

ELECTRONIC STRUCTURE GUIDED INVERSE DESIGN USING GENERATIVE MODELS

Shuyi Jia[†], Panchapakesan Ganesh^{‡§}, Victor Fung^{†§}

[†]Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA, USA

[‡]Center For Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN, USA

[§]Corresponding authors: ganeshp@ornl.gov, victorfung@gatech.edu

ABSTRACT

The electronic structure of a material fundamentally determines its underlying physical, and by extension, its functional properties. Consequently, the ability to identify or generate materials with desired electronic properties would enable the design of tailored functional materials. Traditional approaches relying on human intuition or exhaustive computational screening of known materials remain inefficient and resource-prohibitive for this task. Here, we introduce DOSMatGen, the first instance of a machine learning method which generates crystal structures that match a given desired electronic density of states. DOSMatGen is an $E(3)$ -equivariant joint diffusion framework, and utilizes classifier-free guidance to accurately condition the generated materials on the density of states. Our experiments find this approach can successfully yield materials which are both stable and match closely with the desired density of states. Furthermore, this method is highly flexible and allows for finely controlled generation which can target specific templates or even individual sites within a material. This method enables a more physics-driven approach to designing new materials for applications including catalysts, photovoltaics, and superconductors.

1 INTRODUCTION

The ability to design novel materials with specific functional properties is fundamental to technological breakthroughs, driving innovation in fields such as energy storage, electronics, and healthcare [1, 2, 3, 4, 5]. While traditional approaches to materials design rely heavily on Edisonian trial-and-error testing, or exhaustive screening through experimental or computational methods, these methods are often both time-consuming and resource-intensive [6, 7]. In contrast, recent advances in deep generative models have enabled them to be a potentially transformative approach to materials design. Specifically, this approach to inverse design eliminates the need for exhaustive screening of the materials space by working backwards to create materials from the ground up via a data-driven process [8, 9, 10, 11].

Deep generative models for materials design can be trained to conditionally generate the precise crystal structure of materials, which allows for further computational investigation followed by eventual experimental synthesis and validation. Current state-of-the-art approaches generally operate by generating new candidates within a given materials representation, or by applying an iterative process such as denoising directly on atomic coordinates in the Cartesian space. Examples of the former case include text-based representations for autoregressive language modeling of materials [12, 13, 14], 3D voxelized images for variational autoencoders and generative adversarial networks (GANs) [15, 16, 17], among others [18, 19, 20, 21]. In the latter case, a seminal early example would be the crystal diffusion variational autoencoder (CDVAE) [22], which employs a diffusion model [23] to denoise atomic coordinates while predicting composition and lattice parameters from a latent representation. Building on this foundation, recent diffusion-based and flow-matching approaches have advanced to jointly generate all relevant structural and compositional attributes in a unified process [24, 25, 26]. Using these methods, new materials have been generated with desired properties which have been validated with density functional theory (DFT) [27, 28, 29, 30] and in some cases, experimental studies [25].

In this work, we introduce DOSMatGen, a conditional diffusion model designed to generate stable crystalline materials with targeted atomic-level electronic properties. Extending the periodic $E(3)$ -equivariant denoising framework of DiffCSP [24], DOSMatGen achieves high-quality conditional generation by carefully tuning the balance between the structural validity of the generated materials and their alignment with the target properties. As a concrete demonstration, we focus on conditioning the diffusion process on node-level density of states (DOS). The DOS corresponds to a 1D spectral property that encodes the key electronic-level functionality of a material, from its metal, semiconducting, insulating states, to more ex-

otic quantum phases such as superconductivity and topological quantum states. Prior deep-learning studies have yielded models which can predict the DOS from a given atomic structure [31, 32], but the *inverse process* of generating materials matching a given DOS remains highly challenging. To effectively utilize the DOS information as a condition for structure generation, we implemented and evaluated two distinct guidance strategies for diffusion models: classifier-based guidance [33] and classifier-free guidance [34]. We find classifier-free guidance outperforms classifier-based guidance, offering a more robust mechanism for aligning generated structures with the desired properties. Next, using the classifier-free guidance, we demonstrate DOSMatGen’s capability for fine-grained control over atomic structure generation through three constrained generation tasks that selectively impose distinct constraints on specific subsets of atoms within a single structure. Finally, we demonstrate a post-processing workflow to yield stable, optimized crystal structures which are ready for DFT validation. An overview of the DOSMatGen framework is shown in Fig. 1.

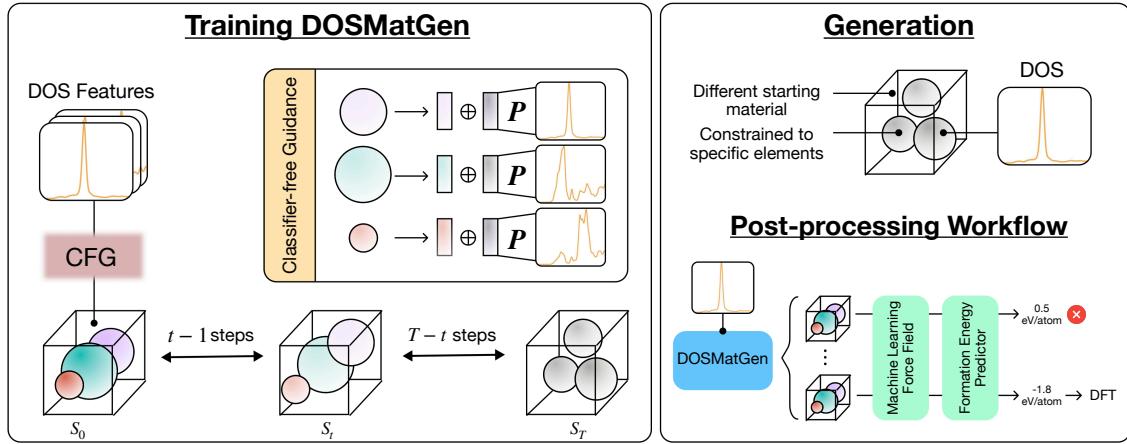


Figure 1: Overview of DOSMatGen. The training process of the denoising neural network in DOSMatGen employs classifier-free guidance (CFG). A projection layer P maps node-level density of states (DOS) features into the same space as the node embeddings, which are then combined through element-wise addition. During training, a structure S_0 is progressively *noised* to S_t at timestep t and passed into the network to predict the noise added. During generation, DOSMatGen can condition one or more atoms on target DOS features, ensuring the denoised structure aligns with the specified conditions. Additional constraints, such as fixing element types or using structural templates as starting points, can be incorporated for finer control over the generation process. Furthermore, DOSMatGen can be integrated as part of larger materials discovery workflows, generating large sets of candidate structures that are subsequently relaxed using a machine learning force field (MLFF) and evaluated with a formation energy predictor. Candidates below a specified formation energy threshold can then be filtered for further analysis, including DFT validation.

2 METHODOLOGY

2.1 REPRESENTATION OF PERIODIC MATERIALS

A periodic material is usually represented by its unit cell—the smallest repeating arrangement of atoms in Cartesian coordinates that, when translated in space, generates the complete crystal structure. Given a structure S with N atoms in its unit cell, it can be fully described by the tuple $S = (\mathbf{A}, \mathbf{X}, \mathbf{L})$, where $\mathbf{L} = (l_1, l_2, l_3) \in \mathbb{R}^{3 \times 3}$ is a matrix of lattice vectors, $\mathbf{X} = (x_1, \dots, x_N) \in \mathbb{R}^{N \times 3}$ represents the Cartesian coordinates of the N atoms within the unit cell, and $\mathbf{A} \in \{0, 1\}^{N \times Z}$ is a matrix of one-hot encoded representations of Z possible elements for each atom. Note that the unit cell can be translated arbitrary in space by \mathbf{L} . In this work, we use fractional coordinates, obtained by transforming Cartesian coordinates x through the inverse lattice matrix: $x_f = \mathbf{L}^{-1}x$. For notational clarity, we omit the subscript f when the coordinate system is contextually clear.

