Density of States-Intermediated Crystal Generation for Material Inverse Design

Sujong Lee [*1 2] **Jaewon Bae** [*1] **Seungyun Shin** [*1 3] **Heesung Choi** [1] **Chanyoung Park** [1 3]

Abstract

Inverse design of crystalline materials often relies on property-conditional generation that directly maps target properties to crystal structures. However, existing approaches ignore physical priors and fail to leverage the intrinsic relationships between physical quantities and crystal structures. To address this, we present *DOS-intermediated crystal generation*, a novel inverse design framework for crystalline materials with a physics-grounded intermediate variable. We construct a two-stage pipeline: (1) property-conditional DOS generation, where a Masked Diffusion Model produces plausible DOS via prior token assignment and sample rejection; and (2) DOS-conditional crystal generation, where a fine-tuned MatterGen backbone produces crystals matching the generated DOS. Empirically, our framework achieves robust performance on multi-property conditional generation and addresses diverse practical materials discovery scenarios beyond the reach of direct property conditioning. We establish the first baseline for DOS-intermediated multi-property inverse design, validating this paradigm and providing a foundation that complements direct-generation models.

1. Introduction

Generative models have emerged as a central tool for the inverse design of crystalline materials (Sanchez-Lengeling & Aspuru-Guzik, 2018). The last several years have produced a substantial body of work on unconditional crystal generation, with models trained to produce atomic structures that are simultaneously stable, novel, and unique (Xie et al., 2022; Jiao et al., 2023; Zeni et al., 2025; Park & Walsh, 2025). A parallel line of work has extended these generative backbones to property-conditional generation, enabling specified target property such as band gap, formation energy, or supply-risk index to be realized as a crystal (Karpovich et al., 2024; Ye et al., 2024; Zeni et al., 2025; Park & Walsh, 2025; Yamazaki et al., 2025). Despite this steady progress, the prevailing paradigm largely treats property-to-crystal mapping as a black-box optimization problem, entirely neglecting the rich physical priors that fundamentally govern material behavior.

Conventional inverse design approaches rely on direct optimization to satisfy property conditions, either independently or jointly. In compositional design, Karpovich et al. (2024) propose a reinforcement learning (RL)-based multi-property generation task, utilizing stacked RNNs to iteratively add elements to a formula. In crystal generation, existing approaches similarly rely on direct property conditioning: Ye et al. (2024) adopts per-property fine-tuning with a separate model for each target, while MatterGen (Zeni et al., 2025) conditions on scalar indices such as HHI supply-risk and magnetic density. Most recently, Yamazaki et al. (2025) scales multi-property conditioning through a Wyckoff-augmented transfer-learning scheme, yet still treats each property as an independent conditioning channel. None of these works attempts to leverage the intrinsic relationships between physical quantities by introducing a unified, physics-grounded intermediate.

In condensed-matter physics, a single observable often encodes multiple electronic properties simultaneously, such as the bandgap and the density of states at the Fermi level (Kohn & Sham, 1965). Such spectral observables can therefore serve as physics-grounded intermediate variables that encode several scalar property descriptors jointly, rather than requiring each target to be provided as a separate conditioning channel. Moreover, the rich information content of these spectra makes them natural targets for neural network predictors, which can learn unified representations capturing these entangled descriptors (Kong et al., 2022; Knøsgaard & Thygesen, 2022). We particularly identify the Density of States (DOS) as such an intermediate, as it provides a global, integrated representation of a crystal's electronic structure that jointly encodes multiple target properties within a single spectrum.

[*]Equal contribution [1]NanoForge AI, Seoul, Republic of Korea [2]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore [3]Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea. Correspondence to: Sujong Lee <LEES0196@e.ntu.edu.sg>, Chanyoung Park <cy.park@kaist.ac.kr>.

Accepted at the AI for Science workshop (ICML 2026), Seoul, Republic of Korea. Copyright 2026 by the author(s).

We propose *DOS-intermediated crystal generation*, a two-stage inverse design framework that routes a multi-property specification through the electronic density of states. In the first stage, *property-conditional DOS generation* maps the user's multi-property specification to a DOS curve that jointly encodes the specified targets. In the second stage, *DOS-conditional crystal generation* produces a crystal structure conditioned on that generated DOS. Under this decomposition, multi-property conditioning on crystal generation reduces to conditioning on a single physics-grounded intermediate, so that joint targeting is handled by one model operating in DOS space rather than by a combination of independently fine-tuned property-conditional generative models.

We instantiate each stage with a specific model. Stage 1, **DOSSampler**, is the masked diffusion model (Sahoo et al., 2024; Ou et al., 2024) over discretized DOS curves, trained to generate plausible DOS, while ensuring the user-specified property values by prior token assignment and sample rejection. Stage 2, **CDOSGen**, fine-tunes the MatterGen backbone (Zeni et al., 2025) into a DOS-conditional crystal generator that consumes the Stage-1 output as its conditioning signal. Our framework naturally extends to multi-property conditioning enhancing conditioning performance and preserves the structural quality (stability, uniqueness, and novelty (S.U.N.) rates) of the MatterGen backbone (Zeni et al., 2025). Empirically, the resulting framework outperforms existing inverse crystal-generation methods in most multi-property settings and generates crystals within the target property range at higher yield rate by well-capturing the highly informative physical prior shown by the reduced property MAE. Furthermore, we empirically validate the effectiveness of our framework by proposing diverse applications, such as non-scalar property generation and crystal-to-crystal translation which provide a practical advantage in real-world applications. Together, these models establish the first baseline for conditional crystal generation via spectral intermediates, broadening the exploration of other spectral quantities for crystal inverse design.

Contributions. To the best of our knowledge, our work is the first to introduce the spectral physical quantity as an intermediate variable for crystal generation to incorporate information-rich physical priors rather than direct scalar property optimization, establishing the first baseline for spectral-intermediate conditional crystal generation.

Our contributions are threefold:

- We propose **DOS-intermediated crystal generation**, a novel two-stage framework in which a property-conditional DOS generator feeds a DOS-conditional crystal generator, using the electronic density of states as a unified, physics-grounded conditioning signal that jointly encodes diverse electronic properties.
- We empirically demonstrate that our approach is highly competitive with state-of-the-art baselines and outperforms them in most tasks as we scale from single- to multi-property conditioning.
- Our method demonstrates empirical success in practical applications where conventional property-conditional generative models cannot be applied without additional training on specific target properties.

2. Related Work

Unconditional generative crystal models. Generative modeling of crystals has advanced rapidly around a family of joint models over (A, X, L), where A denotes the atomic species, X the fractional coordinates, and L the lattice matrix. CDVAE (Xie et al., 2022) pioneered this direction by combining a variational autoencoder with score-based denoising. DiffCSP (Jiao et al., 2023) and its space-group–aware extension DiffCSP++ (Jiao et al., 2024) introduced joint $E(3)$-equivariant diffusion over all three crystal components, and FlowMM (Miller et al., 2024) and CrystalFlow (Luo et al., 2025) provide Riemannian- and continuous-flow alternatives to the diffusion formulation. MatterGen (Zeni et al., 2025) represents the state of the art at industrial scale and forms the pretrained crystal backbone onto which our DOS-to-crystal stage is grafted.

Direct property-conditional crystal generation. A parallel line of work equips these generators with property-conditioning mechanisms. Con-CDVAE (Ye et al., 2024) extends CDVAE with explicit conditioning on formation energy, bandgap, and crystal system. MatterGen (Zeni et al., 2025) supports property-conditional fine-tuning through per-property adapters, and Yamazaki et al. (2025) scales this to multiple targets via Wyckoff-augmented transfer learning. Chemeleon2 (Park & Walsh, 2025) explores a reinforcement-learning alternative, but conditions on a single property (bandgap) alongside structural-quality rewards. Together these define the class we refer to as *direct property-conditional crystal generation*: models that map a property specification directly onto (A, X, L), training a separate conditioning pathway for each target property without exploiting the physical correlations among target properties. Our approach takes a complementary route, routing the multi-property specification through a single spectral intermediate.

DOS-based inverse design. Few prior works have used the electronic DOS as a lever for inverse design. Bang et al. (2024) train a CNN that maps a target DOS pattern onto a chemical composition; the method produces a composition only, not a full crystal structure. DOSMatGen (Jia et al.,

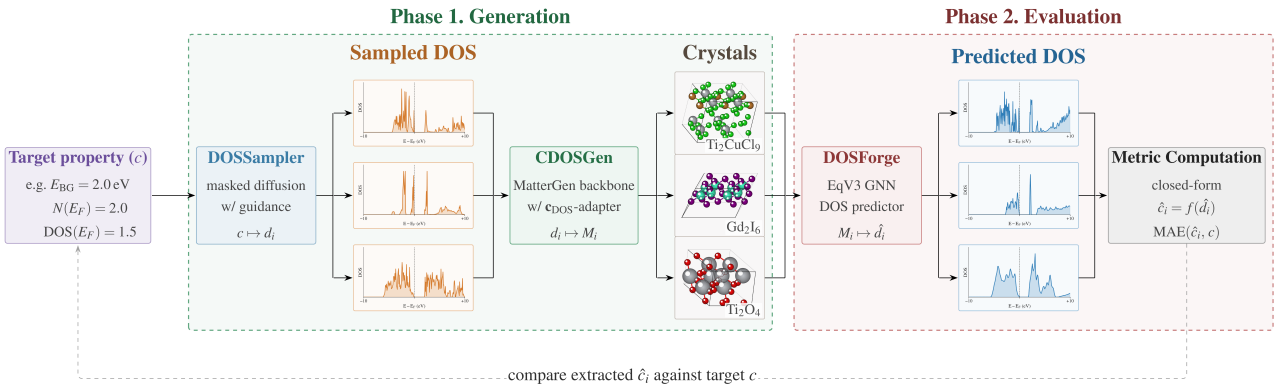

Figure 1. Overview of our DOS-intermediated Crystal Generation framework. A target property specification c, which can encode one or more scalar properties jointly, is mapped to candidate crystals in Phase 1 (DOSSampler → CDOSGen), and the resulting DOS is scored against c in Phase 2 (DOSForge (Jang et al., 2026) → closed-form metric extraction).

2025) is the closest prior work: it conditions a classifier-free-guided crystal diffusion model on atom-level projected DOS (PDOS) to generate full structures. Crucially, however, both works assume that a complete target DOS curve is *given as input*, typically copied from a reference material, and therefore do not address the setting in which the user specifies scalar physical properties and the DOS itself must be inferred from those properties. Bridging that gap is the role of our property-conditional DOS generation stage, which, together with a DOS-conditional crystal generation stage, constitutes our overall framework.

3. Preliminaries

In this section, we briefly review the density of states (DOS), the spectral observable that serves as the physics-grounded intermediate in our framework. In statistical physics and materials science, DOS is a continuous function that describes the number of available electronic states per unit volume per unit energy. The DOS uniquely characterizes a crystalline material, determining key electronic properties and enabling the efficient computation of target metrics directly from the distribution. A critical reference point within the DOS is the Fermi level (E_F), which represents the energy level where the probability of electron occupation is exactly 50% at thermal equilibrium. Among the quantities encoded in the DOS, we choose critical properties that plays a significant role in determining the characteristic of material. We explore three properties in this paper: band gap E_{BG} determines a material's conductive character, DOS at the Fermi level $\mathrm{DOS}(E_F)$ governs metallic and superconducting behavior (Bardeen et al., 1957), and integrated density of states $N(E_F)$ characterizes the total charge carrier density below the Fermi level (Ashcroft & Mermin, 1976).

In practice, we apply per-atom normalization to the total DOS data sourced from the Materials Project (Jain et al., 2013), then discretize the continuous DOS function over a specific energy grid of $[E_F - 10, E_F + 10]$ eV into a fixed grid of 256 discrete bins. The details on computation method of DOS-derived properties is described at appendix B. While properties such as bulk modulus or magnetic density have relatively tractable relationships and can be fine-tuned by the MatterGen model, we hypothesize that DOS-interpretable properties such as band gap or $N(E_F)$ are relatively challenging to condition by simple fine-tuning.

4. Methods

Our pipeline consists of two models trained independently:

1. **DOSSampler**: property-conditional DOS generation Masked Diffusion Model (MDM) with prior token assignment and sample rejection;

2. **CDOSGen**: finetuned DOS-conditional crystal generation model with pretrained MatterGen backbone.

We decompose the step into two separate modeling by neural network: property-conditional DOS generation and DOS-conditional crystal generation. We visualize our full framework at figure 1.

