

MEIDNet: multimodal generative AI framework for inverse materials design

Received: 7 January 2026

Accepted: 20 May 2026

Cite this article as: Babu, A., Gouvêa, R.A., Vanderghenst, P. *et al.* MEIDNet: multimodal generative AI framework for inverse materials design. *npj Comput Mater* (2026). <https://doi.org/10.1038/s41524-026-02153-3>

Anand Babu, Rogério Almeida Gouvêa, Pierre Vanderghenst & Gian-Marco Rignanese

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

MEIDNet: Multimodal generative AI framework for inverse materials design

Anand Babu^{1,*}, Rogério Almeida Gouvêa¹, Pierre Vandergheynst², Gian-Marco Rignanese^{1,3,*}

¹Institute of Condensed Matter and Nanosciences, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

²Signal Processing Laboratory 2, Institute of Electrical and Micro Engineering, School of Engineering, EPFL, Lausanne, Switzerland.

³WEL Research Institute, Avenue Pasteur 6, Wavre, Belgium

*Corresponding authors: anand.babu@uclouvain.be; gian-marco.rignanese@uclouvain.be

Abstract

In this work, we present Multimodal Equivariant Inverse Design Network (MEIDNet), a framework that jointly learns structural information and materials properties through contrastive learning, while encoding structures via an equivariant graph neural network (EGNN). By combining generative inverse design with multimodal learning, our approach accelerates the exploration of chemical-structural space and facilitates the discovery of materials that satisfy predefined property targets. MEIDNet exhibits strong latent-space alignment with cosine similarity ≈ 0.96 by fusion of three modalities through cross-modal learning. Through implementation of curriculum learning strategies, MEIDNet achieves ~ 60 times higher learning efficiency than conventional training techniques. The potential of our multimodal approach is demonstrated by generating low-bandgap perovskite structures at a stable, unique, and novel (SUN) rate of 13.6 %, which are further validated by ab initio methods. Our inverse design framework demonstrates both scalability and adaptability, paving the way for the universal learning of chemical space across diverse modalities.

Introduction

The search for materials with desired properties is crucial for various applications, including energy storage, electronics, optoelectronics, and biomedical devices. However,

conventional trial-and-error approaches are resource-intensive and have a limited scope. AI-enabled computational inverse design provides an efficient way to find candidates that satisfy predefined functional targets [1-3]. It exploits learned structure-property relationships to efficiently navigate complex chemical and structural landscapes. This approach significantly accelerates discovery cycles, guiding experimental efforts toward more targeted explorations [4-6].

Generative AI models, including variational autoencoders (VAEs) [7], generative adversarial networks (GANs) [8], and diffusion models [9-10] have demonstrated promising performance in the design and discovery of new materials. However, most of the proposed frameworks rely on a single mode of information, which limits their effectiveness in fully capturing the complex interplay among multiple property dimensions [10-13]. To address these limitations, multimodal machine learning (ML) has gained traction. By incorporating diverse sources of information, such as structural, electronic, mechanical, and thermodynamic properties, it facilitates the creation of a robust chemical latent space through shared learning [14-18].

Recent efforts have broadened the reach of multimodal frameworks by incorporating techniques such as contrastive learning [19], cross-modal attention mechanisms [20], and constrained-driven materials exploration [21]. For instance, the multimodal foundation model (MultiMat) integrates crystal structures, density of states (DOS), charge densities, and textual descriptions to uncover materials through latent-space fusion [22]. The composition-structure bimodal network (COSNet) improves the prediction accuracy for experimentally measured composition and structural data [23]. SCIGEN integrates geometric lattice constraints into a diffusion-based crystal generator to discover stable motif-guided quantum materials and validates a large subset via prescreening and DFT [21]. Similarly, denoising diffusion techniques coupled with cross-modal contrastive learning have enabled the guided discovery of chemical compositions and crystal structures from textual prompts [24]. Despite these initiatives, reliably generating stable, unique, and novel (SUN) materials with targeted properties and precisely navigating the latent space remains difficult. It requires extensive training and tighter alignment across

modalities [15-16, 25]. Multimodality is clearly the future of materials science, yet only a handful of studies have addressed this challenge.

In this work, we present a multimodal generative AI framework for inverse materials design named MEIDNet, which relies on contrastive learning among three modalities: structural, electronic, and thermodynamic properties. Structural encoding with a state-of-the-art equivariant graph neural network (EGNNs) is tested on three diverse datasets: Perovskite-5, MP-20, and Carbon-24. Latent-space alignment is evaluated for both early and late fusion strategies using the Information Noise-Contrastive Estimation (InfoNCE) loss function. We use the Perovskite-5 dataset (~19k five-atom ABX_3 structures) as a benchmark for our multimodal model. It is simple, widely used, and includes technologically relevant classes of materials, such as photovoltaics, ferroelectrics, and high- κ dielectrics. The cosine similarity and the L2 distances between modality embeddings are ~0.96 and ~0.24, respectively, indicating strong alignment. As a demonstration, 140 perovskite crystal structures are generated targeting thermodynamically stable materials in the low bandgap range (from 0.8 to 1.5 eV). The predicted bandgaps for the generated crystals closely match those determined with the crystal graph convolutional neural network (CGCNN) for single-property prediction (with an MAE of ~0.02). However, subsequent ab initio calculations show a lowering of the prediction. Out of the 140 generated perovskites, 19 are found to be stable, unique and novel, leading to a calculated SUN rate of ~13.6% without any further filtering, which is state of the art for multimodal models for materials to the best of our knowledge. Thus, our multimodal inverse design framework advances multimodal material generation efficiency and shows promise for scalability and adaptability by employing EGNNs for structural encoding and evaluating different fusion strategies. It leads the way toward universal learning of the chemical-structural space, accelerating the discovery of materials with desired properties, and propelling multimodal inverse design in materials science.

Results

E(3)-equivariant graph neural network and Multimodal framework

The crystal encoder transforms 3D crystal structures into an embedded latent representation using an EGNN for the structural encodings [26-27]. It is implemented by message passing [Eq. (i)], feature aggregation/update [Eq. (ii)], and coordinate update [Eq. (iii)]. It is equivariant to translations, rotations/reflections, and node permutations [28].

$$m_{ij} = \varphi_e \left(h_i^l, h_j^l, \|x_i^l - x_j^l\|^2, a_{ij} \right) \quad (1)$$

$$m_i = \Sigma_{j \neq i} m_{ij}; h_i^{(l+1)} = \varphi_h(h_i^l, m_i) \quad (2)$$

$$x_i^{(l+1)} = x_i^l + C \cdot \Sigma_{j \neq i} (x_i^l - x_j^l) \cdot \varphi_x(m_{ij}) \quad (3)$$

$$z_s = \frac{1}{N} \Sigma_{i=1}^N h_i^{(L)}(m_{ij}) \quad (4)$$

Here $x_i^l \in \mathbb{R}^n$ are node coordinates, $h_i^l \in \mathbb{R}^d$ node features, a_{ij} edge attributes; m_{ij} the message (edge embedding) sent from node j to node i , $\varphi_e, \varphi_x, \varphi_h$ multilayer perceptrons (MLPs), and C a normalization constant. Finally, Eq. (iv) represents the global pooling operation where $h_i^{(L)}$ are the node features after the final EGNN layer L , N is the number of atoms in the unit cell, and z_s is the resulting structural latent embedding.

