Diffusion-Based Discovery of Ordered Superconductors

Andrew Valentini[1], Pawan Prakash[1], Richard G. Hennig[1,2]

[1]*Department of Physics, University of Florida*
[2]*Department of Materials Science and Engineering, University of Florida*

Abstract

AI-driven materials discovery pipelines have consistently generated theoretically stable structures which turn out to be disordered solid solutions upon synthesis. This disconnect between computational predictions and experimental realization has been especially prominent in recent attempts in the generation of novel superconductors. To address this, we fine-tune a DiffCSP foundation model on 7,183 superconductors labeled with crystallographic disorder probability, and employ classifier-free guidance to steer generation towards ordered superconductors. We show that generated structures have an 88% chance of having a lower disorder probability than randomly selected structures from an unguided baseline. Candidates are then screened through a multi-stage pipeline which employs machine learning and density functional theory calculations to assess disorder probability, thermodynamic stability, and superconducting properties, with selected structures advancing to experimental synthesis.

Crystalline Disorder & Synthesizability Gap

- Databases used to train ML models, such as the Materials Project, only contain structures with single-occupancy, ordered sites
- Generative models, trained on these databases, implicitly represent candidates as ordered structures
- The A-Lab autonomous synthesis campaign Szymanski *et al.* (2023) claimed 43 novel materials, two-thirds of which were subsequently shown by Leeman *et al.* (2024) to be disordered versions of already-known compounds

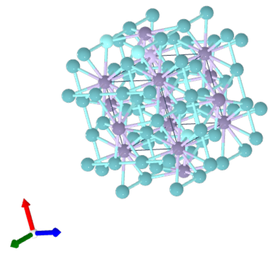

Figure 1. Ordered A15 Nb_3Sn structure

Substitutional Disorder

For an ordered crystal represented by $(\boldsymbol{a}_i, \boldsymbol{f}_i)$ where $\boldsymbol{a}_i \in \{0,1\}^D$ is the one-hot encoding of atomic type and $\boldsymbol{f}_i$ are the fractional coordinates for each site i, **substitutional disorder** replaces $\boldsymbol{a}_i$ with an occupancy vector $\boldsymbol{s}_i = (s_{i,0}, s_{i,1}, \ldots, s_{i,D-1})$ where $s_{i,j}$ is the probability of finding element type j at site i.

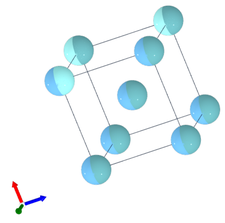

Figure 2. Disordered BCC TiNb structure: Ti and Nb share a single site with 50% fractional occupancy each

Disorder in AI-Driven Superconductor Discovery Pipelines

Discrepancy between prediction and experiment was especially pronounced in a recent superconductor discovery pipeline when all synthesized candidates formed as disordered solid solutions, demonstrated by Fig. 3's enhanced fit of a disordered BCC solution over the model's predicted structure

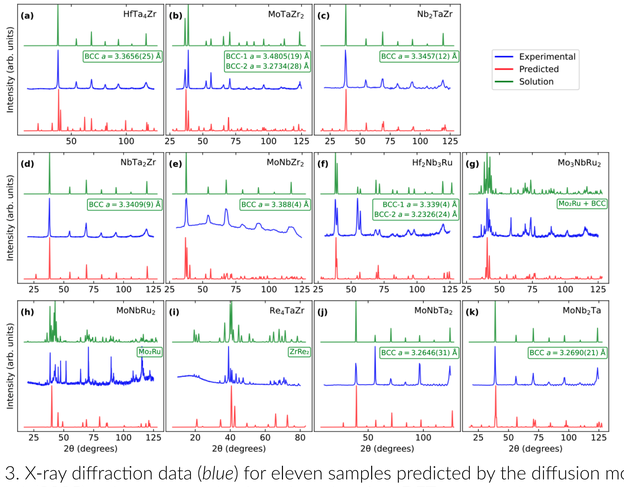

Figure 3. X-ray diffraction data (*blue*) for eleven samples predicted by the diffusion model in Prakash *et al.* (2025) to be superconducting, plotted against DFT-computed spectra computed for the predicted structures (*red*) and disordered BCC structure solutions to each spectra (*green*)

