DISCOVERING OUT-OF-DISTRIBUTION SUPERCONDUCTORS VIA REINFORCEMENT LEARNING AND MODEL MERGING

Po-Yen Tung* **David S. D. Gunn** **Richard Tomsett** **Jonathon F. S. Markanday**
Robert M. Forrest **Jonathan J. Bean**

MatNex, London, UK

{brian.tung, david, richard.tomsett, freddie.markanday,
robert, jonathan}@materialsnexus.com

ABSTRACT

Inverse design of superconducting materials requires generative models that can optimize functional properties without collapsing to narrow regions of known chemistry. We study this problem using diffusion-based generative models fine-tuned via reinforcement learning (RL) with surrogate feedback, where a graph neural network predictor of superconducting critical temperature provides the reward signal. Compared with classifier-free guidance (CFG), RL increases the average surrogate-predicted critical temperature from 10 K to 17 K. To address diversity collapse during feedback-based optimization, we further merge independently fine-tuned models, which increases the fraction of unique chemical systems by over 7% relative to single-model fine-tuning. We quantify exploration using a calibrated out-of-distribution metric, OOD@α, and find that the merged RL model generates $>$92% of samples beyond the 95th percentile of the calibration distribution (OOD@0.05) vs. 21% for CFG-based baselines. This increased exploration does not degrade structural validity, with S.U.N. (stable, unique, and novel) rates improving from 0.19 under CFG to 0.57 under RL. Overall, RL combined with model merging provides a practical, high-potential framework for discovering new series of genuinely out-of-distribution superconductors, and other critical materials.

1 INTRODUCTION

Novel materials discovery, particularly superconductors, remains a longstanding challenge because of the vast combinatorial space of crystal structures and compositions, the scarcity of high-quality labels, and the strong biases present in existing datasets. While recent diffusion-based generative models have demonstrated the ability to produce physically plausible inorganic crystals (Zeni et al., 2025), their use in superconducting materials discovery has largely focused on conditional generation within the support of known compounds (Prakash et al., 2025a;b). As a result, these approaches tend to optimize within established chemical families and struggle to explore genuinely novel regions of the materials space.

Feedback-based learning provides a natural alternative to purely conditional generation. By coupling generative models with surrogate- or simulation-based feedback, reinforcement learning (RL) enables direct optimization of target properties without requiring labeled training data. Prior work (Chen et al., 2025) has shown that RL can improve sample efficiency and property optimization in materials design pipelines. However, its impact on the distributional behavior of generative models, particularly in terms of exploration beyond known superconducting regimes, remains underexplored. In the context of superconductors, such exploration is essential: the discovery of new superconduct-

*Corresponding author

ing families requires controlled departures from the training distribution while preserving structural validity and stability.

In this work, we study RL as a mechanism for distributional reshaping in generative models for superconducting materials. We fine-tune a pretrained diffusion model using RL with a graph neural network (GNN) surrogate to estimate the superconducting critical temperature (T_c), and systematically compare this approach to classifier-free guidance (CFG) (Ho & Salimans, 2022) baselines initialized from the same pretrained model and data. To characterize exploration beyond known superconductors, we introduce a calibrated sample-level out-of-distribution (OOD) metric that quantifies novelty relative to a reference dataset, complementing standard global distributional measures. We further show that independently fine-tuned RL models exhibit complementary exploration behavior, and that simple model-merging provides an effective way to recover compositional diversity without sacrificing property optimization or quality of the generated material structures.

Our results demonstrate that RL-based fine-tuning induces broader yet more controlled exploration of superconducting materials space than CFG-based approaches, leading to substantial gains in predicted T_c, increased discovery of statistically novel candidates, and no degradation in structural validity. These findings suggest that feedback-based learning, when paired with appropriate evaluation and aggregation strategies, offers a practical pathway toward discovering truly new superconducting materials.

Contributions. The main contributions of this work are:

1. We study feedback-based RL as a practical mechanism for steering a pretrained diffusion generator toward superconducting candidates, using a reward-weighted KL-anchored update, with terminal T_c surrogate feedback evaluated on relaxed, quality-filtered structures.
2. We introduce OOD@α, a calibrated sample-level OOD metric in a learned structure-embedding space, and use it alongside standard quality and diversity metrics to disentangle property gains from controlled exploration beyond the fine-tuning distribution.
3. We show that independent RL runs explore complementary regions of structure space, and that simple task-vector averaging consolidates these trajectories into a single generator that broadens exploration signals (OOD@α) while preserving sample quality and optimization behavior.

2 Background and Related Work

Superconductor discovery with machine learning. Early machine learning (ML) efforts in superconductivity focused on predicting T_c from composition-derived descriptors, highlighting both the value of large curated resources (e.g., SuperCon (Materials Database Group, 2022)) and the need for family-aware models to capture heterogeneous mechanisms (Stanev et al., 2018; Jiang & Xu, 2023). Later work explored deeper predictors aimed at extrapolative discovery, including "reading periodic table" style representations (Konno et al., 2021) and hybrid learners such as convolutional gradient boosting decision trees (Dan et al., 2020). Newer screening pipelines incorporate structural information and physically motivated inductive biases via graph models (Gu et al., 2023). More recently, guided diffusion with CFG has been used to propose and filter large candidate sets (Prakash et al., 2025a;b), while integrated "AI-accelerated" stacks combine surrogate modeling, high-throughput computation, and experimental verification (Gibson et al., 2026).

Generative models for crystal structure generation. Generative modeling of periodic crystals has progressed rapidly, with diffusion and transformer architectures now able to generate realistic, stable inorganic structures and guide generation toward desired properties through fine-tuning (Zeni et al., 2025; Joshi et al., 2025). Recent work emphasizes explicit treatment of crystallographic symmetry and periodic geometry. Equivariant diffusion models that jointly generate lattice parameters and fractional coordinates, as well as space-group–constrained generation, reduce geometric artifacts (Jiao et al., 2024a;b). Complementary approaches incorporate symmetry-aware inductive biases through space-group–informed and Wyckoff-based discrete representations (Cao et al., 2024; Kazeev et al., 2025). Symmetry-preserving diffusion jointly generates asymmetric units and symmetry operations, producing crystals with realistic symmetries (Levy et al., 2025). Beyond structure-

native models, language-model paradigms are emerging as flexible interfaces for materials knowledge and targeted design (Mishra et al., 2025; Xu et al., 2025).

Feedback-based learning in materials design. Materials discovery is inherently feedback-driven: candidate proposals are iteratively evaluated and the proposer is updated to improve sample efficiency under expensive, noisy objectives. Representative closed-loop frameworks combine learned surrogates with explicit exploration mechanisms, including active-learning-driven experimental discovery in high-entropy alloys (Rao et al., 2022) and deep active optimization methods for complex systems (Wei et al., 2025). RL provides an alternative route when the "policy" is a generative model and rewards are obtained from surrogates or screening pipelines; recent materials-focused examples include RL for inverse inorganic design with synthesizability constraints (Karpovich et al., 2024), diffusion-model RL fine-tuning for inverse design (Chen et al., 2025), and reinforcement fine-tuning of crystal generative transformers (Cao & Wang, 2025). A practical concern in reinforcement fine-tuning is diversity collapse, motivating explicit regularization and diversity-aware selection in closed-loop updates (Santi et al., 2025; Liu et al., 2026).

3 Framework for Superconductor Discovery

3.1 Problem Formulation and Design Objective

We formulate the inverse design of superconducting materials as a feedback-driven generative modeling problem over periodic crystal structures. Let $\mathcal{X}$ denote the space of candidate crystals (defined by atomic species, lattice parameters, and fractional coordinates). A material $\mathcal{M} \in \mathcal{X}$ is one such crystal. Our objective is to learn a generative distribution $p_\theta(\mathcal{M})$ that proposes *novel*, *physically plausible* superconductors with high critical temperature T_c.

Each generated structure is scored by a surrogate predictor $\hat{T}_c(\mathcal{M})$ trained on first-principles electron–phonon T_c calculations to approximate the superconducting critical temperature. This predicted $\hat{T}_c$ provides the sole scalar reward signal. Rather than conditioning on a fixed target T_c at sampling time, we use $\hat{T}_c$ as feedback in the RL loop to steer updates of the generator itself, thereby reshaping the model distribution toward higher-T_c candidates.

Formally, we aim to solve

$$\max_{p_\theta} \; \mathbb{E}_{\mathcal{M}\sim p_\theta}\left[\hat{T}_c(\mathcal{M})\right], \tag{1}$$

subject to feasibility and novelty constraints reflecting superconductivity physics and synthesis considerations.

