SCALABLE DIFFUSION FOR MATERIALS GENERATION

Sherry Yang,[1,2] **KwangHwan Cho**[2] **Amil Merchant**[1] **Pieter Abbeel**[2]
Dale Schuurmans[1,3] **Igor Mordatch**[1] **Ekin Dogus Cubuk**[1]
[1]Google DeepMind [2]UC Berkeley [3]University of Alberta
sherryy@google.com

ABSTRACT

Generative models trained on internet data are capable of generating novel texts and images. A natural question is whether these models can advance science (e.g., generate novel stable materials). Traditionally, models with explicit structures (e.g., graphs) have been used in modeling structural relationships (e.g., atoms and bonds in crystals), but have faced challenges scaling to large and complex systems. Another challenge in generating materials is the mismatch between standard generative modeling metrics and downstream applications. For instance, common metrics such as the reconstruction error do not correlate well with the downstream goal of discovering *novel* stable materials. In this work, we tackle the scalability challenge by developing a unified crystal representation that can represent *any* crystal structure (UniMat), followed by training a diffusion probabilistic model on the UniMat representations. Despite the lack of explicit structure modeling, UniMat can generate high fidelity crystals from larger and more complex chemical systems, outperforming previous approaches. To better connect generation quality to downstream applications, we propose additional metrics for evaluating generative models of materials, including per-composition formation energy and stability with respect to convex hulls from Density Function Theory (DFT). Lastly, we show that conditional generation with UniMat can scale to up to millions of crystal structures, outperforming random structure search (the current leading method) in discovering new stable materials. See website at https://unified-materials.github.io.

1 INTRODUCTION

Large generative models trained on internet-scale vision and language data have demonstrated exceptional abilities in synthesizing highly realistic texts (OpenAI, 2023; Anil et al., 2023), images (Ramesh et al., 2021; Yu et al., 2022), and videos (Ho et al., 2022a; Singer et al., 2022). The need for novel synthesis, however, goes far beyond conversational agents or generative media, which mostly impact the digital world. In the physical world, technological applications such as catalysis (Nørskov et al., 2009), solar cells (Green et al., 2014), and lithium batteries (Mizushima et al., 1980) are enabled by the discovery of novel materials. The traditional trial-and-error approach that discovered these materials can be highly inefficient and take decades (e.g., blue LEDs (Nakamura, 1998) and high-Tc superconductors (Bednorz & Müller, 1986)). Generative models have the potential to dramatically accelerate materials discovery by generating and evaluating material candidates with desirable properties more efficiently in silico.

One of the difficulties in materials generation lies in characterizing the structural relationships between atoms, which scales quadratically with the number of atoms. While representations with explicit structures such as graphs have been extensively studied (Schütt et al., 2017; Xie & Grossman, 2018; Batzner et al., 2022; Xie et al., 2021), explicit characterization of inter-atomic relationships becomes increasingly challenging as the number of atoms increases, which can prevent these methods from scaling to large materials datasets with complex chemical systems. On the other hand, given that generative models are designed to discover patterns from data, it is natural to wonder if material structures can automatically arise from data through generative modeling, similar to how natural language structures arise from language modeling, so that large system sizes becomes more of a benefit than a roadblock.

Existing generative models that directly model atoms without explicit structures are largely inspired by generative models for computer vision, such as learning VAEs or GANs on voxel images (Noh

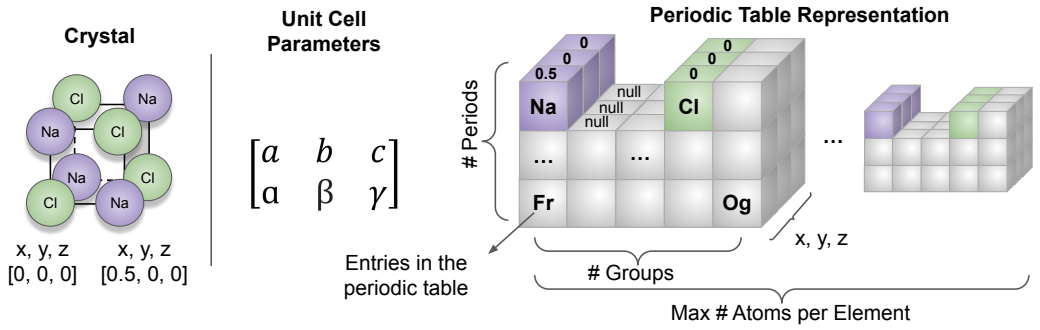

Figure 1: UniMat representation of crystal structures. Crystals are represented by the atom locations stored at the corresponding elements in the periodic table (and additional unit cell parameters if coordinates are fractional). For instance, the bottom right atom Na in the crystal is located at $[1, 0, 0]$, hence the periodic table has value $[1, 0, 0]$ at the Na entry.

et al., 2019; Hanakata et al., 2020) or point cloud representations of materials (Kim et al., 2020). VAEs and GANs have known drawbacks such as posterior collapse (Lucas et al., 2019) and mode collapse (Srivastava et al., 2017), potentially making scaling difficult (Dhariwal & Nichol, 2021). More recently, diffusion models (Song & Ermon, 2019; Ho et al., 2020) have been found particularly effective in generating diverse yet high fidelity image and videos, and have been applied to data at internet scale (Saharia et al., 2022; Ho et al., 2022a). However, it is unclear whether diffusion models are also effective in modeling structural relationships between atoms in crystals that are neither images nor videos.

In this work, we investigate whether diffusion models can capture inter-atomic relationships effectively by directly modeling atom locations, and whether such an approach can be scaled to complex chemical systems with a larger number of atoms. Specifically, we propose a unified representation of materials (UniMat) that can capture *any* crystal structure. As shown in Figure 1, UniMat represents atoms in a material's unit cell (the smallest repeating unit) by storing the continuous value x, y, z atom locations at the corresponding element entry in the periodic table. This representation overcomes the difficulty around joint modeling of discrete atom types and continuous atom locations, while introducing prior knowledge from the periodic table (e.g., elements in the same group have similar chemical properties). With such a unified representation of materials, we train diffusion probabilistic models by treating the UniMat representation as a 4-dimensional tensor and applying interleaved attention and convolution layers, similar to Saharia et al. (2022), across periods and groups of the periodic table. This allows UniMat to capture inter-atom relationships while preserving any inductive bias from the periodic table, such as elements in the same group having similar chemical properties.

We first evaluate UniMat on a set of proxy metrics proposed by Xie et al. (2021), and show that UniMat generally works better than the previous state-of-the-art graph based approach and a recent language model (Flam-Shepherd & Aspuru-Guzik, 2023) and diffusion model (Pakornchote et al., 2023) baseline. However, we are ultimately interested in whether the generated materials are physically valid and can be synthesized in a laboratory (e.g., low-energy materials). We found proxy metrics based on learning a separate energy network either saturate or fall short in evaluating generated materials reliably under the context of material discovery (i.e., generating materials that have not been seen by the energy prediction network). In answering this question, we run DFT relaxations (Hafner, 2008) to compute the formation energy of the generated materials, which is more widely accepted in material science than learned proxy metrics in Bartel et al. (2020). We then use per-composition formation energy and stability with respect to convex hull through decomposition energy as more reliable metrics for evaluating generative models for materials. UniMat drastically outperforms previous state-of-the-art according to these DFT based metrics.