Workflow

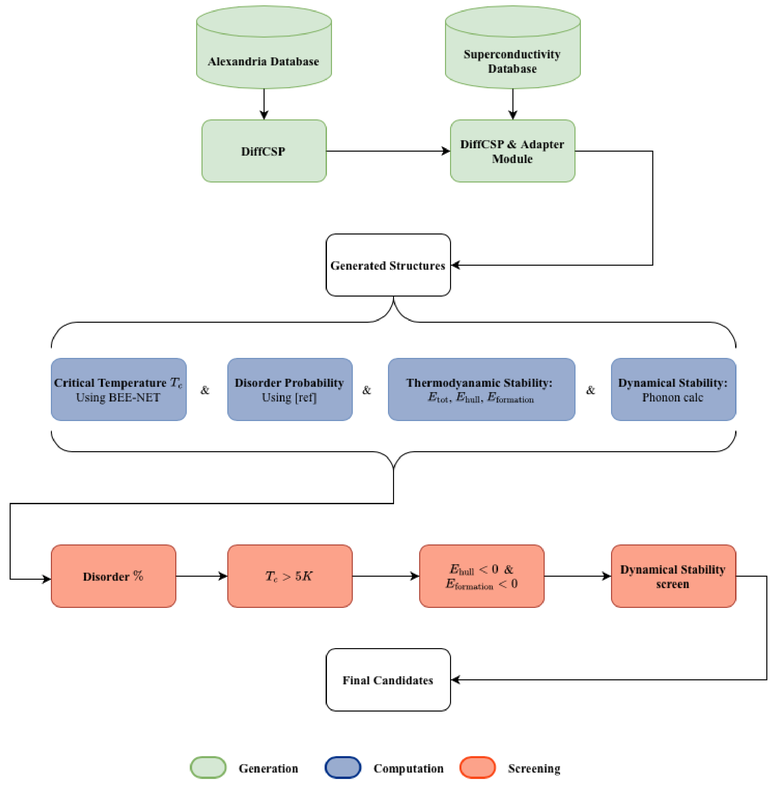

Figure 4. Workflow for diffusion-based discovery of ordered superconductors. Note that all computations are not necessarily carried out before screening as indicated, as this figure is not intended for exact implementation accuracy.

Generation Layer

Foundation Model

- DiffCSP jointly diffuses over lattice parameters $\boldsymbol{\ell}$, fractional coordinates $\boldsymbol{f}$, and atomic types $\boldsymbol{a}$ using an SE(3) equivariant GNN
- Pretrained on $> 2 \times 10^6$ structures from Alexandria database, learning broad prior over plausible crystal structures

Adapter Module

- Adapter is fine-tuned on DiffCSP frozen weights x_t according to scalar target c
- During inference, CFG steers towards desired c conditioned by weight w:

$$\hat{\epsilon}_\theta(x_t, c) = \epsilon_\theta(x_t, \varnothing) + w\big(\epsilon_\theta(x_t, c) - \epsilon_\theta(x_t, \varnothing)\big)$$

Computation Layer

Disorder Probability Predictions

- Computed using classification model developed by Jakob *et al.* (2025)
- Trained on the ICSD, model uses the composition of a material to estimate the presence of *substitutional* disorder with a reported accuracy of 90%

T_c Predictions

- Computed using a Bootstrapped Ensemble of Equivariant Neural Networks (BEE-NET) which learns the Eliashberg spectral function $\alpha^2F(\omega)$ from 7,000 superconductors

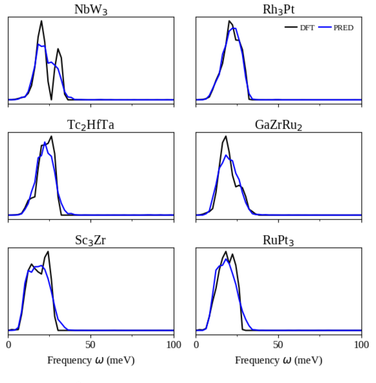

Figure 5. $\alpha^2F(\omega)$ functions predicted by BEE-NET compared to DFT calculations

E_{hull} Predictions

- Relaxed structure and E_{tot} computed using UMA MLIP
- E_{hull} estimated by comparing E_{tot} to competing stable phases found on the LeMaterial database.