The reconstruction fidelity of the autoencoder increases with the latent-space dimensionality, but so does the computational cost. As a result, a balance needs to be found. To assess this, we consider three latent dimensions, namely 64, 128, and 256, using the same number of training epochs on the Perovskite-5 dataset [29] as a benchmark. The 128-dimensional latent space emerges as the sweet spot, yielding a higher structure-matching (SM) rate (see Methods) at roughly the same computational cost (see Supplementary Figure S1). Therefore, from here on, we use a 128-dimensional equivariant crystal autoencoder for multimodal alignment.

To validate the generalizability of our EGNN encoder, we also determine the reconstruction fidelity on two other diverse datasets: MP-20 [30] and Carbon-24 [31]. Thanks to the EGNN integrated autoencoder architecture, MEIDNet outperforms the unimodal Fourier transformed crystal properties (FTCP) [32] and crystal diffusion variational autoencoder (CDVAE) [7] for all three datasets in the SM rate (Table 1).

Table 1. Comparison of the reconstruction performance of MEIDNet, FTCP [32], and CDVAE [7] for the Perovskite-5 [29], MP-20 [30], and Carbon-24 [31] datasets.

Method	Structure-matching rate (%)		
	Perovskite-5	Carbon-24	MP-20
FTCP	99.34	62.28	69.89
CDVAE	97.52	55.22	45.43
MEIDNet	99.85	66.4	72.35

To address the latent search bottleneck in multimodality, we learn an aligned latent representation in which structural (crystal structure), electronic (bandgap), and thermodynamic (formation enthalpy) embeddings co-exist. As summarized in Figure 1a, the structural information is encoded through our EGNN. In parallel, we define a property encoder to map the scalar material properties (bandgap and formation enthalpy) into the shared latent space. This encoder is implemented as a Multilayer Perceptron (MLP) that projects the input property onto a 128-dimensional embedding. This ensures that both structural and property representations possess the same dimensionality, facilitating their alignment via contrastive learning. MEIDNet uses a multimodal autoencoder architecture in which the crystal latent is decoded by an $E(3)$ -equivariant decoder (SE3Decoder) to reconstruct lattice, species, coordinates, and auxiliary adjacency information (Supplementary Figure S2). For detailed discussion see Supplementary Note: Multimodal Generative Framework: From Bi- to Tri-Modality.

For robust joint learning, we investigate the effect of early and late fusion approaches on multimodal alignment (Figure 1b, more details are given in Supplementary Note: Early fusion and Late fusion). In early fusion, modality-specific features are merged at the input feature level, and a shared network learns a joint representation. In late fusion, each modality is encoded separately, and the resulting embeddings are combined at the alignment stage. Contrastive learning unifies structural and property encodings in a joint latent space, which facilitates interactions between distinct modalities and optimizes the alignment of their information [19-20].

The alignment between modalities is achieved via contrastive training using the InfoNCE loss [33-34]

$$\mathcal{L}_{\text{InfoNCE}} = -\frac{1}{B} \sum_{k=1}^B \log \frac{\exp(\text{sim}(z_s^{(k)}, z_p^{(k)})/\tau)}{\sum_{l=1}^B \exp(\text{sim}(z_s^{(k)}, z_p^{(l)})/\tau)} \quad (5)$$

where B is the batch size, and indices (k) and (l) denote samples within the mini batch. $z_s^{(k)}$ and $z_p^{(k)}$ are the aligned structural and property embeddings for the k -th crystal, $\text{sim}(u, v) = u^T v / (\|u\| \|v\|)$ denoting cosine similarity. τ is the temperature hyperparameter, and the denominator sums over all l samples in the batch to normalize the probability.

We implement an inverse design pipeline for target-led navigation in the aligned latent space. An iterative optimization is performed until the predicted properties converge to the targeted values for the generated crystal structure (Figure 1c), detailed information is provided in Supplementary Note: Latent-space optimization objective, solver configuration, and robustness safeguards. Information. Thus, MEIDNet offers a more compact latent space than diffusion models, facilitating interpretability and navigation. In addition to this, it is scalable to numerous modalities toward a unified latent representation. While InfoNCE implicitly assumes unique positive pairs, the risk of false negatives arising from distinct structures with similar properties in the same batch is reduced in our framework by the small batch size ($B = 16$), the use of a continuous multi-property representation rather than a single scalar, and the concurrent MSE-based property loss, which anchors the latent space to physically consistent targets. In addition, batch-size stress tests over $B = 16, 64, 128$ yielded comparable held-out latent property-coherence metrics (Supplementary Figures S3 and S4).

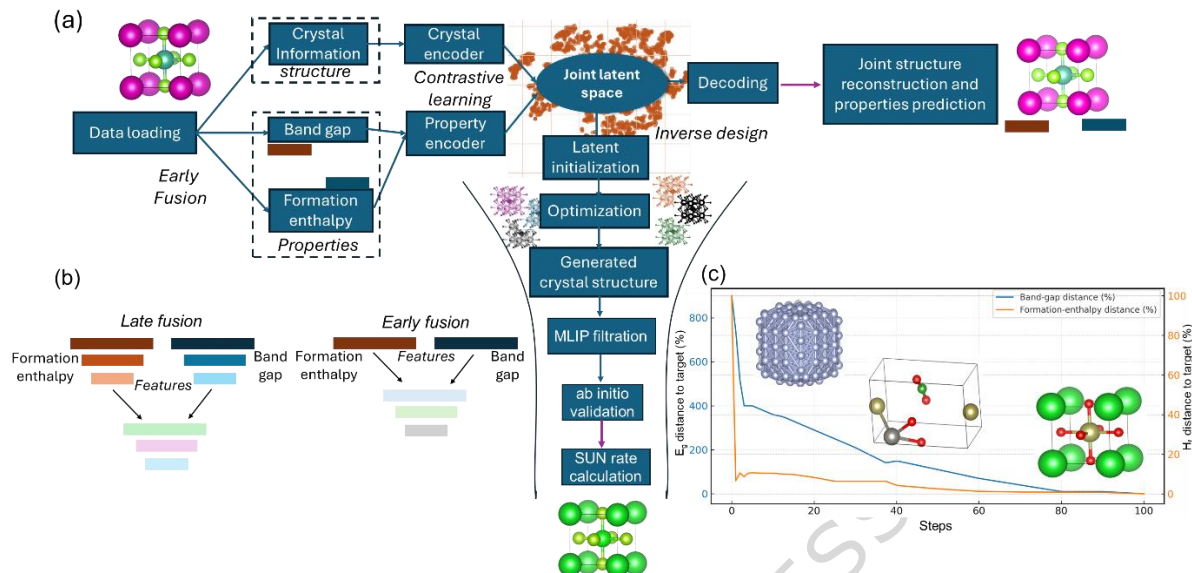


Figure 1. Workflow schematics of the MEIDNet framework for inverse materials design: (a) The process starts with data loading, extracting crystal information and corresponding properties (here, the bandgap and formation enthalpy) which are fused into a joint latent space through contrastive learning, where iterative optimization takes place with respect to the target properties. The optimized latent representations are then decoded into crystal structures, which are then filtered using MLIPs and validated ab initio. (b) Different fusion approaches are considered: early fusion merges modality-specific features at input so a shared network learns a joint representation, whereas late fusion encodes modalities separately and combines their embeddings during alignment. (c) Generation of structures through navigation on the aligned latent space, where the y-axis represents the distance from the target properties.

