliary task; the model simultaneously predicts atom types and their oxidation states, then enforces balance before accepting a sample. Principle 6 replaces continuous coordinate denoising with multinomial sampling for atom counts, guaranteeing integer stoichiometry by construction and eliminating fractional occupancies at the source [4, 8, 13].

For stability loss, Principle 7 adopts stability-aware training: the loss function weights each training example by its distance below the convex hull, so the model pays greater attention to truly stable structures [18]. Principle 8 applies convex hull projection as a post-processing step that gently adjusts generated coordinates toward the nearest stable configuration without destroying chemical identity. Principle 9 introduces phonon-guided generation by injecting a lightweight surrogate phonon stability score into the sampling guidance, steering the reverse process away from imaginary-mode regions [3, 5, 15].

For all three failure modes simultaneously, Principle 10 implements multi-objective generation: the diffusion sampler optimizes a joint objective that balances diversity, compositional validity, and stability through Pareto weighting or sequential guidance. This unified approach prevents trade-offs where fixing one failure exacerbates another.

Collectively, these principles transform diffusion models from black-box samplers into constraint-aware generators [10, 11]. Reviews of generative pipelines for inorganic materials emphasize that such targeted interventions—rather than larger models or more data—yield the largest gains in reliability [4, 5, 7]. Implementation requires only modest architectural changes and can be layered onto existing codebases, making widespread adoption feasible.

Table 2 maps each failure mode onto its corresponding diagnostic logic and mitigation pathway, showing that reliable crystal generation depends on coordinated rather than isolated correction.

Table 2. Analytical alignment between failure modes, detection principles, and mitigation logic in diffusion-based crystal generation

Failure mode	Most direct detection principles	What each detection principle reveals conceptually	Most relevant mitigation principles	Core mitigation logic	Residual trade-off that remains
Mode collapse	Diversity measurement; prototype coverage; pairwise descriptor distance analysis	These diagnostics show whether the model is reproducing only high-density structural modes rather than representing the intended discovery space	Class-conditional generation; diversity-promoting sampling; balanced or augmented training distributions	Force or incentivize the generator to represent rare prototypes and lower-symmetry structure families rather than defaulting to dominant modes	Greater diversity may expose more unstable or compositionally fragile candidates
Compositional violation	Charge-balance check; stoichiometry check; plausibility screening of oxidation-state combinations and site occupancy	These checks distinguish superficial visual plausibility from chemically admissible crystal composition	Charge-neutrality penalties in the denoising objective; auxiliary oxidation-state prediction;	Move compositional validity from post-hoc rejection into the generative mechanism itself	Stronger compositional constraint can reduce flexibility and limit exploration of unusual chemistries

			discrete or multinomial atom-count generation		
Stability loss	Convex-hull distance; phonon check; relaxation test	These diagnostics reveal whether an apparently valid crystal is actually viable as an energy-minimizing periodic structure	Stability-aware training; convex-hull-informed post-processing; phonon-guided sampling	Introduce explicit energetic or dynamical feasibility signals into training or sampling	Enforcing stability too strongly can bias the model toward already-common low-energy families
Collapse plus violation together	Diversity metrics combined with validity checks	A structurally diverse generator is not sufficient if the expanded space is chemically inadmissible	Multi-condition guidance that jointly constrains prototype class and composition	Coordinate diversity and validity objectives instead of optimizing either in isolation	Joint conditioning increases design complexity and may require careful weighting
Violation plus instability together	Stoichiometry/charge checks combined with hull and phonon diagnostics	A chemically balanced crystal may still be unrealizable if it occupies an unstable region of structure space	Constraint-aware generation paired with stability-aware objectives	Ensure that chemical admissibility does not masquerade as true material viability	More constraints can reduce generative freedom and novelty yield
Full system integration	Application of all six detection principles as a mandatory post-generation gate	Reliability depends on evaluating diversity, validity, and viability together rather than reporting only sample quality	Multi-objective generation with Pareto or sequential guidance	Convert diffusion from an unconstrained sampler into a physically disciplined inverse-design engine	Optimization must manage persistent tension among novelty, validity, and stability

Relation to Other Failure Modes

The three failure modes of diffusion models do not exist in isolation; they intersect with well-documented weaknesses of earlier generative architectures and expose broader evaluation gaps in the field.

