OXIGEN: OXIDATION-STATE-AWARE CRYSTAL GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Generative models have the potential to accelerate inorganic materials discovery by enabling inverse design, but generating experimentally realisable crystals remains challenging. Oxidation states are widely used to assess the compositional validity of crystals and guide inorganic materials discovery. While existing generative models for crystals can generate materials with charge-neutral oxidation-state assignments, they poorly reproduce the distributions of oxidation states observed in synthesised materials. To address this limitation, we propose OxiGen, an oxidation-state-aware crystal diffusion model that explicitly represents oxidation states during generation. OxiGen enforces global charge neutrality by construction using a structured output layer with exact inference over a finite-state automaton. Empirically, OxiGen substantially improves oxidation-state fidelity, generates the highest rate of stable, unique, and novel crystals among evaluated methods, and maintains high compositional validity even under property conditioning.

1 INTRODUCTION

The discovery of new inorganic materials is central to the development of technologies including solid-state batteries, photovoltaic cells and electrocatalysis (Yao et al., 2023). Generative models have the potential to accelerate materials discovery by enabling *inverse design*, in which candidate crystal structures with desired properties are generated directly rather than selected by screening existing databases (Park et al., 2024). Recent approaches evaluate the stability of generated crystals using their predicted energy above the convex hull (Zeni et al., 2025; Miller et al., 2024; Veljković et al., 2026), but a low predicted energy above hull does not guarantee that a crystal is experimentally realisable (Park et al., 2026a). Chemical heuristics derived from theory and empirical observation therefore remain important for guiding materials discovery.

Oxidation states are charges assigned to atoms through simple electron-counting rules (Karen, 2015), and are a key heuristic for understanding the chemical bonding and properties of inorganic crystals (Walsh et al., 2018). The oxidation states assigned to a crystal must satisfy *charge neutrality*, meaning that they sum to zero. Each element can adopt a specific set of oxidation states, determined by its electronic configuration and orbital energies, and the observed frequencies of these oxidation states in synthesised materials vary widely. *Compositional validity* evaluates whether the chemical composition of a crystal has at least one valid charge-neutral oxidation-state assignment (Davies et al., 2019), but does not account for the observed frequencies of the required oxidation states. Rare oxidation states often occur only in particular bonding environments, under specific external conditions, or through specialised synthesis routes (Mastej et al., 2026). Consequently, crystals with uncommon oxidation-state assignments may be less likely to be experimentally realisable.

Recent generative models for crystals achieve high compositional validity, but many generated crystals can only be charge balanced using rare oxidation states (Mastej et al., 2026). Filtering based on predicted energy above hull does not necessarily address this issue: analysis of crystals generated by GNoME (Merchant et al., 2023) identified structures predicted to lie on the hull that nevertheless exhibit improbable oxidation-state assignments (Cheetham & Seshadri, 2024). This highlights a limitation of current stability metrics, which may not capture whether the required oxidation states are chemically plausible. We therefore argue that effective generative models for crystals should not only generate stable and compositionally valid crystals, but also reproduce the distributions of oxidation states observed in synthesised materials.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

In this work, we propose a novel approach to oxidation-state-aware crystal generation that explicitly represents oxidation states during generation, allowing the model to learn oxidation-state frequencies directly from the training data. However, explicitly generating oxidation states requires that the predicted oxidation states for all atoms jointly satisfy charge neutrality. To address this, we introduce a *structured output layer* that exploits the simple additive structure of the charge-neutrality constraint for tractable structured inference. We then integrate this layer into a crystal diffusion model to enforce charge neutrality during sampling. Our main contributions are:

- We propose a structured output layer that uses dynamic programming over a finite-state automaton for exact charge-neutral inference.
- We introduce OxiGen, an oxidation-state-aware crystal diffusion model that explicitly represents oxidation states and enforces charge neutrality during generation.
- We show empirically that OxiGen substantially improves oxidation-state fidelity over existing crystal generative models, generates the highest rate of stable, unique, and novel crystals among evaluated methods, and maintains high compositional validity even under property conditioning.