Curriculum learning-enabled training and multimodal fusion strategies

In order to improve the learning efficiency, we employ a progressive weighting on the contrastive loss: its weight α_t is ramped up from 0 to 1 over the first two thirds of training epochs and then held at 1, improving reconstruction, latent alignment, and convergence [35-36]. The pacing follows an exponential schedule:

$$\alpha_t = 1 - \exp(-\gamma t) \quad (6)$$

where t is the epoch index and γ controls the ramp rate. This training strategy leads to 60-times faster learning with respect to traditional training (Figures 2a and 2b). The 60-times faster learning refers to optimization/sample-efficiency at a fixed epoch budget, with Figure 2b comparing curriculum learning (CL) and conventional training (CT) at the same number of training epochs. Thus, the reported advantage means higher structure-matching performance is reached within the same training budget, enabled by an epoch-

dependent curriculum that progressively increases the contrastive-loss weight during training.

We thoroughly investigate the effect of different fusion strategies on the latent space alignment: late fusion (LF), early fusion (EF), and early fusion with curriculum learning (EF+CL). We adopt three key metrics: cosine similarity (CS), structure-matching (SM) rate, and L2 distance (L2D) on the scale of 0-1, 0-100 % and 0-2, respectively. To determine the best strategy, we train MEIDNet on the Perovskite-5 dataset using 200 epochs. The results are illustrated in Figure 2a. LF demonstrates a good SM performance (~66%) but performs badly in latent space alignment (CS~0.31, L2D~1.06). In contrast, EF displays a good alignment (CS~0.89, L2D~0.45), but a lower SM performance (~32%). This can likely be explained by the premature mixing of heterogeneous descriptors causing feature interference/over-smoothing, which yields a measurable drop versus LF, especially in reconstruction fidelity. EF+CL outperforms the previous two fusion strategies: it has a SM performance (~66%) similar to LF and a better alignment (CS~0.91, L2D~0.4) than EF. This promising performance is due to the initial stage of training CL giving more weight to learning the crystal structures, while their corresponding properties are scalars, which are easier to learn at a later stage [37]. Cosine similarity is used here only as an auxiliary latent-alignment diagnostic and not as a standalone success metric. Model performance is therefore evaluated primarily through task-grounded measures, including structure matching, conditional generation outcomes, stability/novelty screening, independent property prediction, and first-principles validation.

For production, MEIDNet is then trained on the same Perovskite-5 dataset but using 2000 epochs and adopting the EF+CL strategy. It demonstrates a high cosine similarity (CS~0.97), which suggests that the shared latent space effectively captures the inherent relationships between structural motifs and their corresponding properties [38-39]. In addition to the cosine similarity, the latent representations exhibit an average L2 distance (L2D) of ~0.24, which indicates that the multimodal embeddings are closely aligned in the latent space [40]. To assess joint learning, we perform regression analyses showing that the jointly reconstructed SM score (Figure 2c), bandgap (E_g) (Figure 2d), and formation enthalpy (H_f) (Figure 2e) track the ground truth with near-linear trends ($R^2 \approx 0.996$). This

consistency indicates that the shared representation captures cross-modal correlations and reduces prediction variance. Figures 2d and 2e demonstrate that the shared latent space preserves sufficient information to reconstruct bandgaps and formation enthalpy in the held-out test set from structure-conditioned embeddings in the Perovskite-5 dataset. These results should, therefore, be interpreted as evidence of joint reconstruction consistency and latent-space alignment. To complement this in-dataset reconstruction evidence, Supplementary Table S1 reports an external structure-only evaluation on 46 unseen perovskites using the frozen crystal encoder pathway without retraining or fine-tuning. Thereby providing a more direct assessment of transferable structure-property information beyond the Perovskite-5 training distribution.