Compared with GANs, diffusion models exhibit milder mode collapse because their likelihood-based training avoids the adversarial feedback loop that can cause complete distributional collapse in GANs [22, 23]. Yet diffusion models still systematically under-sample rare crystal types, inheriting a partial version of the same problem when training data are biased toward high-symmetry prototypes [12, 16]. GANs suffer more catastrophically from mode collapse but, once trained, can sometimes be steered toward greater diversity through architectural tricks that diffusion models have yet to adopt fully [26].

In contrast, VAEs frequently generate invalid structures because their variational bottleneck encourages blurry, averaged outputs that violate stoichiometry and charge rules [25]. Diffusion models generally outperform VAEs in raw sample quality and are less prone to compositional violation, yet they remain imperfect; the absence of explicit constraints still allows charge imbalances and fractional stoichiometries to appear [4, 13]. The iterative nature of diffusion gives it an advantage over the one-shot decoding of VAEs, but that advantage evaporates without enforcement mechanisms.

Current evaluation practices in generative materials design further compound the problem. Most benchmark studies focus narrowly on reconstruction error or property prediction accuracy while omitting systematic tests for mode collapse, compositional validity, or stability loss. This creates an evaluation gap: models appear successful on superficial metrics yet fail when subjected to the detection principles outlined above [5, 24]. The framework presented here directly addresses that gap by providing standardized diagnostics that every generative crystal paper should report.

Finally, these failure modes integrate naturally with active-learning loops. Once unstable or invalid candidates are flagged by the detection principles, active learning can query DFT calculations selectively on the most promising survivors, closing the loop between generation and verification. This synergy turns the identified weaknesses into actionable feedback rather than dead ends [3, 15].

Recognizing these relations clarifies that diffusion models are not uniquely flawed; they simply inherit and sometimes amplify limitations present across the generative modeling landscape [21, 25, 27]. Targeted mitigation therefore benefits the entire ecosystem of AI-driven materials discovery.

Implications for Generative Materials Design

The failure mode analysis carries direct consequences for three stakeholder groups: model developers, practitioners, and benchmark designers.

For model developers the central lesson is clear: diffusion architectures cannot be transplanted from image generation to crystal generation without substantial modification. Assuming that scaling data or network size will automatically solve physical constraints is a fundamental error. Developers must embed charge neutrality, integer stoichiometry, and stability objectives into the core training and sampling pipeline rather than treating them as optional post-processing filters. Failure to do so guarantees that the majority of generated structures will remain unsynthesizable [4, 5, 8, 17,18].

Practitioners who deploy these models for inverse design should adopt a rigorous filtering workflow. Every generated candidate must pass the full suite of detection principles before any DFT validation or experimental consideration. This means expecting rejection rates of 90 % or higher in early iterations. Such filtering is not a sign of model failure but a necessary cost of working with physically grounded materials data. Practitioners should also maintain an internal diversity dashboard that tracks space-group distributions in real time, allowing early intervention when mode collapse begins to dominate [15, 24].

Benchmark designers bear a special responsibility. Existing evaluation suites for generative crystal models are incomplete because they ignore the three failure modes identified here. Future benchmarks must mandate reporting of diversity metrics, compositional validity rates, and stability fractions alongside traditional quality scores. Without these metrics, progress cannot be measured accurately and the community risks mistaking engineering artifacts for genuine scientific advance [5, 7]. Standardized test sets that deliberately include rare prototypes, charge-imbalanced edge cases, and high-energy configurations will drive more robust model development.