Physical and chemical constraints. Optimization is restricted to a feasible subset $\mathcal{X}_{\text{phys}} \subset \mathcal{X}$ defined by the following requirements. (i) *Structural validity*: $\mathcal{M}$ must correspond to a well-formed periodic crystal with valid atomic geometry and stoichiometry. (ii) *Thermodynamic plausibility*: structures should lie within a bounded energy window above the convex hull, consistent with commonly used metastability tolerances in computational screening. (iii) *Electronic feasibility*: candidate materials must be metallic, as insulating phases cannot host conventional superconductivity. (iv) *Chemical plausibility*: generated compositions must respect realistic valence chemistry and exclude physically implausible element combinations.

Explicit novelty and OOD design. Discovering new superconductors requires exploration beyond the structural and compositional regimes represented in existing datasets. Let $\mathcal{D}_{\text{sc}}$ denote the set of known superconducting structures used to train the surrogate model. We therefore specifically target structures whose representations lie outside the high-density region of $\mathcal{D}_{\text{sc}}$ in a learned structure–composition embedding space. In this sense, novelty is defined distributionally rather than combinatorially: a candidate is considered OOD if it is statistically unlikely under the reference distribution of known superconductors, rather than merely absent from the training set. This formulation treats OOD generation as an intended and quantifiable outcome of feedback-based optimization, rather than an incidental artifact of stochastic sampling.

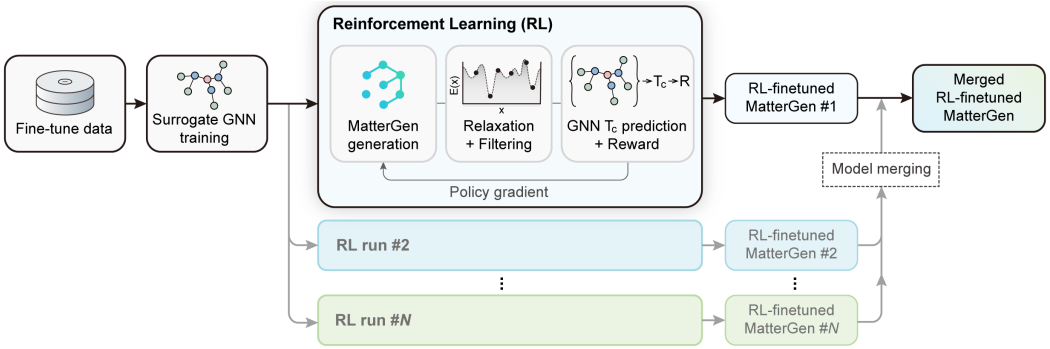

Figure 1: Schematic workflow of the RL-based fine-tuning of generative models, followed by model merging. Fine-tuning data from the superconductivity dataset (Prakash et al., 2025a) are first used to train a surrogate GNN for critical temperature (T_c) prediction. The generative model (MatterGen) is then fine-tuned via RL: candidate structures are generated, relaxed and filtered, and scored using a GNN-based T_c predictor to provide rewards for feedback. Multiple independent RL runs yield diverse fine-tuned models, which are subsequently merged to produce a single RL-finetuned MatterGen.

3.2 Feedback-Driven RL Fine-Tuning

We fine-tune a pretrained MatterGen diffusion generator using a feedback-based RL procedure adapted from MatInvent (Chen et al., 2025). Specifically, we optimize a reward-weighted policy-gradient objective with KL anchoring (a penalty that keeps the policy close to the pretrained generator) and use a terminal reward computed from a superconductivity surrogate on relaxed, filtered structures (both defined below). As shown in Figure 1, the resulting closed loop relaxes and filters MatterGen samples, scores survivors with a surrogate GNN predictor $\hat{T}_c$, and updates the generator via reward-weighted policy gradients, enabling sparse structure-level feedback to reshape the model distribution while maintaining a conservative prior over physically and chemically valid crystals.

Reverse diffusion as an episodic policy with terminal reward. Let $\mathcal{M}_{0:T}$ denote a reverse-diffusion trajectory with T denoising steps, where $\mathcal{M}_T$ is drawn from the noise prior and $\mathcal{M}_0$ is the final generated crystalline material. We treat the conditional transition $p_\theta(\mathcal{M}_{t-1} \mid \mathcal{M}_t)$ as a stochastic policy over intermediate states, and assign reward only to the terminal structure:

$$R_t = \begin{cases} r(\mathcal{M}_0), & t = T-1, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

This episode-level supervision matches our setting: superconductivity feedback is only meaningful after the full structure is generated and processed through the physical evaluation pipeline.

Reward from relaxed, S.U.N.-filtered candidates. For each generated sample, we (i) relax the structure using MatterSim (Yang et al., 2024), (ii) apply validity checks and the S.U.N. (stable, unique, and novel) filter, as defined in Section 4, and (iii) evaluate a pretrained GNN surrogate $\hat{T}_c$ on the relaxed structure. We convert surrogate predictions into bounded rewards by applying a fixed linear rescaling, based on a predefined upper-bound on physically plausible critical temperatures, followed by clipping to the $[0, 1]$ interval. Structures that fail relaxation, validity, or S.U.N. filtering are discarded and do not contribute to the RL update.

Reward-weighted policy gradient with KL anchoring. With terminal reward, the expected-reward objective is

$$\mathcal{J}_{\mathrm{RL}}(\theta) = \mathbb{E}_{p_\theta(\mathcal{M}_{0:T})}[r(\mathcal{M}_0)]\,. \tag{3}$$

We define two per-trajectory gradient components:

$$g_{\mathrm{pg}}(\mathcal{M}_{0:T}) \;=\; \sum_{t=1}^{T} \nabla_\theta \log p_\theta(\mathcal{M}_{t-1} \mid \mathcal{M}_t), \tag{4}$$

$$g_{\mathrm{kl}}(\mathcal{M}_{0:T}) \;=\; \sum_{t=1}^{T} \nabla_\theta D_{\mathrm{KL}}\Big(p_\theta(\mathcal{M}_{t-1} \mid \mathcal{M}_t) \,\Big\|\, p_{\mathrm{pre}}(\mathcal{M}_{t-1} \mid \mathcal{M}_t)\Big). \tag{5}$$

Here p_{pre} is the frozen pretrained MatterGen model. Following Chen et al. (2025), we use a reward-weighted KL anchor with λ slightly larger than the maximum reward. The combined gradient used for fine-tuning can then be written compactly as

$$\nabla_\theta \mathcal{L}(\theta) \;\propto\; -r(\mathcal{M}_0)\, g_{\mathrm{pg}}(\mathcal{M}_{0:T}) + \beta\big(\lambda - r(\mathcal{M}_0)\big)\, g_{\mathrm{kl}}(\mathcal{M}_{0:T}). \tag{6}$$

which preserves the pretrained prior when rewards are low, while allowing higher-reward trajectories to more readily shift probability mass. Technical details of the RL fine-tuning procedure are provided in Appendix A.6.

3.3 Classifier-free guidance baselines

We benchmark RL fine-tuning against CFG baselines following prior guided-diffusion superconductor work Prakash et al. (2025a;b) and MatterGen (Zeni et al., 2025). CFG (Ho & Salimans, 2022) performs inference-time steering by mixing conditional and unconditional scores during reverse diffusion,

$$s_{\mathrm{cfg}}(M_t; c) = \gamma\, s_\theta(M_t; c) + (1-\gamma)\, s_\theta(M_t; \varnothing), \tag{7}$$

where c is the target property value (here T_c), $\varnothing$ denotes a null condition, and γ controls guidance strength. For CFG-MatterGen, we follow MatterGen's property-guided setup and use their released checkpoint where the conditional score network is obtained by fine-tuning the full pretrained score model with property conditioning. Technical details of the CFG fine-tuning procedure are provided in Appendix A.5.

3.4 Task-vector model merging to aggregate RL exploration trajectories

RL fine-tuning from the same pretrained generator can yield complementary solutions across random seeds, reflecting different exploration trajectories under identical surrogate feedback. As shown in Figure 1, merging is a post-hoc aggregation step applied *after* completing n independent closed-loop RL runs from the same θ_{base}; each run produces a fine-tuned checkpoint θ_i, and task-vector averaging yields a single merged generator θ_{merged} used for downstream sampling and evaluation. To consolidate these trajectories into one model without ensembling, we merge independently fine-tuned checkpoints in *task-vector* space, following the task-arithmetic view of fine-tuning as a parameter delta from a shared base (Ilharco et al., 2023) and recent diffusion model merging formulations (Biggs et al., 2024; Ma et al., 2025). Additional details of formulation is provided in Appendix A.7

4 Metrics for Superconductor Discovery

We evaluate each generative model using a set of complementary metrics designed to assess property improvement, physical feasibility, and controlled exploration beyond known superconductors. Rather than relying on a single objective, these metrics jointly characterize a model's ability to generate candidates that are both scientifically viable and diverse in structure and composition.