ARTICLE IN PRESS

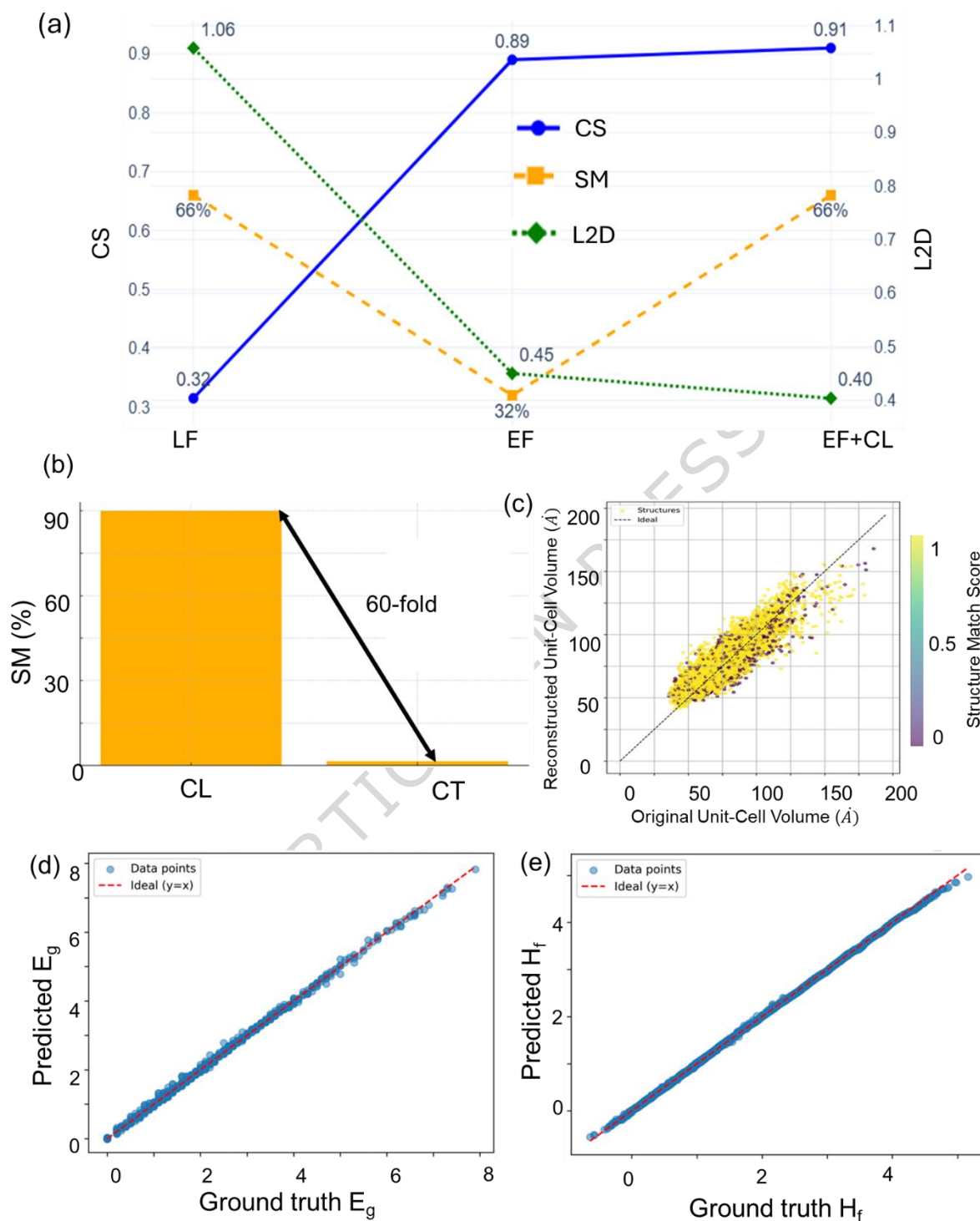


Figure 2. Evaluation of multimodal machine learning model efficiency and joint reconstruction accuracy: (a) Comparative evaluation of multimodal machine learning performance metrics **on the test set**: Performance analysis of late fusion (LF), early fusion (EF) and EF with curriculum learning (EF+CL). The performance metrics are through cosine similarity (CS), structure matching (SM), L2 distance (L2D) and R^2 metrics, which highlights the advantages of EF+CL in enhancing latent alignment and reducing prediction errors (b) 60-fold improvement in training efficiency with curriculum learning (CL) versus conventional training (CT), as evidenced by higher SM (%) at a fixed number of epochs. (c) Scatter plot

demonstrating structure reconstruction and structure matching score prediction of (d) bandgap (E_g) and (e) formation enthalpy (H_f) for MEIDNet training.

Inverse design-led targeted structure generation

To demonstrate the inverse-design pipeline, we test MEIDNet for generating perovskite structures (details provided in Methods). To evaluate target-led generation, we perform conditional sampling in the low bandgap range ($E_g \in \{0.8, 1.0, 1.2, 1.5\}$ eV) under a fixed negative formation enthalpy of -25 meV. For each target, a batch of 35 candidate materials are produced by the model. To reduce sensitivity to local minima in the non-convex latent landscape, we use a multi-start strategy in which 16 independently initialized latent vectors are optimized in parallel for each target, followed by validity, relaxation, and filtering steps to identify viable candidates (see Supplementary Note: Physical-Constraint Objective Used During Latent-Space Optimization for more details).

The thermodynamic stability of the generated candidates is first assessed using the eSEN-30M-OAM machine learning interatomic potentials developed by Facebook AI Research (FAIR) [41] at Meta (Figure 3a). One of the key metrics to quantify the success of generative models in materials science is the fraction of stable, unique, and novel structures, the so-called SUN rate. We adopt the same conventional thermodynamic-stability criterion used in previous generative models, which consider that structures with energy above hull (E_{hull}) below 100 meV, as calculated from the Materials Project database, are stable [7,9,32,43]. Out of 140 generated structures, 19 meet this criterion and are not present in the training dataset, yielding a SUN rate of approximately 13.6% (Figure 3b, Supplementary Figure S5, Table S2, and S3,). These findings are further confirmed by ab initio calculations using the Vienna Ab initio Simulation Package (VASP) (see Methods). We highlight that, although the model is trained mainly on unstable high E_f perovskites (Figure 2e), it still reliably leads to thermodynamically stable structures.

This places our model at the forefront of inverse-design-based multimodal learning approaches with respect to its counterparts [32, 42]. To the best of our knowledge, our model trained on multiple modalities is state-of-the-art in terms of the SUN rate (without

using any filtration) for multimodal models for materials. It is even competitive with generative models trained on single modality including MatterGen ~39% [9], CDVAE \approx 18% [7], FTCP < 5% [32], and Matra-Genoa~16 % [43]. A comparison table is provided in Supplementary Table S4. Besides SUN, we have also calculated the polymorphism-aware and distribution-level evaluation of crystal reconstruction; detailed information has been provided in Supplementary Note: Polymorphism-aware and distribution-level evaluation of crystal reconstruction [44-48].

Turning to the validation of the target bandgap, we first use the CGCNN predictor and obtain an MAE of \sim 0.02 eV as detailed in the Information (Supplementary Figure S6 and Table S2). This shows that the predictive capabilities of MEIDNet are comparable to those of the leading GNN models for predictions. The bandgaps are also computed at the PBE level (without +U or SOC). The calculated values are often lower than the requested target. This downward shift likely reflects (i) a modest training-set skew toward smaller gaps in the Perovskite-5 dataset, (ii) the expected low coverage in the training set for the novel perovskite compositions, and (iii) the fact that band structures were computed at the PBE level without +U.

In terms of the generated structures, we first would like to highlight four novel structures with a bandgap in the low range and very interesting properties (see Figure 3c): (i) YbScSe_3 , (ii) LaScSe_3 , (iii) BaHfSe_3 and (iv) KTaSe_3 . Their corresponding E_{hull} values are 0, 91.88, 66.78, and 88.6 meV, respectively, when calculated with eSEN and 0, 95.6, 98.7, and 0 meV, respectively, when calculated within DFT. Other promising perovskites found using our workflow are presented in the Supplementary Figures S8, S9, and Table S3).

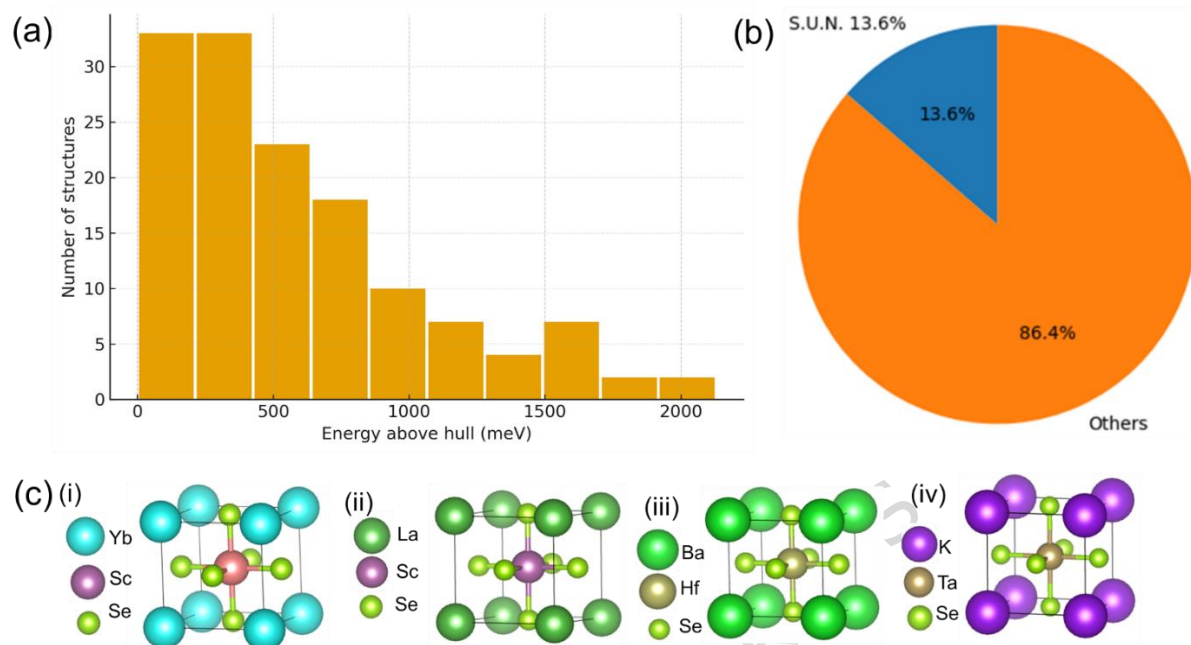


Figure 3. Analysis and validation of optimized structures generated by MEIDNet: (a) Distribution of energies above the convex hull for generated structures. (b) Stable, unique and novel (SUN) rate indicating the potential of target led materials discovery with MEIDNet. (c) Crystal structure of SUN perovskite materials generated by MEIDNet at the low bandgap range and formation enthalpy -25 meV/atom (c) (i) YbScSe₃ (ii) LaScSe₃ (iii) BaHfSe₃ (iv) KTaSe₃.