Taken together, these implications shift the field from optimistic adoption of diffusion models toward disciplined, constraint-aware engineering [10, 18]. The goal is no longer merely to generate large numbers of candidates but to generate candidates that survive the scrutiny of physical laws. Only then can inverse materials design deliver on its promise of accelerating discovery in batteries, catalysts, and quantum materials.

Conclusion

Diffusion models for inverse materials design fail systematically in three interlocking ways: mode collapse that under-samples rare crystal types, compositional violation that produces invalid stoichiometry and charge imbalances, and stability loss that yields thermodynamically or dynamically unviable structures. These failures stem from a core architectural mismatch—applying a continuous, unconstrained denoising process engineered for images to the discrete, constrained, periodic, and physically governed domain of crystals.

The detection principles—diversity measurement, charge balance and stoichiometry checks, convex hull distance, phonon analysis, and relaxation testing—provide an immediate, practical way to diagnose these issues in any existing pipeline. The mitigation principles—class-conditional generation, charge-neutrality constraints, stability-aware training, and multi-objective optimization—offer concrete strategies to overcome them without sacrificing the generative power that first attracted the community to diffusion models.

The field now stands at a crossroads. Continued uncritical application of off-the-shelf diffusion models will produce ever-larger catalogs of invalid candidates, wasting computational resources and eroding confidence in AI-driven materials discovery. Conversely, embracing the failure mode analysis and implementing the proposed detection and mitigation framework will transform diffusion models into reliable tools that respect the fundamental laws of chemistry and physics.

Generative materials engineering will only fulfill its potential when evaluation standards routinely include the three failure modes and when model architectures explicitly enforce physical constraints. The conceptual roadmap presented here supplies the necessary foundation. The next step is adoption: every new diffusion-based crystal generator must be designed, tested, and reported against these criteria. Only then can the materials community move beyond hype toward genuinely trustworthy inverse design.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 29 May 2024

Revised: 09 Oct 2024

Accepted: 13 Dec 2024

Published online: 18 January 2025

Rights and permissions

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

References

- Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. In: Advances in neural information processing systems. 2020;33:6840-51.
- Corvi R, Cozzolino D, Poggi G, Nagano K, Verdoliva L. Intriguing properties of synthetic images: From generative adversarial networks to diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops; 2023. p. 973-82.
- Song Z, Meng Z, King I. A diffusion-based pre-training framework for crystal property prediction. *Proc AAAI Conf Artif Intell.* 2024;38(8):8993-9001.
<https://doi.org/10.1609/aaai.v38i8.28748>.
- Chen L, Zhang W, Nie Z, Li S, Pan F. Generative models for inverse design of inorganic solid materials. *J Mater Inf.* 2021;1(1):4.
<https://doi.org/10.20517/jmi.2021.07>.
- De Breuck PP, Wang HC, Rignanese GM, Botti S, Marques MAL. Generative AI for crystal structures: A review. *npj Comput Mater.* 2025;11(1):370.
<https://doi.org/10.1038/s41524-025-01881-2>.
- Song Y, Siriwardane EMD, Zhao Y, Hu J. Computational discovery of new 2D materials using deep learning generative models. *ACS Appl Mater Interfaces.* 2021;13(45):53303-13.
<https://doi.org/10.1021/acsami.1c01044>.
- Zeni C, Pinsler R, Zügner D, Fowler A, Horton M, Fu X, et al. A generative model for inorganic materials design. *Nature.* 2025;639(8055):624-32.
<https://doi.org/10.1038/s41586-025-08628-5>.
- Luo X, Wang Z, Wang Q, Shao X, Lv J, Wang L, et al. CrystalFlow: A flow-based generative model for crystalline materials. *Nat Commun.* 2025;16(1):9267.
<https://doi.org/10.1038/s41467-025-64364-4>.
- Park H, Onwuli A, Walsh A. Exploration of crystal chemical space using text-guided generative artificial intelligence. *Nat Commun.* 2025;16(1):4379.
<https://doi.org/10.1038/s41467-025-59636-y>.
- Manduchi L, Pandey K, Bamler R, Cotterell R, Däubener S, Fellenz S, et al. On the challenges and opportunities in generative AI. *arXiv Preprint.* 2024.
<https://doi.org/10.48550/arXiv.2403.00025>.
- Ruthotto L, Haber E. An introduction to deep generative modeling. *GAMM Mitt.* 2021;44(2):e202100008.
<https://doi.org/10.1002/gamm.202100008>.
- Barsha FL, Eberle W. An in-depth review and analysis of mode collapse in generative adversarial networks. *Mach Learn.* 2025;114(6):141.
<https://doi.org/10.1007/s10994-025-06772-7>.
- Long T, Fortunato NM, Opahle I, Zhang Y, Samathrakris I, Shen C, et al. Constrained crystals deep convolutional generative adversarial network for the inverse design of crystal structures. *npj Comput Mater.* 2021;7(1):66.