Lastly, we scale UniMat to train on all experimentally verified stable materials as well as additional stable / semi-stable materials found through search and substitution (over 2 million structures in total). We show that predicting material structures conditioned on element type can generalize (in a zero-shot manner) to predicting more difficult structures that are not a neighboring structure to the training set, achieving better efficiency than the predominant random structure search. This allows for the possibility of discovering new materials with desired properties effectively. In summary, our work contributes the following:

- We develop a novel representation of materials that enables diffusion models to scale to large and complex materials datasets, outperforming previous methods on previous proxy metrics.
- We conduct DFT calculations to rigorously verify the stability of generated materials, and propose to use per-composition formation energy and stability with respect to convex hull for evaluating generative models for materials.
- We scale conditional generation to all known stable materials and additional materials found by search and substitution, and observe zero-shot generalization to generating harder structures, achieving better efficiency than random structure search in discovering new materials.

2 Scalable Diffusion for Materials Generation

We start by proposing a novel crystal representation that can represent any material with a finite number of atoms in a unit cell (the smallest repeating unit of a material). We then illustrate how to learn both unconditional and conditional denoising diffusion models on the proposed crystal representations. Lastly, we explain how we can verify generated materials rigorously using quantum mechanical methods.

2.1 Scalable Representation of Crystal Structures

An ideal representation for crystal structures should not introduce any intrinsic errors (unlike voxel images), and should be able to support both up scaling to large sets of materials on the internet and down scaling to a single compound system that a particular group of scientists care about (e.g., silicon carbide). We develop such a scalable and flexible representation below.

Periodic Table Based Material Representation. We first observe that periodic table captures rich knowledge of chemical properties. To introduce such prior knowledge to a generative model as an inductive bias, we define a 4-dimensional material space, $\mathcal{M} := \mathbb{R}^{L \times H \times W \times C}$, where $H = 9$ and $W = 18$ correspond to the number of periods and groups in the periodic table, L corresponds to the maximum number of atoms per element in the periodic table, and $C = 3$ corresponds to the `x,y,z` locations of each atoms in a unit cell. We define a *null* location using special values such as $\mathtt{x} = \mathtt{y} = \mathtt{z} = -1$ to represent the absence of this atom. A visualization of this representation is shown in Figure 1. To account for invariances in order, rotation, translation, and periodicity, we incorporate data augmentation through random shuffling and rotations similar to Hoffmann et al. (2019); Kim et al. (2020); Court et al. (2020). We also include unit cell parameters $(a, b, c) \in \mathbb{R}^3$ and $(\alpha, \beta, \gamma) \in \mathbb{R}^3$ as shown in Figure 1. We denote this representation UniMat, as it is a unified representation of crystals, and has the potential to represent broader chemical structures (e.g., drugs, molecules, and proteins).

Flexibility for Smaller Systems. While UniMat can represent any crystal structure, sometimes one might only be interested in generating structures with one specific element (e.g., carbon in graphene) or two-chemical compounds (e.g., silicon carbide). Instead of setting H and W to the full periods and groups of the periodic table, one can set $H = 1, W = 1$ (for one specific element) or $H = 9, W = 2$ (for elements from two groups) to model specific chemical systems of interest. L can also be adjusted according to the number of elements expected to exist in the system.

2.2 Learning Diffusion Models with UniMat Representation

With the UniMat representation above, we now illustrate how effective training of diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020) on crystal structures can be enabled, followed by how to generate crystal structures conditioned on compositions or other types of material properties. Details of the model architecture and training procedure can be found in Appendix A.

Diffusion Model Background. Denoising diffusion probablistic models (DDPM) are a class of probabilistic generative models initially designed for images where the generation of an image $x \in \mathbb{R}^d$ is formed by iterative denoising. That is, given an image x sampled from a distribution of images $p(x)$, a randomly sampled Gaussian noise variable $\epsilon \sim \mathcal{N}(0, I_d)$, and a set of T different noise levels $\beta_t \in \mathbb{R}$, a denoising model ϵ_θ is trained to denoise the noise corrupted image x at each specified noise level $t \in [1, T]$ by minimizing:

$$\mathcal{L}_{\text{MSE}} = \|\epsilon - \epsilon_\theta(\sqrt{1-\beta_t}x + \sqrt{\beta_t}\epsilon, t))\|^2.$$

Given this learned denoising function, new images may be generated from the diffusion model by initializing an image sample x_T at noise level T from a Gaussian $\mathcal{N}(0, I_d)$. This sample x_T is then iteratively denoised by following the expression:

$$x_{t-1} = \alpha_t(x_t - \gamma_t \epsilon_\theta(x_t, t)) + \xi, \quad \xi \sim \mathcal{N}\big(0, \sigma_t^2 I_d\big), \tag{1}$$

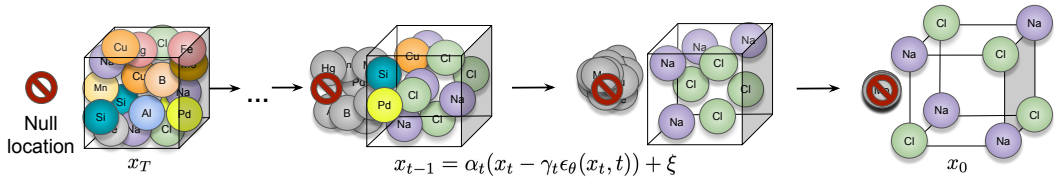

Figure 2: Illustration of the denoising process for unconditional generation with UniMat. The denoising model learns to move atoms from random locations back to their original locations. Atoms not present in the crystal are moved to the null location during the denoising process, allowing crystals with an arbitrary number of atoms to be generated.

where γ_t is the step size of denoising, α_t is a linear decay on the currently denoised sample, and σ_t is some time varying noise level that depends on α_t and β_t. The final sample x_0 after T rounds of denoising corresponds to the final generated image.

Unconditional Diffusion with UniMat. Now instead of an image $x \in \mathbb{R}^d$, we have a material $x \in \mathbb{R}^d$ with $d = L \times H \times W \times 3$ tensor as described in Section 2.1, where the inner-most dimension of x represents the atom locations (`x`, `y`, `z`). The denoising process in Equation 1 now corresponds to the process of moving atoms from random locations back to their original locations in a unit cell as shown in Figure 2. Note that the set of null atoms (i.e., atoms that do not exist in a crystal) will have random locations initially (left-most structure in Figure 2), and are gradually moved to the special null location during the denoising process. The null atoms are then filtered when the final crystals are extracted. The inclusion of null atoms in the representation enables UniMat to generate crystals with an arbitrary number of atoms (up to a maximum size). Since the denoising process of a DDPM nicely corresponds to the process of gradually moving atoms in space until they reach their target location, we choose DDPM over other diffusion models (e.g., denoising score matching). We parametrize $\epsilon_\theta(x_t, t)$ using interleaved convolution and attention operations across the L, H, W dimensions of x_t similar to Saharia et al. (2022), which can capture inter-atom relationships in a crystal structure. When atom locations are represented using fractional coordinates, we treat unit cell parameters as additional inputs to the diffusion process by concatenating the unit cell parameters with the crystal locations.