Critical temperature improvement. We evaluate property optimization using the mean surrogate-predicted critical temperature over generated samples. We intentionally avoid tail-based metrics (e.g., top-percentile T_c) for two reasons. First, the surrogate model is trained on known superconductors with relatively low critical temperatures (T_c<30 K), making extreme extrapolations in the high-T_c regime unreliable and prone to reward exploitation during RL. Second, our objective is not to identify a small number of extreme candidates, but to generate a broad and diverse

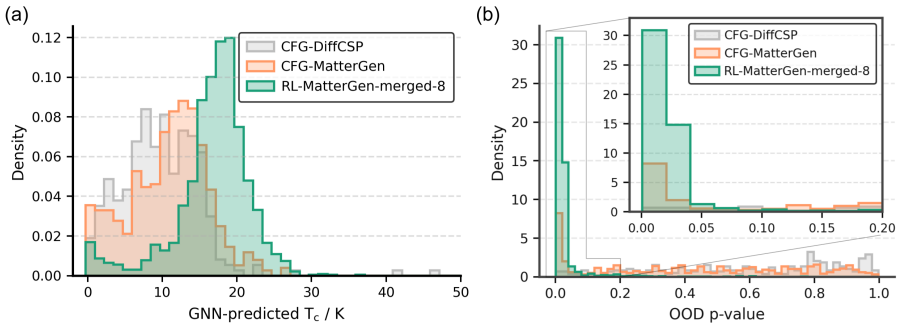

Figure 2: (a) Distributions of GNN-predicted critical temperatures (T_c) for structures generated by CFG-DiffCSP (Prakash et al., 2025a), CFG-MatterGen, and the merged RL-MatterGen-merged-8 model. (b) Distributions of calibration-based OOD conformal p-values, where lower values indicate that they are statistically less likely to originate from the fine-tuning data distribution.

set of structurally and compositionally distinct materials with consistently improved superconducting potential. Therefore, mean T_c serves as a conservative, population-level indicator of systematic property improvement, while diversity and novelty are assessed separately through complementary metrics.

Structural and physical validity (S.U.N.). S.U.N. (Zeni et al., 2025) measures the fraction of generated structures that are simultaneously *stable*, *unique*, and *novel*, and acts as a discovery-oriented sample-quality metric defined in terms of downstream density-functional theory (DFT) calculations and feasibility. Rather than assessing generic generative realism, S.U.N. reflects a minimum viability bar for candidate materials: structures that are unstable, redundant, or already known are unlikely to survive subsequent first-principles screening regardless of predicted superconducting performance. More detailed definitions and criteria can be found in Appendix A.8.

OOD@α: sample-level departure from known superconductors. We quantify sample-level departure from known superconductors using OOD@α, a new calibrated OOD metric introduced in this work, defined in a learned structure-embedding space. Each structure is embedded using UPET-OAM-L graph representations (Mazitov et al., 2025), concatenated across layers. As the reference distribution we use the superconductivity structure dataset (Prakash et al., 2025a), which is split into fit and calibration subsets (80/20). Embeddings from the fit subset are used to fit a preprocessing pipeline consisting of standardization, PCA reduction to 128 dimensions, and whitening. The resulting transformation is then held fixed and applied to both the calibration set and all generated samples. For a given sample j, we define an outlier score s_j as the Euclidean distance to its k-th nearest neighbor (k=20) in the fit set. Applying the same transformation and nearest-neighbor model to the calibration set yields outlier scores $\mathrm{s}^{\mathrm{cal}}$, which we use to compute an empirical OOD p-value:

$$p_j = \frac{1 + \left|\left\{i : s_i^{\mathrm{cal}} \geq s_j\right\}\right|}{1 + n_{\mathrm{cal}}}, \tag{8}$$

where $|\cdot|$ denotes number of elements in the set. We then report the OOD rate at level α over m generated samples as

$$\mathrm{OOD@}\alpha \;=\; \frac{1}{m}\sum_{j=1}^{m} \mathbf{1}\{p_j \leq \alpha\}. \tag{9}$$

OOD@α measures the fraction of generated samples whose structural representations are statistically unlikely under known superconductors at level α. Smaller values of α correspond to stricter OOD criteria; in particular, OOD@0.05 captures samples lying beyond the 95% tail of the calibrated reference distribution. As an intentionally simple, non-parametric detector, OOD@α is used to characterize the quality of exploration at the sample level and is interpreted jointly with S.U.N. to ensure that increased departure from known structures is accompanied by physical plausibility.

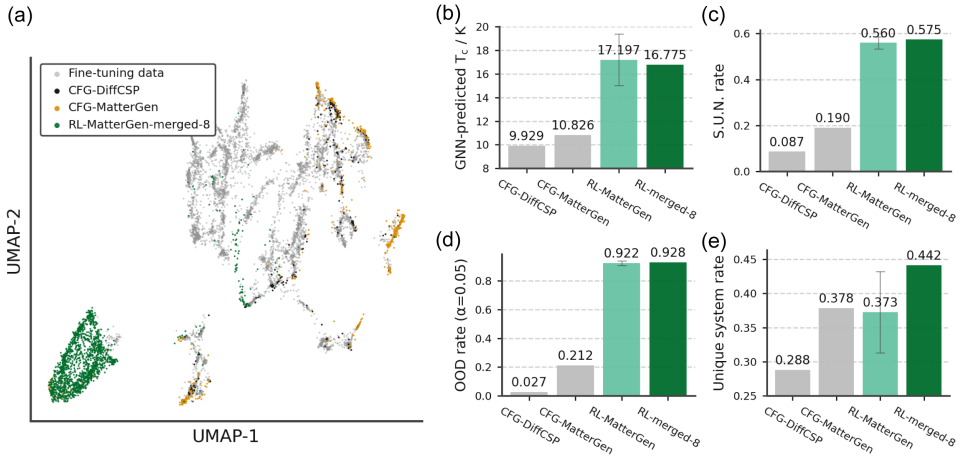

Figure 3: (a) UMAP projection of structure embeddings. CFG-DiffCSP (Prakash et al., 2025a) and CFG-MatterGen samples remain near the fine-tuning manifold, while the merged model (RL-MatterGen-merged-8) explores additional regions, consistent with higher OOD@0.05 and chemical-system diversity. (b–e) Quantitative comparison of CFG-DiffCSP, CFG-MatterGen, RL-MatterGen, and RL-MatterGen-merged-8 in terms of (b) GNN-predicted critical temperature (T_c), (c) structure generation quality (S.U.N. rate), (d) OOD rate at α=0.05, and (e) unique chemical system rate. The merged RL model achieves improved performance while maintaining broad distributional coverage.

Distribution shift (FID). We use Fréchet Inception Distance (FID) as a model-level proxy for exploration, quantifying the global distributional shift between generated structures and known superconductors. FID is computed in a chemically and structurally informed, task-agnostic embedding space using UPET (Mazitov et al., 2025) representations, which are standardized, reduced to 128 dimensions via PCA, and whitened. We then measure the Fréchet distance between Gaussian fits (mean and covariance) of the generated and reference superconductivity datasets. Higher FID indicates a larger global departure from known superconductor families, while lower FID reflects closer alignment with the training distribution. In this work, FID is a *post hoc* diagnostic used to characterize emergent exploration behavior. Because FID does not distinguish between meaningful exploration and unphysical extrapolation, we interpret it jointly with S.U.N. and OOD@α to identify distributional shifts that correspond to viable expansion of the design space.

Unique system rate. We measure compositional diversity using the unique system rate, defined as $\frac{|\mathcal{S}|}{m}$, where m is the number of generated structures and $\mathcal{S}$ is the set of distinct chemical systems, defined by unordered element combinations (e.g., Fe–Ni). This metric serves as a coarse diversity check, capturing whether a model generates candidates across different elemental combinations rather than repeatedly sampling variants within a small number of known systems. In superconductor discovery, where identifying new chemical families is often more impactful than rediscovering variants within established ones, a higher unique system rate indicates broader compositional coverage and is interpreted jointly with S.U.N. to ensure physical plausibility.

Prototype novelty. Unlike OOD@α, which quantifies statistical departure in a learned embedding space, we assess novelty at the crystallographic level by measuring whether generated crystals realize previously unseen prototypes, using a database-referenced *prototype novelty* metric defined as the fraction absent from the AFLOW prototype encyclopedia (Eckert et al., 2024). Details of the AFLOW-XtalFinder matching procedure and the resulting prototype novelty criterion are given in Appendix A.3.