2.2 JOINT DIFFUSION MODEL

Given a structure $S = (\mathbf{A}, \mathbf{X}, \mathbf{L})$, DOSMatGen simultaneously diffuses the atom types \mathbf{A} , the fractional coordinate matrix \mathbf{X} and the lattice \mathbf{L} . Given a timestep $t \in [0, T]$, $S_t = (\mathbf{A}_t, \mathbf{X}_t, \mathbf{L}_t)$ denotes the intermediate *noisy* structure. In particular, the forward diffusion process deterministically adds noise to S_0 , while the backward denoising process iteratively samples from the prior distribution S_T to recover the original structure S_0 . If the prior distribution $p(S_T)$ is invariant and the Markov transition $p(S_{t-1}|S_t)$ is equivari-

ant, then the recovered distribution from S_T will be periodic E(3)-invariance [35]. The overall objective for training the joint diffusion model on \mathbf{A} , \mathbf{X} and \mathbf{L} is

$$\mathcal{L}_S = \lambda_{\mathbf{A}} \mathcal{L}_{\mathbf{A}} + \lambda_{\mathbf{X}} \mathcal{L}_{\mathbf{X}} + \lambda_{\mathbf{L}} \mathcal{L}_{\mathbf{L}}, \quad (1)$$

where λ represents the corresponding weight given to each individual loss component. The full diffusion objectives can be found in Appendix A.

2.3 CLASSIFIER-BASED GUIDANCE

Guided diffusion models typically generate structures S from the conditional distribution $p(S|y)$, where y represents the target label or condition. In classifier-based guidance [33], this is achieved by leveraging Bayes' Rule to guide the diffusion process:

$$\nabla_{S_t} \log p_t(S_t|y) = \nabla_{S_t} \log p_t(S_t) + \nabla_{S_t} \log p_t(y|S_t), \quad (2)$$

where the gradient of the log conditional probability $\nabla_{S_t} \log p_t(S_t|y)$ at timestep t given the condition y can be decomposed into two terms: $\nabla_{S_t} \log p_t(S_t)$ and $\nabla_{S_t} \log p_t(y|S_t)$. The first term corresponds to the unconditional generative model, which guides the generation of S_t and is generally approximated via a neural network $f_{\theta}(S_t, t)$. The second term represents the gradient from a time-dependent classifier, which estimates how likely S_t is to match the condition y and is approximated by a separate neural network $g_{\phi}(y|S_t, t)$. In practice, because the unconditional model and the classifier are separate neural networks, we train them independently.

2.4 CLASSIFIER-FREE GUIDANCE

In contrast to classifier-based guidance, classifier-free guidance [34] integrates the conditional generation directly into the diffusion model by training a single neural network for both conditional and unconditional models. Instead of learning the unconditional score function $\nabla_{S_t} \log p_t(S_t)$ independently, classifier-free guidance learns the unconditional and conditional distributions jointly:

$$\nabla_{S_t} \log p_t(S_t, c), \quad c = \begin{cases} \emptyset & \text{with probability } p_{\text{uncond}} \\ y & \text{else} \end{cases} \quad (3)$$

where p_{uncond} is some probability for unconditional generation, set as a hyper-parameter. During sampling, the mixed score function becomes $(1 + \omega) \nabla_{x_t} \log p_t(x_t, y) - \omega \nabla_{x_t} \log p_t(x_t, \emptyset)$, where ω is the guidance strength. In practice, we project y into the same space as the node representations, and combine them through element-wise addition. In our experiments, we let $p_{\text{uncond}} = 0.2$ and $\omega = 1$.

2.5 DATASET

Our training dataset for the denoising neural network, which we henceforth refer to as "MP DOS," comprises of 50,790 inorganic crystal structures curated from the Materials Project repository [36]. The target property is the atom-level density of states (DOS), with each atom in every structure labeled by its corresponding DOS. The MP DOS dataset is divided into training, testing, and validation splits in a $0.6 : 0.2 : 0.2$ ratio, with only the training split used to train the denoising network, as well as the forward model in classifier-based guidance.

3 RESULTS

3.1 CONDITIONAL GENERATION

We first evaluate the two diffusion model guidance strategies—classifier-based and classifier-free using the testing split of the MP DOS dataset. At the initial timestep $t = T$ of generation, we specify the number of atoms and the DOS while randomly sampling atom types \mathbf{A}_T , fractional coordinates \mathbf{X}_T , and lattice matrix \mathbf{L}_T from their respective prior distributions $p(\mathbf{A}_T)$, $p(\mathbf{X}_T)$ and $p(\mathbf{L}_T)$. In other words, our goal is to reconstruct the entire testing split of the MP DOS dataset using the atom-level DOS as the conditional input. In addition to $T = 1000$, which represents generation from fully randomized initialized structures, we perform additional experiments by generating structures from timesteps $T = 200$ and $T = 500$. These cases involve partially *noising* the original testing split structures using their forward noising processes (see Appendix A). As the timestep decreases, more structural information from the original data is retained, leading to generated structures that more closely resemble those in the testing split.

To compare the DOS of generated structures with the ground truth conditioning data, we trained a separate surrogate model that predicts atom-level DOS from input crystal structures. We first evaluate the conditional generation by calculating the mean absolute errors (MAEs) between generated structures’ DOS and the ground truth DOS. In particular, we let y denote the ground truth DOS used to condition the generation process. We define \hat{y} as the predicted DOS from the ground truth structures, and \hat{y}_{gen} as the predicted DOS from the generated structures. Table 1 presents the MAEs comparing: (1) y with \hat{y}_{gen} , and (2) \hat{y} with \hat{y}_{gen} . We observe that both $\text{MAE}(\hat{y}_{\text{gen}}, y)$ and $\text{MAE}(\hat{y}_{\text{gen}}, \hat{y})$ increase as the starting timestep T of the diffusion process becomes greater. This aligns with the fact that as T increases, less information of the original crystal structures is retained, making the diffusion process more reliant on generating high-quality structures from pure noise. Consequently, the generated structures are more prone to deviate further from the ground truth, leading to higher errors in the predicted DOS.

When comparing classifier-based and classifier-free guidance, we observe from Table 1 that classifier-free guidance consistently achieves better performance. The difference in error between the two strategies is relatively small at $T = 200$, which is expected, as the model retains enough information from the original ground truth structures to enable both models to denoise more effectively to the ground truth structures, and resulting in lower values of both $\text{MAE}(\hat{y}_{\text{gen}}, y)$ and $\text{MAE}(\hat{y}_{\text{gen}}, \hat{y})$. However, at the starting timestep $T = 1000$, where no information from the original structures is present, the performance of classifier-based guidance deteriorates significantly. Specifically, when compared to that at $T = 200$, $\text{MAE}(\hat{y}_{\text{gen}}, y)$ at $T = 1000$ increases from 0.114 to 0.268, representing a 135% increase, while $\text{MAE}(\hat{y}_{\text{gen}}, \hat{y})$ rises from 0.068 to 0.256, marking a 276% increase. In contrast, classifier-free guidance exhibits significantly lower MAE increases from timestep $T = 200$ to $T = 1000$: a 17.6% rise in $\text{MAE}(\hat{y}_{\text{gen}}, y)$ and a 79.6% increase in $\text{MAE}(\hat{y}_{\text{gen}}, \hat{y})$. In particular, the forward surrogate model’s own error, $\text{MAE}(\hat{y}, y)$, is 0.096. This indicates that even at $T = 1000$, classifier-free guidance generates structures that align with the target DOS reasonably well with an MAE of 0.120. Additionally, at $T = 1000$, 314 out of 10,158 structures generated with classifier-based guidance fail to even map to a valid crystal graph for the forward surrogate model, highlighting its inferior performance compared to classifier-free guidance.