<https://doi.org/10.1038/s41524-021-00526-4>.

Lalena JN, Cleary DA, Hardouin Duparc OBM. Principles of inorganic materials design. 3rd ed. Hoboken (NJ): Wiley; 2020. 720 p.

Osting B, Wang D. A diffusion generated method for orthogonal matrix-valued fields. *Math Comput.* 2020;89(322):515-50.
<https://doi.org/10.1090/mcom/3473>.

Alverson M, Baird SG, Murdock R, Ho ESH, Johnson J, Sparks TD. Generative adversarial networks and diffusion models in material discovery. *Digit Discov.* 2024;3(1):62-80.
<https://doi.org/10.1039/D3DD00137G>.

Schubert US, Hüsing N. Synthesis of inorganic materials. 4th ed. Weinheim: Wiley-VCH; 2019. 424 p.

Zhu R, Tian SIP, Ren Z, Li J, Buonassisi T, Hippalgaonkar K. Predicting synthesizability using machine learning on databases of existing inorganic materials. *ACS Omega.* 2023;8(9):8210-8.
<https://doi.org/10.1021/acsomega.2c04856>.

Zhao Y, Siriwardane EMD, Wu Z, Fu N, Al-Fahdi M, Hu M, et al. Physics guided deep learning for generative design of crystal materials with symmetry constraints. *npj Comput Mater.* 2023;9(1):38.
<https://doi.org/10.1038/s41524-023-00987-9>.

Lyngby P, Thygesen KS. Data-driven discovery of 2D materials by deep generative models. *npj Comput Mater.* 2022;8(1):232.
<https://doi.org/10.1038/s41524-022-00923-3>.

Cui H, Krzakala F, Vanden-Eijnden E, Zdeborová L. Analysis of learning a flow-based generative model from limited sample complexity. *arXiv Preprint.* 2023.
<https://doi.org/10.48550/arXiv.2310.03575>.

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. *Commun ACM.* 2020;63(11):139-44.
<https://doi.org/10.1145/3422622>.

Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative adversarial networks: An overview. *IEEE Signal Process Mag.* 2018;35(1):53-65.
<https://doi.org/10.1109/MSP.2017.2765202>.

Chen Z, Meng Z, He T, Li H, Cao J, Xu L, et al. Crystal structure prediction meets artificial intelligence. *J Phys Chem Lett.* 2025;16(10):2581-91.
<https://doi.org/10.1021/acs.jpclett.4c03727>.

Hu Z, Yang Z, Salakhutdinov R, Xing EP. On unifying deep generative models. *arXiv Preprint.* 2017.
<https://doi.org/10.48550/arXiv.1706.00550>.

Jahani A, Chai L, Isola P. On the "steerability" of generative adversarial networks. *arXiv Preprint.* 2019.
<https://doi.org/10.48550/arXiv.1907.07171>.

Shin I, Yang C, Chen LC. Deeply supervised flow-based generative models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2025. p. 16535-44.
<https://doi.org/10.1109/ICCV51701.2025.01535>.