Table 1: Quantitative comparison of CFG-based and RL-based diffusion models. The first three columns report the fraction of S.U.N.-filtered samples exceeding surrogate-predicted critical temperature (T_c) thresholds, followed by global and sample-level distributional metrics.

MODEL	>5 K	>10 K	>20 K	S.U.N.	$\text{FID}_{\text{FINETUNE}}$	OOD@0.01	OOD@0.05	OOD@0.2	PROTOTYPE NOVELTY
CFG-DIFFCSP	0.81	0.48	0.02	0.087	91.02	0.007	0.027	0.108	0.861
CFG-MATTERGEN	0.83	0.61	0.04	0.190	173.15	0.100	0.212	0.325	0.862
RL-MATTERGEN	**0.95**	**0.92**	**0.29**	0.560	**351.17**	**0.182**	0.922	0.975	**0.989**
RL-MERGED-8	**0.95**	0.91	0.21	**0.575**	328.54	0.158	**0.928**	**0.980**	0.983

5 RESULTS

5.1 RL VS CFG: ESCAPING KNOWN REGIMES

Figure 2 compares RL-based fine-tuning and model merging with CFG baselines in terms of property optimization and distributional exploration. As shown in Figure 2a, the merged RL-MatterGen-merged-8 model shows a pronounced rightward shift in the distribution of GNN-predicted T_c, with an average predicted T_c of approximately 16.8 K, compared to 10.8 K for CFG-MatterGen and 9.9 K for CFG-DiffCSP (see Figure 3b). This shift indicates substantially more frequent generation of high-T_c candidates under RL fine-tuning. Table 1 reports the fraction of S.U.N.-filtered samples whose surrogate-predicted T_c exceeds 5, 10, and 20 K; RL increases the yield of high-T_c candidates, and merging largely preserves these gains.

To assess whether these gains are accompanied by broader exploration beyond the fine-tuning data, we measure OOD p-values using a calibrated conformal test in structure-embedding space (Section 4). Figure 2b shows that the merged RL model produces a substantially higher fraction of samples with OOD p-values below 0.05. These samples are statistically less likely under the fine-tuning distribution, therefore indicating increased exploration into less represented regions of structure-embedding space. Overall, when trained on identical fine-tuning data, RL combined with model merging simultaneously improves T_c optimization and promotes controlled OOD exploration, whereas CFG-based methods (particularly, CFG-DiffCSP) remain largely concentrated in high p-value regions closer to the training distribution.

5.2 DISTRIBUTIONAL SHIFT AND PHYSICAL VALIDITY

Table 1 reports complementary distribution-level and sample-level measures of distributional shift. We use FID to quantify global deviation of the generated distribution from the fine-tuning data, complementing the sample-level OOD@α metric, which measures the fraction of generated structures lying in the calibrated tail of the reference distribution. We also report prototype novelty, a database-referenced crystallographic metric that serves as a supporting indicator of structural exploration beyond known prototypes.

RL-based models exhibit substantially higher FID than CFG baselines (e.g., 328.5 for RL-MatterGen-merged-8 vs. 173.2 for CFG-MatterGen), indicating a coherent shift of the overall generated distribution rather than isolated outliers. Consistent with this global shift, RL models also achieve markedly higher OOD rates across multiple thresholds, with the merged model reaching OOD@0.01, OOD@0.05, and OOD@0.2 of 0.158, 0.928, and 0.980, respectively. Prototype novelty is also higher for RL-based models (0.983 for the merged model), indicating that the RL-induced distributional shift extends beyond known crystallographic prototypes rather than repeated sampling within established ones. Importantly, this broader exploration of the generative space is achieved without degrading structural generation quality: the merged RL model attains a S.U.N. rate of 0.575, compared to 0.190 for CFG-MatterGen and 0.087 for CFG-DiffCSP, indicating that increased exploratory coverage does not come at the expense of validity or diversity constraints.

5.3 MODEL MERGING EXPANDS CHEMICAL-SYSTEM COVERAGE

Figure 3 summarizes how model merging affects property optimization, sample quality, and compositional coverage. Comparing to CFG baselines, RL fine-tuning shifts the generated set toward

higher surrogate-predicted T_c (Figure 3b). Merging eight independently fine-tuned RL models produces a pronounced increase in the unique chemical system rate (Figure 3e), indicating substantially broader superconducting chemistry coverage. This diversity gain is not accompanied by a collapse in sample quality: the merged model matches the S.U.N. rate of single RL models (Figure 3c) and achieves the highest OOD@0.05 rate (Figure 3d). While merging slightly reduces mean predicted T_c (17.2 K $\rightarrow$ 16.8 K), the property profile remains comparable, suggesting that merging primarily trades a small amount of surrogate score for a large expansion in chemical-system breadth.

Merged RL model vs. CFG baselines. Figure 3a visualizes the UMAP projections of structure embeddings from different data sources (see A.2 for technical details of the embedding analysis). In Figure 3a, CFG samples (black/orange) largely lie in the high-density regions of the fine-tuning manifold (gray), with most points lying within existing branches and clusters. In contrast, RL-MatterGen-merged-8 populates a clearly separated, dense cluster in the lower-left of the embedding. This separation is qualitatively consistent with the merged model's higher OOD@0.05 and its increased unique chemical system rate, indicating exploration into regions underrepresented in the fine-tuning set.

Consolidation of complementary RL trajectories. Figure A1 resolves how merging achieves this broader coverage. The eight RL-fine-tuned models exhibit substantial overlap on the main manifold while emphasizing partially distinct sub-regions (Figure A1a). The merged model recovers samples across these sub-regions, including the detached island, yielding a composite distribution that reflects the union of exploration modes rather than a collapse to any single RL trajectory (Figure A1b). In the next section, we highlight representative OOD superconducting candidates revealed by RL exploration and model merging.

5.4 First-principles validation of RL candidates

To quantify whether RL-driven exploration yields *both* distributional novelty and physically competitive superconducting response, we validate a set of candidates with DFT, density-functional perturbation theory (DFPT), and electron-phonon Wannier (EPW) calculations (see Appendix A.9).

The selected set comprises n=10 candidates that are simultaneously high-scoring under the surrogate and strongly OOD in structure-embedding space (11 candidates were initially selected for EPW validation, but one candidate was removed due to identification of a match in the AFLOW database). The mean surrogate critical temperature is $T_c^{\mathrm{GNN}} = 26.70 \pm 4.34$ K, while the mean OOD p-value is $p_{\mathrm{OOD}} = 0.0163 \pm 0.0072$ (smaller values indicate stronger departure from the reference set). Despite these low p-values, EPW calculations yield mean $T_c^{\mathrm{EPW}} = 22.2 \pm 10.89$ K, suggesting that superconductivity under EPW remains competitive even for strongly OOD structures.
Figure 4 summarizes the two signals together: the validated candidates lie at low p_{OOD} (strongly OOD) while achieving high EPW-computed T_c values. These results demonstrate that RL fine-tuning can identify promising superconducting candidates in regions that are poorly represented in the fine-tuning data.

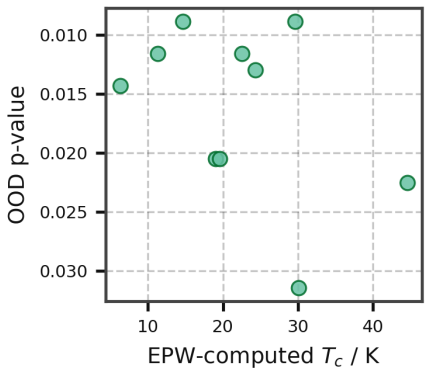

Figure 4: First-principles validation set (n=10): EPW-computed critical temperature (T_c) vs. OOD p-value in structure-embedding space (lower indicates stronger OOD).

6 Discussion

This work presents a practical route to broaden and accelerate the search for superconductors beyond the existing chemical and structural regimes that dominate the literature. By fine-tuning a pretrained crystal diffusion model via RL, using a surrogate T_c predictor as the reward, we steer generation towards candidates with higher predicted critical temperatures, while maintaining physical feasibility.

Compared with CFG baselines trained on the same data, RL induces a stronger distributional shift, yielding substantially more statistically OOD samples.

Beyond superconductors, the methodology we demonstrate in this work constitutes a mechanism for turning general-purpose pretrained generative models into goal-directed discovery tools in specific areas of interest, while under realistic data constraints. Many material domains face similar issues to those seen in superconductor research, including scarcity of accurate, reliable, and unbiased data, and expensive, time-consuming, experiments and simulations. The framework presented here aims to provide a means of AI-assisted materials discovery while accounting for these issues; by using surrogate or quick simulation feedback as the reinforcement-learning signal, the generator can be steered toward target properties without the expense of full simulations or experiments, enabling rapid iteration. Simultaneously, the newly proposed OOD@α metric provides a calibrated sample-level measure of distributional novelty relative to a reference dataset, allowing improved evaluation of potential new discoveries.