	$T = 200$		$T = 500$		$T = 1000$	
	$\text{MAE}(\hat{y}_{\text{gen}}, y)$	$\text{MAE}(\hat{y}_{\text{gen}}, \hat{y})$	$\text{MAE}(\hat{y}_{\text{gen}}, y)$	$\text{MAE}(\hat{y}_{\text{gen}}, \hat{y})$	$\text{MAE}(\hat{y}_{\text{gen}}, y)$	$\text{MAE}(\hat{y}_{\text{gen}}, \hat{y})$
Classifier	0.114	0.068	0.200	0.182	0.268 [†]	0.256 [†]
Classifier-free	0.102	0.049	0.109	0.065	0.120	0.088

Table 1: Mean absolute errors (MAEs) of density of states (DOS) comparing structures generated using classifier-based and classifier-free guidance. \hat{y}_{gen} is the DOS predicted from generated structures, \hat{y} is the DOS predicted from ground truth structures and y is the target DOS from test set. Comparisons are made across three different denoising timesteps, $T \in \{200, 500, 1000\}$, with $T = 1000$ representing generation from fully random initializations. [†]: 314 out of 10,158 generated structures were excluded due to errors in constructing the crystal graph. The forward surrogate model’s intrinsic error, $\text{MAE}(\hat{y}, y)$, is 0.096.

	$T = 200$		$T = 500$		$T = 1000$	
	Struct. Match %	Comp. Match %	Struct. Match %	Comp. Match %	Struct. Match %	Comp. Match %
Classifier	63.5	73.8	4.23 [†]	6.01	0.07 [‡]	2.95
Classifier-free	81.6	94.4	42.0	73.0	14.7	58.7

Table 2: Structure match rate and composition match rate for structures generated using classifier-based and classifier-free guidance, compared to the test set. The structure match rate is the percentage of structures satisfying the tolerances (`stol=0.5`, `angletol=10`, `ltol=0.3`) using `pymatgen`’s `StructureMatcher`. The composition match rate is the percentage of structures whose compositions perfectly match those of the test set. Comparisons are made across three different denoising timesteps, $T \in \{200, 500, 1000\}$, with $T = 1000$ representing generation from fully random initializations. ^{†‡}: 9/372 out of 10,158 generated structures were excluded due to timeouts during metric calculation.

In Table 2, we evaluate the performance of the conditional generation using (1) structure match rate and (2) composition match rate. The structure match rate quantifies the percentage of generated structures that are geometrically similar to the ground truth structures, as determined by the `StructureMatcher` algorithm from `pymatgen` within specified tolerances [22, 37]. The composition match rate represents the percentage of generated structures that exactly match the ground truth structures in terms of chemical composition. From Table 2, we observe that as T increases, both structure and composition match rates decreases, highlighting that higher amount of noise during the generation process result in structures that

deviate further from the ground truth. Comparing the two guidance strategies, classifier-free guidance consistently demonstrates superior performance. At timestep $T = 1000$, it maintains a structure match rate of 14.7% and a composition match rate of 58.7%, significantly outperforming classifier-based guidance’s match rates of 0.07% and 2.95%, respectively.

In Fig. 2, we present the distribution of formation energies per atom for the structures generated in Table 1 at $T = 1000$. Structure relaxation and formation energy calculations for all structures are performed using `Orb-v2` [38], an accurate and efficient universal machine learning force field (MLFF). The ground truth distribution corresponds to the original 10,158 structures from the test set. The results in Fig. 2 show that structures generated with classifier-free guidance closely match the ground truth distribution of formation energies, exhibiting significant overlap. Notably, the majority of these structures—relaxed or unrelaxed—have formation energies below 0 eV/atom, indicating a reasonable range of thermodynamic stability. In contrast, structures generated with classifier-based guidance are noticeably shifted toward higher formation energies and deviate further from the ground truth distribution. The distribution plots for structures generated at $T = 200$ and $T = 500$ are provided in Appendix B.

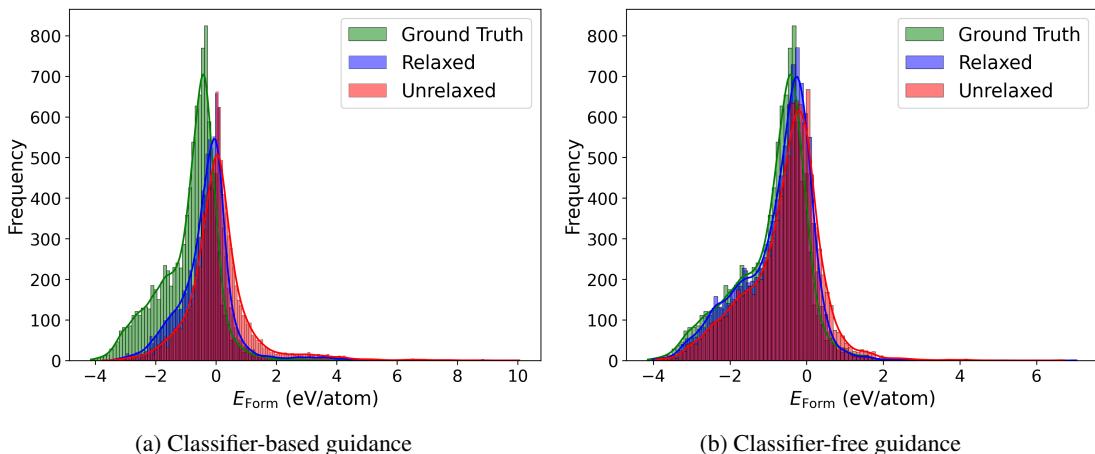


Figure 2: Distribution of formation energies per atom (eV/atom) for structures conditionally generated at $T = 1000$ using (a) classifier-based guidance and (b) classifier-free guidance. All generated structures are relaxed using the `Orb-v2` interatomic potential. The ground truth distribution corresponds to the formation energies of the original test set structures, with all energies predicted using `Orb-v2`.

From Table 1 and 2, it is clear that classifier-free guidance consistently outperforms classifier-based guidance. It achieves lower MAEs across all starting timesteps, and maintains significantly higher structure and composition match rates at $T = 1000$. While classifier-free guidance excels in generating structures that align closely with the ground truth both structurally and compositionally, it is important to recognize that such strong alignment can be a double-edged sword. On one hand, it naturally leads to structures which match better to the desired DOS, which is central to our conditional generation approach. On the other hand, this convergence towards the ground truth structures may come at the cost of reduced diversity in the generated structures, which would be detrimental to the goal of finding novel materials. Based on our evaluation, we find the classifier-free version of DOSMatGen yields good agreement with the desired DOS while having a moderate match rate which is neither too high nor too low, making it suitable for materials discovery tasks. Next, we will apply DOSMatGen to three different types of constrained generation tasks to further assess the model.

3.2 CONSTRAINED GENERATION

Here, we evaluate DOSMatGen’s ability to achieve fine-grained control over atomic structure generation by applying targeted constraints to a specific subset of atoms within a single structure. Classifier-free guidance is utilized in all constrained generation. The three constrained generation tasks are 1) masked generation, 2) template guided generation and 3) fixed atom type generation. A visual representation of the three constrained generation tasks is shown in Fig. 3. For demonstration, we employ the nickel DOS from Au_3Ni^1 as the target DOS for conditional generation across all three tasks. Au_3Ni is chosen as a characteristic example of an alloy where the minority element, Ni in this case, exhibits “free-atom like” (d-)states which are sharply localized in energy. Materials with sites exhibiting free-atom like states can have unique physical and chemical properties, including in applications towards catalysis [39, 40, 41, 42]. Prior studies have shown that free-atom like states can induce strong chemisorption and facile bond activation, though materials which exhibit this property can be relatively rare. Therefore, using DOSMatGen to generate other

¹Materials Project ID: mp-976806.

instances of materials with free-atom like states serves as an illuminating test for the capability of this model to generalize well to rare cases in the training dataset.