YbScSe₃ shows broadly dispersive bands with a shallow indirect separation between the valence and conduction bands and a relatively smooth density of states (DOS) around the gap (Figure 4a). This suggests good baseline mobility and temperature-activated transport. Additionally, strain or epitaxial constraints may offer a potential pathway to tune gap directness or drive band inversion in these materials. LaScSe₃ (Figure 4b) exhibits a narrow, weakly dispersive conduction manifold just above the Fermi level, producing a sharp DOS onset and implying heavy electron effective masses in certain directions, whereas the valence edge is more dispersive and consistent with lighter holes. The indirect gap with a flat plus dispersive topology is attractive for further investigations in its thermoelectric properties and for strain tunable band edge engineering. The band structure of BaHfSe₃ (Figure 4c) indicates an indirect very small-gap semiconductor. The valence band structure near R implies light holes, equally the dispersive CBM at Γ signals light electrons. KTaSe₃ (Figure 4d) shows a metallic band structure, with finite DOS at Fermi energy and steep dispersions, pointing to intrinsically good conductivity. Thus, the results indicate that MEIDNet can generate physically plausible structures within the requested targets of low bandgap and thermodynamic stability.

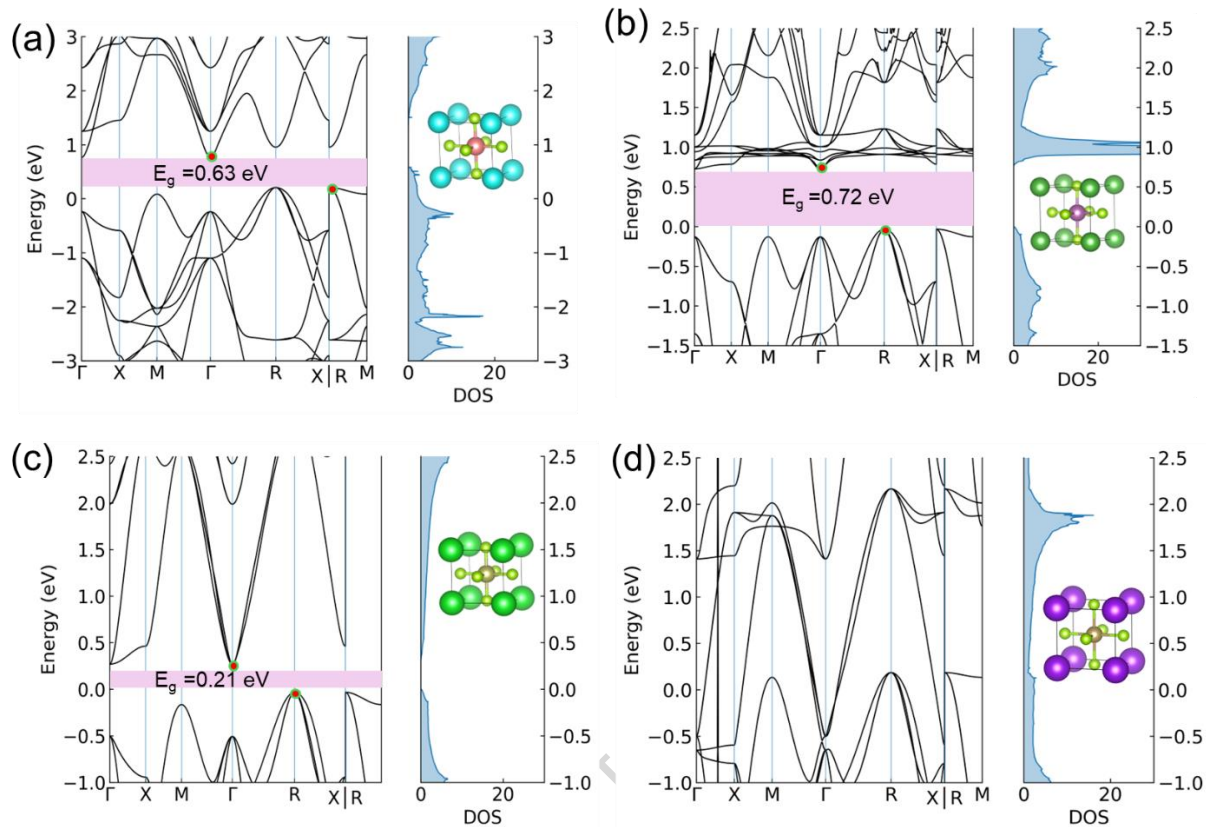


Figure 4. Band structure and density of states (DOS) plots of selected SUN materials generated by MEIDNet: (a) YbScSe₃ (b) LaScSe₃ (c) BaHfSe₃ (d) KTaSe₃ at the low bandgap range target with the constant formation enthalpy of -25 meV/atom.

Our framework addresses key prerequisites by first identifying thermodynamically plausible candidates that satisfy the conventional energy-above-hull criterion against known phases. However, we must also confirm that the material's crystal structure is dynamically stable, meaning that any small thermal vibration or perturbation would not be enough to cause it to spontaneously transform into a different structure.

Therefore, we probe the dynamical stability of the novel perovskites obtained from MEIDNet by computing phonon dispersions using Phonopy with the MACE-OMAT-0 model [48]. Our analysis reveals that these structures, despite their thermodynamic stability, exhibit soft modes in their phonon dispersions, indicating that they are dynamically unstable at 0 K. Although it is possible that these instabilities could be resolved at finite temperatures, we opt to provide a more immediate path to viable materials.