2 Related Work

Crystal Generative Models. Recent work has explored a wide range of generative methods for inorganic crystals, including diffusion models (Xie et al., 2022; Jiao et al., 2023; Zeni et al., 2025; Joshi et al., 2025; Veljković et al., 2026; Park & Walsh, 2026; Yi et al., 2026), conditional flow matching (Miller et al., 2024; Luo et al., 2025), and stochastic interpolants (Höllmer et al., 2025). Several approaches incorporate crystallographic constraints directly into the representation or generative process, particularly space-group symmetry (Jiao et al., 2024; Levy et al., 2025; Chang et al., 2025). Other approaches address compositional validity. Crystal-GFN enforces charge neutrality as a hard constraint within a GFlowNet (Mila AI4Science et al., 2023), but only generates composition, space group and lattice parameters, not atomic coordinates. Mastej et al. (2026) use an oxidation-state-based chemical validity operator as a reinforcement-learning reward, but do not include oxidation states within the generative model. CrysVCD proposes a chemical-formula transformer that represents compositions as sequences of element–oxidation-state pairs, but requires a separate diffusion model to generate the corresponding crystal structure (Cheng et al., 2026). ChargeDIFF instead jointly generates the crystal structure and its continuous charge density, but focuses on modelling electronic structure rather than addressing compositional validity (Park et al., 2026b). In contrast to these approaches, OxiGen incorporates oxidation states directly within a crystal diffusion model, jointly generating composition and structure while enforcing charge neutrality by construction.

Structured Prediction. Structured prediction layers are a common approach for enforcing hard constraints on discrete neural network predictions (Dragone et al., 2021; Hoernle et al., 2022; Ahmed et al., 2022). In particular, Semantic Probabilistic Layers (Ahmed et al., 2022) construct a structured distribution by restricting the support of a factorised predictive distribution to assignments satisfying a symbolic constraint. More recently, constraints have been incorporated into the sampling process of discrete diffusion models (Cardei et al., 2025; Christopher et al., 2025; Suresh et al., 2025; Tomasi et al., 2026). DINGO (Suresh et al., 2025) represents regular-language constraints with finite automata and uses dynamic programming for exact constrained maximum a posteriori (MAP) decoding in diffusion language models. In this work, we represent charge neutrality as a finite-state automaton and show that the resulting dynamic program supports both exact MAP decoding and exact constrained sampling over oxidation-state assignments. Concurrent work by Dang & Ermon (2026) independently develops exact constrained sampling over finite automata for diffusion language models.

3 Preliminaries

Crystal representation. A crystal is a periodic arrangement of atoms, represented by a repeating unit cell whose translations tile 3D space. We refer to each indexed atom in the unit cell as an *atomic site*. A unit cell containing N atomic sites can be represented by the triple $\boldsymbol{M} = (\boldsymbol{A}, \boldsymbol{F}, \boldsymbol{L})$. Here, $\boldsymbol{A} = (a_1, a_2, \ldots, a_N) \in \mathcal{A}^N$ gives the element identity of each site, where $\mathcal{A}$ is the set of chemical

elements; $\boldsymbol{F} = [\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_N] \in [0,1)^{3\times N}$ gives the fractional coordinates of each site within the unit cell; and $\boldsymbol{L} = [\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3] \in \mathbb{R}^{3\times 3}$ is called the *lattice matrix*, whose columns are the three basis vectors defining the unit cell. The corresponding Cartesian coordinates are $\boldsymbol{X} = \boldsymbol{L}\boldsymbol{F}$.

Oxidation states. Formal oxidation states are integer-valued charges associated with the atoms in a crystal. For each element $a \in \mathcal{A}$, we denote its set of admissible oxidation states by $\mathcal{Z}(a) \subset \mathbb{Z}$. We define a *species* as an element–oxidation-state pair $s = (a, z)$, and let $\mathcal{S} = \{(a, z) : a \in \mathcal{A}, z \in \mathcal{Z}(a)\}$ denote the species vocabulary. Charge neutrality requires that the oxidation states of the atoms in a crystal sum to zero, $\sum_{i=1}^{N} z_i = 0$. While the majority of synthesised inorganic crystals have a charge-neutral oxidation-state assignment, formal integer oxidation states can become ambiguous in certain classes of materials. In this work we assume all crystals have a charge-neutral integer oxidation-state assignment. We discuss the limitations of this assumption in Section 6.