4.1. Property-Conditional DOS Generation

Our first model DOSSampler generates "plausible" DOS that satisfy a certain condition that are complex and difficult to directly apply finetuning in crystal structure. Among various options for the generative model, we chose Masked Diffusion Model (MDM) (Sahoo et al., 2024; Ou et al., 2024) for our DOS generation model. We chose MDM to explicitly condition on the partially-unmasked sequence of tokens which enables specifically enforcing constraints. We apply tokenization of the DOS within the range of $[0, 1]$ into

1025 values and auxiliary band gap token. The band gap token was introduced at the band gap as the frequency of zero token is significantly higher than that of other tokens, which causes a strong bias towards predicting zero too frequently. Furthermore, we allow the prediction of the MDM with per-atom normalized full DOS within $[0, K]$ range by incorporating a scaling constant at the beginning of the unmasking process such that the score network $q_\theta(\hat{x}_t^i \mid x_t^{\text{UM}}, K)$ receives the scaling constant K as condition with the unmasked partial observation x_t^{UM} to predict the unmasked token $\hat{x}_t^i$.

To support property-conditioning in DOSSampler, we provide a prior DOS which the subinterval of the band gap is fixed to band gap token [BG] at initialization for band gap condition. For other constraints, we apply classifier-guidance by learning a regressor which receives the soft embeddings of unmasked tokens to compute the MSE between target property. Based on the classifier which regresses the scalar property, we update the logit $\boldsymbol{\ell}$ to the direction which minimizes the MSE between the predicted $\hat{y}$

$$\boldsymbol{\ell}' = \boldsymbol{\ell} - \gamma \nabla_{\boldsymbol{\ell}} (\hat{y} - y^*)^2. \tag{1}$$

Here, γ dictates the guidance strength, effectively steering the generative trajectory toward the desired property y^*. After the guided generation, we reject the DOS samples that the condition does not satisfy the property conditioning to ensure the DOS samples that will be provided as condition satisfies the given properties. The training algorithm follows the structure of MDM, particularly the Reparameterized Absorbing Discrete Diffusion (RADD)'s λ denoising loss (Ou et al., 2024). The full training algorithm and details of RADD are summarized in section C.1.

4.2. DOS-Conditional Crystal Generation

We use MatterGen (Zeni et al., 2025) for base generation process and follow their definitions. We apply an extension to the DOS conditional generation on MatterGen backbone by modifying the MLP-layer from $\mathbb{R} \to \mathbb{R}^{512}$ to the $\mathbb{R}^{256} \to \mathbb{R}^{512}$. We refer the details on the MatterGen at appendix D.

Raw DOS values span a large dynamic range (0 to ~ 3000 states/eV/atom), which we compress via a logarithmic transformation $\tilde{d}_i = \log(1 + d_i)$ that maps the range to $[0, 8.0]$ while preserving inter-bin relationships and avoiding the singularity at zero. The compressed spectrum is projected to a vector embedding $\mathbf{g}_{\text{DOS}} \in \mathbb{R}^{512}$ by a two-layer MLP:

$$\mathbf{g}_{\text{DOS}} = \text{GELU}(W_2 \cdot \text{GELU}(\text{LN}(W_1 \tilde{d} + b_1)) + b_2) \tag{2}$$

where $W_1 \in \mathbb{R}^{512\times256}$, $W_2 \in \mathbb{R}^{512\times512}$, LN denotes layer normalization, and GELU is the Gaussian Error Linear Unit activation. For classifier-free guidance, the conditioning vector is replaced with $\mathbf{g}_{\text{DOS}} = 0$ with probability p_{uncond} during training.

We follow (Zeni et al., 2025) for injection of the condition into the crystal generation backbone. The per-crystal conditioning vector is first broadcast to each atom i belonging to crystal, then concatenated with the current atom representation $h_i^{(L)}$ and transformed by a two-layer adapter network:

$$h_{\text{adapt},i}^{(L)} = f_{\text{mixin}}^{(L)} \cdot f_{\text{adapt}}^{(L)}([h_i^{(L)}; \mathbf{g}_{\text{DOS},k}]) \tag{3}$$

$$h_i^{(L)} \leftarrow h_i^{(L)} + m_i \cdot h_{\text{adapt},i}^{(L)} \tag{4}$$

where $[\cdot;\cdot]$ denotes concatenation, $f_{\text{adapt}}^{(L)}$ maps from $\mathbb{R}^{1024}$ to $\mathbb{R}^{512}$, $f_{\text{mixin}}^{(l)} \in \mathbb{R}^{512\times512}$ is a linear projection initialized to zero, and $m_i \in \{0, 1\}$ is the classifier-free guidance mask.

4.3. Pipeline Construction

After the two stage models, DOSSampler and CDOSGen, are trained separately, we connect the models end-to-end to create a full inference pipeline. The inference process initiates by receiving a user-specified property condition that is considered to be physically meaningful. We use the DOSSampler to generate a property-conditional DOS of 256-dimension vector by MDM. We consider the MDM to model $F_{DOS}^{\phi} : \mathbb{R}^k \to \mathcal{T}^{256} \times \mathbb{R}$ which receives k input properties as scalar and maps them to the space of DOS which satisfies the property where $\mathcal{T}$ is the token space.

By definition, the output of DOSSampler is a token between $[0, 1]$. We therefore decode the tokens into a float of $[0, 1]$ range which is then multiplied by K to construct the absolute DOS. The scaling factor K is sampled from the distribution obtained by the Materials Project dataset, and provided to DOSSampler as condition to let the model be aware of the scaling factor to allow prediction of the absolute value of physical quantities. Subsequently, DOS is passed to the CDOSGen model $G_{Crystal}^{\psi} : \mathbb{R}^{256} \to \mathcal{M}$. which is a fine-tuned MatterGen with DOS conditions to output the crystal structure with $M = (A, X, L)$. Here, we represent $A \in \mathbb{A}^n$, the atomic species (atom types) of the n atoms, $X \in [0, 1)^{3\times n}$, the fractional atomic coordinates relative to the lattice vectors, $L \in \mathbb{R}^{3\times3}$, the lattice matrix representing the unit cell, where columns are the lattice vectors (l^1, l^2, l^3) and n is the number of atoms in the crystal unit cell. Therefore, the full pipeline can be interpreted as a mapping $f_\theta : \mathbb{R}^k \to \mathcal{T}^{256} \times \mathbb{R} \to \mathbb{R}^{256} \to \mathcal{M}$.

While our CDOSGen model modifies the MatterGen's adapter's MLP layer to process a 256-dimensional vector ($\mathbb{R}^{256} \to \mathbb{R}^{512}$) rather than a scalar ($\mathbb{R} \to \mathbb{R}^{512}$) as the embedding dimension remains the same. Therefore, the crystal generation stage itself incurs virtually no computational overhead compared to baseline MatterGen models. While the DOSSampler does require inference time to generate the conditioning DOS, our decoupled pipeline allows this step to be executed offline. By pre-computing a dictionary of

target DOS profiles, CDOSGen can subsequently generate crystals on-demand with latency equivalent to that of the baseline.

5. Results

We train the DOSSampler model for 10k steps and CDOSGen model for 100 epochs on two Nvidia H100 GPUs for 12 hours. MatterGen suggests to train the model until no validation loss improvement over 100 iterations which is strictly more than our setting. The crystal is passed to DOSForge (Jang et al., 2026), the state-of-the-art model that learns the forward crystal-to-DOS mapping, the inverse of our generative pipeline's DOS-to-crystal mapping, as an alternative to the expensive Density Functional Theory (DFT) simulation with iConFormer (Jo et al., 2025) encoder replaced by EquiformerV3 (Liao et al., 2026). Refer to appendix E for details about DOSForge.

5.1. Multi-Property Inverse Design

To validate the efficiency of our DOS-intermediated generation framework, we evaluate our model across various conditions measured in mean absolute error (MAE) and mean $\pm 10\%$ accuracy (hereafter, 10% accuracy) across 1024 samples. Additionally, we tracked the S.U.N. metric to assess the stability, uniqueness, and novelty of the generated structures. We evaluated our property-conditional approach against two baselines—unconditional DOS and DOS from Materials Project (MP DOS)—using 1,024 random samples per method. Across all evaluated conditioning settings, our conditional method consistently outperforms both baselines, confirming that our model successfully adheres to specified target properties. Table 1 reports the results for our framework, and Table 2 reports those for the MatterGen baseline under the same conditioning settings.

Our method outperforms the MatterGen (Zeni et al., 2025) baseline across most multi-property conditioning settings, achieving higher 10% accuracy in the majority of cases and lower MAE in over half of cases. MatterGen exhibits higher performance in the single-property setting compared to our model, but its performance deteriorates when conditions are added. In contrast, our method remains robust and the performance improves when simultaneously conditioned on $E_{BG} = 2.0$ eV and $N(E_F) = 2.0$ reducing the E_{BG} MAE to 1.397 eV (down from 1.460 in the single-condition setup) and the $N(E_F)$ MAE to 2.483 (down from 2.638). Similarly, conditioning simultaneously on DOS$(E_F) = 1.5$ eV and $N(E_F) = 2.0$ reduces the respective MAE simultaneously and increases the 10% accuracy from 7.2% to 9.4%, 8.3% to 9.9% and 2.6% to 4.8% for multi-property conditional generation. In contrast, MatterGen struggles in multi-property setting where the band gap MAE for target 2.0 eV increases from 1.223 eV to 1.352 eV and DOS(E_F) MAE increases from 0.734 to 1.763 exceeding the value obtained from arbitrary DOS. The flexibility of multi-property conditioning allows us to guide the model toward specific classes of materials with high precision without deterioration of conditioning performance.

Furthermore, our empirical results show that our method achieves better structural quality measured in S.U.N. metric for DOS-conditional model compared to the direct-optimization method by MatterGen. The Unconditional model achieves a S.U.N. score of 24.4% and by adding moderate physical priors, the S.U.N. remains within a similar range without significant degradation.

5.2. Property-Conditional DOS Generation

In this section, we evaluate the DOSSampler, the first stage of our generation pipeline. Since our full pipeline feeds the generated DOS as input to CDOSGen, it is essential to verify that the generated DOS faithfully captures the target electronic structure. We analyze the performance of the DOSSampler from three complementary perspectives: plausibility, physical validity, and diversity.

Our experiments show that the DOSSampler produces realistic DOS, and that this realism is preserved as we progressively impose additional conditioning targets. As shown in Figure 2, generated DOS samples closely resemble DFT-computed MP DOS in both the unconditional and conditional settings. Further details are deferred to Appendix F.

5.3. DOS-Conditional Crystal Generation

In this section, we evaluate the performance of the CDOSGen model, the second model of our generation pipeline, which generates a crystal structure based on the conditional DOS. For the evaluation, we use 1024 samples from the test dataset composed of 256 metals and 768 non-metals (1:3 ratio) for stronger focus on the mean-absolute error of non-zero property.

Table 3 reports the performance of the conditioning mechanism of DOS in various distributional metric and MAE of the target properties such as band gap, DOS(E_F), and $N(E_F)$. Given MP DOS $\{d_i\}_{i=1}^N$ and the predicted DOS of the corresponding generated crystals $\{\hat{d}_i\}_{i=1}^N$, "Ours" reports each metric averaged over the matched pairs $(d_i, \hat{d}_i)$, while the random baseline reports the same metric averaged over mismatched pairs $(d_i, \hat{d}_{\sigma(i)})$, where σ is a permutation ($\sigma(i) \neq i$ for all i). This baseline reflects the metric value expected under random DOS pairing, isolating the contribution of the conditioning mechanism beyond the marginal distribution of generated DOS.

We see that the cosine similarity of the DOS between arbitrary MP DOS is 0.557 while our model demonstrates

Table 1. The performance of our DOS-intermediate generation framework across various conditioning settings.

Conditioning			±10% (↑)			MAE (↓)			S.U.N. (↑)
E_{BG}	DOS(E_F)	$N(E_F)$	E_{BG}	DOS(E_F)	$N(E_F)$	E_{BG}	DOS(E_F)	$N(E_F)$	
Unconditional			2.9%	4.6%	2.7%	1.798	1.011	2.815	24.4%
MP DOS			4.0%	1.6%	3.9%	1.692	1.290	2.781	N/A
2.0	–	–	7.2%	–	–	1.460	–	–	29.1%
–	1.5	–	–	8.3%	–	–	0.792	–	19.8%
–	–	2.0	–	–	2.6%	–	–	2.638	24.0%
2.0	–	2.0	9.4%	–	2.7%	1.397	–	2.483	27.7%
–	1.5	2.0	–	9.9%	4.8%	–	0.786	2.055	21.3%
2.0	0.0	2.0	9.4%	N/A	2.7%	1.397	0.335	2.483	27.7%
0.0	1.5	2.0	N/A	9.9%	4.8%	0.407	0.786	2.055	21.3%

Table 2. The performance of MatterGen baseline across various conditioning settings.

Conditioning			±10% (↑)			MAE (↓)			S.U.N. (↑)
E_{BG}	DOS(E_F)	$N(E_F)$	E_{BG}	DOS(E_F)	$N(E_F)$	E_{BG}	DOS(E_F)	$N(E_F)$	
2.0	–	–	8.9%	–	–	1.360	–	–	23.5%
–	1.5	–	–	13.7%	–	–	0.711	–	17.7%
–	–	2.0	–	–	2.8%	–	–	2.349	22.7%
2.0	–	2.0	8.7%	–	7.4%	1.405	–	1.564	25.1%
–	1.5	2.0	–	6.3%	3.6%	–	1.186	1.875	19.5%
2.0	0.0	2.0	8.1%	N/A	8.8%	1.413	0.281	1.530	26.2%
0.0	1.5	2.0	N/A	7.4%	3.9%	0.309	1.097	1.752	18.1%

Table 3. Comparison of DOS evaluation metrics with randomized baseline.

