CSX Framework for Synthesis-Oriented Generative Materials Discovery

Shuya Yamazaki[1] **Zekun Ren**[2] **Martin Hoffmann Petersen**[2] **Dai Haiwen**[1] **Kedar Hiappalgaonkar**[1,2,3]
[1]*School of Materials Science and Engineering, Nanyang Technological University, Singapore 639798*
[2]*Berkeley Education Alliance for Research in Singapore, CREATE Tower, Singapore 138602*
[3]*Institute for Functional Intelligent Materials, National University of Singapore, Singapore 117544*

Correspondence to: Kedar Hiappalgaonkar kedar@ntu.edu.sg

1. Introduction

Property-directed generative design remains the holy grail of materials science, yet a formidable gap persists between *in silico* discovery and experimental realization. This disconnect arises because existing generative models often treat synthesis as a post-hoc validation rather than a foundational constraint, neglecting the nuanced realities of the **wet-lab environment** [1]. In practice, discovery is bounded by precursor accessibility, which governs feasible kinetic pathways. Experimental protocols are also highly system-specific, making broad, "scattershot" generative exploration across the periodic table logistically inefficient and costly for focused experimental campaigns. Moreover, real synthesis cannot achieve perfect structural control [2]. Functional success therefore may depend not on isolated point solutions, but on local phase regions where properties remain robust against deviations in stoichiometry, structure, or disorder. Actionable discovery thus requires identifying such property-stable local phase regions within chemical systems.

To bridge this gap, we introduce the **Chemical System eXploration (CSX) framework**, which reframes materials discovery around wet-lab design constraints. CSX shifts the objective from isolated point-structure generation to the systematic mapping of precursor-constrained chemical systems via three integrated components: the **CSX-Generator**, a synthesis-oriented generative backbone for sampling within precursor-defined element sets; the **Energy Oracle**, which evaluates thermodynamic stability; and the **Property Oracle**, which assesses target functional properties, each using multi-fidelity verifiers. By mapping the energy-property landscape of a given chemical system (**Figure 1**), CSX transforms abstract structural predictions into experimentally actionable insights. This approach reveals robust basins where target properties remain stable against experimental deviations, thereby identifying the chemical systems and their local phase space where high-cost simulations and experimental efforts should be prioritized under the precursor constraints. Ultimately, the CSX framework rethinks generative materials discovery from first principles, establishing a new paradigm that anchors computational exploration within the realities of experimental design.

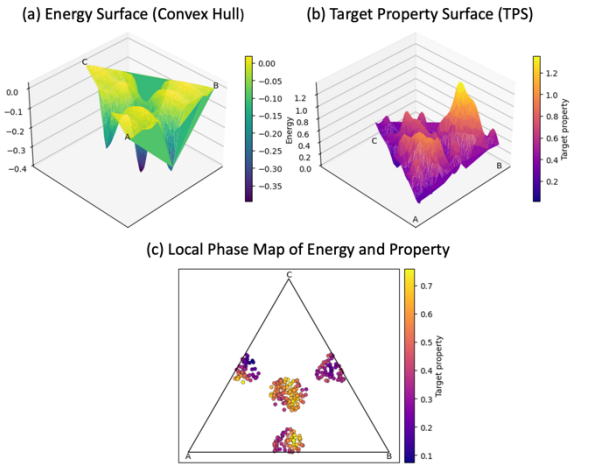

Figure 1: CSX framework maps local phase space of energy and property. (a) The Energy Oracle constructs the thermodynamic energy surface of a given chemical system; (b) The Property Oracle predicts the corresponding target-property landscape; (c) Their integration yields a local phase map jointly describing energy and property within the chemical system.

2. Results and discussion

2.1 Related work

Few existing models support element-conditioned generation tailored to specific precursor sets. Although MatterGen introduces soft conditioning at the chemical-system level, it does not provide a holistic, synthesis-oriented framework [1]. Most current efforts to bridge the synthesis gap therefore rely on post-hoc screening, using synthesizability classifiers or precursor predictors to filter candidates after generation [3], [4]. Such strategies treat experimental feasibility as an external constraint, rather than as a core principle of the generative framework. In contrast, the CSX framework restructures this hierarchy by placing wet-lab constraints at the foundation of the pipeline.