Conditioned Diffusion with UniMat. While the unconditional generation procedure described above allows generation of materials from random noise, naïvely sampling from the unconditional model can lead to samples that largely overlap with the training set. This is undesirable in the context of materials discovery, where the goal is to discover *novel* materials that do not exist in the training set. Futhermore, practical applications such as material synthesis often focus on specific types of materials, but one do not have much control over what compound gets generated during an unconditional denoising process. This suggests that conditional generation may be more relevant for materials discovery.

We consider conditioning generation on compositions (types and ratios of chemical elements) $c \in \mathbb{R}^{H \times W}$ when only the composition types are specified (e.g., carbon and silicon), or on $c \in \mathbb{R}^{L \times H \times W}$ when the exact composition (number of atoms per element) is given (e.g., Si4C4). We denote the conditional denoising model as $\epsilon_\theta(x_t, t|c)$. Since the input to the unconditional denoising model $\epsilon_\theta(x_t, t)$ is a noisy material of dimensions $(L, H, W, 3)$, we concatenate the conditioning variable c with the noisy material along the last dimension before inputting the noisy material into the denoising model, so that the denoising model can easily condition on compositions as desired.

To condition on auxiliary information such as energy, we can leverage classifier-free guidance (Ho & Salimans, 2022) and use

$$\hat{\epsilon}_\theta(x_t, t|c, \texttt{aux}) = (1 + \omega)\epsilon_\theta(x_t, t|c, \texttt{aux}) - \omega\epsilon_\theta(x_t, t|c) \tag{2}$$

as the denoising model in the reverse process for sampling materials conditioned on auxiliary information `aux`, where ω controls the strength of auxiliary information conditioning.

2.3 Evaluating Generated Materials

Different from generative models for vision and language where the quality of generation can be easily assessed by humans, evaluating generated crystals rigorously requires calculations from Density Functional Theory (DFT) (Hohenberg & Kohn, 1964), which we elaborate in detail below.

Drawbacks of Learning Based Evaluations. One way to evaluate generative models for materials is to compare the distributions of formation energy E_f between a generated and reference set, $D(p(E_f^{\text{gen}}), p(E_f^{\text{ref}}))$, where D is a distance measure over distributions, such as earth mover's dis-

tance (Xie et al., 2021). Since using DFT to compute E_f is computationally demanding, previous work has relied on a learned network to predict E_f from generated materials (Xie et al., 2021). However, predicting E_f can have intrinsic errors, particularly in the context of materials discovery where the goal is to generate *novel* materials beyond the training manifold of the energy prediction network.

Even when E_f can be predicted with reasonable accuracy, a low E_f does not necessarily reflect ground-truth (DFT) stability. For example, Bartel et al. (2020) reported that a model that can predict E_f with an error of 60 meV/atom (a 16-fold reduction from random-guessing) does not provide any predictive improvement over random guessing for stable material discovery. This is because most variations in E_f are between different chemical systems, whereas for stability assessment, the important comparison is between compounds in a single chemical system. When materials generated by two different models contain different compounds, the model that generated materials with a lower E_f could have simply generated compounds from a lower E_f system without enabling efficient discovery (Merchant et al., 2023).

The property that captures *relative* stabilities between different compositions is known as decomposition energy (E_d). Since E_d depends on the formation energy of other compounds from the same system, predicting E_d directly using machine learning models has been found difficult (Bartel et al., 2020).

Evaluating via Per-Composition Formation Energy. Different from learned energy predictors, DFT calculations provide more accurate and reliable E_f values. When two models each generate a structure of the same composition, we can directly compare which structure has a lower DFT computed E_f (and is hence more stable). We call this the *per-composition* formation energy comparison. We define average difference in per-composition formation energy between two sets of materials A and B as

$$\Delta E_f(A, B) = \frac{1}{|C|} \sum_{(x,x') \in C} \left(E^A_{f,x} - E^B_{f,x'} \right), \tag{3}$$

where $C = \{(x, x') \mid x \in A, x' \in B, \text{comp}(x) = \text{comp}(x')\}$ denotes the set of structures from A and B that have the same composition. We also define the E_f Reduction Rate between set A and B as the rate where structures in A have a lower E_f than the structures in B of the corresponding compositions, i.e.,

$$E_f \text{ Reduction Rate}(A, B) = \frac{1}{|C|} |\{(x, x') \mid (x, x') \in C \wedge E^A_{f,x} < E^B_{f,x'}\}|, \tag{4}$$

where C is the same as in Equation 3. We can then use ΔE_f and the E_f Reduction Rate to compare a generated set of structures to some reference set, or to compare two generated sets. $\Delta E_f(A, B)$ measures how much lower in E_f (on average) the structures in a set A are compared to the structures of correponding compositions in a set B, while E_f Reduction Rate(A, B) reflects how many structures in A have lower E_f than the corresponding structures in B. We use these metrics to evaluate generated materials in Section 3.2.1.

Evaluating Stability via Decomposition Energy We also want to compare generated materials that differ in composition. To do so, we can use DFT to compute decomposition energy E_d. E_d measures a compound's thermodynamic decomposition enthalpy into its most stable compositions on a convex hull phase diagram, where the convex hull is formed by linear combinations of the most stable (lowest energy) phases for each known composition (Jain et al., 2013). As a result, decomposition energy allows us to compare compounds from two generative models that differ in composition by separately computing their decomposition energy with respect to the convex hull formed by a larger materials database. The distribution of decomposition energies will reflect a generative model's ability to generate relatively stable materials. We can further compute the number of *novel* stable ($E_d < 0$) materials from set A with respect to convex hull as

$$\text{\# Stable}(A) = |\{x \in A \mid E^A_{d,x} < 0\}|, \tag{5}$$

and compare this quantity to some other set B. We apply this metric to evaluate generative models for materials in Section 3.2.

Evaluating against Random Search Baseline. For structure prediction given compositions, one popular non-learning based approach is Ab initio random structure search (AIRSS) (Pickard & Needs, 2011). AIRSS works by initializing a set of sensible structures given the composition and a target volume, relaxing randomly initialized structures via soft-sphere potentials, followed by DFT relaxations to minimize the total energy of the system. However, discovering structures (especially

		Validity % ↑		COV % ↑		Property Statistics ↓		
Method	Dataset	Structure	Composition	Recall	Precision	Density	Energy	# Elements
CDVAE	Perov5	**100**	98.5	99.4	98.4	0.125	0.026	0.062
	Carbon24	**100**	–	99.8	83.0	0.140	0.285	–
	MP20	**100**	86.7	99.1	99.4	0.687	0.277	1.432
DP-CDVAE	Perov5	**100**	98.0	99.5	97.2	0.102	0.026	**0.021**
	Carbon24	99.9	–	100	77.98	0.097	0.259	–
	MP20	99.9	85.4	99.4	99.3	0.179	0.052	0.567
LM	Perov5	**100**	98.7	**99.6**	**99.4**	**0.071**	–	0.036
	MP20	95.8	88.8	99.6	98.5	0.696	–	0.092
UniMat	Perov5	**100**	**98.8**	99.2	98.2	0.076	**0.022**	0.025
	Carbon24	**100**	–	**100**	**96.5**	**0.013**	**0.207**	–
	MP20	97.2	**89.4**	**99.8**	**99.7**	**0.088**	**0.034**	**0.056**

Table 1: Proxy evaluation of unconditional generation using CDVAE (Xie et al., 2021), language model (Flam-Shepherd & Aspuru-Guzik, 2023), diffusion baseline (Pakornchote et al., 2023), and UniMat. UniMat generally performs better in terms of property statistics, and achieves the best coverage on more difficult dataset (MP-20). We note the limitation of these proxy metrics, and defer more rigorous evaluation to DFT calculations.