Metric	Ours	Random baseline
Cosine similarity	**0.722**	0.557
Pearson	**0.499**	0.197
JS divergence	**0.184**	0.307
E_{BG} MAE	**1.107**	1.513
DOS(E_F) MAE	**0.359**	0.488
$N(E_F)$ MAE	**2.442**	2.942

stronger correlation of 0.722. Other distributional measures show similar gains: Pearson correlation increased by $2.5\times$ and JS divergence which decreased to 0.184 demonstrate that our method effectively learns the density of states in distributional perspective. Furthermore, the MAE for physical quantities has reduced compared to the MP DOS. The band gap MAE has significantly decreased from 1.513 to 1.107 by 27% and similarly for DOS(E_F) indicating that the DOS-conditional model learns the encoded properties in DOS. Our model inherently generates the crystal that matches condition on three different physical quantities which may be extended to more properties at the cost of additional DOS inference. We provide further details and the visualization of the crystal and DOS comparison at appendix G and I.2.

6. Applications

The physics-grounded DOS intermediate unlocks a broader class of materials-discovery tasks that direct property-conditional generators cannot address without per-target retraining. We validate this practical advantage on two representative materials-discovery scenarios: transparent conducting material generation via non-scalar DOS conditions (Section 6.1) and thermoelectric material discovery via crystal-to-crystal translation (Section 6.2).

6.1. Transparent Conducting Materials via Non-Scalar Conditioning

Although existing property-conditional inverse design frameworks can generate crystals satisfying scalar-valued properties, they struggle to generalize to boolean or categorical conditions. In contrast, our DOS-conditional generation model generalizes to categorical properties by rejecting DOS samples that do not match the target conditions, leveraging efficient MDM-based DOS sampling. We validate this by experiment on Transparent Conducting Material (TCM) generation task.

TCM is a class of materials that simultaneously conduct electricity and remain transparent to visible light, with applications in solar cells and LEDs (Edwards et al., 2004). TCMs are commonly known to satisfy the following spectral conditions: (1) a wide band gap ($\geq$ 3.1 eV) located below the Fermi level, (2) non-zero DOS(E_F), and (3) a

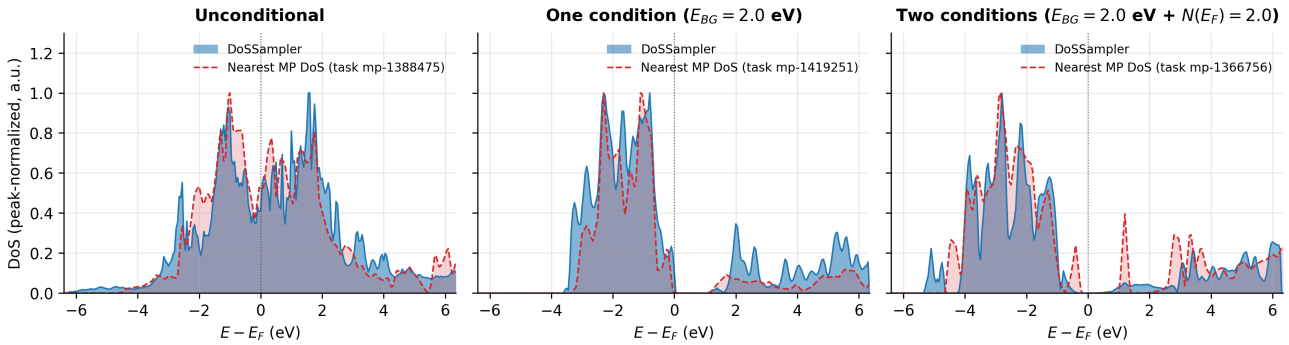

Figure 2. Representative DOSSampler samples vs. nearest Materials Project (MP) match by cosine similarity. generated DOS (blue solid) overlaid on the nearest-MP DOS (red dotted).

conduction bandwidth in a specific range that determines the TCM type—Type 1 (degenerate semiconductors such as ITO, with broad s-character conduction band of width > 4 eV) (Edwards et al., 2004) and Type 2 (gapped metals such as $SrNbO_3$, with narrow d-band of width < 3 eV) (Zhang et al., 2016; Park et al., 2020). We therefore frame the TCM generation task as generating crystals whose DOS satisfies these conditions, and report TCM yield and type-correct rate alongside the unconditional baseline in Table 4.

For conditional generation, we produced 3,072 samples per type. In contrast, the unconditional baseline utilized 6,144 samples, representing the combined total of the 3,072 samples from both Type 1 and Type 2. For unconditional generation, the pretrained MatterGen checkpoint provides a total TCM yield of 0.034%. In contrast, we applied our DOS-conditional generation framework and rejected the DOSSampler-generated sampled DOS that does not satisfy the TCM criteria while providing the sub-interval of the band gap as prior. For Type 1 at 99 epochs, the total TCM yield increased to 0.31% ($9.1\times$ over unconditional) and the type-correct yield to 0.25% ($15.6\times$); for Type 2, the corresponding rates were 0.64% ($18.8\times$) and 0.11% ($6.9\times$). Furthermore, we have tested our result on a further trained CDOSGen checkpoint at 299 epochs. While the Type 2 TCM yield remained approximately the same at 0.70%, the Type 1 type-correct yield increased to 0.90%, which is $3.6\times$ that of the 99 epoch and 56.3× the unconditional baseline.

We provide a visualization for 3 samples of TCM crystal corresponding conditioning and predicted DOS at Figure 3. For all cases, both the conditioning DOS and the predicted DOS exhibit a wide band gap (> 3.1 eV) below the Fermi level and a strictly positive DOS(E_F).

6.2. Thermoelectric Materials via Crystal-to-Crystal Translation

We introduce crystal-to-crystal translation, a setting analogous to image-to-image translation in computer vision, where the conditioning signal is the DOS of a reference crystal. This setting reflects a common scenario in materials discovery known as analogical discovery, in which researchers seek new materials that share the desirable functional behavior of a reference compound, even when the full set of physical descriptors driving that behavior is difficult to enumerate as a small collection of scalar targets (Isayev et al., 2015; Ihalage & Hao, 2021). While conventional direct optimization methods fail in these scenarios because they require explicitly defined target properties, our model takes the DOS of a reference crystal as the conditioning signal for CDOSGen and generates novel crystals whose DOS closely matches that of the reference.

We assess the generated crystals along two axes reported in Table 5. First, we measure the spectral similarity between the conditioning DOS and the predicted DOS of each generated crystal following the protocol in Section 5.3. Second, since determining whether a material is thermoelectric typically requires expensive transport measurements (e.g., the figure-of-merit ZT), we instead apply a TE-filter and report the fraction of generated crystals passing it: (1) band gap in $[0.1, 1.0]$ eV (Sofo & Mahan, 1994), (2) DOS(E_F)=0, and (3) cosine similarity above 0.9 with a reference thermoelectric DOS.

Table 5 reports the crystal-to-crystal generation results across various metrics for thermoelectric materials. Figure 9 provides the visualization of 4 generated thermoelectric materials for each conditioning crystal. We observe a significant improvement on all metrics measuring the similarity between the DOS of the conditioning crystal and the DOS of the translated crystal, with the cosine similarity reaching up to 0.835. Most importantly, the TE-filter pass rate increases from 1.46% to 11.29% at 99 epochs, and further to 12.26% at the 299-epoch checkpoint—corresponding to $7.7\times$ and $8.4\times$ more crystals passing the filter, respectively. While we have validated this approach on thermoelectric materials, the same crystal-to-crystal framework extends naturally to discovering other materials with special spectral signatures,

Table 4. TCM generation performance. Our DOS-conditional framework generates up to 56× more type-correct TCMs than the unconditional MatterGen baseline.

TCM Type	epoch	TCM yield	Type-correct rate	lift over uncond yield	lift over uncond Type-correct
Type 1	uncond.	0.034%	0.016%	1×	1×
	99	0.31%	0.25%	9.1×	15.6×
	299	0.97%	0.90%	28.5×	56.3×
Type 2	uncond.	0.034%	0.016%	1×	1×
	99	0.64%	0.11%	18.8×	6.9×
	299	0.70%	0.10%	20.6×	6.3×

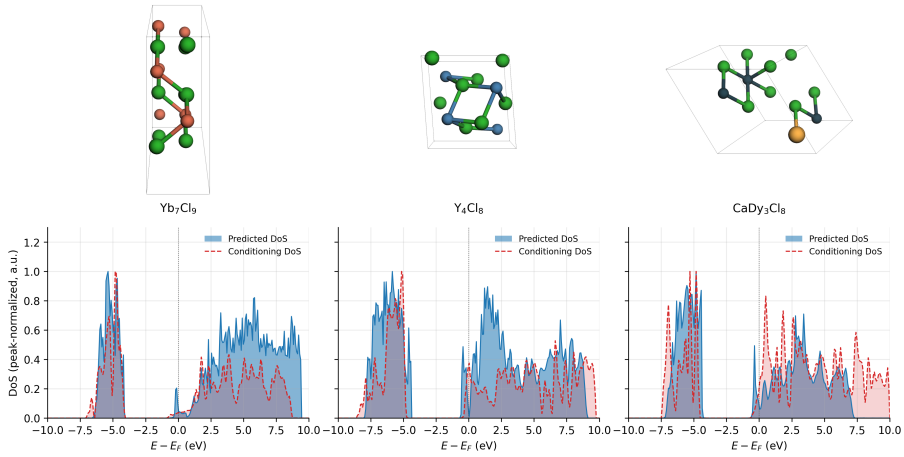

Figure 3. Examples of Transparent Conducting Materials generated by CDOSGen. Top: generated crystal structures. Bottom: predicted DOS of each generated crystal (blue solid) overlaid on the conditioning DOS (red dotted).

Table 5. Comparison of DOS evaluation metrics for thermoelectric materials with randomized baseline. For explanation of random baseline, refer to Section 5.3

Metric	99 epoch	299 epoch	Random baseline
Cosine similarity ↑	0.835	**0.858**	0.557
Pearson ↑	0.617	**0.669**	0.197
JS divergence ↓	0.065	**0.055**	0.307
E_{BG} MAE (eV) ↓	0.253	**0.225**	1.513
DOS(E_F) MAE ↓	0.261	**0.248**	0.488
TE-filter pass rate ↑	11.29%	**12.26%**	1.46%

which we leave for future work.

7. Conclusion

Summary. In this work, we proposed **DOS-intermediated crystal generation**, a novel framework for crystalline-material multi-property inverse design that routes scalar property targets through the electronic density of states as a physics-grounded intermediate. The framework decomposes inverse design into two stages: a masked diffusion model (DOSSampler) synthesizes a DOS from scalar property specifications, and a DOS-conditional generator (CDOSGen) maps it onto a crystal structure on top of the MatterGen backbone. Our approach is competitive with state-of-the-art baselines and outperforms them in most settings as conditioning scales from single- to multi-property targets, while preserving the structural quality (stability, uniqueness, novelty) of the unconditional backbone — establishing DOS-intermediated conditioning as a practical path forward for multi-property inverse design. Beyond the standard multi-property setting, the physics-grounded DOS intermediate unlocks downstream tasks that direct property-conditional generators cannot address without per-target retraining, including non-scalar property generation and crystal-to-crystal translation, demonstrating practical value for real-world materials discovery.

Limitations and future work. Our framework is tied to properties derivable from the electronic DOS; targets governed by phonon or magnetic degrees of freedom, or transport coefficients requiring additional electronic-structure information, fall outside the current scope. Furthermore, as the first instantiation of a spectral-intermediate paradigm, our approach leaves room for further refinement in some single-property settings where highly optimized direct-generation models remain advantageous. Addressing these limitations involves further extending the property-to-DOS stage to jointly condition on auxiliary physical channels such as phonon DOS or magnetic descriptors, alongside architectural and training refinements to close the remaining performance gap. More fundamentally, our work establishes physics-grounded intermediate-conditioned crystal generation as a paradigm that provides a foundation for future refinement against highly optimized direct-generation models, with natural extensions to richer spectral observables, including projected or k-resolved DOS and X-ray absorption near-edge structure (XANES).

Acknowledgements

This work was supported by the Technology development Program (RS-2025-25462152) funded by the Ministry of SMEs and Startups (MSS, Korea).

References

Ashcroft, N. W. and Mermin, N. D. *Solid State Physics*. Holt, Rinehart and Winston, New York, 1976.