Discrete diffusion. Discrete diffusion models generate categorical data by corrupting clean samples through a fixed forward process q and learning a reverse process that progressively denoises them. Let $\boldsymbol{S}_0 = (s_{0,1}, \ldots, s_{0,N}) \in \mathcal{S}^N$ denote a clean *species assignment* over the finite vocabulary $\mathcal{S}$. Absorbing-state diffusion independently replaces each species with a mask token m according to a schedule $\alpha_t \in [0,1]$ (Austin et al., 2021; Sahoo et al., 2024):

$$q(s_{t,i} \mid s_{0,i}) = \begin{cases} 1-\alpha_t, & s_{t,i} = m, \\ \alpha_t, & s_{t,i} = s_{0,i}, \end{cases} \qquad q(\boldsymbol{S}_t \mid \boldsymbol{S}_0) = \prod_{i=1}^{N} q(s_{t,i} \mid s_{0,i}). \tag{1}$$

Under the x_0-parameterisation, a denoiser predicts a factorised distribution over the clean state,

$$p_\theta(\boldsymbol{S}_0 \mid \boldsymbol{S}_t, t) = \prod_{i=1}^{N} p_{\theta,i}(s_{0,i} \mid \boldsymbol{S}_t, t). \tag{2}$$

The clean-state prediction can be combined with the tractable posterior of the forward process to obtain a reverse transition,

$$p_\theta(\boldsymbol{S}_r \mid \boldsymbol{S}_t, t) = \prod_{i=1}^{N} \sum_{s\in\mathcal{S}} q(s_{r,i} \mid s_{t,i}, s_{0,i} = s)\, p_{\theta,i}(s \mid \boldsymbol{S}_t, t), \qquad r < t. \tag{3}$$

4 Proposed Method: OxiGen

4.1 Species representation

Previous generative models for crystals typically represent a crystal as $\boldsymbol{M} = (\boldsymbol{A}, \boldsymbol{F}, \boldsymbol{L})$, where $\boldsymbol{A}$ specifies the element identity of each atomic site. In these models, oxidation states are implicit, and many generated crystals can only be charge balanced using uncommon oxidation states (Mastej et al., 2026). We instead characterise each site by a species $s = (a, z)$ that jointly specifies its element identity and oxidation state. OxiGen therefore generates a categorical species token at each site, drawn from a vocabulary $\mathcal{S}$ of species observed in synthesised materials. The element assignment $\boldsymbol{A}$ is replaced by a species assignment $\boldsymbol{S} = (s_1, \ldots, s_N) \in \mathcal{S}^N$, where each $s_i \in \mathcal{S}$ has element $a(s_i)$ and oxidation state $z(s_i)$, and the overall crystal is represented as $\boldsymbol{M} = (\boldsymbol{S}, \boldsymbol{F}, \boldsymbol{L})$. For the species assignment to be chemically meaningful, $\boldsymbol{S}$ must satisfy charge neutrality, requiring the oxidation states across all sites to sum to zero.

4.2 Diffusion Model

OxiGen is a crystal diffusion model that jointly generates the species assignment, fractional coordinates, and lattice matrix. Starting from a clean crystal $\boldsymbol{M}_0 = (\boldsymbol{S}_0, \boldsymbol{F}_0, \boldsymbol{L}_0)$, the forward diffusion process independently corrupts each component to obtain a noisy crystal $\boldsymbol{M}_t = (\boldsymbol{S}_t, \boldsymbol{F}_t, \boldsymbol{L}_t)$,

$$q(\boldsymbol{M}_t \mid \boldsymbol{M}_r) = q(\boldsymbol{S}_t \mid \boldsymbol{S}_r)\, q(\boldsymbol{F}_t \mid \boldsymbol{F}_r)\, q(\boldsymbol{L}_t \mid \boldsymbol{L}_r) \qquad 0 \le r < t \le 1. \tag{4}$$