ACKNOWLEDGMENTS

The authors acknowledge support from the IRCAI & AWS Compute for Climate Fellowship, in the form of compute credits and technical guidance.

REFERENCES

P. B. Allen and R. C. Dynes. Transition temperature of strong-coupled superconductors reanalyzed. *Phys. Rev. B*, 12:905–922, Aug 1975. doi: 10.1103/PhysRevB.12.905. URL https://link.aps.org/doi/10.1103/PhysRevB.12.905.

Stefano Baroni, Stefano de Gironcoli, Andrea Dal Corso, and Paolo Giannozzi. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.*, 73:515–562, Jul 2001. doi: 10.1103/RevModPhys.73.515. URL https://link.aps.org/doi/10.1103/RevModPhys.73.515.

Benjamin Biggs, Arjun Seshadri, Yang Zou, Achin Jain, Aditya Golatkar, Yusheng Xie, Alessandro Achille, Ashwin Swaminathan, and Stefano Soatto. Diffusion soup: Model merging for text-to-image diffusion models, 2024. URL https://arxiv.org/abs/2406.08431.

Zhendong Cao and Lei Wang. Crystalformer-rl: Reinforcement fine-tuning for materials design, 2025. URL https://arxiv.org/abs/2504.02367.

Zhendong Cao, Xiaoshan Luo, Jian Lv, and Lei Wang. Space group informed transformer for crystalline materials generation, 2024. URL https://arxiv.org/abs/2403.15734.

Junwu Chen, Jeff Guo, Edvin Fako, and Philippe Schwaller. Accelerating inverse materials design using generative diffusion models with reinforcement learning, 2025. URL https://arxiv.org/abs/2511.03112.

Yabo Dan, Rongzhi Dong, Zhuo Cao, Xiang Li, Chengcheng Niu, Shaobo Li, and Jianjun Hu. Computational prediction of critical temperatures of superconductors based on convolutional gradient boosting decision trees. *IEEE Access*, 8:57868–57878, 2020. doi: 10.1109/ACCESS.2020.2981874.

Hagen Eckert, Simon Divilov, Michael J. Mehl, David Hicks, Adam C. Zettel, Marco Esters, Xiomara Campilongo, and Stefano Curtarolo. The aflow library of crystallographic prototypes: Part 4. *Computational Materials Science*, 240:112988, May 2024. ISSN 0927-0256. doi: 10.1016/j.commatsci.2024.112988. URL http://dx.doi.org/10.1016/j.commatsci.2024.112988.

José A. Flores-Livas, Lilia Boeri, Antonio Sanna, Gianni Profeta, Ryotaro Arita, and Mikhail Eremets. A perspective on conventional high-temperature superconductors at high pressure: Methods and materials. *Physics Reports*, 856:1–78, 2020. ISSN 0370-1573. doi: https://doi.org/10.1016/j.physrep.2020.02.003. URL https://www.sciencedirect.com/science/article/pii/S0370157320300363. A perspective on conventional high-temperature superconductors at high pressure: Methods and materials.

Johannes Gasteiger, Florian Becker, and Stephan Günnemann. Gemnet: Universal directional graph neural networks for molecules, 2024. URL https://arxiv.org/abs/2106.08903.

P Giannozzi, O Andreussi, T Brumme, O Bunau, M Buongiorno Nardelli, M Calandra, R Car, C Cavazzoni, D Ceresoli, M Cococcioni, N Colonna, I Carnimeo, A Dal Corso, S de Gironcoli, P Delugas, R A DiStasio, A Ferretti, A Floris, G Fratesi, G Fugallo, R Gebauer, U Gerstmann, F Giustino, T Gorni, J Jia, M Kawamura, H-Y Ko, A Kokalj, E Küçükbenli, M Lazzeri, M Marsili, N Marzari, F Mauri, N L Nguyen, H-V Nguyen, A Otero-de-la Roza, L Paulatto, S Poncé, D Rocca, R Sabatini, B Santra, M Schlipf, A P Seitsonen, A Smogunov, I Timrov, T Thonhauser, P Umari, N Vast, X Wu, and S Baroni. Advanced capabilities for materials modelling with quantum espresso. *Journal of Physics: Condensed Matter*, 29(46):465901, oct 2017. doi: 10.1088/1361-648X/aa8f79. URL https://doi.org/10.1088/1361-648X/aa8f79.

Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiarotti, Matteo Cococcioni, Ismaila Dabo, Andrea Dal Corso, Stefano de Gironcoli, Stefano Fabris, Guido Fratesi, Ralph Gebauer, Uwe Gerstmann, Christos Gougoussis, Anton Kokalj, Michele Lazzeri, Layla Martin-Samos, Nicola Marzari, Francesco Mauri, Riccardo Mazzarello, Stefano Paolini, Alfredo Pasquarello, Lorenzo Paulatto, Carlo Sbraccia, Sandro Scandolo, Gabriele Sclauzero, Ari P Seitsonen, Alexander Smogunov, Paolo Umari, and Renata M Wentzcovitch. Quantum espresso: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter*, 21(39): 395502, sep 2009. doi: 10.1088/0953-8984/21/39/395502. URL https://doi.org/10.1088/0953-8984/21/39/395502.

Jason B. Gibson, Ajinkya C. Hire, Pawan Prakash, Philip M. Dee, Benjamin Geisler, Jung Soo Kim, Zhongwei Li, James J. Hamlin, Gregory R. Stewart, P. J. Hirschfeld, and Richard G. Hennig. Developing a complete AI-accelerated workflow for superconductor discovery. 2026. ISSN 2057-3960. doi: 10.1038/s41524-026-01964-8. URL https://doi.org/10.1038/s41524-026-01964-8.

Feliciano Giustino. Electron-phonon interactions from first principles. *Rev. Mod. Phys.*, 89:015003, Feb 2017. doi: 10.1103/RevModPhys.89.015003. URL https://link.aps.org/doi/10.1103/RevModPhys.89.015003.

Liang Gu, Yang Liu, Pin Chen, Haiyou Huang, Ning Chen, Yang Li, Yutong Lu, and Yanjing Su. Searching high temperature superconductors with the assistance of graph neural networks, 2023. URL https://arxiv.org/abs/2308.11160.

Christoph Heil, Samuel Poncé, Henry Lambert, Martin Schlipf, Elena R. Margine, and Feliciano Giustino. Origin of superconductivity and latent charge density wave in nbs_2. *Phys. Rev. Lett.*, 119:087003, Aug 2017. doi: 10.1103/PhysRevLett.119.087003. URL https://link.aps.org/doi/10.1103/PhysRevLett.119.087003.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance, 2022. URL https://arxiv.org/abs/2207.12598.

Gabriel Ilharco, Marco Tulio Ribeiro, Mitchell Wortsman, Suchin Gururangan, Ludwig Schmidt, Hannaneh Hajishirzi, and Ali Farhadi. Editing models with task arithmetic, 2023. URL https://arxiv.org/abs/2212.04089.

Juntao Jiang and Renjun Xu. Critical Temperature Prediction of Superconductors Based on Machine Learning: A Short Review, 2023.

Rui Jiao, Wenbing Huang, Peijia Lin, Jiaqi Han, Pin Chen, Yutong Lu, and Yang Liu. Crystal structure prediction by joint equivariant diffusion, 2024a. URL https://arxiv.org/abs/2309.04475.

Rui Jiao, Wenbing Huang, Yu Liu, Deli Zhao, and Yang Liu. Space group constrained crystal generation, 2024b. URL https://arxiv.org/abs/2402.03992.

Chaitanya K. Joshi, Xiang Fu, Yi-Lun Liao, Vahe Gharakhanyan, Benjamin Kurt Miller, Anuroop Sriram, and Zachary W. Ulissi. All-atom diffusion transformers: Unified generative modelling of molecules and materials, 2025. URL https://arxiv.org/abs/2503.03965.

Christopher Karpovich, Elton Pan, and Elsa A. Olivetti. Deep reinforcement learning for inverse inorganic materials design. 10(1):287, 2024. ISSN 2057-3960. doi: 10.1038/s41524-024-01474-5. URL https://doi.org/10.1038/s41524-024-01474-5.

Nikita Kazeev, Wei Nong, Ignat Romanov, Ruiming Zhu, Andrey Ustyuzhanin, Shuya Yamazaki, and Kedar Hippalgaonkar. Wyckoff transformer: Generation of symmetric crystals, 2025. URL https://arxiv.org/abs/2503.02407.