Bai, J., Du, Y., Wang, Y., Kong, S., Gregoire, J., and Gomes, C. P. Xtal2dos: Attention-based crystal to sequence learning for density of states prediction. In *NeurIPS 2022 AI for Science: Progress and Promises*, 2022.

Bang, K., Kim, J., Hong, D., Kim, D., and Han, S. S. Inverse design for materials discovery from the multidimensional electronic density of states. *Journal of Materials Chemistry A*, 12:6004–6013, 2024. doi: 10.1039/D3TA06491C.

Bardeen, J., Cooper, L. N., and Schrieffer, J. R. Theory of superconductivity. *Physical Review*, 108(5):1175–1204, 1957. doi: 10.1103/PhysRev.108.1175.

Edwards, P. P., Porch, A., Jones, M. O., Morgan, D. V., and Perks, R. M. Basic materials physics of transparent conducting oxides. *Dalton Transactions*, (19):2995–3002, 2004. doi: 10.1039/B408864F.

Gasteiger, J., Becker, F., and Günnemann, S. Gemnet: Universal directional graph neural networks for molecules. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 6790–6802. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/35cf8659cfcb13224cbd47863a34fc58-Paper.pdf.

Ihalage, A. and Hao, Y. Analogical discovery of disordered perovskite oxides by crystal structure information hidden in unsupervised material fingerprints. *npj Computational Materials*, 7(1):75, 2021. doi: 10.1038/s41524-021-00536-2.

Isayev, O., Fourches, D., Muratov, E. N., Oses, C., Rasch, K., Tropsha, A., and Curtarolo, S. Materials cartography: Representing and mining materials space using structural and electronic fingerprints. *Chemistry of Materials*, 27 (3):735–743, 2015. doi: 10.1021/cm503507h.

Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., and Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, 2013. doi: 10.1063/1.4812323.

Jang, K., Park, D., Bae, J., and Park, C. Synergistic multitask learning for electronic density of states prediction. In *AI for Accelerated Materials Design - ICLR 2026*, 2026. URL https://openreview.net/forum?id=80O5IRJ8Yb.

Jia, S., Ganesh, P., and Fung, V. Electronic structure guided inverse design using generative models. *arXiv preprint arXiv:2504.06249*, 2025.

Jiao, R., Huang, W., Lin, P., Han, J., Chen, P., Lu, Y., and Liu, Y. Crystal structure prediction by joint equivariant diffusion. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Jiao, R., Huang, W., Liu, Y., Zhao, D., and Liu, Y. Space group constrained crystal generation. In *International Conference on Learning Representations (ICLR)*, 2024.

Jo, H., Choi, H., Cho, M., and Min, D. iconformer: Dynamic parameter-efficient tuning with input-conditioned adaptation, 2025. URL https://arxiv.org/abs/2409.02838.

Karpovich, C., Pan, E., and Olivetti, E. A. Deep reinforcement learning for inverse inorganic materials design. *npj Computational Materials*, 10(1):287, 2024.

Knøsgaard, N. R. and Thygesen, K. S. Representing individual electronic states for machine learning GW band structures of 2D materials. *Nature Communications*, 13: 468, 2022. doi: 10.1038/s41467-022-28122-0.

Kohn, W. and Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133–A1138, 1965. doi: 10.1103/PhysRev.140.A1133.

Kong, S., Ricci, F., Guevarra, D., Neaton, J. B., Gomes, C. P., and Gregoire, J. M. Density of states prediction for materials discovery via contrastive learning from probabilistic embeddings. *Nature Communications*, 13:949, 2022. doi: 10.1038/s41467-022-28543-x.

Lee, N., Noh, H., Kim, S., Hyun, D., Na, G. S., and Park, C. Density of states prediction of crystalline materials via prompt-guided multi-modal transformer. *Advances in Neural Information Processing Systems*, 36:61678–61698, 2023.

Liao, Y.-L., Hoffman, A. J., Shen, S. C., Duval, A., Norwood, S. W., and Smidt, T. Equiformerv3: Scaling efficient, expressive, and general se(3)-equivariant graph attention transformers, 2026. URL https://arxiv.org/abs/2604.09130.

Luo, X., Wang, Z., Wang, Q., Shao, X., Lv, J., Wang, L., Wang, Y., and Ma, Y. CrystalFlow: A flow-based generative model for crystalline materials. *Nature Communications*, 16:9267, 2025. doi: 10.1038/s41467-025-64364-4.

Miller, B. K., Chen, R. T. Q., Sriram, A., and Wood, B. M. FlowMM: Generating materials with riemannian flow matching. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, volume 235 of *PMLR*, pp. 35664–35686, 2024.

Ou, J., Nie, S., Xue, K., Zhu, F., Sun, J., Li, Z., and Li, C. Your absorbing discrete diffusion secretly models the conditional distributions of clean data, 2024.

Park, H. and Walsh, A. Guiding generative models to uncover diverse and novel crystals via reinforcement learning. *arXiv preprint arXiv:2511.07158*, 2025.

Park, Y., Roth, J., Oka, D., Hirose, Y., Hasegawa, T., Paul, A., Pogrebnyakov, A., Gopalan, V., Birol, T., and Engel-Herbert, R. $SrNbO_3$ as a transparent conductor in the visible and ultraviolet spectra. *Communications Physics*, 3(1):102, 2020. doi: 10.1038/s42005-020-0372-9.

Sahoo, S. S., Arriola, M., Gokaslan, A., Marroquin, E. M., Rush, A. M., Schiff, Y., Chiu, J. T., and Kuleshov, V. Simple and effective masked diffusion language models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=L4uaAR4ArM.

Sanchez-Lengeling, B. and Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science*, 361(6400):360–365, 2018. doi: 10.1126/science.aat2663.

Sofo, J. O. and Mahan, G. D. Optimum band gap of a thermoelectric material. *Physical Review B*, 49(7):4565–4570, 1994. doi: 10.1103/PhysRevB.49.4565.

Xie, J., Zhang, Z., and Cao, X.-M. Density of states prediction by a metric-optimized equivariant graph neural network. *Chemistry of Materials*, 37(16):6313–6322, 2025. doi: 10.1021/acs.chemmater.5c01359. URL https://doi.org/10.1021/acs.chemmater.5c01359.

Xie, T., Fu, X., Ganea, O.-E., Barzilay, R., and Jaakkola, T. Crystal diffusion variational autoencoder for periodic material generation. In *International Conference on Learning Representations (ICLR)*, 2022.

Yamazaki, S., Nong, W., Zhu, R., Novoselov, K. S., Ustyuzhanin, A., and Hippalgaonkar, K. Multi-property directed generative design of inorganic materials through Wyckoff-augmented transfer learning. *arXiv preprint arXiv:2503.16784*, 2025.

Ye, C.-Y., Weng, H.-M., and Wu, Q.-S. Con-CDVAE: A method for the conditional generation of crystal structures. *Computational Materials Today*, 1:100003, 2024. doi: 10.1016/j.commtd.2024.100003. Also arXiv:2403.12478.

Zeni, C., Pinsler, R., Zügner, D., Fowler, A., Horton, M., Fu, X., Wang, Z., Shysheya, A., Crabbé, J., Ueda, S., Sordillo, R., Sun, L., Smith, J., Nguyen, B., Schulz, H., Lewis, S., Huang, C.-W., Lu, Z., Zhou, Y., Yang, H., Hao, H., Li, J., Yang, C., Li, W., Tomioka, R., and Xie, T. A generative model for inorganic materials design. *Nature*, 2025. doi: 10.1038/s41586-025-08628-5.

Zhang, L., Zhou, Y., Guo, L., Zhao, W., Barnes, A., Zhang, H.-T., Eaton, C., Zheng, Y., Brahlek, M., Haneef, H. F., et al. Correlated metals as transparent conductors. *Nature Materials*, 15(2):204–210, 2016. doi: 10.1038/nmat4493.

A. Dataset

We use the available DOS from the Materials Project dataset (MP) which in total composes of more than 700k materials with corresponding supplementary information such as the DOS and formation energy. We selectively filter some materials for stable training of both DOSSampler and CDOSGen model.

For DOSSampler, we initially apply preprocessing of the MP DOS to the range of ± 10 from the fermi level E_F. For CDOSGen, the filtering becomes more critical. The MatterGen pre-training was achieved with selective subset of the MP dataset. The criteria for elimination from the dataset include the number of atoms and elements that composes the crystal. In this paper, we apply the same filtering on MP dataset to prevent catastrophic forgetting caused by training on out-of-distribution data samples from those eliminated in the MatterGen.

We validate the performance of the model using DOSForge, which is trained to predict the DOS of crystals with fewer than 100 atoms. Crystals with more than 100 atoms may yield unreliable predictions, since DOSForge treats such crystals as out-of-distribution and may extrapolate incorrectly. Furthermore, we specifically eliminate the crystals with more than 20 atoms, which even narrows down the DOSForge constraint. The pretrained MatterGen base model is trained on Alex-MP-20 which does not include any crystal with more than 20 atoms. Subsequently, we eliminate the crystal which includes the elements: Tc(43), Pm(61) and those with atomic number greater than or equal to 84.

Another important preprocessing that we apply is the per-atom normalization of the DOS in the materials project dataset. DOS represents the number of possible states per unit energy per unit volume. Naturally, increasing the number of atoms results in a larger number of states and hence a higher DOS. To prevent this unnecessary bias, we train the model after normalizing the DOS dataset with respect to the number of atoms.

After the preprocessing, the number of data samples we finally reach to 272 658 samples for training. We summarize the filtering step at table 6.

Table 6. Dataset statistics across different filtering stages

Stage	Train	Validation	Total
Full MP Dataset	**705509**	NA	**705509**
Filter DOS/Fermi Level Unavailable	**445649**	55706	**501355**
Filter max atoms > 20	286861	35792	322653
Filter excluded elements ($z \notin \{43, 61\}$ and $z \leq 83$)	429011	53635	482646
All three filters combined	**272658 (61.2 %)**	**34002 (61.0 %)**	**306660 (61.2 %)**

B. DOS-Derived Properties

Here, we describe the target properties that we use for conditioning. We focus on following quantities, E_{BG}, DOS(E_F) (unit: states/eV, omitted hereafter), and $N(E_F)$ (unit: states, omitted hereafter), as target properties for multi-property inverse design.

Band Gap Band gap is defined as the width of the zero-DOS region between the valence band and conduction band around the Fermi level. The band gap E_{BG} determines material's conductive character whether a material is classified as metal, semiconductor or insulator and many other properties. To compute the band gap, instead of directly computing the width between the first non-zero values on either side of the Fermi level (which is included in the band gap), we follow Jang et al. (2026) for the computation.

We compute the cumulative sum of the DOS value outward starting from the Fermi level. The cumulative sum continues until the value reaches $1/K$ where K is a parameter that is chosen as a threshold. We set $K = 10$ such that negligible fluctuations in the band gap region do not affects the band gap computation but sufficiently low to acknowledge the deviation from the zero at valence and conduction band. We summarize the computation of band gap at algorithm 1.

DOS(E_F) The DOS value at Fermi level, DOS(E_F) is the representation which governs the metallic and superconducting behavior of a generated crystal (Bardeen et al., 1957). The DOS value at Fermi level is computed relatively easily. As we compute the grid of $[E_F - 10, E_F + 10]$ of 256 grid, we have the Fermi level at the center which is at the interpolation of 127th and 128th bin of the 256 dimensional vector.

Algorithm 1 Band Gap Extraction Algorithm from DOS

Require: Fermi-centered DOS spectrum $\mathbf{d} \in \mathbb{R}^T$, sharpness k ($k = 10$), energy spacing $\Delta E \approx 0.078\,\text{eV}$
Ensure: Band gap $\hat{E}_g$
$\mathbf{d} \leftarrow \text{ReLU}(\mathbf{d})$ ▷ Enforce $D(E) \geq 0$
$\mathbf{d}_L \leftarrow \text{flip}(\mathbf{d}_{1:T/2}), \quad \mathbf{d}_R \leftarrow \mathbf{d}_{T/2+1:T}$ ▷ Split at Fermi level, order outward
for $s \in \{L, R\}$ **do**
  $\mathbf{c}_s \leftarrow \text{cumsum}(\mathbf{d}_s)$ ▷ Cumulative sum from Fermi level
  $\mathbf{m}_s \leftarrow \text{ReLU}(1 - k \cdot \mathbf{c}_s)$ ▷ Soft gap mask
end for
$\hat{E}_g \leftarrow (\sum_i m_{L,i} + \sum_i m_{R,i}) \cdot \Delta E$
return $\hat{E}_g$

$\mathbf{N(E_F)}$ The integrated density of states $N(E_F)$ is the cumulative sum of DOS up to the Fermi level, i.e.

$$\int_{-\infty}^{E_F} g(E')dE'. \tag{5}$$

Theoretically, $N(E_F)$ determines the total charge carrier density below the Fermi level (Ashcroft & Mermin, 1976). The computation of this on the discretized bin is achieved by computing the cumulative sum up to 127th bin of the 256 dimensional vector by

$$N(E_F) = \sum \text{dos}[:128] \tag{6}$$

C. Training Details

C.1. DOSSampler Training

Reparameterized Absorbing Discrete Diffusion (RADD) (Ou et al., 2024) is a simple but effective modification of the Masked Diffusion Model by reparameterization of the concrete score as a conditional distribution of the denoised data. This interpretation is well-aligned with our framework as well which constructs the band gap region as prior and computes the score which are conditional distribution of the denoised band gap region.