The reverse process is parameterised by a shared SE(3)-equivariant graph neural network that takes the complete noisy crystal as input. Here, we adapt MatterGen (Zeni et al., 2025), retaining the network architecture and continuous diffusion processes for the lattice matrix and fractional coordinates,

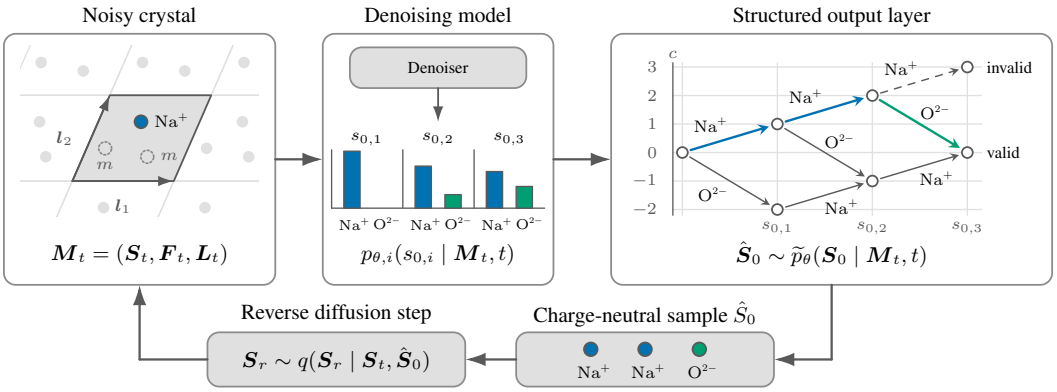


Figure 1: Example of OxiGen discrete diffusion sampling over a species vocabulary of $\{\mathrm{Na}^+, \mathrm{O}^{2-}\}$. The noisy crystal state $\boldsymbol{M}_t$ is processed by the denoiser, which predicts site-wise species probabilities $p_{\theta,i}(s_{0,i} \mid \boldsymbol{M}_t, t)$. These are passed through the structured output layer to sample a charge-neutral assignment $\hat{\boldsymbol{S}}_0$, which is used in the reverse diffusion step.

and modify only the discrete representation and diffusion process over species. We focus on these modifications here; further architectural and training details are provided in Appendix C.

For the species component, we use an absorbing-state discrete diffusion process based on Sahoo et al. (2024). Given a noisy crystal state $\boldsymbol{M}_t$, the denoiser defines a factorised conditional distribution over the clean species assignment,

$$p_\theta(\boldsymbol{S}_0 \mid \boldsymbol{M}_t, t) = \prod_{i=1}^{N} p_{\theta,i}(s_{0,i} \mid \boldsymbol{M}_t, t). \tag{5}$$

Although each factor $p_{\theta,i}$ may depend on the complete noisy crystal $\boldsymbol{M}_t$, independently sampling from the site-wise distributions can in general produce assignments that violate charge neutrality. Enforcing charge neutrality as a global constraint therefore requires coupling the site-wise predictions during inference.

4.3 Structured Inference

We define a structured joint distribution by conditioning the factorised distribution on charge neutrality,

$$\widetilde{p}_\theta(\boldsymbol{S}_0 \mid \boldsymbol{M}_t, t) = \frac{1}{Z_\theta(\boldsymbol{M}_t, t)} \mathbb{1}\left[\sum_{i=1}^{N} z(s_{0,i}) = 0\right] \prod_{i=1}^{N} p_{\theta,i}(s_{0,i} \mid \boldsymbol{M}_t, t), \tag{6}$$

where $Z_\theta(\boldsymbol{M}_t, t) = \sum_{\boldsymbol{S}_0 \in \mathcal{S}^N} \mathbb{1}\left[\sum_{i=1}^{N} z(s_{0,i}) = 0\right] \prod_{i=1}^{N} p_{\theta,i}(s_{0,i} \mid \boldsymbol{M}_t, t)$ is the normalising constant. The support of $\widetilde{p}_\theta$ is restricted to charge-neutral assignments, so the constraint is satisfied by construction. However, computing Z_θ naively requires summing over an exponentially large space of species assignments. We instead represent the charge-neutrality constraint as a finite-state automaton, enabling exact inference by dynamic programming.