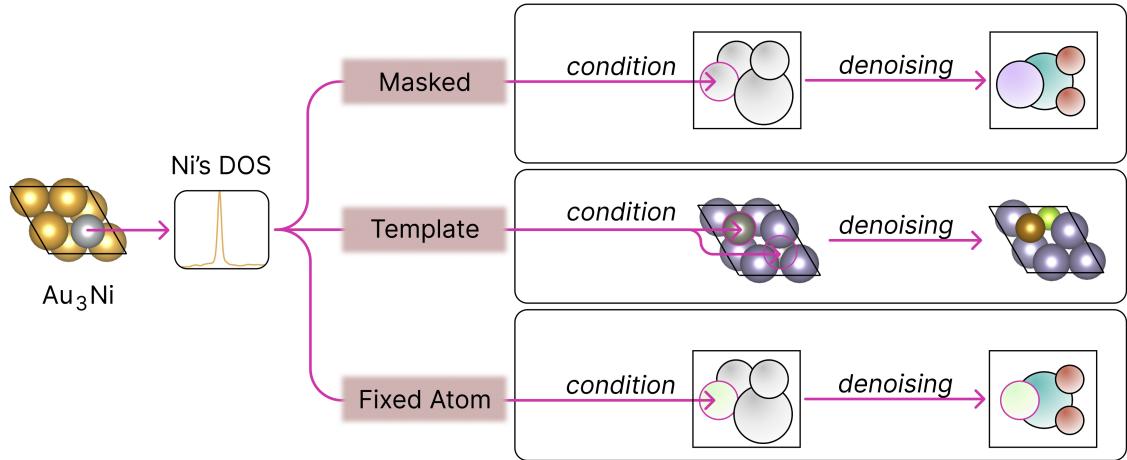


Figure 3: A visual representation of the three constrained generation tasks. (1) Masked generation: a single atom of the structure is conditioned on Ni’s DOS, while the remaining atoms are left unconditioned. (2) Template guided generation: Instead of complete random initialization, use an existing material as a template. A subset of the template material’s atoms is then conditioned on Ni’s DOS. (3) Fixed atom type generation: a single atom of the structure is conditioned on Ni’s DOS, but with its atom type held fixed (e.g., pre-defined as Co).

3.2.1 MASKED GENERATION

In the masked generation task, we seek to match the target DOS for an arbitrary subset of atoms, while allowing the DOS of the unconditioned atoms to vary freely, thereby enabling the generation of diverse structures that still meet the specified constraints. Effectively, we are generating new local atomic environments which yield the same desired DOS, which still come together to form a stable crystal structure. Constraining only a subset of atoms rather than the entire material also allows for greater diversity in the generated structures, without compromising on the overall goal of finding materials with free-atom like states. We begin the generation procedure at starting timestep $T = 1000$, where \mathbf{A}_T , \mathbf{X}_T and \mathbf{L}_T are sampled from their respective prior distributions. In the masked generation case, we constrain the number of atoms to the set $N_{\text{atoms}} \in \{4, 5, 6, 7\}$, though this can be readily changed as needed. For each value of N_{atoms} , we condition the first atom using Ni’s DOS from Au_3Ni , while leaving the remaining atoms unconditioned. We then generate 50 candidate structures for each N_{atoms} and select a representative structure to visualize its first atom’s DOS with the ground truth in Fig. 4.

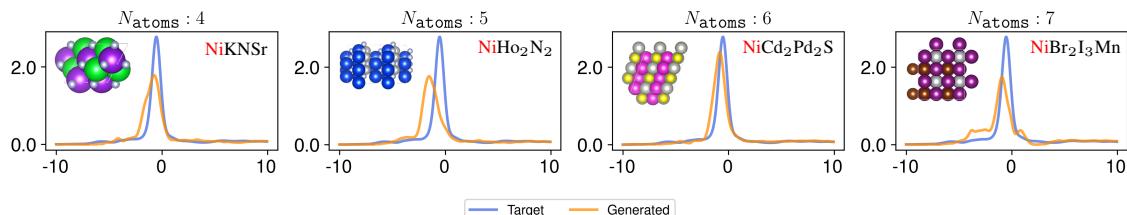


Figure 4: Density of states (DOS) for generated materials with $N_{\text{atoms}} \in \{4, 5, 6, 7\}$. The orange curve represents the DOS of the first atom in each generated material, conditioned on Ni’s DOS from Au_3Ni (shown in blue). The element type of the first atom is indicated in red. Visualizations of all structures are of a $2 \times 2 \times 2$ supercell.

As shown in Fig. 4, the ground truth DOS of the nickel atom in Au_3Ni features a centralized sharp peak. Similarly, the four selected generated materials, each with a different N_{atoms} , also exhibit a centralized sharp peak near the ground truth. While the first atom in each generated material is explicitly conditioned on the reference Ni DOS, the other atoms are free to adopt different element types and locations. Consequently, the majority of the generated structures end up with Ni for the conditioned atom (even though it is not explicitly specified), while the other atoms in the unit cell end up with many diverse compositions. This test therefore successfully demonstrates the ability of DOSMatgen to recreate the Ni electronic structure within Au_3Ni in novel materials are dramatically different from the original reference material.

3.2.2 TEMPLATE GUIDED GENERATION

In this task, we investigate the use of an existing material to serve as a template for the diffusion process. Instead of sampling \mathbf{A}_T , \mathbf{X}_T , and \mathbf{L}_T from their prior distributions at $T = 1000$ —representing the generation from complete noise—we initiate the diffusion process at $T < 1000$. This approach involves adding noise to a specified template material at the chosen timestep. Similar to the masked generation task, only a subset of the atoms of the template material is conditioned on a target DOS, while the DOS of the unconditioned atoms can vary freely. This type of task can be useful when performing inverse design within specific materials classes such as perovskites or two-dimensional materials. For illustration, we use Zn_2Sn_6 ² as the template material and initialize the diffusion process at $T \in \{200, 350, 500\}$. In this task, the Zn atoms in Zn_2Sn_6 are conditioned to match Ni’s DOS. We generate 50 candidate structures for each T and select a representative structure to visualize its conditioned atoms’ DOS with the ground truth in Fig. 5.

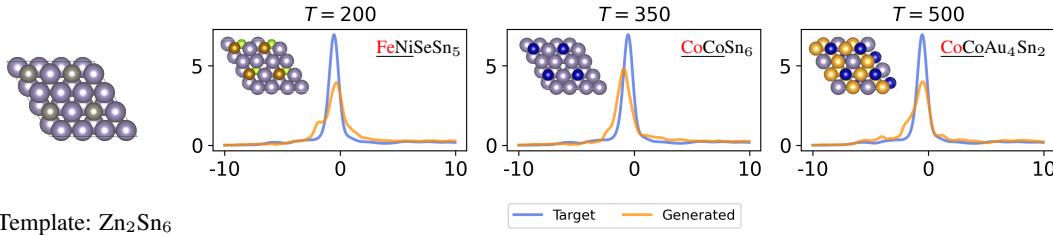


Figure 5: Density of states (DOS) for generated materials from starting timestep $T \in \{200, 350, 500\}$. The template material Zn_2Sn_6 is shown on the far left. The orange curve represents the DOS of the first atom in each generated material, conditioned on Ni’s DOS from Au_3Ni (shown in blue). The conditioned atoms are underlined, with the first atom highlighted in red to indicate that the plotted DOS corresponds to it. Visualizations of all structures are of a $2 \times 2 \times 2$ supercell.