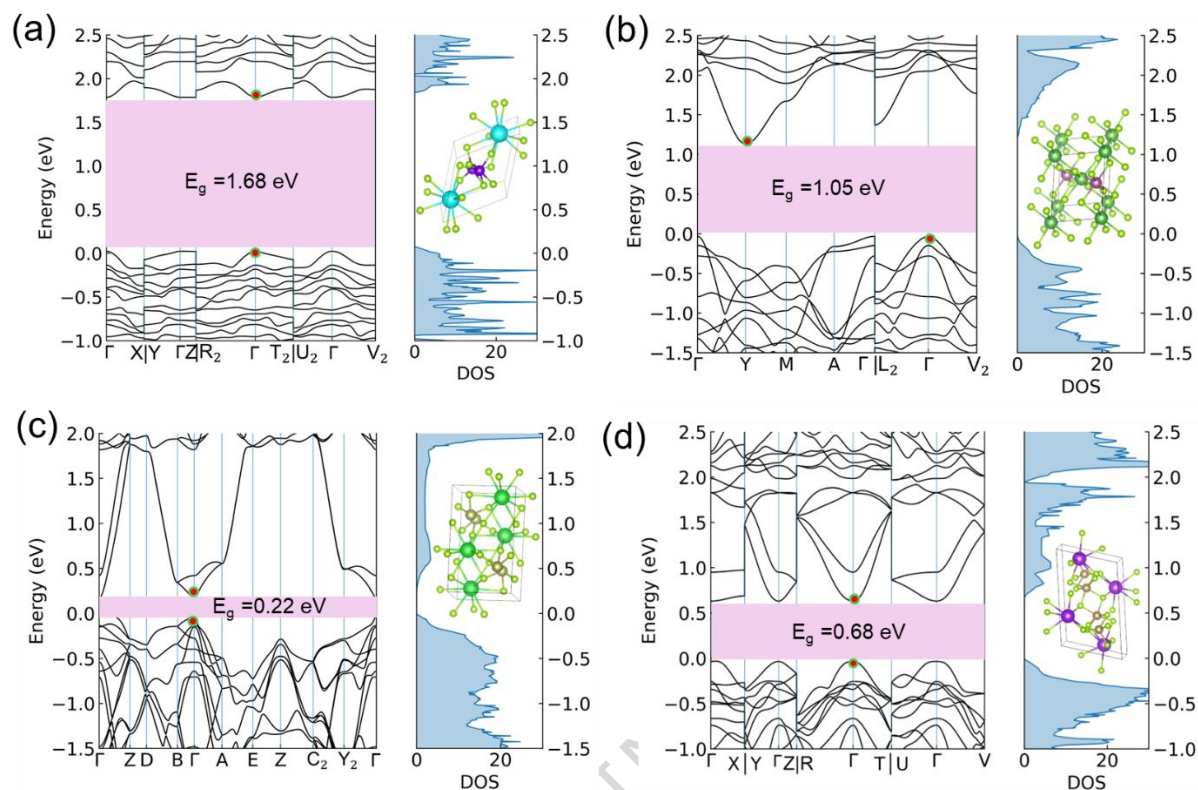


Figure 5. Band structure of the VibroML-treated SUN materials at low bandgap energy: (a) YbScSe₃ (b) LaScSe₃ (c) BaHfSe₃ (d) KTaSe₃.

We utilize the VibroML toolkit (see Methods) to follow these soft modes, systematically mapping the potential energy surface to identify the most stable, corresponding polymorphs. The verified convex hull distances for each of the lowest energy polymorphs of VibroML treated YbScSe₃, LaScSe₃, BaHfSe₃, KTaSe₃ are 0, 58, 0 and 0 meV, respectively, indicating the reduction in energy as shown in Supplementary Table S3. The lower-energy polymorphs found show larger bandgaps than the original pristine perovskite structure, as shown in Figure 5. This is expected, as these polymorphs exhibit a reduced electronic dimensionality compared to the 3D-connected pristine phase. The conduction and valence bands of the low-energy polymorphs of YbScSe₃, LaScSe₃, and BaHfSe₃ are generally less dispersive than their pristine perovskite counterparts, but given their lower energy and small bandgaps, these materials remain promising for further investigation for diverse optoelectronic applications. Compared to the metallic pristine KTaSe₃ structure, the corresponding polymorph found presents a small direct bandgap of 0.68 eV with fairly dispersive conduction and valence bands at Γ , which makes this a very promising material for optoelectronic applications.

Furthermore, the band structures are calculated for other MEIDNet-generated structures that pass the thermodynamic and dynamic stability screening (see Supplementary Figure Table S3 and Figure S10). These are found to show similar promising results, as presented in Supplementary Figures S8 and S9. To facilitate further investigation and collaboration within the community, the CIFs of all the novel structures found with the MEIDNet and VibroML workflow are provided in our Support Data repository.

Discussion

This work establishes that integrating multimodal generative models like MEIDNet with dynamic instability remediation and ab initio validation is a highly promising strategy. Our combined workflow (MEIDNet-VibroML-DFT) effectively delivers a curated set of physically grounded materials, providing the research community with a curated set of promising, physically grounded materials for experimental consideration that can be easily extended for other systems of interest.

We introduced MEIDNet, a multimodal inverse-design framework that jointly learns structural information and materials properties through latent space alignment. We here focused on electronic and thermodynamic properties, but the model can be easily adapted to other properties. Our autoencoder-based multimodal framework offers a more interpretable, easily navigable learning space, compared with diffusion models, and it is naturally extensible to additional modalities toward a universal latent representation. By coupling an equivariant graph neural network with curriculum-guided contrastive training, the model achieves faster convergence and improved latent alignment.

We demonstrated the utility of our end-to-end framework by generating novel perovskite candidates with pre-determined properties with a competitive SUN rate. More broadly, this success demonstrates a powerful and generalizable framework poised to accelerate the computational discovery of stable materials with targeted properties across diverse chemical classes.

Methods

Dataset Preparation and Representation

The training and evaluation datasets consist of crystal structures originating from the Perovskite-5 (~19k), MP-20 (~45k), and Carbon-24 (~10k) datasets. The crystal structures and properties are converted into a dense numerical representation suitable for machine learning using the Python Materials Genomics (pymatgen) library. Lattice parameters are encoded by EGNN, while atomic species are represented using one-hot encoding across a comprehensive list of 56 elements (detailed information is provided in Supplementary Note: Physical constraint objective used during latent space optimization).

Model Architecture and Training

We develop a dual- and tri-auto encoder architecture integrating a crystal graph autoencoder with a property encoder-decoder network. The crystal autoencoder is employed through EGNN layers capable of encoding spatial geometry and atomic species information. The model utilizes three sequential EGNN layers, enabling accurate encoding and decoding of both spatial coordinates and adjacency matrices representing atomic interactions.

To train the model, a contrastive learning approach is adopted to align the crystal structure and scalar property modalities in a shared latent space. The training loss function combines reconstruction losses for the structural descriptors (lattice parameters, adjacency matrices, atomic species, and atomic coordinates) with mean squared error (MSE) for scalar property predictions. All training procedures employ the Adam optimizer with a learning rate of $1e^{-3}$, a batch size of 16, and training extends up to 2000 epochs. Training is carried out on GPU-equipped high-performance computing clusters to facilitate rapid convergence. All computational workflows are implemented using Python, with extensive use of deep learning frameworks (PyTorch), materials informatics libraries (pymatgen), and numeric analysis tools (numpy, pandas). Structure-matching (SM) metric was computed using pymatgen Structure Matcher with the following tolerances: lattice tolerance (ltol) = 0.3, species tolerance (stol) = 0.5, and angle tol (tolerance) = 10° . We use an 80% train / 20% test split for each dataset in MEIDNet, with model selection performed only within the training portion using an internal validation split/cross-validation. To prevent data leakage, train and test sets are ID-disjoint, all preprocessing

steps are fitted only on the training data, and hyperparameter tuning, early stopping, and model selection never access the test set. Code execution is carried out via GPU-based high-performance computational environments.