Tomohiko Konno, Hodaka Kurokawa, Fuyuki Nabeshima, Yuki Sakishita, Ryo Ogawa, Iwao Hosako, and Atsutaka Maeda. Deep learning model for finding new superconductors. *Phys. Rev. B*, 103:014509, Jan 2021. doi: 10.1103/PhysRevB.103.014509. URL https://link.aps.org/doi/10.1103/PhysRevB.103.014509.

Daniel Levy, Siba Smarak Panigrahi, Sékou-Oumar Kaba, Qiang Zhu, Kin Long Kelvin Lee, Mikhail Galkin, Santiago Miret, and Siamak Ravanbakhsh. Symmcd: Symmetry-preserving crystal generation with diffusion models, 2025. URL https://arxiv.org/abs/2502.03638.

Jinmei Liu, Haoru Li, Zhenhong Sun, Chaofeng Chen, Yatao Bian, Bo Wang, Daoyi Dong, Chunlin Chen, and Zhi Wang. Beyond the dirac delta: Mitigating diversity collapse in reinforcement fine-tuning for versatile image generation, 2026. URL https://arxiv.org/abs/2601.12401.

Qianli Ma, Xuefei Ning, Dongrui Liu, Li Niu, and Linfeng Zhang. Decouple-then-merge: Finetune diffusion models as multi-task learning, 2025. URL https://arxiv.org/abs/2410.06664.

E. R. Margine and F. Giustino. Anisotropic migdal-eliashberg theory using wannier functions. *Phys. Rev. B*, 87:024505, Jan 2013. doi: 10.1103/PhysRevB.87.024505. URL https://link.aps.org/doi/10.1103/PhysRevB.87.024505.

Materials Database Group. Mdr supercon datasheet ver.220808, 2022.

Arslan Mazitov, Filippo Bigi, Matthias Kellner, Paolo Pegolo, Davide Tisi, Guillaume Fraux, Sergey Pozdnyakov, Philip Loche, and Michele Ceriotti. PET-MAD as a lightweight universal interatomic potential for advanced materials modeling. 16(1):10653, 2025. ISSN 2041-1723. doi: 10.1038/s41467-025-65662-7. URL https://doi.org/10.1038/s41467-025-65662-7.

Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2020. URL https://arxiv.org/abs/1802.03426.

W. L. McMillan. Transition temperature of strong-coupled superconductors. *Phys. Rev.*, 167:331–344, Mar 1968. doi: 10.1103/PhysRev.167.331. URL https://link.aps.org/doi/10.1103/PhysRev.167.331.

Vaibhav Mishra, Somaditya Singh, Dhruv Ahlawat, Mohd Zaki, Vaibhav Bihani, Hargun Singh Grover, Biswajit Mishra, Santiago Miret, Mausam, and N. M. Anoop Krishnan. Foundational large language models for materials research, 2025. URL https://arxiv.org/abs/2412.09560.

P. Morel and P. W. Anderson. Calculation of the superconducting state parameters with retarded electron-phonon interaction. *Phys. Rev.*, 125:1263–1271, Feb 1962. doi: 10.1103/PhysRev.125.1263. URL https://link.aps.org/doi/10.1103/PhysRev.125.1263.

Arash A. Mostofi, Jonathan R. Yates, Young-Su Lee, Ivo Souza, David Vanderbilt, and Nicola Marzari. wannier90: A tool for obtaining maximally-localised wannier functions. *Computer Physics Communications*, 178(9):685–699, 2008. ISSN 0010-4655. doi: https://doi.org/10.1016/j.cpc.2007.11.016. URL https://www.sciencedirect.com/science/article/pii/S0010465507004936.

Arash A. Mostofi, Jonathan R. Yates, Giovanni Pizzi, Young-Su Lee, Ivo Souza, David Vanderbilt, and Nicola Marzari. An updated version of wannier90: A tool for obtaining maximally-localised wannier functions. *Computer Physics Communications*, 185(8):2309–2310, 2014. ISSN 0010-4655. doi: https://doi.org/10.1016/j.cpc.2014.05.003. URL https://www.sciencedirect.com/science/article/pii/S001046551400157X.

John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996. doi: 10.1103/PhysRevLett.77.3865. URL https://link.aps.org/doi/10.1103/PhysRevLett.77.3865.

Chris J. Pickard, Ion Errea, and Mikhail I. Eremets. Superconducting hydrides under pressure. *Annual Review of Condensed Matter Physics*, 11(1):57–76, March 2020. ISSN 1947-5462. doi: 10.1146/annurev-conmatphys-031218-013413. URL http://dx.doi.org/10.1146/annurev-conmatphys-031218-013413.

S. Poncé, E.R. Margine, C. Verdi, and F. Giustino. Epw: Electron–phonon coupling, transport and superconducting properties using maximally localized wannier functions. *Computer Physics Communications*, 209:116–133, 2016. ISSN 0010-4655. doi: https://doi.org/10.1016/j.cpc.2016.07.028. URL https://www.sciencedirect.com/science/article/pii/S0010465516302260.

Pawan Prakash, Jason B. Gibson, Zhongwei Li, Gabriele Di Gianluca, Juan Esquivel, Eric Fuemmeler, Benjamin Geisler, Jung Soo Kim, Adrian Roitberg, Ellad B. Tadmor, Mingjie Liu, Stefano Martiniani, Gregory R. Stewart, James J. Hamlin, Peter J. Hirschfeld, and Richard G. Hennig. Guided diffusion for the discovery of new superconductors, 2025a. URL https://arxiv.org/abs/2509.25186.

Pawan Prakash, Jason B. Gibson, Zhongwei Li, Gabriele Di Gianluca, Juan Esquivel, Eric Fuemmeler, Benjamin Geisler, Adrian Roitberg, Ellad B. Tadmor, Mingjie Liu, Stefano Martiniani, Gregory R. Stewart, James Hamlin, Peter Hirschfeld, and Richard Hennig. Inverse design of novel superconductors via guided diffusion. In *AI for Accelerated Materials Design - NeurIPS 2025*, 2025b. URL https://openreview.net/forum?id=tY7MTfuqfB.

Ziyuan Rao, Po-Yen Tung, Ruiwen Xie, Ye Wei, Hongbin Zhang, Alberto Ferrari, T.P.C. Klaver, Fritz Körmann, Prithiv Thoudden Sukumar, Alisson Kwiatkowski da Silva, Yao Chen, Zhiming Li, Dirk Ponge, Jörg Neugebauer, Oliver Gutfleisch, Stefan Bauer, and Dierk Raabe. Machine learning–enabled high-entropy alloy discovery. *Science*, 378(6615):78–85, 2022. doi: 10.1126/science.abo4940. URL https://www.science.org/doi/abs/10.1126/science.abo4940.

Riccardo De Santi, Marin Vlastelica, Ya-Ping Hsieh, Zebang Shen, Niao He, and Andreas Krause. Flow Density Control: Generative Optimization Beyond Entropy-Regularized Fine-Tuning, 2025. URL http://arxiv.org/abs/2511.22640.

Valentin Stanev, Corey Oses, A. Gilad Kusne, Efrain Rodriguez, Johnpierre Paglione, Stefano Curtarolo, and Ichiro Takeuchi. Machine learning modeling of superconducting critical temperature. 4(1):29, 2018. ISSN 2057-3960. doi: 10.1038/s41524-018-0085-8. URL https://doi.org/10.1038/s41524-018-0085-8.

M.J. van Setten, M. Giantomassi, E. Bousquet, M.J. Verstraete, D.R. Hamann, X. Gonze, and G.-M. Rignanese. The pseudodojo: Training and grading a 85 element optimized norm-conserving pseudopotential table. *Computer Physics Communications*, 226:39–54, 2018. ISSN 0010-4655. doi: https://doi.org/10.1016/j.cpc.2018.01.012. URL https://www.sciencedirect.com/science/article/pii/S0010465518300250.

Ye Wei, Bo Peng, Ruiwen Xie, Yangtao Chen, Yu Qin, Peng Wen, Stefan Bauer, Po-Yen Tung, and Dierk Raabe. Deep active optimization for complex systems. 5(9):801–812, 2025. ISSN 2662-8457. doi: 10.1038/s43588-025-00858-x. URL https://doi.org/10.1038/s43588-025-00858-x.

Andy Xu, Rohan Desai, Larry Wang, Gabriel Hope, and Ethan Ritz. Plaid++: A preference aligned language model for targeted inorganic materials design, 2025. URL https://arxiv.org/abs/2509.07150.

Han Yang, Chenxi Hu, Yichi Zhou, Xixian Liu, Yu Shi, Jielan Li, Guanzhi Li, Zekun Chen, Shuizhou Chen, Claudio Zeni, Matthew Horton, Robert Pinsler, Andrew Fowler, Daniel Zügner, Tian Xie, Jake Smith, Lixin Sun, Qian Wang, Lingyu Kong, Chang Liu, Hongxia Hao, and Ziheng Lu. Mattersim: A deep learning atomistic model across elements, temperatures and pressures, 2024. URL https://arxiv.org/abs/2405.04967.