From Fig. 5, we can observe that the DOS of the first atom in the generated materials closely aligns with the target DOS, but still maintains the 3D structure of the original template. Due to the diffusion process starting from a template material *noisified* at $T < 1000$, the final composition exhibits a strong bias towards the original Sn, though at higher T , the composition moves farther away from Sn to become other elements. Interestingly, the original Zn atoms, which are conditioned on Ni’s DOS, are replaced by other element types such as Fe and Co rather than Ni. This suggests that DOSMatGen can propose chemically viable substitutions for the conditioned atoms by considering the overall composition, even when the target DOS is directly associated with that of Ni.

3.2.3 FIXED ATOM TYPE GENERATION

In this task, we explore the feasibility of fixing the first atom’s element type to an arbitrary specified element. Similar to the masked generation task, the diffusion process begins at $T = 1000$, meaning the materials are generated entirely from noise. The key difference in this task is that the first atom is constrained to a predefined element type, whereas in the masked generation task, the model can freely diffuse the first atom into any element type. The first atom in the generated structure is still conditioned on the DOS of Ni from Au_3Ni . In our experiments, we define a constraint on the element type to be Ti, Co, Cu, or Pd. Co and Cu are selected because their proximity to Ni in the same period, and similarly Pd is chosen for being in the same group as Ni. Meanwhile, Ti is included as an element which bears little chemical similarity to Ni, to test if the model can nevertheless generate an atomic environment to enable Ti to behave electronically like Ni. For each element type, we generate 50 candidate structures and select a representative structure to visualize its first atom’s DOS with the ground truth in Fig. 6.

Fig. 6 reveals that when the first atom is constrained to neighboring transition metals such as Co, Cu, and Pd, the resulting density of states (DOS) demonstrates remarkable alignment with the target DOS. This is to be expected due to the proximity of Pd, Co and Cu to Ni in the periodic table. Meanwhile, the DOS of the Ti-constrained case is not as similar to target DOS, though it still exhibits a peak near the ground truth peak location, suggesting that this is a harder task, but one which DOSMatGen is still able to follow the guidance to some extent. In addition, in all four generated structures shown in Fig. 6, we observe high compositional diversity in the unconditioned atoms. These are exciting observations, as it suggests that DOSMatGen can provide viable elemental substitutions while maintaining a similar electronic structure. This concept can conceivably be expanded to tackle pressing materials design challenges such as finding earth-abundant and low-toxicity substitutions to existing functional materials such as noble metal catalysts or lead-based halide perovskites.

²Materials Project ID: mp-971919.

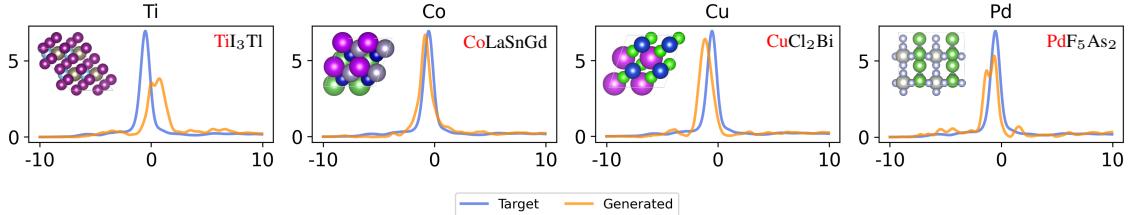


Figure 6: Density of states (DOS) for generated materials with the first atom constrained to Ti, Co, Cu or Pd. The orange curve represents the DOS of the first atom in each generated material, conditioned on Ni’s DOS from Au_3Ni (shown in blue). Visualizations of the corresponding structures are of a $2 \times 2 \times 2$ supercell.

3.3 POST-PROCESSING WORKFLOW

The results provided so far provide illustrative examples of the capabilities of DOSMatGen for generating well-conditioned crystal structures under various constraints. Here, we show how DOSMatGen can be further incorporated into a more comprehensive materials discovery workflow to generate large quantities of high-quality materials. The workflow involves using DOSMatGen to first conditionally generate many candidate structures (i.e. 10,000 or more), followed by structural relaxation using a universal machine learning force field (MLFF) to yield local minimum structures. After relaxation, the formation energy per atom (in eV/atom) is calculated for all structures based on MLFF energies. Structures with formation energies (or optionally, energy below hull) below a desired threshold are then filtered. From this subset, the predicted DOS of the generated materials is compared against the ground truth DOS, and the best-matched structures are then selected for further evaluation with DFT calculations. These post-processing steps help ensure the generated materials will all be locally minimum, stable, and well-matched in terms of the DOS.

We select the DOS of Ir from rutile IrO_2 as the target, which presents more a complex target DOS than the previous example. IrO_2 is widely considered to be a state-of-the-art electrocatalyst for oxygen evolution [43, 44, 45]. Designing new materials which exhibit the electronic structure of Ir could help lead to similarly well-performing catalysts. Here we first generate 10,000 candidate structures, with number of atoms in the unit cell within the range $N_{\text{atoms}} \in \{4, 5, 6, 7, 8\}$. For each value of N_{atoms} , 2,000 structures are generated, ensuring an even distribution across the possible values. We then use Orb-v2 [38] as the MLFF for structure relaxation and for calculating the formation energy per atom for all structures. Structures with formation energies ≤ -1.5 eV/atom are then selected.

Fig. 7 shows the distribution of formation energies per atom for the 10,000 unrelaxed and relaxed structures. As expected, the distribution for the relaxed structures is shifted to the left compared to the unrelaxed structures, demonstrating that structural relaxation with an MLFF effectively lowers the formation energies. Notably, both distributions closely resemble Gaussian distributions centered around zero, with significant overlap between them. This indicates that DOSMatGen can conditionally generate relatively stable crystal structures even without structural optimization, and the the MLFF is mainly used for providing further refinement.

Applying a formation energy threshold of ≤ -1.5 eV/atom yields a total of 108 candidate structures, from which 8 are selected for further DFT validation based on closeness to the target DOS. Fig. 8 compares the target DOS with the generated DOS, obtained using either the surrogate GNN model or computed from DFT. While the 8 selected structures do not perfectly reproduce all features of the target DOS, they generally show strong resemblance, capturing key peaks and overall trends. This remains true under DFT validation as well. Visualizations of the eight selected structures are provided in Appendix C. Here, we note that the final materials all tend towards containing Ir in some rutile or rutile-like configuration. It is an unsurprising result, given the target DOS is derived from Ir in rutile IrO_2 , though it fails to yield the desired earth-abundant alternative materials. This may be attributed to the fact that matching the DOS of Ir exactly may be too strong a constraint leading to low diversity, whereas it may only be necessary to match certain regions in the DOS (i.e. near the Fermi level). At the same time, it may also suggest the intrinsic difficulty of this particular problem, which we leave for future investigation.

4 CONCLUSION

In this work, we introduced DOSMatGen, an $E(3)$ -equivariant joint diffusion framework for the conditional generation of crystalline materials guided by quantities from electronic structure such as the DOS. DOSMatGen excels in generating compositionally diverse and stable structures while maintaining strong

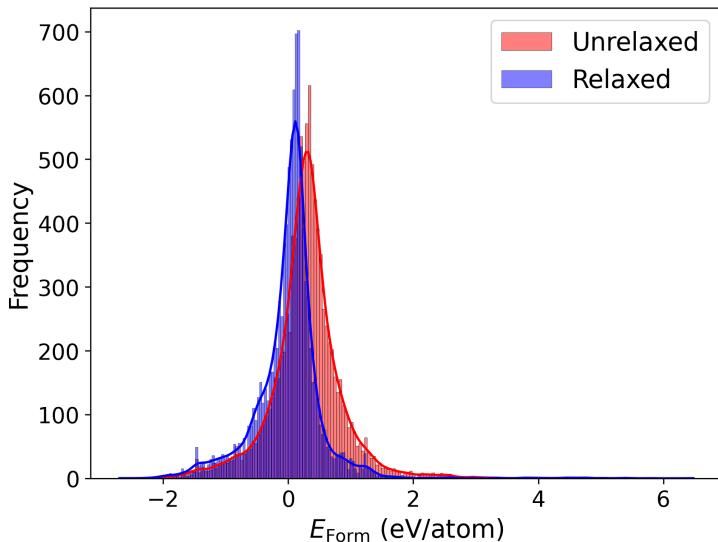


Figure 7: Distribution of formation energies per atom (eV/atom) for 10,000 generated structures before (unrelaxed) and after (relaxed) structural optimization using Orb-v2 . The histogram shows the frequency of structures across different formation energies.