Mathematically, the concrete score is factorized as a noise-dependent scalar term and unmasked data distribution via

$$\begin{aligned}\underbrace{\frac{p_t(x_t^1 \dots \widehat{x}_t^i \dots x_t^d)}{p_t(x_t^1 \dots x_t^i \dots x_t^d)}}_{\text{concrete score}} &= \underbrace{\frac{e^{-\bar{\sigma}(t)}}{1 - e^{-\bar{\sigma}(t)}}}_{\text{scalar}} \underbrace{p_0(\hat{x}_t^i | \boldsymbol{x}_t^{UM})}_{\substack{\text{clean data}\\ \text{distribution}}} \\ &= \underbrace{\frac{e^{-\bar{\sigma}(t)}}{1 - e^{-\bar{\sigma}(t)}}}_{\text{scalar}} \underbrace{q_\theta(\hat{x}_t^i \mid x_t^{\text{UM}})}_{\text{score network}}\end{aligned} \tag{7}$$

Here, q_θ is a score network which approximates the clean data distribution p_0. This allows implementation of time-independent score-prediction network that approximates the conditional distribution given the unmasked data.

The score network q_θ is parameterized by a bidirectional Transformer encoder. The unmasked tokens and mask tokens are first projected into a continuous embedding space. To inject the global conditioning variable K and the timestep t, Adaptive Layer Normalization (AdaLN) is used across the Transformer blocks. This allows the network to globally modulate its feature maps based on the desired DOS scaling and the current noise level.

We use a classifier guidance for 1) its simple implementation and 2) interpretable implementation in the embedding space. Classifier guidance steers the masked diffusion sampler by biasing the current logits through a small MLP regressor that predicts a target property from the expected token embedding $\mathbf{pE}$, where $\mathbf{p}$ is the softmax over logits and $\mathbf{E}$ is the diffusion model's own token embedding matrix. At every denoising step the logits are updated as

$$\boldsymbol{\ell}' = \boldsymbol{\ell} - \gamma \nabla_{\boldsymbol{\ell}} (\hat{y} - y^*)^2. \tag{8}$$

with a residual-scaled γ, and the next token is drawn from $\mathrm{softmax}(\ell^*)$, so the sample is nudged toward — rather than hard-constrained to — the desired property. A predictor–corrector loop that re-masks and re-samples a fraction of committed tokens under the same guided logits, optionally combined with rejection sampling, turns this soft steering mechanism into a calibrated sampler of property-conditioned DOS.

Using the algorithms provided above, we make sampling via algorithm 2.

Algorithm 2 Guided Masked Diffusion Sampling

Require: Diffusion f_θ, regressor r_ϕ, embeddings $\mathbf{E}$, target y^*, guidance γ, steps N, corrector steps K, re-mask fraction α

$\mathbf{x} \leftarrow [\texttt{MASK}]^L$
for $i = 0$ **to** $N-1$ **do**
  $\boldsymbol{\ell} \leftarrow f_\theta(\mathbf{x}, t_i)$
  $\boldsymbol{\ell}' \leftarrow \boldsymbol{\ell} - \gamma \nabla_{\boldsymbol{\ell}} \left(r_\phi(\text{softmax}(\boldsymbol{\ell})\mathbf{E}) - y^*\right)^2$
  $\mathcal{U}_i \leftarrow \{l : x_l = \texttt{MASK} \text{ and } \text{Uniform}(0,1) < r(t_i)\}$
  if $i = N-1$ **then** $\mathcal{U}_i \leftarrow \{l : x_l = \texttt{MASK}\}$
  end if
  $\mathbf{x}_{\mathcal{U}_i} \sim \text{Categorical}(\text{softmax}(\boldsymbol{\ell}'_{\mathcal{U}_i}))$
  if $i < N-1$ **then**
    for $k = 1$ **to** K **do**
      Sample $\mathcal{R} \subset \{l : x_l \neq \texttt{MASK}\}$ where $|\mathcal{R}| = \lfloor \alpha \cdot |x \neq \texttt{MASK}| \rfloor$
      $\mathbf{x}_{\mathcal{R}} \leftarrow \texttt{MASK}$
      $\boldsymbol{\ell}^c \leftarrow f_\theta(\mathbf{x}, t_i)$
      $\boldsymbol{\ell}^{c\prime} \leftarrow \boldsymbol{\ell}^c - \gamma \nabla_{\boldsymbol{\ell}^c} \left(r_\phi(\text{softmax}(\boldsymbol{\ell}^c)\mathbf{E}) - y^*\right)^2$
      $\mathbf{x}_{\mathcal{R}} \sim \text{Categorical}(\text{softmax}(\boldsymbol{\ell}^{c\prime}_{\mathcal{R}}))$
    end for
  end if
end for
return $\mathbf{x}$

C.2. CDOSGen Training

Our CDOSGen model finetunes the pretrained MatterGen checkpoint using the DOS as the input condition. For the finetuning of pretrained MatterGen, there exist two potential methods: adapter-only finetuning and full-finetuning. We apply full-finetuning which does not freeze the original MatterGen checkpoint following the MatterGen while training the adapter module concurrently.

The adapter modules are injected into each message-passing layer of the base model to augment the atom embeddings with target property information. Each adapter consists of a property embedding layer, a two-layer multi-layer perceptron (MLP), and a zero-initialized mix-in linear layer. The fine-tuning loss is identical to the training objective used during the unconditional pre-training stage, simply with the conditional property labels incorporated into the inputs. This objective is a combined loss function comprising three parts: a score matching loss for the atomic coordinates, a score matching loss for the periodic lattice, and a cross-entropy loss for the discrete atom types.

$$L_{cell} = \sum_{t=1}^{T} (1 - \overline{\alpha}_t)\sigma_t(n)^2 \mathbb{E}_{q(L_0)} \mathbb{E}_{q(L_t|L_0)} \left[||s_{L,\theta}(X_t, L_t, A_t, t) - \nabla_{L_t} \log q(L_t|L_0)||_2^2 \right] \tag{9}$$

$$\begin{aligned} L_{vb} = \mathbb{E}_{q(a_0)} \Big[& - \mathbb{E}_{q(a_1|a_0)} \log p_\theta(a_0|a_1, 1) \\ & + D_{KL}[q(a_T|a_0)||q(a_T)] \\ & + \sum_{t=2}^{T} \mathbb{E}_{q(a_t|a_0)} D_{KL}[q(a_{t-1}|a_t, a_0)||p_\theta(a_{t-1}|a_t)] \Big] \end{aligned} \tag{10}$$

$$L_{CE} = -\mathbb{E}_{q(a_0)}\left[\sum_{t=2}^{T}\mathbb{E}_{q(a_t|a_0)}\log p_\theta(a_0|a_t,t)\right] \tag{11}$$

Then, the final loss is computed by their weighted sum.

C.3. Implementation

We train the model with Adam optimizer with learning rate 5×10^{-6}. The learning rate reduces to 60% at plateau to 1×10^{-6} at minimum. We set the batch size as 64 with gradient clipping of 0.5 for each GPU in the multi-GPU training. MatterGen allows two finetuning modes: full-finetuning and adapter-only-finetuning. In this experiment we use full-finetuning which the MatterGen claims to perform better and our result supports this empirically. We train the model for 100 epoch which is less than the suggested MatterGen finetuning training epoch for fair comparison as DOS naturally is more complex $\mathbb{R}^{256}$ dimensional vector than $\mathbb{R}$ scalar.

D. MatterGen

MatterGen (Zeni et al., 2025) is a foundational diffusion model for crystal generation, jointly modeling atomic species, fractional coordinates, and the unit cell lattice.

The forward corruption process adds noise to each component independently: a discrete diffusion with an absorbing state for categorical atom types, a variance-exploding wrapped normal distribution for periodic fractional coordinates, and a variance-preserving diffusion restricted to symmetric matrices for the unit cell lattice. To generate a novel structure, the framework reverses this process by starting from a random distribution and iteratively applying an SE(3)-equivariant score network based on the GemNet-T backbone (Gasteiger et al., 2021) to jointly predict the denoising scores for all three structural components.

For property-conditional generation, the score network is augmented with a lightweight adapter that injects a property embedding into each message-passing layer; the adapter mechanism natively supports joint conditioning on multiple scalar properties, and we adopt this same adapter interface for our DOS conditioning signal in CDOSGen. The property is injected by the equation

$$h_j'^{(L)} = h_j^{(L)} + m_i \cdot f_{\text{mixin}}^{(L)} \cdot f_{\text{adapter}}^{(L)}(\boldsymbol{g}). \tag{12}$$

E. DOSForge

Although the Density of States (DOS) encodes crucial information about a crystal's underlying characteristics, computing it relies on solving the Schrödinger equation—a notoriously time-consuming process. To enable efficient evaluation across a statistically meaningful number of generated samples, we employ DOSForge (Jang et al., 2026), a state-of-the-art (SOTA) DOS prediction model. DOSForge uses 62 964 structures from the Materials Project (MP) as dataset diverse element compositions after discretization by 256 bins over $[E_F - 10, E_F + 10]$ eV energy level with additional filtering of unstable materials.

Table 7. Comparison of DOS prediction model.

Method	MAE	MSE	R^2	WD
Xtal2DoS	3.24	83.2	.475	6.82
DOSTransformer	3.20	82.2	.463	6.45
EqDOS	3.26	97.3	.342	10.4
DOSForge	**2.77**	**82.1**	**.495**	**5.70**

Compared to the conventional DOS prediction model (Bai et al., 2022; Lee et al., 2023; Xie et al., 2025), its prediction outperforms in various metrics of MAE, Mean-Squared Error, R^2 and Wasserstein Distance (WD). Furthermore, it better predicts the important properties such as the band gap that we have previously described ((Jang et al., 2026); Table 2). We therefore assume that the evaluation result via DOSForge is sufficiently accurate and fair.

F. DOSSampler

Audit overview. The DOSSampler is the first stage of our two-stage pipeline; its output is consumed by CDOSGen as the conditioning signal for crystal generation, so any failure here propagates downstream. To our knowledge no prior work evaluates a generative DOS sampler in this role. We audit it from three angles—spectral plausibility against real Materials Project (MP) DOS, physical validity of the spectrum, and within-batch diversity. The audit asks whether the generated spectrum is itself MP-like; how closely a sampled DOS hits the numerical conditioning targets is the responsibility of CDOSGen at the next pipeline stage and is outside the scope of this section.

What we mean by realistic. We call a generated DOS *realistic* when (i) it lies within the natural similarity scale of MP DOS under a 4-metric percentile criterion, (ii) it satisfies the physical validity checks defined below—per-bin non-negativity and no cliff-edge artifact at the gap edge—and (iii) it belongs to a sample set that does not exhibit catastrophic mode collapse and whose within-batch spread matches or exceeds that of the matched MP reference.

Settings. For each of three conditioning settings—unconditional, single-condition E_{BG}=2.0 eV, and two-condition E_{BG}=2.0 eV with the additional integrated-density constraint $N(E_F)$=2.0—we generate 1,024 DOS samples and run the same audit pipeline. All numbers come from Table 8.

Table 8. Spectral audit of DOSSampler outputs across three conditioning settings (1,024 samples each). Conditional settings are scored against the matched MP subset with $E_{BG} \in [1.8, 2.2]$ eV (N=6,446); the unconditional setting is scored against a fixed random subsample of 5,000 MP entries (seed 0). The **MP (ref.)** column applies each metric to MP entries themselves: percentile-based metrics by leave-one-out on the same subset, absolute-valued metrics directly on the subset. Edge jump is computed only on samples with `measured_gap` > 0.1 eV. "–" marks metrics that have no self-reference analog.

Category	**Metric**	Uncond.	E_{BG}=2.0	E_{BG}=2.0 $+ N(E_F)$=2.0	**MP (ref.)**
Plausibility	Plausibility rate (%) ↑	51.5	44.0	51.6	~ 83
	4-metric agreement (%) ↑	91.1	90.7	91.4	–
Physical validity	Edge jump (gapped) ↓	0.025	0.024	0.014	~ 0.020
	Per-bin non-negativity (%) ↑	100	100	100	100
Diversity	Pairwise mean cosine ↓	0.42	0.40	0.46	~ 0.51
	Unique nearest MP (/1,024) ↑	793	813	724	911 ± 9

Reference set per setting. We score each generated DOS against MP DOS entries within $\pm 10\%$ of the conditioning targets. For the two conditional settings this gives a matched MP subset of 6,446 entries with $E_{BG} \in [1.8, 2.2]$ eV; for unconditional generation we use a fixed random subsample of 5,000 MP entries (seed 0) to keep self-similarity computations tractable. LOO percentiles are computed independently within each setting, so plausibility rate and 4-metric agreement are not comparable between the unconditional and conditional columns. The two conditional columns share the same MP reference and are directly comparable to each other; edge jump, per-bin non-negativity, pairwise cosine, and unique-nearest-MP are absolute statistics and remain comparable across all columns.