Charge neutrality as a finite-state automaton. A *finite-state automaton* consists of a finite set of states and a transition rule between them, with valid sequences corresponding to paths from an initial state to an accepting state. Charge neutrality has a simple additive structure that can be represented by a finite-state automaton. After assigning the first i sites, whether the remaining sites can satisfy charge neutrality depends only on the cumulative charge

$$c_i = \sum_{j=1}^{i} z(s_{0,j}). \tag{7}$$

Charge-neutral assignments therefore correspond to paths through a layered finite-state automaton with initial state $c_0 = 0$, transition rule $c_i = c_{i-1} + z(s_{0,i})$, and accepting state $c_N = 0$, as illustrated

by the example in Figure 1. Dynamic programming over this automaton enables exact, tractable inference without summing over complete assignments. We adapt the standard forward–backward procedure for hidden Markov models (Rabiner, 1989). Let $\mathcal{C}_i$ denote the set of reachable cumulative charges after assigning the first i sites. The forward weight $F_i(c)$ gives the total probability mass of assignments to the first i sites with cumulative charge c,

$$F_i(c) = \sum_{\substack{s_{0,1:i} \in \mathcal{S}^i \\ \sum_{j=1}^{i} z(s_{0,j}) = c}} \prod_{j=1}^{i} p_{\theta,j}(s_{0,j} \mid \boldsymbol{M}_t, t), \tag{8}$$

while the backward weight $B_i(c)$ gives the probability mass of assignments to the remaining sites that complete charge c to zero,

$$B_i(c) = \sum_{\substack{s_{0,i+1:N} \in \mathcal{S}^{N-i} \\ c + \sum_{j=i+1}^{N} z(s_{0,j}) = 0}} \prod_{j=i+1}^{N} p_{\theta,j}(s_{0,j} \mid \boldsymbol{M}_t, t). \tag{9}$$

These quantities satisfy the recursions

$$F_0(c) = \mathbb{1}[c = 0], \qquad F_i(c) = \sum_{s \in \mathcal{S}} p_{\theta,i}(s \mid \boldsymbol{M}_t, t) F_{i-1}(c - z(s)), \tag{10}$$

$$B_N(c) = \mathbb{1}[c = 0], \qquad B_{i-1}(c) = \sum_{s \in \mathcal{S}} p_{\theta,i}(s \mid \boldsymbol{M}_t, t) B_i(c + z(s)). \tag{11}$$

Although the recursions are sequential, inference is invariant to the ordering of the atomic sites because charge neutrality depends only on the sum of their oxidation states, which is commutative. After all N sites have been assigned, $F_N(0)$ is by definition the total probability mass of charge-neutral assignments, which is the normalising constant $Z_\theta(\boldsymbol{M}_t, t)$. Similarly, $B_0(0)$ sums the same mass from the backward direction, so

$$Z_\theta(\boldsymbol{M}_t, t) = F_N(0) = B_0(0). \tag{12}$$

To obtain the marginal probability that site i is assigned species s, we sum the probability mass of all charge-neutral assignments consistent with $s_{0,i} = s$. For any cumulative charge c before site i, $F_{i-1}(c)$ gives the mass of compatible prefix assignments, $p_{\theta,i}(s \mid \boldsymbol{M}_t, t)$ gives the site-wise probability of assigning s, and $B_i(c + z(s))$ gives the mass of suffix assignments that return the resulting charge to zero. Summing over all reachable c and normalising by Z_θ therefore gives

$$\widetilde{p}_{\theta,i}(s \mid \boldsymbol{M}_t, t) = \frac{p_{\theta,i}(s \mid \boldsymbol{M}_t, t)}{Z_\theta(\boldsymbol{M}_t, t)} \sum_{c \in \mathcal{C}_{i-1}} F_{i-1}(c)\, B_i(c + z(s)). \tag{13}$$

A complete forward or backward pass has time complexity $O(N^2|\mathcal{S}|(z_{\max} - z_{\min}))$, where $z_{\max}$ and $z_{\min}$ are the maximum and minimum oxidation states in the species vocabulary. We report empirical computational overhead in Appendix C.3.