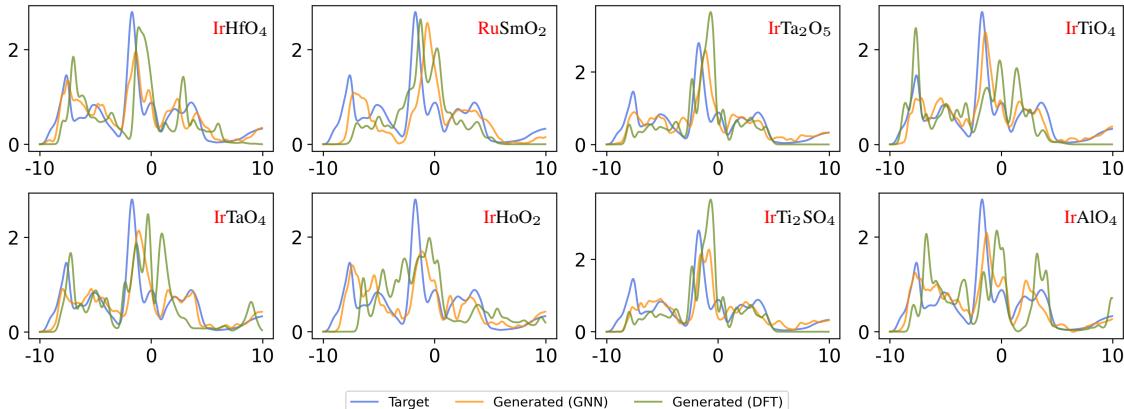


Figure 8: Density of states (DOS) for generated materials with the first atom conditioned on the DOS of Ir from IrO_2 . The blue curve represents the ground truth target DOS. The orange curve represents the DOS of the first atom in each generated material, predicted using the forward GNN model. The green curve represents the DOS of the first atom computed via DFT. The first atom’s element type is highlighted in red in the composition.

alignment with target properties. The model demonstrates remarkable versatility across constrained generation tasks, including masked generation, initialization from template materials, and atom-type constraints, enabling precise and fine-grained control over the generated structures. Additionally, through a large-scale screening approach, DOSMatGen effectively identifies stable candidate materials with complex electronic properties such as those of IrO_2 . The ability to balance structural validity, property alignment, and compositional diversity positions DOSMatGen as a valuable tool for a wide range of materials discovery tasks, guided by the physically important objective of the electronic structure. Finally, it is worth noting that DOSMatGen is property-agnostic and can be readily adapted to other electronic and even spectroscopic properties, including 1D X-ray diffraction (XRD) spectra, angle-resolved photoemission spectra (ARPES), X-ray absorption spectra (XAS) among others, as well as 2D images such as scanning transmission electron microscopy (STEM) images.

5 DATA AVAILABILITY

The authors declare that all data, materials, and code supporting the results reported in this study are available at the following repository: <https://github.com/Fung-Lab/DOSMatGen>.

6 ACKNOWLEDGEMENTS

This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231 using NERSC award BES-ERCAP0022842. Part of this research is sponsored by the INTERSECT Initiative as part of the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT- Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR22725, via the QCAD project (PI:P.Ganesh). Part of this research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 using NERSC award BES-ERCAP0031261, as part of a user project at the Center for Nanophase Materials Sciences (CNMS), which is a US Department of Energy, Office of Science User Facility at Oak Ridge National Laboratory.

REFERENCES

- [1] Deringer, V. L. Modelling and understanding battery materials with machine-learning-driven atomistic simulations. *Journal of Physics: Energy* **2**, 041003 (2020).
- [2] Liu, C. *et al.* Two-dimensional materials for next-generation computing technologies. *Nature Nanotechnology* **15**, 545–557 (2020).
- [3] Lou, Z., Wang, L., Jiang, K., Wei, Z. & Shen, G. Reviews of wearable healthcare systems: Materials, devices and system integration. *Materials Science and Engineering: R: Reports* **140**, 100523 (2020).
- [4] Iqbal, S. M., Mahgoub, I., Du, E., Leavitt, M. A. & Asghar, W. Advances in healthcare wearable devices. *NPJ Flexible Electronics* **5**, 9 (2021).
- [5] Ling, C. A review of the recent progress in battery informatics. *npj Computational Materials* **8**, 33 (2022).
- [6] Pyzer-Knapp, E. O., Suh, C., Gómez-Bombarelli, R., Aguilera-Iparraguirre, J. & Aspuru-Guzik, A. What is high-throughput virtual screening? a perspective from organic materials discovery. *Annual Review of Materials Research* **45**, 195–216 (2015).
- [7] Liu, Y., Zhao, T., Ju, W. & Shi, S. Materials discovery and design using machine learning. *Journal of Materomics* **3**, 159–177 (2017).
- [8] Bilodeau, C., Jin, W., Jaakkola, T., Barzilay, R. & Jensen, K. F. Generative models for molecular discovery: Recent advances and challenges. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **12**, e1608 (2022).
- [9] Noh, J., Gu, G. H., Kim, S. & Jung, Y. Machine-enabled inverse design of inorganic solid materials: promises and challenges. *Chemical Science* **11**, 4871–4881 (2020).
- [10] Fuhr, A. S. & Sumpter, B. G. Deep generative models for materials discovery and machine learning-accelerated innovation. *Frontiers in Materials* **9**, 865270 (2022).
- [11] Park, H., Li, Z. & Walsh, A. Has generative artificial intelligence solved inverse materials design? *Matter* **7**, 2355–2367 (2024).
- [12] Flam-Shepherd, D. & Aspuru-Guzik, A. Language models can generate molecules, materials, and protein binding sites directly in three dimensions as xyz, cif, and pdb files. *arXiv preprint arXiv:2305.05708* (2023).
- [13] Antunes, L. M., Butler, K. T. & Grau-Crespo, R. Crystal structure generation with autoregressive large language modeling. *arXiv preprint arXiv:2307.04340* (2023).
- [14] Gruver, N. *et al.* Fine-tuned language models generate stable inorganic materials as text. *arXiv preprint arXiv:2402.04379* (2024).
- [15] Hoffmann, J. *et al.* Data-driven approach to encoding and decoding 3-d crystal structures. *arXiv preprint arXiv:1909.00949* (2019).
- [16] Court, C. J., Yildirim, B., Jain, A. & Cole, J. M. 3-d inorganic crystal structure generation and property prediction via representation learning. *Journal of Chemical Information and Modeling* **60**, 4518–4535 (2020).
- [17] Long, T. *et al.* Constrained crystals deep convolutional generative adversarial network for the inverse design of crystal structures. *npj Computational Materials* **7**, 66 (2021).
- [18] Fung, V. *et al.* Atomic structure generation from reconstructing structural fingerprints. *Machine Learning: Science and Technology* **3**, 045018 (2022).
- [19] Yang, M. *et al.* Scalable diffusion for materials generation. *arXiv preprint arXiv:2311.09235* (2023).
- [20] Sinha, A., Jia, S. & Fung, V. Representation-space diffusion models for generating periodic materials. *arXiv preprint arXiv:2408.07213* (2024).
- [21] Joshi, C. K. *et al.* All-atom diffusion transformers: Unified generative modelling of molecules and materials. *arXiv preprint arXiv:2503.03965* (2025).
- [22] Xie, T., Fu, X., Ganea, O.-E., Barzilay, R. & Jaakkola, T. Crystal diffusion variational autoencoder for periodic material generation. *arXiv preprint arXiv:2110.06197* (2021).
- [23] Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. *Advances in neural information processing systems* **33**, 6840–6851 (2020).
- [24] Jiao, R. *et al.* Crystal structure prediction by joint equivariant diffusion. *Advances in Neural Information Processing Systems* **36** (2024).
- [25] Zeni, C. *et al.* A generative model for inorganic materials design. *Nature* 1–3 (2025).
- [26] Miller, B. K., Chen, R. T., Sriram, A. & Wood, B. M. Flowmm: Generating materials with riemannian flow matching. In *Forty-first International Conference on Machine Learning* (2024).