Plausibility under a 4-metric percentile criterion. We compare each generated DOS to its matched MP reference under four distance measures: cosine and Tanimoto similarity, the 1-Wasserstein distance W_1, and MSE.[1] Cosine and Tanimoto read spectral shape, W_1 is sensitive to mass shifts along the energy axis, and MSE captures pointwise deviations. For each generated DOS we take its best match within the reference under each metric (max cosine, max Tanimoto, min W_1, min MSE) and convert the four values to percentiles within the reference's LOO self-similarity distribution under that metric. Each metric selects its own best-match MP entry, since the four distances do not necessarily agree on which reference entry is closest.

We label the sample *plausible* when all four percentiles exceed 10: equivalently, for every one of the four distances, the sample's best-match similarity to MP must exceed the 10th percentile of the MP-to-MP self-similarity distribution under that distance. The decile cutoff was fixed before computing per-setting statistics. The *plausibility rate* in Table 8 is the fraction of samples meeting this condition. We separately report the *4-metric agreement rate*: the fraction of samples whose

[1]Definitions: $\cos\theta = \langle x, y\rangle / (\|x\|\,\|y\|)$; Tanimoto $= \langle x, y\rangle / (\|x\|^2 + \|y\|^2 - \langle x, y\rangle)$, which unlike cosine penalizes magnitude differences even when shapes are aligned; $W_1(x, y) = \Delta E \sum_i |F_x(i) - F_y(i)|$ with $F.$ the cumulative distribution along the energy axis and ΔE the bin width; MSE is bin-wise squared error.

four percentile placements agree to within two adjacent decile buckets. This is a robustness diagnostic for the plausibility verdict, asking whether the verdict depends on the choice of distance.

Plausibility rates fall in the range 44%–52%, against an MP self-consistency value of $\sim 83\%$ on the same LOO criterion—roughly 53%–62% of MP's own rate. The MP self-consistency is below 100% because the LOO percentile criterion is strict by design (a 10-percentile threshold under all four distances jointly), so a substantial gap from 100% is expected even on real MP entries. The 4-metric agreement rate exceeds 90% in every setting, so the verdict for any individual sample does not depend on which of the four distances is taken as primary. This addresses clause (i) of our *realistic* definition: a substantial fraction of the generated samples lies within the MP-to-MP similarity range under all four distances jointly, and that verdict is robust to the choice of distance.

Physical validity. Plausibility metrics already capture spectral-shape realism in aggregate, but they depend on best-match comparisons and average over the spectrum. This section adds two checks that act on each spectrum on its own—per-bin non-negativity and gap-edge sharpness—so that local artifacts at single bins are visible independently of the per-sample MP comparison.

Per-bin non-negativity. A real DOS is non-negative everywhere by definition; a generated spectrum with negative bins is not a valid DOS. The DOSSampler tokenizes DOS values into positive values, which we report as the metric for completeness. The rate is 100% in all three settings (and trivially in the MP reference).

Edge jump. For samples with a measurable band gap (measured_gap > 0.1 eV) we report the *normalized edge jump* $\mathrm{DOS}(e)/\max(\mathrm{DOS})$, where e is the first bin to the right of E_F exceeding 1% of the row maximum. DFT-smeared edges have small jumps that arise from Gaussian tails; a step-function (cliff) edge characteristic of hard-pinning has a jump close to 1. Median normalized edge jumps fall in the range 0.014–0.025 across the three settings, all within the natural DFT-smeared range of MP itself (median ~ 0.020); values below the MP median indicate softer Gaussian-like edges, well within the DFT-smeared regime and far from the cliff-edge (≈ 1) failure mode we are screening against. The generated samples therefore show no cliff-edge artifact.

Together, these two checks address clause (ii) of our *realistic* definition: the generated samples are well-formed bin-wise and show no cliff-edge artifact at the gap.

Diversity. We use two statistics. The *within-batch pairwise cosine* is the mean cosine similarity across distinct sample pairs in the $1{,}024$-sample set; we treat it as the primary diagnostic because it directly measures the spread of the generated samples. The *unique-nearest-MP count* maps each generated sample to its cosine-nearest entry in the matched MP subset and reports the number of distinct MP entries appearing in this list; we use it as a secondary check against catastrophic mode collapse. This count is a derived quantity that depends on both the generated samples and the density of the MP reference, which is why we treat it as secondary.[2]

Within-batch pairwise mean cosines fall in 0.40–0.46, all below the matched MP reference's pairwise mean (~ 0.51): under the primary diagnostic the generated samples span a similarity scale at least as broad as MP itself. Under the secondary diagnostic, unique-nearest-MP counts are 724–813 out of $1{,}024$. The MP self-baseline—random subsamples of $1{,}024$ MP entries from the matched conditional reference, with each entry's cosine-nearest neighbor sought in the same reference (self-excluded), averaged over 100 trials—yields 911 ± 9 unique entries (the unconditional reference yields 833 ± 10 under the same procedure).[3] The generated counts reach 79%–95% of the corresponding MP self-baseline, indicating mild concentration relative to MP's own mode density and well above the catastrophic-collapse regime (which would yield counts of order 100 or below). This addresses clause (iii): the generated samples match or exceed MP's spectral spread under the primary diagnostic, with only mild concentration under the secondary.

Effect of adding the second rejection condition. A central design question is whether adding $N(E_F)$=2.0 on top of the E_{BG}=2.0 eV target erodes spectral statistics. We use *proximity to the MP reference* as the criterion: a metric is favorable when it moves toward, or stays within, the MP value. Comparing the two conditional columns of Table 8 (which share

[2] A within-batch cosine-threshold deduplication analogous to the SUN metric's `StructureMatcher.fit` in Section 5.1 does not differentiate the three settings here: the maximum within-batch pairwise cosine is below 0.95 in all three settings, so every setting yields $1{,}024$ unique samples under any threshold above 0.95.

[3] For reference, the random-uniform expectation $E[\text{unique}] = N(1-(1-1/N)^K) \approx 947$ for N=6,446 and ≈ 926 for N=5,000, with K=1,024, verified by 1,000 Monte Carlo trials. This is an idealized upper bound that ignores MP's natural mode density; the MP self-baseline is the appropriate comparison.

the same MP reference), three of the four headline metrics move favorably—plausibility rate rises ($44.0 \rightarrow 51.6\%$), edge jump stays within the MP-smeared range ($0.024 \rightarrow 0.014$, MP $\sim$0.020), and within-batch pairwise cosine moves toward MP ($0.40 \rightarrow 0.46$, MP $\sim$0.51). The single unfavorable move is unique-nearest-MP ($813 \rightarrow 724$), which we read as tighter conditioning narrowing the anchor set rather than a regression on realism: the count remains at 79% of the MP self-baseline (911 ± 9), well above the catastrophic-collapse regime, and the primary diversity diagnostic still shows a broader spread than MP itself. The additional rejection condition tightens conditioning at no cost to the headline realism statistics audited here.

G. CDOSGen

Comparison with DOSMatGen (Jia et al., 2025) While we first introduce DOS as an intermediate variable which encodes number of properties into a form of physical prior, Jia et al. (2025) attempts to provide an inverse design framework by matching the electronic DOS. Our work differ from this work in the following aspects

1. Our work generates the DOS via DOSSampler as an intermediate variable for property condition. However, DOSMatGen assumes the DOS (DFT-computed) is given and aims to model the DOS-to-crystall mapping only.
2. Our work computes the total DOS structure with per-atom normalization instead of atomic DOS
3. Our model is based on the MatterGen architecture while DOSMatGen uses DiffCSP (Jiao et al., 2023) as the backbone.

Further training Zeni et al. (2025) suggests the training to iterate for a sufficiently large iteration and stops when there is no improvement in the validation loss over 100 epochs. However, to achieve a fair comparison between MatterGen and our method, we fix the training iteration to 100 epochs which is significantly less than the suggested number. We believe this number is at most fair or beneficial for MatterGen which at most provide 3 scalar inputs as condition which is relatively simple compared to the DOS.

Table 9 shows that the MAE has decreased for all properties with concurrent increase in the distributional metrics when trained longer. We have trained the model for 300 epoch for verification which has demonstrated an improved performance in the DOS-conditioning. The validation loss indicates that the reported performance has not yet converged to its optimum. In particular, the band gap MAE decreases significantly to 0.970 eV, suggesting that with an optimally constructed pipeline and longer training, DOS-intermediated generation may achieve even better MAE. We believe longer training is capable of further improving the performance in the respective metric which we leave for further studies.

Table 9. Performance comparison of "Ours" results between the 99-epoch and 299-epoch checkpoints.

Metric	99 Epochs	299 Epochs	Δ (Absolute (%))
Cosine similarity	0.722	**0.726**	0.004 (0.6%)
Pearson	0.499	**0.516**	0.017 (3.4%)
JS divergence	0.184	**0.178**	-0.006 (3.3%)
E_{BG} MAE	1.107	**0.970**	-0.137 (12.4%)
$N(E_F)$ MAE	2.442	**2.351**	-0.091 (3.7%)
DOS(E_F) MAE	0.359	**0.349**	-0.010 (2.8%)

Furthermore we evaluate on randomly sampled (instead of random within fixed 1:3 ratio) 1024 ground truth DOS at table 10, to see the general performance of our model on arbitrary DOS. We observe an increase in the evaluation result in most cases as expected. A larger fraction of metals in the conditioning DOS reduces the E_{BG} prediction error (MAE), since predicting metallic behavior is a safer default."

H. Varying the Number of Crystals Generated per DOS Condition

Our CDOSGen model is inherently a property-conditioned generative model which can produce multiple crystals from a single DOS as condition. While this problem may not be well-defined in material science perspective as we assume there exists one crystal corresponding to the DOS, we perform evaluation by generating different number of crystals using the same DOS. By generating multiple crystals conditioned on the same DOS, we attempt to investigate whether the model explores diverse minima what may not exactly match but similar to the given DOS.

Table 10. Comparison of completely random sampled DOS evaluation metrics between 99 and 299 epoch checkpoints with randomized baseline

Metric	99 Epochs	299 Epochs	Random
Cosine similarity	0.713	**0.726**	0.557
Pearson	0.493	**0.518**	0.197
JS divergence	0.189	**0.179**	0.307
E_{BG} MAE	1.003	**0.937**	1.513
$N(E_F)$ MAE	2.414	**2.319**	2.942
DoS(E_F) MAE	**0.376**	0.382	0.488

We choose 256, 64 and 16 DOS that are most plausible from the conditionally generated samples based on the evaluation metric as in table 11. While there are subtle variance across different numbers of crystals per DOS, we see that the model's performance are invariant with respect to the crystal per DOS.

I. Sample Results

I.1. DOSSampler

Here we provide the cherry-picked visualization of sampled DOS result of the DOSSampler. We visualize the result on 4 different conditions on single-property conditioning case at figure 4, 5, 6 and 7. We see that the resulting DOS satisfies the specified conditioning well in all cases while remaining plausible based on the discussion in section F.