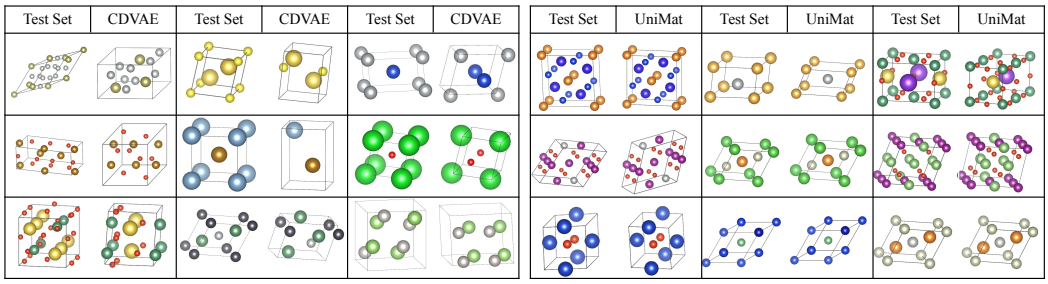

Figure 3: Qualitative evaluation of materials generated by CDVAE (Xie et al., 2021) (left) and UniMat (right) trained on MP-20 in comparison to the test set materials of the same composition. Materials generated by UniMat generally align better with the test set.

if done in a high-throughput framework) requires a large number of initializations and relaxations which can often fail to converge (Cheon et al., 2020; Merchant et al., 2023).

One practical use of conditional UniMat is to propose initial structures given compositions, with the hope that the generated structures will result in a higher convergence rate for DFT calculations compared to structures proposed by AIRSS, which are based on manual heuristics and random guessing of initial volumes.

3 Experimental Evaluation

We now evaluate UniMat using previous proxy metrics from Xie et al. (2021) and metrics derived from DFT calculations from Section 2.3. UniMat is able to generate orders of magnitude more stable materials verified by DFT calculations compared to the previous state-of-the-art generative model. We further demonstrate UniMat's ability in accelerating random structure search through conditional generation.

3.1 Evaluating Unconditional Generation Using Proxy Metrics

Datasets, Metrics, and Baselines. We begin the evaluation following the same setup as CDVAE Xie et al. (2021) using Perov-5, Carbon-24, and MP-20 materials datasets. We report metrics on structural and composition validity determined by atom distances and SMACT, coverage metrics based on CrystalNN fingerprint distances, and property distributions in density, learned formation energy, and number of atoms (e.g., earth mover's distance between the distribution of number of elements in generated materials versus test materials.). We include a recent language model baseline (Flam-Shepherd & Aspuru-Guzik, 2023) and a diffusion baseline (Pakornchote et al., 2023).

Results. Evaluation results on UniMat and baselines are shown in Table 5. All four models perform similarly in terms of structure and composition validity on the Perov-5 dataset due to its simplicity. UniMat performs slightly worse on the coverage based metrics on Perov-5, but achieves better distributions in energy and number of unique elements. On Carbon-24, UniMat outperforms CDVAE in

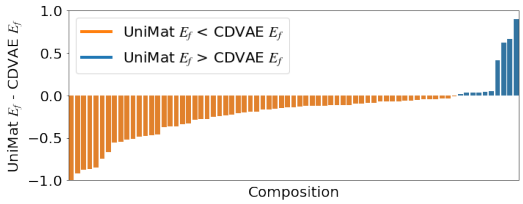

Figure 5: Difference in E_f for each composition generated by UniMat and CDVAE, i.e., $E^A_{f,x} - E^B_{f,x'}$, where A and B are sets of structures generated by UniMat and CDVAE, respectively. UniMat generates more structures with lower E_f.

A, B	ΔE_f (eV/atom)	E_f Reduc. Rate
CDVAE, MP-20 test	0.279	0.083
UniMat, MP-20 test	**0.061**	**0.254**
UniMat, CDVAE	**-0.216**	**0.863**

Table 2: ΔE_f (Equation 3) and E_f Reduction Rate (Equation 4) between CDVAE and MP-20 test, between UniMat and MP-20 test, and between UniMat and CDVAE. UniMat generates structures with an average of -0.216 eV/atom lower E_f than CDVAE. 86.3% of the overlapping (in composition) structures generated by UniMat and CDVAE has a lower energy in UniMat.

all metrics. On the more realistic MP-20 dataset, UniMat achieves the best property statistics, coverage, and composition validity, but worse structure validity than CDVAE. Results on full coverage metrics from CDVAE are in Appendix D. We note that some of these metrics have been saturated with close to 100% performance. We defer more rigorous evaluations with DFT calculations to Section 3.2.

In addition, we qualitatively evaluate the generated materials from training on MP-20 in Figure 3. We select generated materials that have the same composition as the test set from MP-20, and use the VESTA crystal visualization tool (Momma & Izumi, 2011) to plot both the test set materials and the generated materials. The range of fractional coordinates in the VESTA settings were set from -0.1 to 1.1 for all coordinates to represent all fractional atoms adjacent to the unit cell. In general, we found that UniMat generates materials that are visually more aligned with the test set materials than CDVAE.

Ablation on Model Size. In training on larger datasets with more diverse materials such as MP-20, we found benefits in scaling up the model as shown in Table 4, which suggests that the UniMat representation and the UniMat training objective can be further scaled to systems larger than MP-20, which we elaborate more in Section 3.3.

	Validity % ↑		COV % ↑	
Model size	Struct.	Comp.	Recall	Precision
Small (64)	95.7	86.0	99.8	99.3
Medium (128)	96.8	86.7	99.8	99.5
Large (256)	**97.2**	**89.4**	**99.8**	**99.7**

Figure 4: UniMat trained with a larger feature dimension results in better validity and coverage.

3.2 Evaluating Unconditional Generation Using DFT Calculations

As discussed in Section 2.3, proxy-based evaluation in Section 3.1 should be backed by DFT verifications similar to Noh et al. (2019). In this section, we evaluate stability of generated materials using metrics derived from DFT calculations in Section 2.3.

3.2.1 Per-Composition Formation Energy

Setup. We start by running DFT relaxations using the VASP software (Hafner, 2008) to relax both atomic positions and unit cell parameters on generated materials from models trained on MP-20 to compute their formation energy E_f (see details of DFT in Appendix B). We then compare average difference in per-composition formation energy (ΔE_f in Equation 3) and the formation energy reduction rate (E_f Reduction Rate in Equation 4) between materials generated by CDVAE and the MP-20 test set, between UniMat and the test set, and between UniMat and CDVAE.