Inverse Design Framework, SUN screening and ab initio validation

The inverse-design pipeline is implemented as follows. Randomly initialized vectors in the aligned latent space are gradually improved via gradient descent using the Adam optimizer. During this navigation in the latent space, the vectors move toward structures with the target properties (bandgap and formation enthalpy). Optimization is guided by weight losses incorporating physical constraints, such as lattice volume, minimum interatomic distances, and charge neutrality.

We also probe how adding domain constraints affects inverse design by comparing models trained with early constraints against models trained without them (Supplementary Figure S7). These constraints, such as enforcing charge neutrality, ensuring valid oxidation states, maintaining minimum interatomic distances, and using tolerance factor windows, steer the model effectively minimizing the impact in the exploration of novel structures. As shown in Supplementary Figure S5, introducing constraints early leads to faster convergence and better reconstruction quality with fewer epochs, however, it narrows down the exploratory breadth [49]. Regardless of the training regime, applying constraints during generation helps guide the decoded structures toward physically plausible crystals [50].

For *de novo* generation, the initial latent vectors $z^{(0)}$ are sampled from a standard multivariate Gaussian distribution, such that $z^{(0)} \sim \mathcal{N}(0, I)$. Because our joint latent space is L2-normalized to a unit hypersphere, these sampled vectors are subsequently projected onto the unit sphere. The continuous optimization is then formulated as finding a latent vector z^* that minimizes the objective function:

$$z^* = \operatorname{arg\,min}_z [\mathcal{L}_{target}(z) + \lambda \mathcal{L}_{physical}(z)] \quad (7)$$

where \mathcal{L}_{target} is the mean squared error between the decoded properties and the desired target values, $\mathcal{L}_{physical}$ represents penalty terms for constraint violations (e.g., minimum

interatomic distances, charge neutrality), and λ is a weighting hyperparameter. This optimization is performed iteratively via gradient descent using the Adam optimizer.

The generated crystal structures undergo structural relaxation using the eSEN-30M-OAM machine learning interatomic potential (MLIP) to minimize internal forces and energies, achieving physically plausible equilibrium configurations. The relaxation convergence criteria include a force threshold of 0.01 eV/Å and a maximum optimization step count of 300. Using the Materials Project API, the energy above hull is evaluated to find the targeted stable generated structures. Stable structures are further validated by the Ab Initio calculations using VASP with PBE PAW pseudopotentials. For the latter calculations, we adopt the Materials Project default parameters: a 520 eV plane-wave cutoff and a k-point density of 1000 k-points per reciprocal atom (KPPRA). These standardized settings place total energies on a consistent scale, enabling reliable convex-hull construction and thermodynamic stability assessment [51-53].

Dynamical stability evaluation

The dynamical stability is investigated using Phonopy [54] and VibroML [55]. The latter is an automated toolkit for identifying and stabilizing dynamically unstable crystalline phases using machine-learned interatomic potentials (MLIPs) [56]. The algorithm employs a two-stage optimization approach. First, a parameter sweep optimization identifies optimal phonon calculation settings (supercell size, atomic displacement, force tolerance) that minimize imaginary phonon frequencies. When soft modes persist, the toolkit automatically triggers a genetic-algorithm-driven structure exploration that generates new atomic configurations by displacing atoms along unstable phonon eigenmodes with varying displacement scales, mode coupling ratios, and cell transformations.

The genetic algorithm (GA) evolves a population of structural variants through selection, crossover, and mutation operations, where fitness is determined by the relaxed energy per atom of each generated structure. Each GA generation produces offspring structures that are relaxed using MLIPs, and the algorithm iteratively refines the search space by performing phonon analysis on the most promising candidates to identify new soft modes for subsequent iterations. This approach enables systematic exploration of the potential

energy landscape to discover lower-energy, dynamically stable phases from initially unstable starting structures.

For the generated perovskites, a total of 19 generations are evaluated, with all calculations performed using the MACE-OMAT-0 model [50-57]. From the three structures with the lowest energy and no soft modes generated by the workflow, we select the one with the highest symmetry for subsequent calculations.

Data availability

The crystal structures generated via MEIDNet and their associated data are available at <https://github.com/ABnano/MEIDNet>.

Code availability

The implementation code for MEIDNet has been provided at <https://github.com/ABnano/MEIDNet>.

Acknowledgements

We gratefully acknowledge high-performance computing support from the Université Catholique de Louvain (CISM/UCL) and the Consortium des Équipements de Calcul Intensif en Fédération Wallonie-Bruxelles (CÉCI). This work also relied on the use of Lucia, the Tier-1 supercomputer of the Walloon Region. This study was financed by Université catholique de Louvain (Ref. No.: ARH/MKK/01155003).

Author contributions

A.B., R.G., and G.M.R. conceptualized the study. A.B. performed the experiments and data analysis and wrote the main manuscript. R.G. carried out vibrational analysis. P.V., and G.M.R. supervised the investigation. A.B., R.G., P.V., and G.M.R. reviewed and approved the final manuscript.

Competing interests

The authors declare no competing financial or non-financial interests.

References

1. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).
2. Sanchez-Lengeling, B. & Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **361**, 360-365 (2018).
3. Jha, D. et al. ElemNet: Deep learning the chemistry of materials from only elemental composition. *Sci. Rep.* **8**, 17593 (2018).
4. Noh, J. et al. Inverse design of solid-state materials via a continuous representation. *Matter* **1**, 1370-1384 (2019).
5. Ong, S. P. et al. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314-319 (2013).
6. Curtarolo, S., Hart, G. L. W., Nardelli, M. B., Mingo, N. & Levy, O. The high-throughput highway to computational materials design. *Nat. Mater.* **12**, 191-201 (2013).
7. Xie, T., Fu, X., Ganea, O.-E., Barzilay, R. & Jaakkola, T. Crystal diffusion variational autoencoder for periodic material generation. *In Proc. Int. Conf. Learn. Represent.* (2022).
8. Noura, A., Sokolovska, N. & Crivello, J.-C. CrystalGAN: Learning to discover crystallographic structures with generative adversarial networks. Preprint at <https://arxiv.org/abs/1810.11203> (2018).
9. Zeni, C. et al. A generative model for inorganic materials design. *Nature* **639**, 624-632 (2025).
10. Jiao, R. et al. Crystal structure prediction by joint equivariant diffusion. Preprint at <https://doi.org/10.48550/arXiv.2309.04475> (2023).
11. Zhuo, Y., Mansouri Tehrani, A., Brgoch, J. & Ong, S. P. Predicting the band gaps of inorganic solids by machine learning. *J. Phys. Chem. Lett.* **9**, 1668-1673 (2018).
12. Choudhary, K., Garrity, K. F. & Tavazza, F. High-throughput density functional perturbation theory and machine learning predictions of infrared, piezoelectric, and dielectric responses. *Phys. Rev. Mater.* **3**, 083801 (2019).