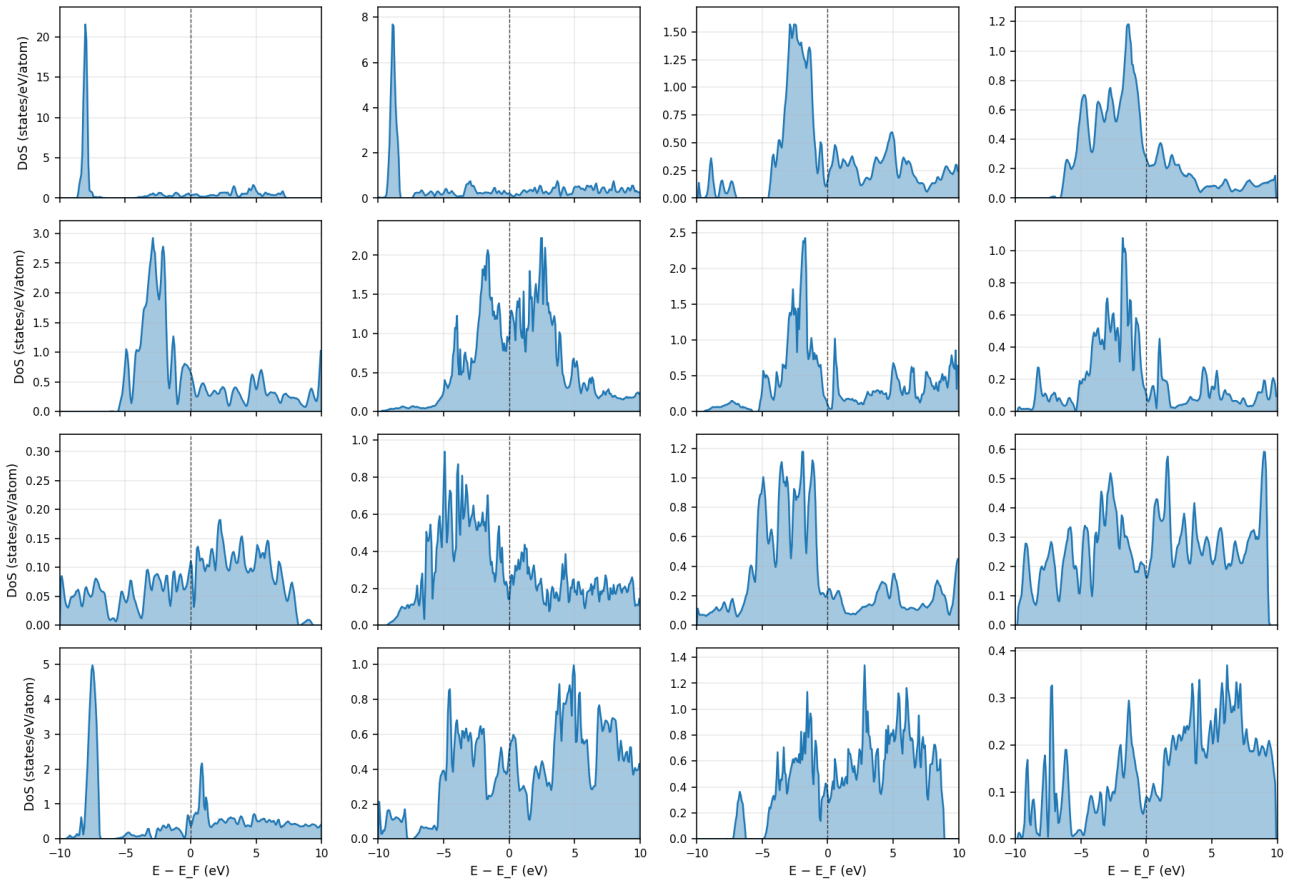

Figure 4. 16 samples from DOSSampler with out any conditioning (unconditional generation).

Table 11. Detailed performance metrics across generation repetitions and conditioning setups.

Setup	**Conditioning**			**±10% (↑)**			**MAE (↓)**			**S.U.N.**	**Stable**	**Novel**
(unique×reps)	E_{BG}	DOS(E_F)	$N(E_F)$	E_{BG}	DOS(E_F)	$N(E_F)$	E_{BG}	DOS(E_F)	$N(E_F)$			
Conditioning: $E_{BG} = 2.0$												
1024×1	2.0	–	–	**7.2%**	–	–	1.460	–	–	**29.1%**	48.6%	76.4%
256×4	2.0	–	–	5.7%	–	–	**1.444**	–	–	27.3%	51.1%	72.9%
64×16	2.0	–	–	6.3%	–	–	1.482	–	–	27.2%	49.1%	75.7%
16×64	2.0	–	–	4.9%	–	–	1.628	–	–	25.7%	51.3%	72.0%
Conditioning: DOS$(E_F) = 1.5$												
1024×1	–	1.5	–	–	8.3%	–	–	0.792	–	19.8%	35.5%	80.1%
256×4	–	1.5	–	–	9.5%	–	–	0.698	–	18.0%	38.2%	74.5%
64×16	–	1.5	–	–	**9.6%**	–	–	0.690	–	20.3%	40.4%	74.9%
16×64	–	1.5	–	–	9.2%	–	–	**0.643**	–	**20.9%**	43.8%	71.7%
Conditioning: $N(E_F) = 2.0$												
1024×1	–	–	2.0	–	–	**2.6%**	–	–	**2.638**	24.0%	42.9%	77.8%
256×4	–	–	2.0	–	–	1.4%	–	–	2.961	21.9%	43.4%	73.7%
64×16	–	–	2.0	–	–	1.4%	–	–	3.026	23.6%	46.6%	73.1%
16×64	–	–	2.0	–	–	1.3%	–	–	2.570	**24.3%**	49.3%	70.4%
Conditioning: $E_{BG} = 2.0, N(E_F) = 2.0$												
1024×1	2.0	–	2.0	**9.4%**	–	**2.7%**	**1.397**	–	**2.483**	27.7%	52.1%	74.0%
256×4	2.0	–	2.0	7.7%	–	1.5%	1.449	–	2.610	**30.2%**	**55.9%**	71.4%
64×16	2.0	–	2.0	7.0%	–	1.4%	1.495	–	2.668	28.0%	54.1%	71.7%
16×64	2.0	–	2.0	6.0%	–	1.1%	1.563	–	2.730	26.8%	53.3%	70.9%
Conditioning: DOS$(E_F) = 1.5, N(E_F) = 2.0$												
1024×1	–	1.5	2.0	–	9.9%	4.8%	–	**0.786**	2.055	21.3%	30.7%	85.7%
256×4	–	1.5	2.0	–	**12.5%**	4.4%	–	0.821	1.990	**23.7%**	36.1%	84.2%
64×16	–	1.5	2.0	–	12.0%	**5.3%**	–	0.804	**1.986**	22.2%	37.3%	80.2%
16×64	–	1.5	2.0	–	8.5%	3.5%	–	1.010	2.142	19.7%	33.4%	82.4%
Insulator (DOS$(E_F) = 0$**) derived from** $E_{BG} = 2.0, N(E_F) = 2.0$												
1024×1	2.0	0.0	2.0	**9.4%**	67.8%*	**2.7%**	**1.397**	0.335	**2.483**	27.7%	52.1%	74.0%
256×4	2.0	0.0	2.0	7.7%	**68.5%***	1.5%	1.449	0.137	2.610	**30.2%**	55.9%	71.4%
64×16	2.0	0.0	2.0	7.0%	66.2%*	1.4%	1.495	0.160	2.668	28.0%	54.1%	71.7%
16×64	2.0	0.0	2.0	6.0%	67.1%*	1.1%	1.563	**0.134**	2.730	26.8%	53.3%	70.9%
Metallic ($E_{BG} = 0$**) derived from DOS**$(E_F) = 1.5, N(E_F) = 2.0$												
1024×1	0.0	1.5	2.0	84.4%*	9.9%	4.8%	0.407	**0.786**	2.055	21.3%	30.7%	85.7%
256×4	0.0	1.5	2.0	93.7%*	**12.5%**	4.4%	0.131	0.821	1.990	**23.7%**	36.1%	84.2%
64×16	0.0	1.5	2.0	97.0%*	12.0%	**5.3%**	**0.034**	0.804	**1.986**	22.2%	37.3%	80.2%
16×64	0.0	1.5	2.0	**97.5%***	8.5%	3.5%	0.037	1.010	2.142	19.7%	33.4%	82.4%
Unconditional												
1024×1	–	–	–	2.9%	4.6%	2.7%	1.798	1.011	2.815	24.4%	42.1%	78.8%
256×4	–	–	–	–	–	–	–	–	–	21.2%	41.1%	75.3%
64×16	–	–	–	–	–	–	–	–	–	**26.0%**	47.0%	74.7%
16×64	–	–	–	–	–	–	–	–	–	23.7%	52.1%	67.3%

*For derived target=0, hit fraction indicates samples within ≤ 0.1 eV absolute tolerance.

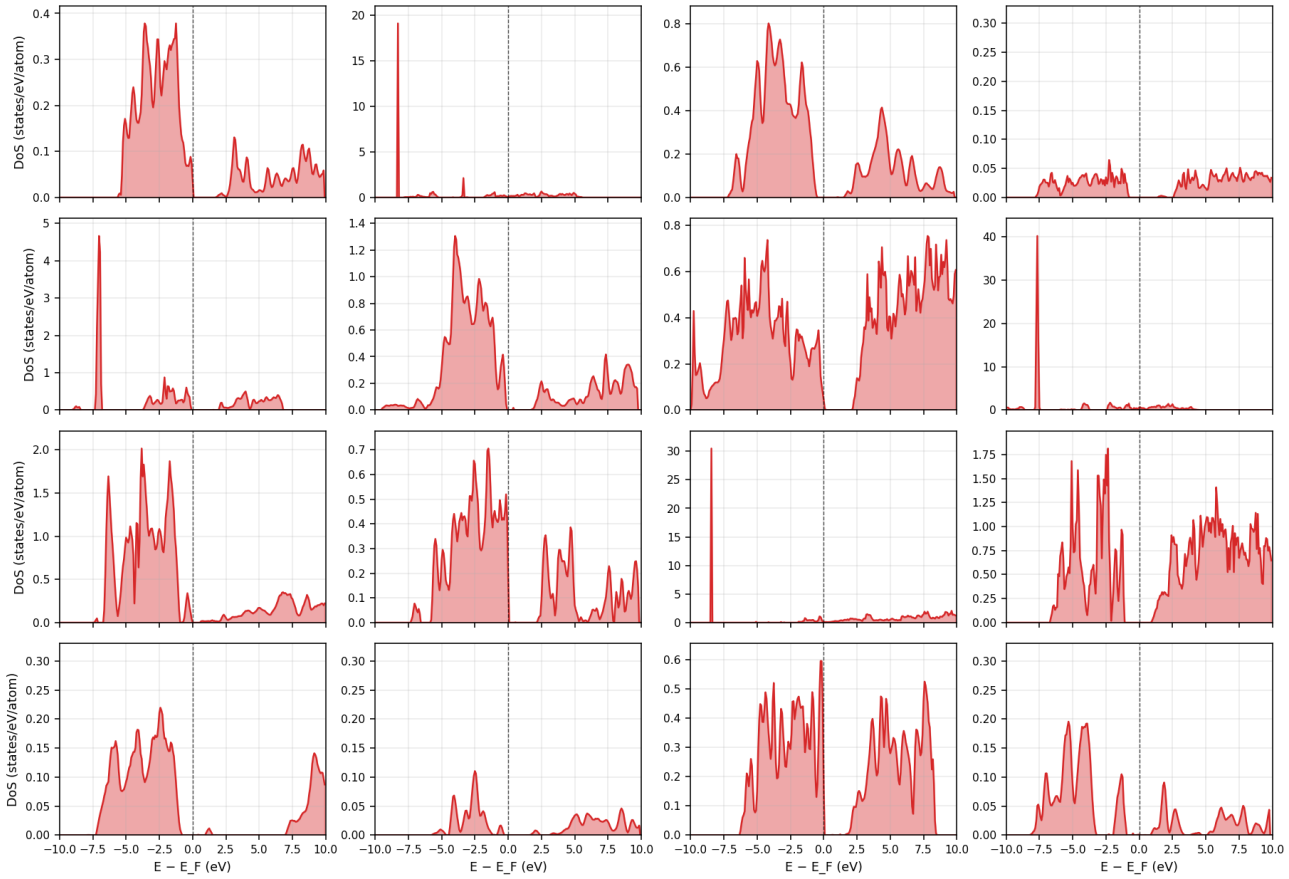

Figure 5. 16 samples from DOSSampler conditioned on target band gap = 2.0 eV.