-
- [27] Zhao, Y. *et al.* Physics guided deep learning for generative design of crystal materials with symmetry constraints. *npj Computational Materials* **9**, 38 (2023).
 - [28] Ren, Z. *et al.* An invertible crystallographic representation for general inverse design of inorganic crystals with targeted properties. *Matter* **5**, 314–335 (2022).
 - [29] Wines, D., Xie, T. & Choudhary, K. Inverse design of next-generation superconductors using data-driven deep generative models. *The Journal of Physical Chemistry Letters* **14**, 6630–6638 (2023).
 - [30] Kwon, H. *et al.* Spectroscopy-guided discovery of three-dimensional structures of disordered materials with diffusion models. *Machine Learning: Science and Technology* **5**, 045037 (2024).
 - [31] Kong, S. *et al.* Density of states prediction for materials discovery via contrastive learning from probabilistic embeddings. *Nature communications* **13**, 949 (2022).
 - [32] Fung, V., Ganesh, P. & Sumpter, B. G. Physically informed machine learning prediction of electronic density of states. *Chemistry of Materials* **34**, 4848–4855 (2022).
 - [33] Dhariwal, P. & Nichol, A. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems* **34**, 8780–8794 (2021).
 - [34] Ho, J. & Salimans, T. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598* (2022).
 - [35] Xu, M. *et al.* Geodiff: A geometric diffusion model for molecular conformation generation. *arXiv preprint arXiv:2203.02923* (2022).
 - [36] Jain, A. *et al.* Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL materials* **1** (2013).
 - [37] Ong, S. P. *et al.* Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* **68**, 314–319 (2013).
 - [38] Neumann, M. *et al.* Orb: A fast, scalable neural network potential. *arXiv preprint arXiv:2410.22570* (2024).
 - [39] Greiner, M. T. *et al.* Free-atom-like d states in single-atom alloy catalysts. *Nature chemistry* **10**, 1008–1015 (2018).
 - [40] Thirumalai, H. & Kitchin, J. R. Investigating the reactivity of single atom alloys using density functional theory. *Topics in Catalysis* **61**, 462–474 (2018).
 - [41] Fung, V., Hu, G. & Sumpter, B. Electronic band contraction induced low temperature methane activation on metal alloys. *Journal of Materials Chemistry A* **8**, 6057–6066 (2020).
 - [42] Rosen, A. S., Vijay, S. & Persson, K. A. Free-atom-like d states beyond the dilute limit of single-atom alloys. *Chemical Science* **14**, 1503–1511 (2023).
 - [43] Seitz, L. C. *et al.* A highly active and stable iro x/sriro3 catalyst for the oxygen evolution reaction. *Science* **353**, 1011–1014 (2016).
 - [44] Song, J. *et al.* A review on fundamentals for designing oxygen evolution electrocatalysts. *Chemical Society Reviews* **49**, 2196–2214 (2020).
 - [45] Wei, C. *et al.* Recommended practices and benchmark activity for hydrogen and oxygen electrocatalysis in water splitting and fuel cells. *Advanced Materials* **31**, 1806296 (2019).
 - [46] Nichol, A. Q. & Dhariwal, P. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, 8162–8171 (PMLR, 2021).

SUPPLEMENTARY INFORMATION

A ADDITIONAL DETAILS ON JOINT EQUIVARIANT DIFFUSION

Diffusion on Lattice \mathbf{L} . Inspired by DiffCSP [24], the forward noising process that progressively adds noise to \mathbf{L} is defined as

$$q(\mathbf{L}_t | \mathbf{L}_0) = \mathcal{N}(\mathbf{L}_t | \sqrt{\bar{a}_t} \mathbf{L}_0, (1 - \bar{a}_t) \mathbf{I}), \quad (\text{A1})$$

where \mathbf{L}_t is the noisy version of \mathbf{L}_0 at timestep t , and $\bar{a}_t = \Pi_s^t (1 - \beta_t)$ determines how much of the original data, \mathbf{L}_0 , is preserved at t given $\beta_t \in (0, 1)$ [46]. During the diffusion process, we start with $\mathbf{L}_T \sim \mathcal{N}(0, \mathbf{I})$, and use the trained denoising neural network to obtain \mathbf{L}_{t-1} from \mathbf{L}_t :

$$p(\mathbf{L}_{t-1} | S_t) = \mathcal{N}(\mathbf{L}_{t-1} | \mu(S_t), \sigma^2(S_t) \mathbf{I}), \quad (\text{A2})$$

$$\mu(S_t) = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{L}_t - \frac{\beta_t}{\sqrt{1 - \bar{a}_t}} \hat{\epsilon}_{\mathbf{L}}(S_t, t) \right), \quad (\text{A3})$$

$$\sigma^2(S_t) = \beta_t \left(\frac{1 - \bar{a}_{t-1}}{1 - \bar{a}_t} \right). \quad (\text{A4})$$

Here, $\hat{\epsilon}_{\mathbf{L}}$ is predicted by the denoising network given the intermediate noisy structure $S_t = (\mathbf{A}_t, \mathbf{X}_t, \mathbf{L}_t)$.

Diffusion on Fractional Coordinates \mathbf{X}_f . Following DiffCSP, we use wrapped normal distribution as the prior for \mathbf{X}_f with the following forward noising process:

$$q(\mathbf{X}_t | \mathbf{X}_0) \propto \sum_{\mathbf{Z} \in \mathbb{Z}^{3 \times N}} \exp \left(-\frac{\|\mathbf{X}_t - \mathbf{X}_0 + \mathbf{Z}\|^2}{2\sigma_t^2} \right), \quad (\text{A5})$$

which ensures the probability distribution over $[z, z+1]^{3 \times N}$ for any integer z is the same. Here, $\sigma_t = \sigma_1 \left(\frac{\sigma_T}{\sigma_1} \right)^{\frac{t-1}{T-1}}$ for $t > 0$ and $\sigma_0 = 0$. During the diffusion process, we start with $\mathbf{X}_T \sim \mathcal{U}(0, 1)$, and use the trained denoising neural network to obtain \mathbf{X}_{t-1} from \mathbf{X}_t using the denoising term $\hat{\epsilon}_{\mathbf{X}}$.

Diffusion on Atom Types \mathbf{A} . We consider the atomic numbers of a crystal structure S to be a continuous variable in real space $\mathbb{R}^{h \times N}$, where h is the latent dimension of the atoms and N is the total number of atoms. The atom types are initially one-hot encoded. Similar to the diffusion process for \mathbf{L} , the forward noising process and the backward diffusing process are

$$q(\mathbf{A}_t | \mathbf{A}_0) = \mathcal{N}(\mathbf{A}_t | \sqrt{\bar{a}_t} \mathbf{A}_0, (1 - \bar{a}_t) \mathbf{I}), \quad (\text{A6})$$

$$p(\mathbf{A}_{t-1} | S_t) = \mathcal{N}(\mathbf{A}_{t-1} | \mu_{\mathbf{A}}(S_t), \sigma_{\mathbf{A}}^2(S_t) \mathbf{I}), \quad (\text{A7})$$

where $\mu_{\mathbf{A}}$ and $\sigma_{\mathbf{A}}^2$ are defined in a similar way as Eqs. A3 and A4.