13. Goodall, R. E. A. & Lee, A. A. Predicting materials properties without crystal structure: Deep representation learning from stoichiometry. *Nat. Commun.* **10**, 3569 (2019).
14. Baltrušaitis, T., Ahuja, C. & Morency, L.-P. Multimodal machine learning: A survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**, 423-443 (2019).
15. Ramachandram, D. & Taylor, G. W. Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Process. Mag.* **34**, 96-108 (2017).
16. Gao, J., Li, P., Chen, Z. & Zhang, J. A survey on deep learning for multimodal data fusion. *Neural Comput.* **32**, 829-864 (2020).
17. Zhang, C., Yang, Z., He, X. & Deng, L. Multimodal intelligence: Representation learning, information fusion, and applications. Preprint at <https://arxiv.org/abs/1911.03977> (2019).
18. Li, S. & Tang, H. Multimodal alignment and fusion: A survey. Preprint at <https://arxiv.org/abs/2411.17040> (2024).
19. Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. A simple framework for contrastive learning of visual representations. *In Proc. 37th Int. Conf. Mach. Learn.* **119**, 1597–1607 (2020).
20. Lu, J., Batra, D., Parikh, D. & Lee, S. ViLBERT: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. *Adv. Neural Inf. Process. Syst.* **32**, 13-23 (2019).
21. Okabe, R. et al. Structural constraint integration in a generative model for the discovery of quantum materials. *Nat. Mater.* **25**, 223-230 (2026).
22. Moro, V. et al. Multimodal foundation models for material property prediction and discovery. *Newton* **1**, 100016 (2025).
23. Gong, S., Wang, S., Zhu, T., Shao-Horn, Y. & Grossman, J. C. Multimodal machine learning for materials science: Composition-structure bimodal learning for experimentally measured properties. Preprint at <https://arxiv.org/abs/2309.04478> (2023).

24. Park, H., Onwuli, A. & Walsh, A. Exploration of crystal chemical space using text-guided generative artificial intelligence. *Nat. Commun.* **16**, 4379 (2025).
25. Park, H., Li, Z. & Walsh, A. Has generative artificial intelligence solved inverse materials design? *Matter* **7**, 2355-2367 (2024).
26. Zhao, Y., Wang, L. & Li, Z. Graph neural networks: A review of methods and applications. *IEEE Trans. Neural Netw. Learn. Syst.* **33**, 6413-6434 (2022).
27. Fung, V., Zhang, J., Juarez, E. & Sumpter, B. G. Benchmarking graph neural networks for materials chemistry. *npj Comput. Mater.* **7**, 84 (2021).
28. Satorras, V. G., Hoogeboom, E. & Welling, M. E(n) equivariant graph neural networks. *In Proc. 38th Int. Conf. Mach. Learn.* **139**, 9323-9332 (2021).
29. Castelli, I. E. et al. New cubic perovskites for one- and two-photon water splitting using the computational materials repository. *Energy Environ. Sci.* **5**, 9034-9043 (2012).
30. Jain, A. et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
31. Pickard, C. J. AIRSS data for carbon at 10 GPa and the C+N+H+O system at 1 GPa. Materials Cloud <https://doi.org/10.24435/MATERIALSCLOUD:2020.0026/V1> (2020).
32. Ren, Z. et al. An invertible crystallographic representation for general inverse design of inorganic crystals with targeted properties. *Matter* **5**, 314-332 (2022).
33. Radford, A. et al. Learning transferable visual models from natural language supervision. *In Proc. 38th Int. Conf. Mach. Learn.* **139**, 8748-8763 (2021).
34. Wang, T. & Isola, P. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. *In Proc. 37th Int. Conf. Mach. Learn.* **119**, 9929–9939 (2020).
35. Soviany, P., Ionescu, R. T. & Leordeanu, M. Curriculum learning: A survey. *IEEE Trans. Neural Netw. Learn. Syst.* **33**, 1526-1541 (2022).
36. Bengio, Y., Louradour, J., Collobert, R. & Weston, J. Curriculum learning. *In Proc. 26th Annu. Int. Conf. Mach. Learn.* 41-48 (2009).

37. Hachohen, G. & Weinshall, D. On the power of curriculum learning in training deep networks. *In Proc. 36th Int. Conf. Mach. Learn.* **97**, 2535-2544 (2019).
38. van den Oord, A., Li, Y. & Vinyals, O. Representation learning with contrastive predictive coding. Preprint at <https://arxiv.org/abs/1807.03748> (2018).
39. CrystalCLR paper, use: Koker, T., Quigley, K., Spaeth, W., Frey, N. C. & Li, L. Graph contrastive learning for materials. Preprint at <https://arxiv.org/abs/2211.13408> (2022).
40. Bartel, C. J. et al. A critical examination of compound stability predictions from machine-learned formation energies. *npj Comput. Mater.* **6**, 97 (2020).
41. Fu, X., Wood, B. M., Barroso-Luque, L., Levine, D. S., Gao, M., Dzamba, M. & Zitnick, C. L. Learning smooth and expressive interatomic potentials for physical property prediction. *In Proc. 42nd Int. Conf. Mach. Learn.* **267**, 17875-17893 (2025).
42. Miller, B. K., Chen, R. T. Q., Sriram, A. & Wood, B. M. FlowMM: Generating materials with Riemannian flow matching. *In Proc. 41st Int. Conf. Mach. Learn.* 235, 35664-35686 (2024).
43. De Breuck, P.-P., Piracha, H. A., Rignanese, G.-M. & Marques, M. A. L. A generative material transformer using Wyckoff representation. *npj Comput. Mater.* **12**, 60 (2026).
44. Martirosyan, M. M. et al. All that structure matches does not glitter. Preprint at <https://doi.org/10.48550/arXiv.2509.12178> (2025).
45. Luo, X. et al. CrystalFlow: A flow-based generative model for crystalline materials. *Nat. Commun.* **16**, 9267 (2025).
46. Campbell, C. R., Romero, A. H. & Choudhary, K. AtomBench: A benchmark for generative atomic structure models using GPT, diffusion, and flow architectures. Preprint at <https://doi.org/10.48550/arXiv.2510.16165> (2025).
47. Negishi, M., Park, H., Mastej, K. O. & Walsh, A. Continuous uniqueness and novelty metrics for generative modeling of inorganic crystals. Preprint at <https://doi.org/10.48550/arXiv.2510.12405> (2025).

48. ACESuit. MACE foundations. GitHub <https://github.com/ACESuit/mace-foundations> (2026).
49. Abeer, A. N. M. N., Urban, N. M., Weil, M. R., Alexander, F. J. & Yoon, B.-J. Multi-objective latent space optimization of generative molecular design models. *Patterns* **5**, 101042 (2024).
50. Oviedo, F., Lavista Ferres, J. M., Buonassisi, T. & Butler, K. T. Interpretable and explainable machine learning for materials science and chemistry. *Acc. Mater. Res.* **3**, 1-16 (2022).
51. Ganose, A. M. et al. Atomate2: Modular workflows for materials science. *Digit. Discov.* **4**, 1944-1973 (2025).
52. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758-1775 (1999).
53. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169-11186 (1996).
54. Togo, A. & Tanaka, I. First principles phonon calculations in materials science. *Scr. Mater.* **108**, 1-5 (2015).
55. Goberna Rogerio, R. VibroML package. GitHub <https://github.com/rogeriog/VibroML> (2026).
56. Batatia, I. et al. A foundation model for atomistic materials chemistry. *J. Chem. Phys.* **163**, 184110 (2025).
57. Loew, A. et al. Universal machine learning interatomic potentials are ready for phonons. *npj Comput. Mater.* **11**, 178 (2025).