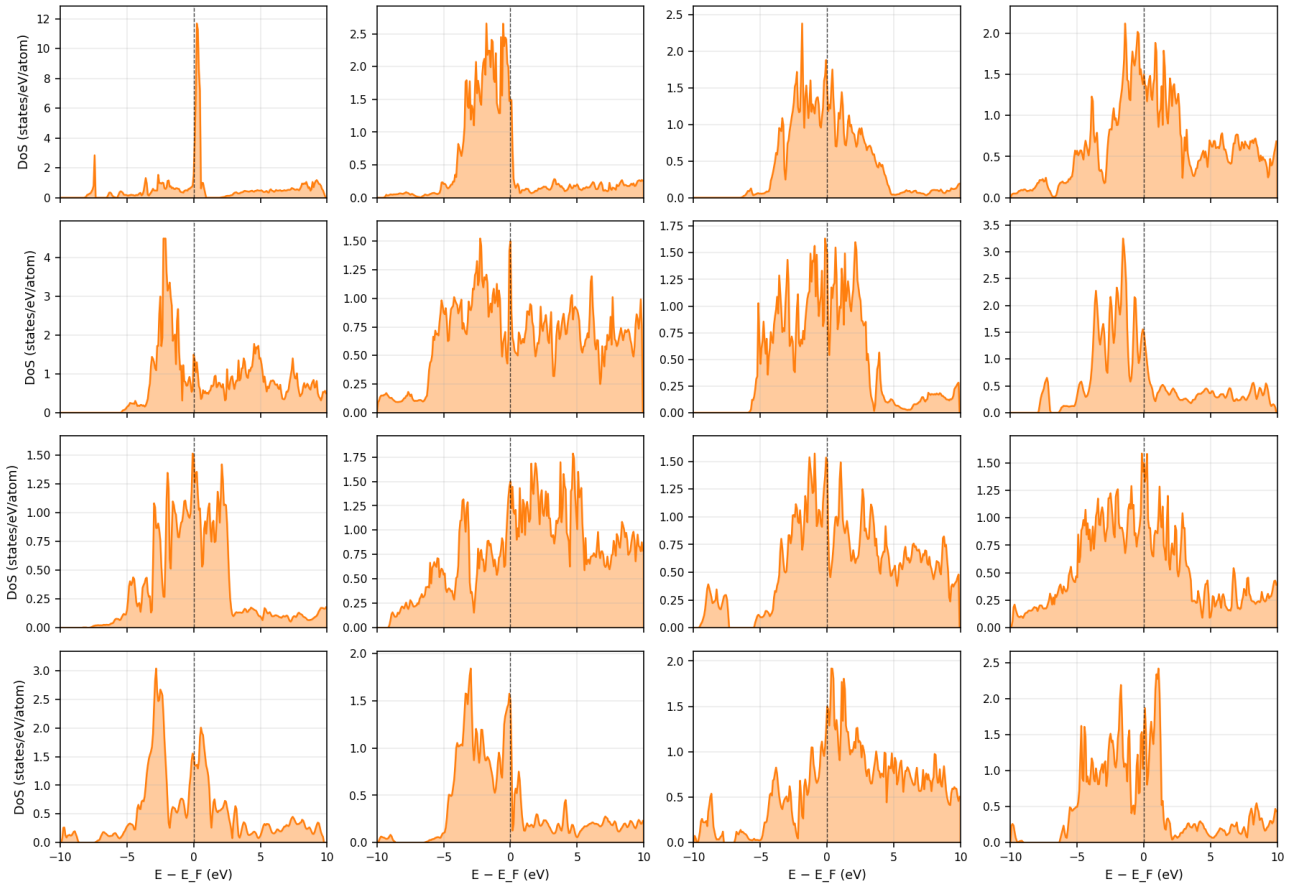

Figure 6. 16 samples from DOSSampler conditioned on target DOS(E_F) = 1.5.

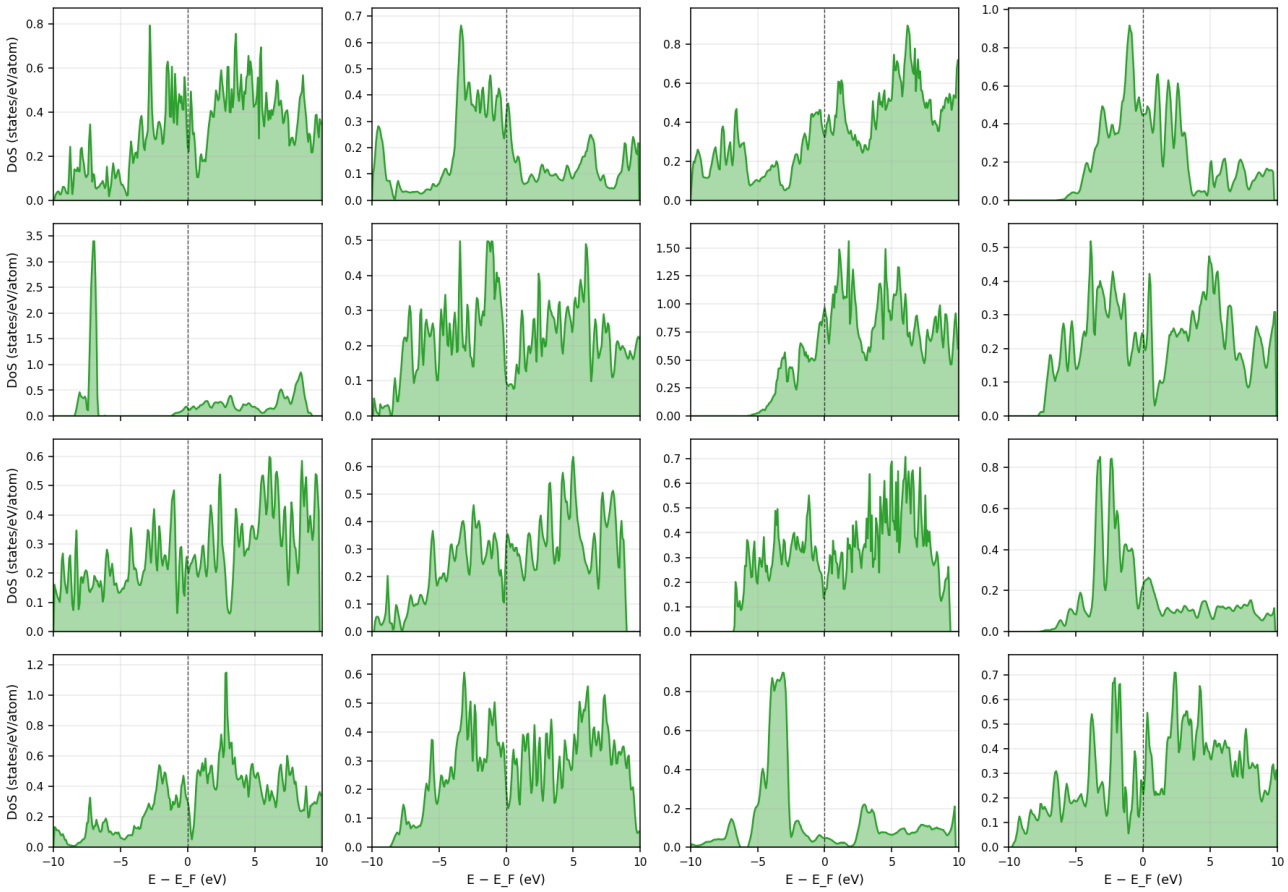

Figure 7. 16 samples from DOSSampler conditioned on target $N(E_F) = 2.0$.

I.2. CDOSGen

Figure 8 shows examples of crystals generated by CDOSGen conditioned on DOS. We select three samples whose predicted DOS attain the highest cosine similarity with the corresponding ground-truth DOS sampled from the Materials Project dataset.

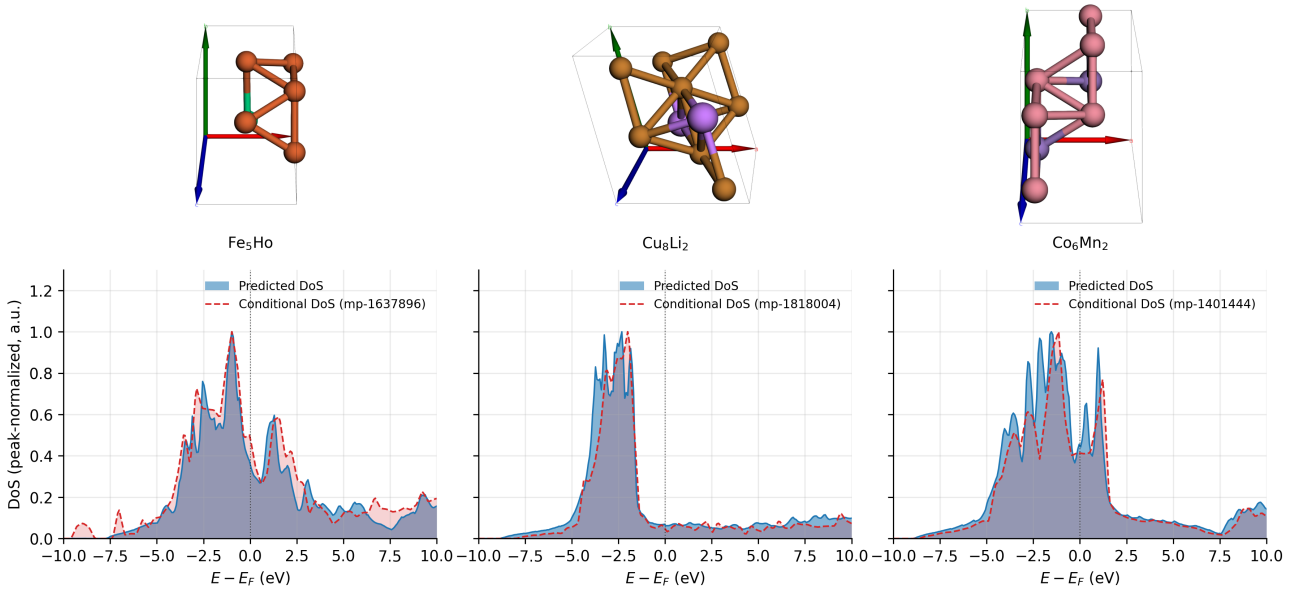

Figure 8. Examples of crystals generated by CDOSGen. Top: generated crystal structures. Bottom: predicted DOS of each generated crystal (blue solid) overlaid on the conditioning DOS sampled from the Materials Project (red dotted).

I.3. Thermoelectric Material

Figure 9 shows examples of Thermoelectric Materials (TEs) generated by CDOSGen conditioned on DOS. For all cases, both the conditioning DOS and the predicted DOS exhibit a band gap of range 0.1 eV to 1.0 eV with other properties such as the DOS gradient at Fermi level.

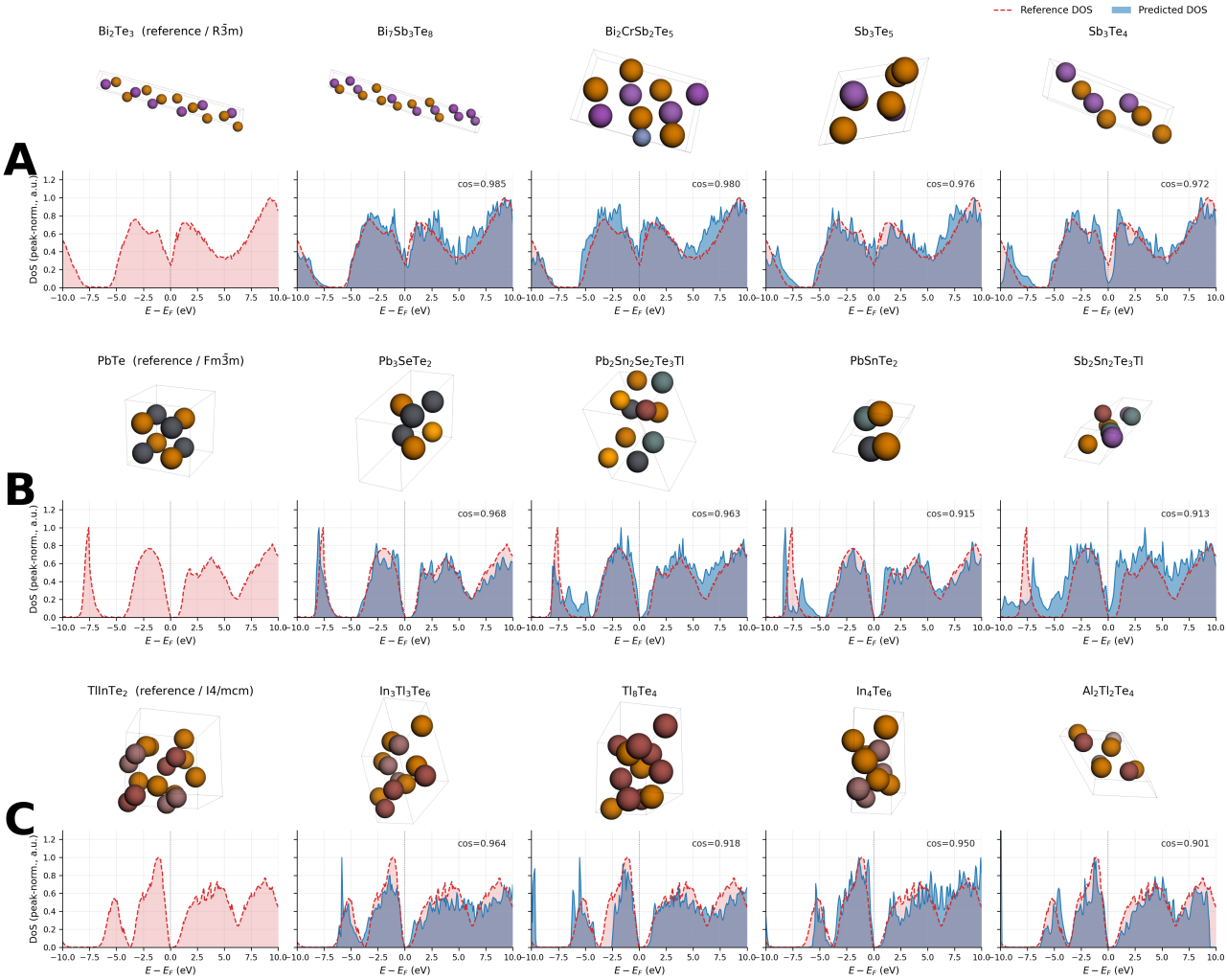

Figure 9. Examples of Thermoelectric Materials generated by Crystal to Crystal Task. Top: generated crystal structures. Bottom: predicted DOS of each generated crystal (blue solid) overlaid on the conditioning DOS (red dotted).