Charge-neutral MAP decoding. We can compute the MAP assignment under the charge-neutral structured distribution in Eq. (6) using the Viterbi algorithm (Forney, 1973). Let $V_i(c)$ denote the maximum probability of any species assignment to the first i sites with cumulative charge c. The recursion is

$$V_i(c) = \max_{s \in \mathcal{S}} p_{\theta,i}(s \mid \boldsymbol{M}_t, t)\, V_{i-1}(c - z(s)), \qquad V_0(c) = \mathbb{1}[c = 0]. \tag{14}$$

For each (i, c), we store the maximising species as a backpointer.

Proposition 1 (Exact charge-neutral MAP decoding)**.** *Backtracking from $V_N(0)$ returns the exact MAP species assignment under $\widetilde{p}_\theta(\boldsymbol{S}_0 \mid \boldsymbol{M}_t, t)$.*

Charge-neutral sampling. While the Viterbi recursion provides exact MAP decoding, the reverse diffusion process requires samples from the full structured distribution. During finite-step generation, a reverse step may unmask multiple sites simultaneously. Sampling these sites independently would

discard the dependencies induced by charge neutrality. We therefore sample a joint clean species assignment from the structured distribution. Given cumulative charge c_{i-1} after the first $i-1$ sites, the conditional distribution for site i is

$$\widetilde{p}_\theta(s_{0,i} = s \mid c_{i-1}, \boldsymbol{M}_t, t) = \frac{p_{\theta,i}(s \mid \boldsymbol{M}_t, t)\, B_i(c_{i-1} + z(s))}{B_{i-1}(c_{i-1})}, \tag{15}$$

where the backward weights $B_i(c)$ are computed using the backward recursion in Eq. (11).

Proposition 2 (Exact charge-neutral sampling)**.** *Sequentially sampling $s_{0,i}$ according to Eq.* (15) *for $i = 1, \ldots, N$ produces an exact sample from $\widetilde{p}_\theta(\boldsymbol{S}_0 \mid \boldsymbol{M}_t, t)$.*

At each OxiGen reverse step, we first sample a complete charge-neutral clean species assignment $\hat{\boldsymbol{S}}_0 \sim \widetilde{p}_\theta(\boldsymbol{S}_0 \mid \boldsymbol{M}_t, t)$ and then sample $\boldsymbol{S}_r \sim q(\boldsymbol{S}_r \mid \boldsymbol{S}_t, \hat{\boldsymbol{S}}_0)$ from the tractable posterior of the forward absorbing process. This procedure avoids explicitly marginalising over clean assignments as in Eq. (3), but is equivalent to sampling from the corresponding marginal reverse transition under the structured distribution. It therefore preserves the dependencies among sites unmasked within the same reverse step and guarantees that the final species assignment is charge neutral. For OxiGen, the structured output layer is applied only during sampling, while training uses the standard factorised masked-diffusion objective. We derive the corresponding structured training objective in Appendix C.1 and evaluate it in Section 5.2.

5 Experiments

5.1 Oxidation-State Prediction

Training OxiGen requires crystal structures labelled with oxidation states, which are not provided for all structures in public crystal-structure databases such as the Materials Project (Jain et al., 2013). We therefore train a crystal transformer adapted from Veljković et al. (2026) for oxidation-state prediction. To guarantee charge-neutral predictions, we use the structured Viterbi decoding introduced in Section 4.3. We additionally modify the dynamic program so that mixed valence, where different sites of the same element have different oxidation states, is only permitted for elements with established mixed-valence chemistry. Further details are provided in Appendix B.3. We compare against `BVAnalyzer` (Ong et al., 2013), a standard method based on empirical bond-valence analysis that assigns oxidation states subject to charge neutrality, and BERTOS (Fu et al., 2023), a composition-only transformer that does not guarantee charge-neutral predictions.