Learning Disorder

We find that our fine-tuned diffusion model *is* able to generate superconducting candidates which are likely to be ordered. The distributional shift towards ordered materials after fine-tuning is applied is show in Fig 6.

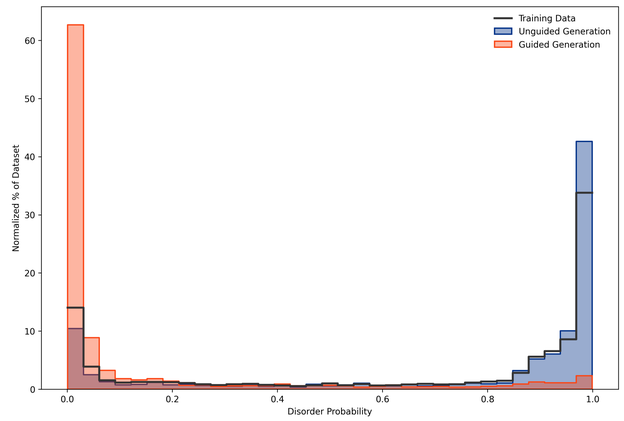

Figure 6. Histograms comparing the disorder probability distributions for the fine-tuning training data, generated materials without any guidance, and generated materials with disorder guidance.

- We apply a Mann-Whitney U test to quantify the distributional shift towards ordered superconductors, finding an 88% chance of randomly drawing a more ordered candidate from the fine-tuned data with respect to an unguided base
- We confirm that an unguided parameter, T_c, follows the training data distribution in Fig. 7

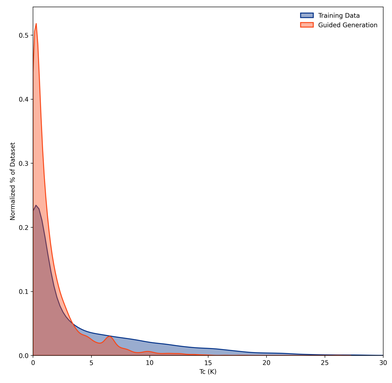

Figure 7. Confirming that the distribution of guided structures' predicted T_c values follows that of the exponential decay found in the training data

Conclusions & Future Work

- We demonstrate that a diffusion model is capable of being steered towards the generation of ordered superconductors
- Experimental validation of candidates which have passed all screens is ongoing
- Upon experimental validation of diffusion-based generation of ordered superconductors, we intend to exploit the high T_c region of training data with reinforcement learning

Acknowledgments

This work is supported by the U.S. National Science Foundation under Grant DMREF-2118718. Computational resources were provided by the University of Florida HiPerGator high-performance computing system.

References

N. J. Szymanski, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, M. J. McDermott, M. Gallant, E. D. Cubuk, A. Merchant, *et al.*, Nature **624**, 86 (2023).

J. Leeman, Y. Liu, J. Stiles, S. B. Lee, P. Bhatt, L. M. Schoop, and R. G. Palgrave, PRX Energy **3**, 011002 (2024).

P. Prakash, J. B. Gibson, Z. Li, G. D. Gianluca, J. Esquivel, E. Fuemmeler, B. Geisler, J. S. Kim, A. Roitberg, E. B. Tadmor, M. Liu, S. Martiniani, G. R. Stewart, J. J. Hamlin, P. J. Hirschfeld, and R. G. Hennig, Guided diffusion for the discovery of new superconductors (2025), arXiv:2509.25186 [cond-mat.supr-con] .

K. Jakob, A. Walsh, K. Reuter, and J. T. Margraf, ChemRxiv **2025**, 10.26434/chemrxiv-2025-f52qs (2025), https://chemrxiv.org/doi/pdf/10.26434/chemrxiv-2025-f52qs .