Results. We plot the difference in formation energy for each pair of generated structures from UniMat and CDVAE with the same composition in Figure 5. We see the majority of the generated compositions from UniMat have a lower formation energy. We further report ΔE_f and the E_f Reduction Rate in Table 2. We see that among the set of materials generated by UniMat and CDVAE with overlapping compositions, 86% of them have a lower energy when generated by UniMat. Furthermore, materials generated by UniMat have an average of -0.21 eV/atom lower E_f than CDVAE. Comparing the generated set against the MP-20 test set also favors UniMat.

3.2.2 Stability Analysis through Decomposition Energy

As discussed in Section 2.3, generated structures relaxed by DFT can be compared against the convex hull of a larger materials database in order to analyze their stability through decomposition energy. Specifically, we downloaded the full Materials Project database (Jain et al., 2013) from

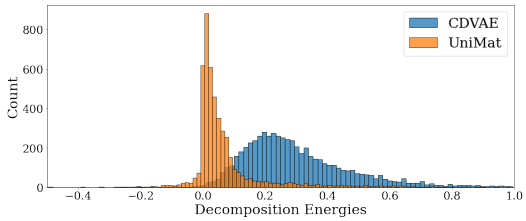

Figure 6: Histogram of decomposition energy E_d of structures generated by CDVAE and UniMat after DFT relaxation. UniMat generates structures with lower decomposition energies.

	# Stable MP 2021	# Metastable MP 2021	# Stable GNoME
CDVAE	56	90	1
UniMat	**414**	**2157**	**32**

Table 3: Number of stable ($E_d < 0$) and metastable ($E_d < 25$meV/atom) materials generated compared against the convex hull of MP 2021, and stability against GNoME with 2 million structures. UniMat generates an order of magnitude more stable / metastable materials than CDVAE.

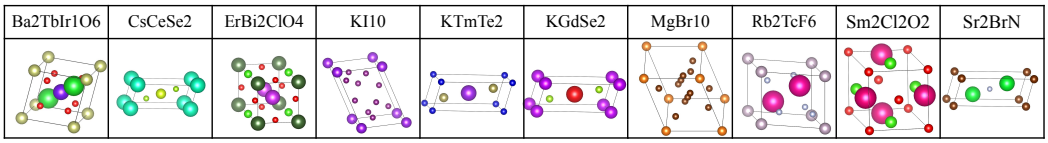

Figure 7: Visualizations of materials generated by UniMat trained on MP-20 before DFT relaxation that have $E_d < 0$ after relaxation compared against the convex hull of MP 2021. We note that these materials require further analysis and verification before they can be claimed to be realistic or stable.

July 2021, and used this to form the convex hull. We then compute the decomposition energy for materials generated by UniMat and CDVAE individually against the convex hull.

Results. We plot the distributions of the decomposition energies after DFT relaxation for the generated materials from both models in Figure 6. Note that only the set of generated materials that converged after DFT calculations are plotted. We see that UniMat generates materials that are lower in decomposition energy after DFT relaxation compared to CDVAE. We further report the number of newly discovered stable / metastable materials (with $E_d < 25$meV/atom) from both UniMat and CDVAE in Table 3. In addition to using the convex hull from Materials Project 2021, we also use another dataset (GNoME) with 2.2 million materials constructed via structure search to construct a more challenging convex hull (Merchant et al., 2023). We see that UniMat is able to discover an order of magnitude more stable materials than CDVAE with respect to convex hulls constructed from both datasets. We visualize examples of newly discovered stable materials by UniMat in Figure 7.

3.3 Evaluating Composition Conditioned Generation

We have verified that some of the unconditionally generated materials from UniMat are indeed novel and stable through DFT calculations. We now assess composition conditioned generation which is often more practical for downstream synthesis applications.

Setup. We use AIRSS to randomly initialize 100 structures per composition followed by relaxation via soft-sphere potentials. We then run DFT relaxations on these AIRSS structures. For conditional generation using UniMat, we train composition conditioned UniMat on the GNoME dataset consisting of 2.2 million stable materials. We then sample 100 structures per composition for the same compositions used by AIRSS. We evaluate the rate of compositions for which at least 1 out of 100 structures converged during DFT calculations. In addition to convergence rate, we also evaluate the $\Delta E_f(\text{UniMat}, \text{AIRSS})$ and the E_f Reduction Rate $(\text{UniMat}, \text{AIRSS})$ on the DFT relaxed structures. Since none of the test compositions exist in the training set, we are evaluating the ability of UniMat to generalize to more difficult structures in a zero-shot manner. See details of AIRSS in Appendix C.

Results. We first observe that AIRSS has an overall convergence rate of 0.55, whereas UniMat has an overall convergence rate of 0.81. We note that both AIRSS and UniMat can be further optimized for convergence rate, so these results are only initial signals on how conditional generative models compare to structure search. Next, we take the relaxed structure with the lowest E_f from both UniMat and AIRSS for each composition, and plot the per-composition E_f difference in Figure 8, and

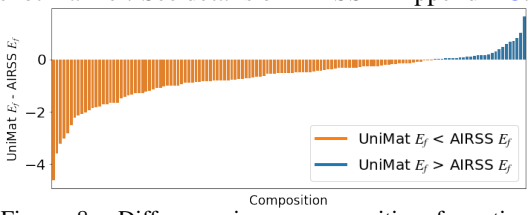

Figure 8: Difference in per-composition formation energy between structures produced by UniMat and AIRSS. More compounds generated by UniMat lead to lower formation energy than AIRSS.

$\Delta E_f(\text{UniMat}, \text{AIRSS}) = -0.68$eV/atom, and E_f Reduction Rate(UniMat, AIRSS) $= 0.8$, which suggests that UniMat is indeed effective in initializing structures that lead to lower E_f than AIRSS.

4 Related Work

Diffusion Models for Structured Data Diffusion models (Song & Ermon, 2019; Ho et al., 2020; Kingma et al., 2021) were initially proposed for generating images from noise of the same dimension through a Markov chain of Gaussian transitions, and have been adopted to structured data such as graphs (Niu et al., 2020; Vignac et al., 2022; Jo et al., 2022; Yim et al., 2023), sets (Giuliari et al., 2023) and point clouds (Qi et al., 2017; Luo & Hu, 2021; Lyu et al., 2021). Diffusion modeling for materials requires modeling continuous atom locations and discrete atom types. Previous approaches either embed discrete quantities into a continuous latent space, risking information loss (Xie et al., 2021), or directly learn discrete-space transformations (Vignac et al., 2022; Austin et al., 2021) on graphs represented by adjacency matrices that scale quadratically in the number of atoms.