Evaluation. We evaluate oxidation-state prediction on 28,492 experimental crystal structures from the Inorganic Crystal Structure Database (ICSD) (Zagorac et al., 2019), a curated, licensed database that provides oxidation-state labels for each atomic site. We use a composition-disjoint 22,879/2,831/2,782 train/validation/test split. We reimplement BERTOS and train it on the same split. For our model, we evaluate both standard argmax decoding and structured Viterbi decoding. Following Fu et al. (2023), we report the charge-neutrality rate, defined as the fraction of crystals with charge-neutral predictions; site accuracy, the fraction of atomic sites assigned the correct oxidation state; and crystal accuracy, the fraction of crystals for which all atomic sites are assigned correctly. `BVAnalyzer` fails for 9.1% of test structures for which no assignment is compatible with the bond-valence model; we count these cases as incorrect for site and crystal accuracy and exclude them when computing charge neutrality. Further details on dataset curation, model architecture, and training are provided in Appendix B. Table 1 shows that our model achieves higher site and crystal accuracy than both baselines. Charge-neutral Viterbi decoding enforces charge neutrality for all predictions and further improves site and crystal accuracy over argmax decoding.

MP-20-OS dataset. We use our predictor to label MP-20, a subset of the Materials Project (Jain et al., 2013) containing crystal structures with up to 20 atoms in the unit cell. We use the splits provided by Zeni et al. (2025) and label each split independently. Following convention, structures composed exclusively of metals are assigned oxidation state zero at every site, while all remaining structures are labelled using our predictor. We remove 5.9% of structures using the filtering procedure described in Appendix B.4, including 3.7% with no charge-neutral oxidation-state assignment. The resulting MP-20-OS dataset contains 25,536/8,495/8,545 train/validation/test structures.

Table 1: Oxidation-state prediction results, reported as mean ± std over five independently seeded training runs; the deterministic BVAnalyzer is reported as a single value. Best values are shown in **bold** and second-best values are underlined.

Method	Decoding	Charge Neutrality (%) ↑	Site Accuracy (%) ↑	Crystal Accuracy (%) ↑
`BVAnalyzer`	–	**100.00**	88.19	86.02
BERTOS	Argmax	92.83±0.22	97.81±0.17	90.93±0.24
Ours	Argmax	94.37±0.26	98.47±0.10	92.79±0.18
	Viterbi	**100.00**±0.00	**98.74**±0.09	**96.00**±0.25

Table 2: *De novo* generation results, reported as mean ± std over five independently seeded training runs, with 10,240 generated structures per seed. Best values are shown in **bold** and second-best values are underlined. MP-20-OS denotes the test set.

Method	Oxidation-State Fidelity			Stability & Relaxation		
	Comp. Val. (%) ↑	f_{OS} (%)	d_{OS} ↓	S.U.N. (%) ↑	E_{hull} (eV/atom) ↓	RMSD (Å) ↓
MP-20-OS	91.73	53.78	–	–	–	–
DiffCSP	78.47±2.55	35.67±1.05	0.1566±0.0037	18.24±1.38	0.2007±0.0101	0.3555±0.0427
MatterGen	81.18±0.94	37.89±1.02	0.1537±0.0039	24.54±0.91	0.1799±0.0085	0.1747±0.0203
FlowMM	79.87±0.20	37.34±0.65	0.1626±0.0037	13.85±0.39	0.2481±0.0104	0.6704±0.0347
OMatG	82.32±0.48	43.08±0.35	0.1284±0.0060	19.03±0.78	0.1831±0.0030	0.5244±0.0278
Crystalite	82.66±0.51	42.45±1.04	0.1350±0.0017	22.83±0.76	**0.1437**±0.0083	0.2496±0.0198
OxiGen	**90.65**±0.94	**53.01**±0.96	**0.0911**±0.0028	**26.38**±1.01	0.1575±0.0080	**0.1679**±0.0117

5.2 Crystal Generation

We evaluate OxiGen on *de novo* crystal generation using MP-20-OS and compare against five strong baseline methods: DiffCSP (Jiao et al., 2023), MatterGen (Zeni et al., 2025), FlowMM (Miller et al., 2024), OMatG (Höllmer et al., 2025), and Crystalite (Veljković et al., 2026). We retrain all methods on MP-20-OS using the authors' original code and reported hyperparameters over five independent random seeds, and sample 10,240 crystals from each method for each seed.