Claudio Zeni, Robert Pinsler, Daniel Zügner, Andrew Fowler, Matthew Horton, Xiang Fu, Zilong Wang, Aliaksandra Shysheya, Jonathan Crabbé, Shoko Ueda, Roberto Sordillo, Lixin Sun, Jake Smith, Bichlien Nguyen, Hannes Schulz, Sarah Lewis, Chin-Wei Huang, Ziheng Lu, Yichi Zhou, Han Yang, Hongxia Hao, Jielan Li, Chunlei Yang, Wenjie Li, Ryota Tomioka, and Tian Xie. A generative model for inorganic materials design. 639(8055):624–632, 2025. ISSN 1476-4687. doi: 10.1038/s41586-025-08628-5. URL https://doi.org/10.1038/s41586-025-08628-5.

A Appendix

A.1 Sensitivity of OOD@α to kNN and embedding dimension

OOD@α is implemented with an intentionally simple kNN distance detector in a PCA-reduced embedding space (Section 4), so it is important to check that our findings are not tied to a single choice of neighborhood size k or retained dimension. We therefore recompute OOD@0.05 while varying $k \in \{5, 10, 20\}$ at fixed $\dim = 128$ and varying $\dim \in \{64, 256\}$ at fixed $k = 20$ (Table A1).

Across all settings, the qualitative picture is unchanged: RL-based models consistently exhibit substantially higher OOD rates than CFG baselines, and model merging preserves the elevated OOD@0.05 of independently fine-tuned policies. Specifically, RL-MatterGen and RL-merged-8 remain in a tight high range (0.896–0.943), whereas CFG-MatterGen stays moderate (0.190–0.217) and CFG-DiffCSP remains near zero (0.027–0.050). While smaller k slightly increases OOD@0.05 for RL models (more local outlier sensitivity) and changing the PCA dimension shifts distances slightly, these effects do not alter the ranking nor the gap between RL and CFG. This stability supports interpreting OOD@α as a robust indicator of sample-level departure from known superconductors, rather than a brittle consequence of a particular k or embedding dimension choice.

Table A1: Robustness of OOD rate (threshold 0.05) to hyperparameters. The common baseline ($k = 20$, $\dim = 128$) is shown explicitly, with sensitivity to k and PCA dimension analyzed independently.

	Baseline	Vary k ($\dim = 128$)		Vary $\dim$ ($k = 20$)	
	$k = 20,\ \dim = 128$	$k = 5$	$k = 10$	$\dim = 64$	$\dim = 256$
CFG-DiffCSP	0.027	0.027	0.030	0.030	0.050
CFG-MatterGen	0.212	0.190	0.209	0.202	0.217
RL-MatterGen	0.922	0.942	0.939	0.880	**0.906**
RL-merged-8	**0.928**	**0.943**	**0.940**	**0.899**	0.896

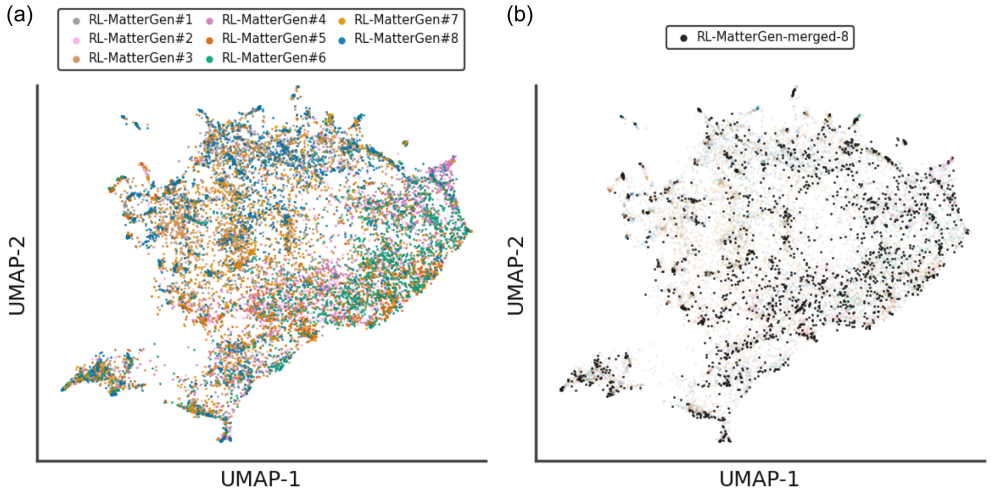

Figure A1: UMAP projection of RL-fine-tuned and merged models. UMAP embedding of structures generated by (a) eight independently RL-fine-tuned MatterGen models, showing overlapping yet complementary coverage of the generative space, and (b) by the merged model (RL-MatterGen-merged-8), showing that merging integrates samples from multiple RL runs while preserving overall distributional coverage.

A.2 TECHNICAL DETAILS: STRUCTURE EMBEDDING ANALYSIS

For visualization, we use Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2020) to further project the processed 128-dimensional embeddings to two dimensions. These embeddings are obtained by concatenating the outputs of the three message passing layers of the UPET-OAM-L model (Mazitov et al., 2025), followed by standardization, PCA reduction to 128 dimensions, and whitening computed on the fit split of the superconductivity structure dataset (Prakash et al., 2025a)). We set $n_{\text{components}} = 2$, $n_{\text{neighbors}} = 40$, and $\min_{\text{dist}} = 0.05$. To ensure comparability, embeddings from different models or datasets are projected jointly using a single UMAP fit (e.g., independently trained models and the merged model in Fig. A1).

A.3 TECHNICAL DETAILS: AFLOW-XTALFINDER MATCHING PROCEDURE

After relaxation and S.U.N. filtering, candidates are analyzed with AFLOW-XtalFinder (Eckert et al., 2024) using (i) prototype matching to the AFLOW library and (ii) isopointal matching, which identifies structures with equivalent symmetry and Wyckoff occupancy independent of species labels. A candidate is considered *prototype-novel* only if it matches neither a known prototype nor an isopointal crystallographic prototype, thereby avoiding "cheap novelty" from elemental redecoration or near-duplicate reproduction. For m evaluated candidates, we define

$$\begin{aligned} \mathrm{PN} &= \frac{1}{m}\sum_{j=1}^{m} z_j, \\ z_j &= \mathbf{1}[x_j \notin \mathcal{P}_{\mathrm{AFLOW}} \wedge x_j \notin \mathcal{I}_{\mathrm{AFLOW}}]\,. \end{aligned} \tag{10}$$

where z_j is an indicator of prototype novelty for candidate x_j, and $\mathcal{P}_{\mathrm{AFLOW}}$ and $\mathcal{I}_{\mathrm{AFLOW}}$ are the sets of structures assigned a prototype match and an isopointal match by AFLOW-XtalFinder (Eckert et al., 2024), respectively.

A.4 TECHNICAL DETAILS: T_c SURROGATE PREDICTOR

We train a GNN-based surrogate model to predict the raw superconducting critical temperature T_c (K). The model uses a GemNet-dT encoder (Gasteiger et al., 2024) with four message-passing layers (r_c=10 Å, hidden dimension 32 for node/edge features), followed by mean pooling and an MLP readout to produce a scalar prediction $\hat{T}_c$.

The surrogate is trained on the superconductivity structure dataset of Prakash et al. (2025a) (N=7217) using an 85/15 train/validation split. We optimize mean squared error with Adam (batch size 64) for 100 epochs at learning rate 5×10^{-4} and weight decay 5×10^{-4}. The resulting model achieves a validation MAE of 1.59 K and is used as the property-feedback signal for RL fine-tuning.

A.5 TECHNICAL DETAILS: CFG FINE-TUNING

For CFG-DiffCSP, we directly use the pretrained guided-diffusion checkpoint from Prakash et al. (2025a) and generate samples conditioned on $T_c = 20\,\mathrm{K}$. In all CFG experiments we fix $c = 20\,\mathrm{K}$ and set $\gamma = 2$ for consistency, and we condition on T_c only (applied to both continuous structure variables and discrete atom types in the model parameterization).

A.6 TECHNICAL DETAILS: RL FINE-TUNING

We fine-tune MatterGen using the MatInvent (Chen et al., 2025) RL framework with terminal surrogate feedback. Each RL iteration samples 64 candidate structures from the current policy, which are relaxed using MatterSim and filtered by structural validity and S.U.N. criteria. If more than 16 samples pass filtering, a random subset of 16 is retained. Rewards are computed using the surrogate T_c predictor with linear rescaling, followed by a composition-based diversity filter. The top 50% highest-reward samples are selected and augmented with two samples from an experience replay buffer before performing a reward-weighted policy-gradient update with KL regularization.