Joint Diffusion Process. The training objectives for \mathbf{A} , \mathbf{X} and \mathbf{L} are:

$$\mathcal{L}_{\mathbf{L}} = \mathbb{E}_{\epsilon_{\mathbf{L}} \sim \mathcal{N}(0, I), t \sim \mathcal{U}(1, T)} \left[\|\epsilon_{\mathbf{L}} - \hat{\epsilon}_{\mathbf{L}}(S_t, t)\|_2^2 \right], \quad (\text{A8})$$

$$\mathcal{L}_{\mathbf{X}} = \mathbb{E}_{\mathbf{X}_t \sim q(\mathbf{X}_t | \mathbf{X}_0), t \sim \mathcal{U}(1, T)} \left[\lambda_t \|\nabla_{\mathbf{X}_t} \log q(\mathbf{X}_t | \mathbf{X}_0) - \hat{\epsilon}_{\mathbf{X}}(S_t, t)\|_2^2 \right], \quad (\text{A9})$$

$$\mathcal{L}_{\mathbf{A}} = \mathbb{E}_{\epsilon_{\mathbf{A}} \sim \mathcal{N}(0, I), t \sim \mathcal{U}(1, T)} \left[\|\epsilon_{\mathbf{A}} - \hat{\epsilon}_{\mathbf{A}}(S_t, t)\|_2^2 \right], \quad (\text{A10})$$

where $\lambda_t = \mathbb{E}_{\mathbf{X}_t}^{-1} \left[\|\nabla_{\mathbf{X}_t} \log q(\mathbf{X}_t | \mathbf{X}_0)\|_2^2 \right]$.

The overall combined objective for training the denoising neural network is then

$$\mathcal{L}_S = \lambda_{\mathbf{A}} \mathcal{L}_{\mathbf{A}} + \lambda_{\mathbf{X}} \mathcal{L}_{\mathbf{X}} + \lambda_{\mathbf{L}} \mathcal{L}_{\mathbf{L}}, \quad (\text{A11})$$

as shown in Eq. 1.

B ADDITIONAL DISTRIBUTION PLOTS OF FORMATION ENERGIES PER ATOM

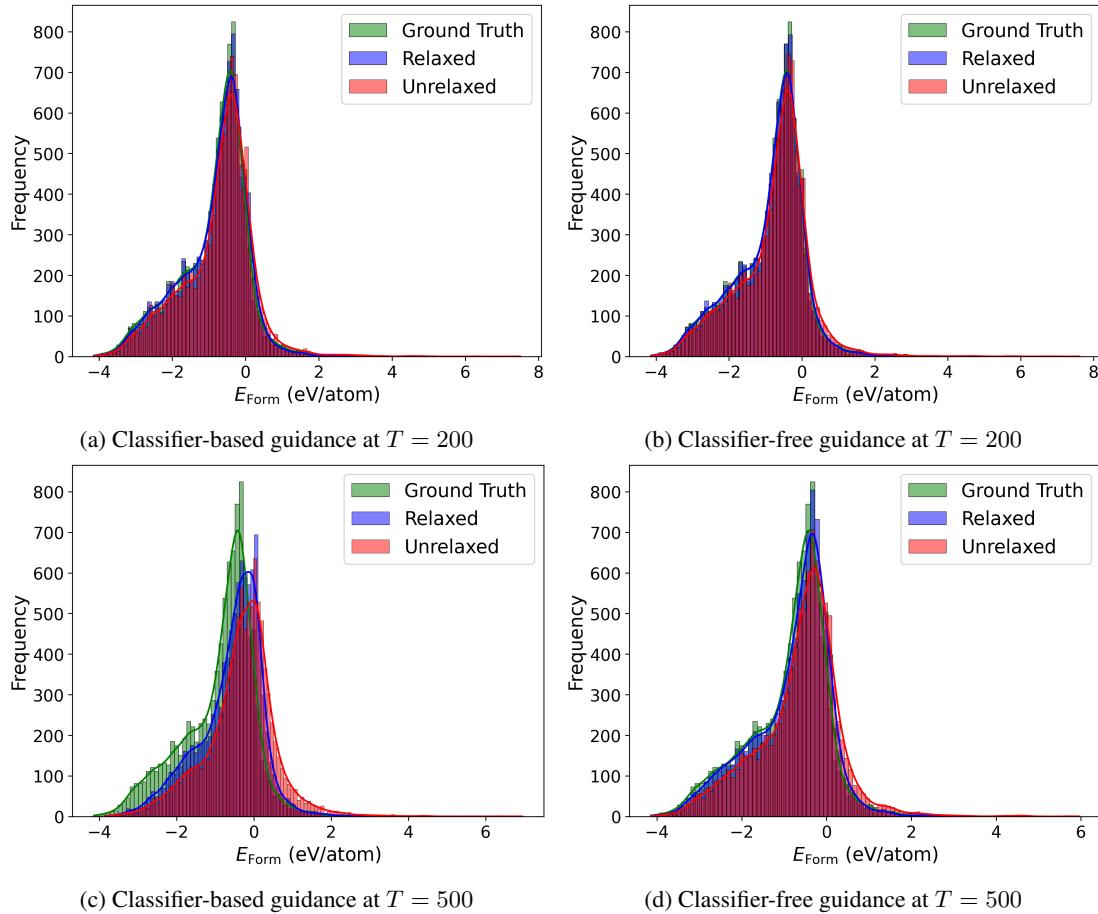


Figure B.1: Distribution of formation energies per atom (eV/atom) for structures conditionally generated at $T \in \{200, 500\}$ using (a) classifier-based guidance and (b) classifier-free guidance. All generated structures are relaxed using the `Orb-v2` interatomic potential. The ground truth distribution corresponds to the formation energies of the original test set structures, with all energies predicted using `Orb-v2`.

C VISUALIZATIONS OF SELECTED STRUCTURES FROM POST-PROCESSING WORKFLOW

The following visualizations correspond to the structures in Fig. 8.

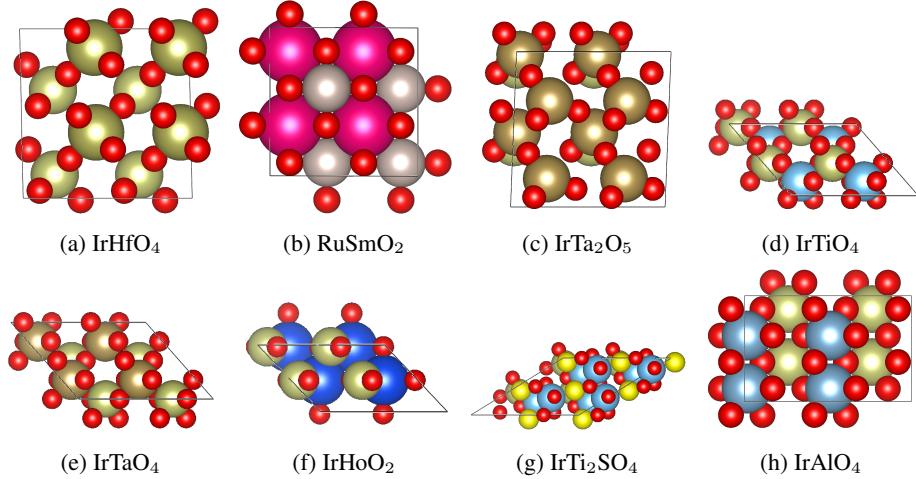


Figure C.1: Visualizations of the crystal structures generated during the post-processing workflow (refer to Fig. 8), shown as $2 \times 2 \times 2$ supercells.