Generative Models for Materials Discovery. Generative models originally designed for images have been applied to generating material structures, such as GANs (Nouira et al., 2018; Kim et al., 2020; Long et al., 2021; 2022), VAEs (Hoffmann et al., 2019; Noh et al., 2019; Ren et al., 2020; Court et al., 2020), and diffusion models (Xie et al., 2021). These methods were developed to work with different materials representations as voxel images (Hoffmann et al., 2019; Noh et al., 2019; Court et al., 2020), graphs (Xie et al., 2021), point clouds (Kim et al., 2020), and phase fields or electron density maps (Vasylenko et al., 2021; Court et al., 2020). However, existing work has mostly focused on simpler materials in binry compounds (Noh et al., 2019; Long et al., 2021), ternary compounds (Nouira et al., 2018; Kim et al., 2020), or cubic systems (Hoffmann et al., 2019). Xie et al. (2021) show that graph neural networks with latent space diffusion guided by gradient of formation energy can scale to Materials Project (Jain et al., 2013). However, the quality of generated materials seems to decrease drastically on complex datasets. Recently, large language models have been applied to generate crystal files (Antunes et al., 2023; Flam-Shepherd & Aspuru-Guzik, 2023). However, the ability of language models to generate files with structural information requires further confirmation, and the generated materials require further DFT verification. Pakornchote et al. (2023) uses diffusion models to model atom locations, but Pakornchote et al. (2023) uses a separate VAE to predict lattice parameters and number of atoms, limiting modeling flexibility.

Evaluation of Materials Discovery The most reliable verification of generated materials is through Density Function Theory (DFT) calculations (Neugebauer & Hickel, 2013), which uses quantum mechanics to calculate thermodynamic properties such as formation energy and energy above the hull, thereby determining the stability of generated structures (Noh et al., 2019; Long et al., 2021; Choubisa et al., 2020; Dan et al., 2020; Korolev et al., 2020; Ren et al., 2022; Long et al., 2021; Kim et al., 2020). However, DFT calculations require extensive computational resources. Alternative proxy metrics such as pairwise atom distances and charge neutrality (Davies et al., 2019) were developed as a sanity check of generated materials (Xie et al., 2021; Flam-Shepherd & Aspuru-Guzik, 2023). Fingerprint distances (Zimmermann & Jain, 2020; Ward et al., 2016) have also been used to measure precision and recall between the generated set and some held-out test set (Ganea et al., 2021; Xu et al., 2022; Xie & Grossman, 2018; Flam-Shepherd & Aspuru-Guzik, 2023). To evaluate properties of generated materials, previous work learns a separate graph neural network, which has intrinsic errors. Furthermore, Bartel (2022) has shown that learned formation energies do not reproduce DFT-calculated relative stabilities, bringing the value of learned property based evaluation into question.

5 Limitations and Conclusion

We have presented the first diffusion model for materials generation that can scale to datasets with millions of materials. To enable effective scaling, we developed a novel representation, UniMat, based on the periodic table, which enables any crystal structure to be effectively represented. The advantage of UniMat lies in modeling flexibility which enables scalability and computational efficiency compared to traditional search methods. UniMat has a few limitations. It does not achieve 100% validity on complex datasets (e.g., MP-20). The UniMat representation is sparse when the chemical system is small, which incurs additional computational cost (e.g., 99% atoms might be null atoms). Despite these limitations, UniMat enables training of diffusion models that results in better generation quality than previous state-of-the-art learned materials generators. We further advocate for using DFT calculations to perform rigorous stability analysis of materials generated by generative models. Expanding UniMat to other materials (e.g., non-crystalline or amorphous) and broader scientific data is an exciting direction of future work.

REFERENCES

Rohan Anil, Andrew M Dai, Orhan Firat, Melvin Johnson, Dmitry Lepikhin, Alexandre Passos, Siamak Shakeri, Emanuel Taropa, Paige Bailey, Zhifeng Chen, et al. Palm 2 technical report. *arXiv preprint arXiv:2305.10403*, 2023.

Luis M Antunes, Keith T Butler, and Ricardo Grau-Crespo. Crystal structure generation with autoregressive large language modeling. *arXiv preprint arXiv:2307.04340*, 2023.

Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow, and Rianne Van Den Berg. Structured denoising diffusion models in discrete state-spaces. *Advances in Neural Information Processing Systems*, 34:17981–17993, 2021.

Christopher J Bartel. Review of computational approaches to predict the thermodynamic stability of inorganic solids. *Journal of Materials Science*, 57(23):10475–10498, 2022.

Christopher J Bartel, Amalie Trewartha, Qi Wang, Alexander Dunn, Anubhav Jain, and Gerbrand Ceder. A critical examination of compound stability predictions from machine-learned formation energies. *npj computational materials*, 6(1):97, 2020.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E Smidt, and Boris Kozinsky. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13(1): 2453, 2022.

J George Bednorz and K Alex Müller. Possible high t c superconductivity in the ba- la- cu- o system. *Zeitschrift für Physik B Condensed Matter*, 64(2):189–193, 1986.

Peter E Blöchl. Projector augmented-wave method. *Physical review B*, 50(24):17953, 1994.

Gowoon Cheon, Lusann Yang, Kevin McCloskey, Evan J Reed, and Ekin D Cubuk. Crystal structure search with random relaxations using graph networks. *arXiv preprint arXiv:2012.02920*, 2020.

Hitarth Choubisa, Mikhail Askerka, Kevin Ryczko, Oleksandr Voznyy, Kyle Mills, Isaac Tamblyn, and Edward H Sargent. Crystal site feature embedding enables exploration of large chemical spaces. *Matter*, 3(2):433–448, 2020.

Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pp. 424–432. Springer, 2016.

Callum J Court, Batuhan Yildirim, Apoorv Jain, and Jacqueline M Cole. 3-d inorganic crystal structure generation and property prediction via representation learning. *Journal of Chemical Information and Modeling*, 60(10):4518–4535, 2020.

Yabo Dan, Yong Zhao, Xiang Li, Shaobo Li, Ming Hu, and Jianjun Hu. Generative adversarial networks (gan) based efficient sampling of chemical composition space for inverse design of inorganic materials. *npj Computational Materials*, 6(1):84, 2020.

Daniel W Davies, Keith T Butler, Adam J Jackson, Jonathan M Skelton, Kazuki Morita, and Aron Walsh. Smact: Semiconducting materials by analogy and chemical theory. *Journal of Open Source Software*, 4(38):1361, 2019.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Daniel Flam-Shepherd and Alán Aspuru-Guzik. 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.

Octavian Ganea, Lagnajit Pattanaik, Connor Coley, Regina Barzilay, Klavs Jensen, William Green, and Tommi Jaakkola. Geomol: Torsional geometric generation of molecular 3d conformer ensembles. *Advances in Neural Information Processing Systems*, 34:13757–13769, 2021.

Francesco Giuliari, Gianluca Scarpellini, Stuart James, Yiming Wang, and Alessio Del Bue. Positional diffusion: Ordering unordered sets with diffusion probabilistic models. *arXiv preprint arXiv:2303.11120*, 2023.

Martin A Green, Anita Ho-Baillie, and Henry J Snaith. The emergence of perovskite solar cells. *Nature photonics*, 8(7):506–514, 2014.

Jürgen Hafner. Ab-initio simulations of materials using vasp: Density-functional theory and beyond. *Journal of computational chemistry*, 29(13):2044–2078, 2008.

Paul Z Hanakata, Ekin D Cubuk, David K Campbell, and Harold S Park. Forward and inverse design of kirigami via supervised autoencoder. *Physical Review Research*, 2(4):042006, 2020.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Jonathan Ho, William Chan, Chitwan Saharia, Jay Whang, Ruiqi Gao, Alexey Gritsenko, Diederik P Kingma, Ben Poole, Mohammad Norouzi, David J Fleet, et al. Imagen video: High definition video generation with diffusion models. *arXiv preprint arXiv:2210.02303*, 2022a.

Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J. Fleet. Video diffusion models, 2022b.

Jordan Hoffmann, Louis Maestrati, Yoshihide Sawada, Jian Tang, Jean Michel Sellier, and Yoshua Bengio. Data-driven approach to encoding and decoding 3-d crystal structures. *arXiv preprint arXiv:1909.00949*, 2019.

Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.

Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL materials*, 1(1), 2013.

Jaehyeong Jo, Seul Lee, and Sung Ju Hwang. Score-based generative modeling of graphs via the system of stochastic differential equations. In *International Conference on Machine Learning*, pp. 10362–10383. PMLR, 2022.

Sungwon Kim, Juhwan Noh, Geun Ho Gu, Alan Aspuru-Guzik, and Yousung Jung. Generative adversarial networks for crystal structure prediction. *ACS central science*, 6(8):1412–1420, 2020.

Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *Advances in neural information processing systems*, 34:21696–21707, 2021.

Vadim Korolev, Artem Mitrofanov, Artem Eliseev, and Valery Tkachenko. Machine-learning-assisted search for functional materials over extended chemical space. *Materials Horizons*, 7 (10):2710–2718, 2020.

Georg Kresse and Jürgen Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational materials science*, 6(1):15–50, 1996a.

Georg Kresse and Jürgen Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical review B*, 54(16):11169, 1996b.

Georg Kresse and Daniel Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical review b*, 59(3):1758, 1999.

Teng Long, Nuno M Fortunato, Ingo Opahle, Yixuan Zhang, Ilias Samathrakis, Chen Shen, Oliver Gutfleisch, and Hongbin Zhang. Constrained crystals deep convolutional generative adversarial network for the inverse design of crystal structures. *npj Computational Materials*, 7(1):66, 2021.

Teng Long, Yixuan Zhang, Nuno M Fortunato, Chen Shen, Mian Dai, and Hongbin Zhang. Inverse design of crystal structures for multicomponent systems. *Acta Materialia*, 231:117898, 2022.

James Lucas, George Tucker, Roger Grosse, and Mohammad Norouzi. Understanding posterior collapse in generative latent variable models. 2019.

Shitong Luo and Wei Hu. Diffusion probabilistic models for 3d point cloud generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2837–2845, 2021.

Zhaoyang Lyu, Zhifeng Kong, Xudong Xu, Liang Pan, and Dahua Lin. A conditional point diffusion-refinement paradigm for 3d point cloud completion. *arXiv preprint arXiv:2112.03530*, 2021.

Kiran Mathew, Joseph H Montoya, Alireza Faghaninia, Shyam Dwarakanath, Muratahan Aykol, Hanmei Tang, Iek-heng Chu, Tess Smidt, Brandon Bocklund, Matthew Horton, et al. Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows. *Computational Materials Science*, 139:140–152, 2017.

Amil Merchant, Simon Batzner, Samuel S. Schoenholz, Muratahan Aykol, Gowoon Cheon, and Ekin Dogus Cubuk. Scaling deep learning for materials discovery. 2023.

KJPC Mizushima, PC Jones, PJ Wiseman, and John B Goodenough. Lixcoo2 (0¡ x¡-1): A new cathode material for batteries of high energy density. *Materials Research Bulletin*, 15(6):783–789, 1980.

Koichi Momma and Fujio Izumi. Vesta 3 for three-dimensional visualization of crystal, volumetric and morphology data. *Journal of applied crystallography*, 44(6):1272–1276, 2011.

Shuji Nakamura. The roles of structural imperfections in ingan-based blue light-emitting diodes and laser diodes. *Science*, 281(5379):956–961, 1998.

Jörg Neugebauer and Tilmann Hickel. Density functional theory in materials science. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 3(5):438–448, 2013.

Chenhao Niu, Yang Song, Jiaming Song, Shengjia Zhao, Aditya Grover, and Stefano Ermon. Permutation invariant graph generation via score-based generative modeling. In *International Conference on Artificial Intelligence and Statistics*, pp. 4474–4484. PMLR, 2020.

Juhwan Noh, Jaehoon Kim, Helge S Stein, Benjamin Sanchez-Lengeling, John M Gregoire, Alan Aspuru-Guzik, and Yousung Jung. Inverse design of solid-state materials via a continuous representation. *Matter*, 1(5):1370–1384, 2019.

Jens Kehlet Nørskov, Thomas Bligaard, Jan Rossmeisl, and Claus Hviid Christensen. Towards the computational design of solid catalysts. *Nature chemistry*, 1(1):37–46, 2009.

Asma Nouira, Nataliya Sokolovska, and Jean-Claude Crivello. Crystalgan: learning to discover crystallographic structures with generative adversarial networks. *arXiv preprint arXiv:1810.11203*, 2018.

Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L Chevrier, Kristin A Persson, and Gerbrand Ceder. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013.

OpenAI. Gpt-4 technical report, 2023.

Teerachote Pakornchote, Natthaphon Choomphon-anomakhun, Sorrjit Arrerut, Chayanon Atthapak, Sakarn Khamkaeo, Thiparat Chotibut, and Thiti Bovornratanaraks. Diffusion probabilistic models enhance variational autoencoder for crystal structure generative modeling. *arXiv preprint arXiv:2308.02165*, 2023.

John P Perdew, Matthias Ernzerhof, and Kieron Burke. Rationale for mixing exact exchange with density functional approximations. *The Journal of chemical physics*, 105(22):9982–9985, 1996.

Chris J Pickard and RJ Needs. Ab initio random structure searching. *Journal of Physics: Condensed Matter*, 23(5):053201, 2011.

Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652–660, 2017.

Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation. In *International Conference on Machine Learning*, pp. 8821–8831. PMLR, 2021.

Zekun Ren, Juhwan Noh, Siyu Tian, Felipe Oviedo, Guangzong Xing, Qiaohao Liang, Armin Aberle, Yi Liu, Qianxiao Li, Senthilnath Jayavelu, et al. Inverse design of crystals using generalized invertible crystallographic representation. *arXiv preprint arXiv:2005.07609*, 3(6):7, 2020.

Zekun Ren, Siyu Isaac Parker Tian, Juhwan Noh, Felipe Oviedo, Guangzong Xing, Jiali Li, Qiaohao Liang, Ruiming Zhu, Armin G Aberle, Shijing Sun, et al. An invertible crystallographic representation for general inverse design of inorganic crystals with targeted properties. *Matter*, 5(1): 314–335, 2022.

Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in Neural Information Processing Systems*, 35:36479–36494, 2022.

Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in neural information processing systems*, 30, 2017.

Uriel Singer, Adam Polyak, Thomas Hayes, Xi Yin, Jie An, Songyang Zhang, Qiyuan Hu, Harry Yang, Oron Ashual, Oran Gafni, et al. Make-a-video: Text-to-video generation without text-video data. *arXiv preprint arXiv:2209.14792*, 2022.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pp. 2256–2265. PMLR, 2015.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

Akash Srivastava, Lazar Valkov, Chris Russell, Michael U Gutmann, and Charles Sutton. Veegan: Reducing mode collapse in gans using implicit variational learning. *Advances in neural information processing systems*, 30, 2017.