The experience replay buffer stores up to 100 samples and is refreshed to remove repeatedly generated compositions, following the MatInvent (Chen et al., 2025) protocol. Diversity filtering is

applied at the chemical-system level using a tolerance of $\mathrm{Tol} = 1$ and buffer size $\mathrm{Buff} = 5$, reducing the reward of samples from overrepresented systems.

All updates use the Adam optimizer with learning rate 1×10^{-5}. Gradient accumulation is used with a batch size of 10 and 50 accumulation steps per update. The KL coefficient is fixed at $\lambda = 0.025$. Each RL run is trained for 200 iterations with no early stopping.

A.7 Additional details: model merging

Let θ_{base} denote the pretrained MatterGen parameters and let θ_i be the final checkpoint after RL fine-tuning run $i \in \{1, \ldots, n\}$ (same setup, different seeds). We define the task vector

$$\Delta\theta_i = \theta_i - \theta_{\mathrm{base}}, \tag{11}$$

and form the merged model by uniformly averaging task vectors and adding the result back to the base:

$$\theta_{\mathrm{merged}} = \theta_{\mathrm{base}} + \frac{1}{n} \sum_{i=1}^{n} \Delta\theta_i, \tag{12}$$

with merge scale fixed to λ=1 throughout. For floating-point parameters, this update is equivalent to checkpoint averaging, but the task-vector form makes the base-anchored update explicit and aligns with prior model-merging formulations for diffusion (Biggs et al., 2024; Ma et al., 2025).

A.8 Additional details: S.U.N. metric

Stable structures are defined as having MatterSim-relaxed energies within 0.1 eV/atom above the convex hull of the reference dataset Alex-MP-20, following standard practice and previous MatterGen evaluations (Zeni et al., 2025). *Unique* requires non-duplication within the generated set, discouraging collapse toward repetitive candidates. *Novel* denotes database novelty, i.e., no match in the reference crystal database, computed using the same ordered–disordered structure matching protocol as in Zeni et al. (2025); this notion of novelty is different from the distributional novelty quantified by OOD metrics below. Aggregating these criteria enables direct comparison of models in terms of their ability to generate physically sensible and practically actionable candidate structures.

A.9 Details of first-principles superconductivity validation

A.9.1 Workflow and numerical settings

Superconducting properties were computed using a first-principles approach combining density functional theory (DFT), density functional perturbation theory (DFPT), and the EPW (Electron–Phonon coupling using Wannier functions) code. This methodology closely follows established workflows for predictive superconductivity calculations in chemically diverse materials systems, including high-throughput and machine-learning-guided studies (Poncé et al., 2016; Flores-Livas et al., 2020).

Ground-state electronic structures were obtained within DFT using the Perdew–Burke–Ernzerhof (PBE) generalised gradient approximation for the exchange–correlation functional (Perdew et al., 1996), as implemented in the *Quantum ESPRESSO* suite (Giannozzi et al., 2009; 2017). Norm-conserving pseudopotentials from the PseudoDojo library were employed to ensure accuracy and transferability across chemically diverse compositions (van Setten et al., 2018). Plane-wave basis sets were used with a kinetic energy cutoff of 100 Ry and a charge density cutoff equal to four times the wavefunction cutoff. Brillouin zone integrations were performed using Monkhorst–Pack k-point meshes corresponding to a uniform spacing of approximately 0.04 Å^{-1}, which was verified to yield converged total energies and Fermi surface properties.

Non-self-consistent field (NSCF) calculations were performed on dense k-point meshes consistent with those used for Wannierisation and electron–phonon interpolation. A sufficient number of unoccupied bands was included to accurately describe electronic states within the energy window relevant for electron–phonon coupling, following best practices established in prior EPW studies (Heil et al., 2017).

Phonon frequencies and electron–phonon matrix elements were computed using DFPT as implemented in the ph.x module of *Quantum ESPRESSO* (Baroni et al., 2001). Dynamical matrices were calculated on uniform coarse q-point grids, with convergence thresholds chosen to ensure numerical stability of phonon dispersions and electron–phonon coupling constants. These calculations provide the input required for subsequent Wannier-based interpolation of phonon and electron–phonon quantities.

Maximally localised Wannier functions (MLWFs) were constructed using the *Wannier90* code (Mostofi et al., 2008; 2014) to represent the low-energy electronic structure. The choice of initial projections and disentanglement windows was guided by the orbital character near the Fermi level, following established protocols for reliable electron–phonon interpolation [11](Giustino, 2017).

Electron–phonon coupling matrix elements were computed and interpolated from coarse k- and q-point grids onto ultra-fine meshes using the EPW code(Poncé et al., 2016; Flores-Livas et al., 2020). This Wannier–Fourier interpolation approach enables accurate Brillouin zone sampling at manageable computational cost and has been extensively validated for superconductivity calculations in both conventional and unconventional materials (Flores-Livas et al., 2020; Heil et al., 2017).

A.9.2 SUPERCONDUCTIVITY CALCULATIONS AND T_c EXTRACTION

Superconducting properties were evaluated within the Migdal–Eliashberg formalism as implemented in the EPW superconductivity module. The anisotropic, multi-band Eliashberg equations were solved self-consistently on the imaginary frequency axis (Matsubara frequencies), allowing for momentum-resolved superconducting gap structures. Analytic continuation to the real frequency axis was performed using Padé approximants, following standard practice in first-principles superconductivity studies (Margine & Giustino, 2013).

The screened Coulomb interaction was treated within the Morel–Anderson formalism through the Coulomb pseudopotential μ^*, which was set to 0.10. This value is commonly adopted for conventional phonon-mediated superconductors and has been shown to yield reliable critical temperature estimates across a wide range of materials (Morel & Anderson, 1962). Given the exploratory nature of the present study and the chemically diverse set of candidate materials generated via machine learning, a uniform μ^* was used to facilitate consistent comparison between systems. The superconducting critical temperature T_{cwas} determined using two complementary approaches. First, an approximate estimate was obtained from the McMillan–Allen–Dynes formula (McMillan, 1968; Allen & Dynes, 1975), as implemented in EPW, using the calculated electron–phonon coupling constant λ and the logarithmic average phonon frequency ω_{log}. This approach provides a useful baseline for assessing trends across materials and is widely used in high-throughput superconductivity screening studies (Flores-Livas et al., 2020).

Second, a more refined estimate of T_{cwas} obtained from the temperature dependence of the superconducting gap $\Delta(T)$ computed directly from the Eliashberg equations. The gap values were extracted at multiple temperatures and fitted using a power-law form,

$$\Delta(T) = \Delta_0 \left(1 - \frac{T}{T_c}\right)^n \tag{13}$$

where Δ_0 is the zero-temperature gap, T_{cis} the critical temperature, and n is a fitting exponent. Least-squares optimisation was employed, with initial values of $T_{cinformed}$ by the Allen–Dynes estimate. This approach has been shown to improve the reliability of critical temperature determination in systems exhibiting strong electron–phonon coupling or anisotropic superconducting gaps (Margine & Giustino, 2013).

All calculations were performed using the *Quantum ESPRESSO* suite (*pw.x* for electronic structure, *ph.x* for phonons, and *epw.x* for electron–phonon and superconductivity calculations), in conjunction with *Wannier90* for Wannierisation. This workflow follows established EPW-based superconductivity studies and is well suited for validating superconducting properties predicted by machine-learning-driven materials discovery frameworks (Flores-Livas et al., 2020; Pickard et al., 2020).

A.9.3 CRITERIA FOR VALIDATION SET SELECTION

We report T_c using (i) Allen–Dynes–McMillan estimates as a screening baseline and (ii) (an)isotropic Migdal–Eliashberg calculations. Unless otherwise noted we use a uniform Coulomb pseudopotential $\mu^* = 0.10$ and report sensitivity over a small range in μ^* in the appendix. To ground the generative results in first-principles evidence, we perform DFT, density-functional perturbation theory (DFPT), and electron-phonon Wannier (EPW) calculations on a small set of high-scoring structures sampled from the RL-MatterGen models. Candidate materials were prioritized by surrogate-predicted T_c: we rank the full set of generative outputs in descending order of T_c and select from the top of this list. To keep the EPW workflow tractable, we further restrict to structures with ≤ 12 atoms in the primitive cell, since the cost of plane-wave DFT, phonons, and electron–phonon coupling increases rapidly with system size. This case study therefore provides concrete exemplars of what the model discovers, while testing whether its highest-surrogate-predicted T_c proposals remain competitive under explicit electron–phonon calculations.