2.2 Objective

As a testbed for the CSX framework, we target lattice thermal conductivity (K_{lat}), motivated by the availability of a high-fidelity computational pipeline and curated dataset for this property [5]. The objective of this work is to discover and synthesize novel compounds with ultra-low lattice thermal conductivity using CSX under laboratory-accessible precursor constraints, thereby demonstrating synthesis-oriented, property-driven materials discovery.

2.3 CSX Generator

We adopt **WyFormer** as the generative backbone of the CSX framework [6], trained on the LeMat-Bulk dataset. **CSX-WyFormer** extends this model to enable inference-time generation under precursor-constrained element sets, allowing targeted sampling within specific chemical systems. As validation, CSX-WyFormer rediscovered 7 of 14 held-out ICSD compounds with calculated K_{lat} below 0.5 W m^{-1} K^{-1} under precursor constraints, whereas 100k unconditional generations recovered none. This demonstrates the importance of chemical-system conditioning for synthesis-oriented generative discovery.

2.4 Energy Oracle

The Energy Oracle operates across multiple fidelity levels to balance accuracy and computational cost. At the high-throughput evaluation stage, generated structures are pre-relaxed and evaluated using the Orb-v3 model. Finite-temperature (FT) free energies are then computed for promising chemical systems using TensorNet trained on the MatPES-r2SCAN dataset, followed by DFT refinement. Stability is assessed against convex hulls referenced to the LeMat-Bulk dataset.

2.5 Property Oracle

The Property Oracle similarly integrates multi-fidelity evaluations. Surrogate models trained on the Phonix database are used for rapid prediction of K_{lat}, which will be reported in separate work. Higher-accuracy evaluations are performed using eSEN and DFT within the AutoKappa workflow, which solves the Boltzmann transport equation (BTE) to compute K_{lat} [5].

2.6 CSX Validation

CSX was validated using a precursor-constrained element set (Ag, Bi, Ge, Al, Na, Co, Cu, Ni, S, Sb, Se, Sn, Te, Fe, Ti) with a two-stage strategy: broad exploration using CSX-WyFormer with LeMat-Bulk sampling, followed by focused generation within selected chemical systems. Systems were ranked by phase-space exploration, robustness of low K_{lat}, with metastability thresholds applied, using the Energy and Property Oracles. After expert screening, Bi–Na–Sn was selected as one of the pilot systems. It exhibits robust low K_{lat} phase landscapes while remaining largely underexplored in ICSD, with no reported ternary compounds to date. **Figure 2** shows the energy and K_{lat} phase maps of the Bi–Na–Sn system, with the Sn–NaBi pseudo-binary line highlighted. Since NaBi is experimentally known, this provides a realistic synthesis pathway, along which predicted K_{lat} remain low and metastable energy minima are observed. FT TensorNet free-energy calculations confirm one candidate on the convex hull. These results demonstrate the capability of CSX to identify experimentally accessible low K_{lat} candidates within unexplored chemical systems. Future work will extend this validation through experimental synthesis and broader system coverage.

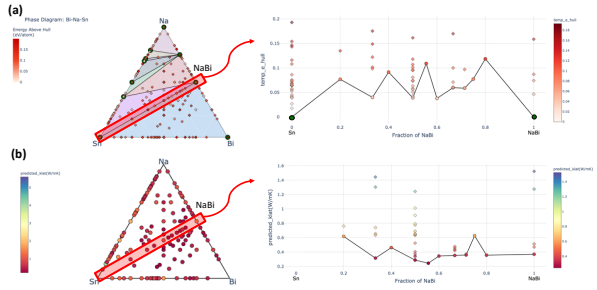

Figure 2: Energy and property phase maps of the Bi–Na–Sn chemical system. (a) Thermodynamic energy phase diagram and Sn-NaBi pseudo-binary line; (b) Lattice thermal conductivity phase map and Sn-NaBi pseudo-binary line

Acknowledgments

K.H. acknowledges funding from the MAT-GDT Program at A*STAR via the AME Programmatic Fund by the Agency for Science, Technology and Research under Grant No. M24N4b0034.

References

[1] Zeni, C. et al. Nature 639, 624–632 (2025).
[2] Szymanski, N.J. et al. Nature 624, 86–91 (2023).
[3] Zhu, R. et al. ACS Omega 8, 8210–8218 (2023).
[4] Prein, T. et al. arXiv:2502.04289 (2025).
[5] Ohnishi, M. et al. arXiv:2504.21245 (2025).
[6] Kazeev, N. et al. arXiv:2503.02407 (2025).