Andrij Vasylenko, Jacinthe Gamon, Benjamin B Duff, Vladimir V Gusev, Luke M Daniels, Marco Zanella, J Felix Shin, Paul M Sharp, Alexandra Morscher, Ruiyong Chen, et al. Element selection for crystalline inorganic solid discovery guided by unsupervised machine learning of experimentally explored chemistry. *Nature communications*, 12(1):5561, 2021.

Clement Vignac, Igor Krawczuk, Antoine Siraudin, Bohan Wang, Volkan Cevher, and Pascal Frossard. Digress: Discrete denoising diffusion for graph generation. *arXiv preprint arXiv:2209.14734*, 2022.

Logan Ward, Ankit Agrawal, Alok Choudhary, and Christopher Wolverton. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Computational Materials*, 2(1):1–7, 2016.

Tian Xie and Jeffrey C Grossman. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters*, 120(14):145301, 2018.

Tian Xie, Xiang Fu, Octavian-Eugen Ganea, Regina Barzilay, and Tommi Jaakkola. Crystal diffusion variational autoencoder for periodic material generation. *arXiv preprint arXiv:2110.06197*, 2021.

Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. Geodiff: A geometric diffusion model for molecular conformation generation. *arXiv preprint arXiv:2203.02923*, 2022.

Jason Yim, Brian L Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. Se (3) diffusion model with application to protein backbone generation. *arXiv preprint arXiv:2302.02277*, 2023.

Jiahui Yu, Yuanzhong Xu, Jing Yu Koh, Thang Luong, Gunjan Baid, Zirui Wang, Vijay Vasudevan, Alexander Ku, Yinfei Yang, Burcu Karagol Ayan, et al. Scaling autoregressive models for content-rich text-to-image generation. *arXiv preprint arXiv:2206.10789*, 2(3):5, 2022.

Nils ER Zimmermann and Anubhav Jain. Local structure order parameters and site fingerprints for quantification of coordination environment and crystal structure similarity. *RSC advances*, 10 (10):6063–6081, 2020.

Appendix

A Architecture and Training

We repurpose the 3D U-Net architecture (Çiçek et al., 2016; Ho et al., 2022b) which originally models the spatial and time dimensions of videos into modeling periods and groups of the periodic table as well as the number of atoms dimension, which can be seen as the time dimension in videos. We apply the spatial downsampling pass followed by the spatial upsampling pass with skip connections to the downsampling pass activations with interleaved 3D convolution and attention layers as in standard 3D U-Net. The hyperparamters in training the UniMat diffusion model are summarized in Table 4.

Hyperparameter	**Value**
Base channels	256
Optimizer	Adam ($\beta_1 = 0.9, \beta_2 = 0.99$)
Channel multipliers	1, 2, 4
Learning rate	0.0001
Blocks per resolution	3
Batch size	512
Attention resolutions	1, 3, 9
EMA	0.9999
Attention head dimension	64
Dropout	0.1
Training hardware	32 TPU-v4 chips
Training steps	200000
Diffusion noise schedule	cosine
Noise schedule log SNR range	[-20, 20]
Sampling timesteps	256
Sampling log-variance interpolation	$\gamma = 0.1$
Weight decay	0.0
Prediction target	ϵ

Table 4: Hyperparameters for training the UniMat diffusion model.

B Details of DFT Calculations

We use the Vienna ab initio simulation package (VASP) (Kresse & Furthmüller, 1996b;a) with the Perdew-Burke-Ernzerhof (PBE) (Perdew et al., 1996) functional and projector-augmented wave (PAW) (Blöchl, 1994; Kresse & Joubert, 1999) potentials in all DFT calculations. Our DFT settings are consistent with Materials Project workflows as encoded in pymatgen (Ong et al., 2013) and atomate (Mathew et al., 2017). We use consistent settings with the Materials Project workflow including the Hubbard U parameter applied to a subset of transition metals in DFT+U, 520 eV plane-wave basis cutoff, magnetization settings and the choice of PBE pseudopotentials, except for Li, Na, Mg, Ge, and Ga. For Li, Na, Mg, Ge, and Ga, we use more recent versions of the respective potentials with the same number of valence electrons. For all structures, we use the standard protocol of two stage relaxation of all geometric degrees of freedom, followed by a final static calculation along with the custodian package (Ong et al., 2013) to handle any VASP related errors that arise and adjust appropriate simulations. For the choice of KPOINTS, we also force gamma centered kpoint generation for hexagonal cells rather than the more traditional Monkhorst-Pack. We assume ferromagnetic spin initialization with finite magnetic moments, as preliminary attempts to incorporate different spin orderings showed computational costs prohibitive to sustain at the scale presented. In AIMD simulations, we turn off spin-polarization and use the NVT ensemble with a 2 fs time step, except for simulations including hydrogen, where we reduce the time step to 0.5 fs.

C Details of AIRSS and Conditional Evaluation

Random structures for conditional evaluation of UniMat are generated through Ab initio random structure search (Pickard & Needs, 2011). Random structures are initialized as "sensible" structures (obeying certain symmetry requirements) to a target volume then relaxed via soft-sphere potentials.

For this paper, we always generate 100 AIRSS structures for every composition, many of which failed to converge as detailed in Section 3.3. We try a range of initial volumes spanning 0.4 to 1.2 times a volume estimated by considering relevant atomic radii, finding that the DFT relaxation fails or does not converge for the whole range for each composition. Note that these settings could be further finetuned to optimize AIRSS for convergence rate.

To compute the convergence rate for AIRSS, we use a total of 57,655 compositions from previous AIRSS runs(Merchant et al., 2023), for which 31,917 converged, and hence the AIRSS convergence is 0.55. When we run conditional generation, we randomly sampled 157 compounds from the 31,917 AIRSS-converged compounds, and 309 compounds from the 25,738 compounds where AIRSS had no structure that converged. Among the 157 compounds where AIRSS converged, 137 from UniMat converged, and among the 309 compounds that AIRSS did not converge, 231 from UniMat converged, resulting in an overall convergence rate $137/157 * 31917/(31917 + 25738) + 231/309 * 25738/(31917 + 25738) = 0.817$ for UniMat.

D Additional Results

Method	Dataset	COV-R ↑	AMSD-R ↓	AMCD-R ↓	COV-P ↑	AMSD-P ↓	AMCD-P ↓
CDVAE	Perov-5	**99.4**	0.048	**0.696**	**98.4**	**0.059**	**1.27**
	Carbon-24	99.8	**0.048**	0.00	83.0	**0.134**	0.00
	MP-20	99.15	0.154	3.62	99.49	**0.1883**	4.014
UniMat	Perov5	99.2	**0.046**	0.711	98.2	0.074	1.399
	Carbon24	**100**	**0.018**	0.0	**96.5**	**0.052**	0.0
	MP20	**99.8**	**0.097**	**2.41**	**99.7**	**0.119**	**2.41**

Table 5: Full proxy coverage metrics from CDVAE. UniMat performs better on larger datasets such as MP-20.