Evaluation metrics. Compositional validity, computed using `SMACT` (Davies et al., 2019), measures whether a composition has a valid charge-neutral oxidation-state assignment. Although every crystal in MP-20-OS is labelled with a charge-neutral assignment, the test-set compositional validity is not perfect because the `SMACT` evaluation additionally requires electronegativity balance and does not permit mixed-valence assignments. Since compositional validity does not account for how common the required oxidation states are, we report two additional metrics. The oxidation-state frequency score f_{OS} is the geometric mean of the ICSD frequencies of distinct species in an assignment (Mastej et al., 2026), and measures how commonly those oxidation states occur in synthesised materials. We report the mean across crystals, with the goal of matching the test-set value rather than simply favouring the most common oxidation states. The oxidation-state distribution distance d_{OS} measures the Jensen–Shannon distance between the oxidation-state distributions of the generated and test sets for each element, reported as the mean across elements. For a consistent comparison across methods, we enumerate all valid assignments for each crystal using `SMACT`, rather than using oxidation states generated by OxiGen or inferred by our predictor. For f_{OS} we take the maximum score across these assignments and for d_{OS} we construct the distributions by marginalising over assignments weighted by their ICSD occurrence frequencies. Metallic alloys and elemental structures are excluded from all three metrics because oxidation states are not meaningfully defined for these systems. Full definitions and evaluation details are provided in Appendix E.2.

Oxidation-state fidelity. OxiGen substantially improves oxidation-state fidelity over the baselines, closely matching the MP-20-OS test-set values for compositional validity and oxidation-state frequency score while achieving the lowest oxidation-state distribution distance (Table 2). As shown in Figure 2, the baselines rely more heavily on uncommon oxidation states, while OxiGen closely reproduces both the test-set compositional validity curve and oxidation-state distributions.

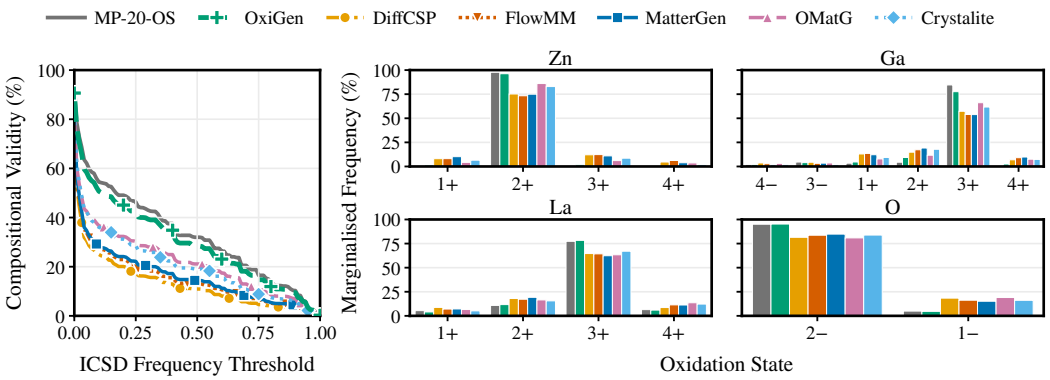


Figure 2: **Left**: Compositional validity as a function of the ICSD frequency threshold required for an oxidation state to be considered admissible. **Right**: Marginal oxidation-state distributions for selected elements. MP-20-OS denotes the test set in both panels.

Table 3: *De novo* generation ablation results, reported as mean ± std over five independently seeded training runs, with 10,240 generated structures per seed. Best values are shown in **bold** and second-best values are underlined. The shaded row is OxiGen.

Representation	Structured Output Layer		Charge Neutrality	Comp. Val.	f_{OS}	d_{OS}
	Training	Sampling	(%) ↑	(%) ↑	(%)	↓
Element	✗	✗	–	81.12±0.46	37.78±1.16	0.1589±0.0050
Species	✗	✗	39.08±2.60	82.31±0.77	40.54±1.29	0.1368±0.0047
Species	✗	✓	**100.00**±0.00	90.65±0.94	**53.01**±0.96	0.0911±0.0028
Species	✓	✓	**100.00**±0.00	**90.70**±0.67	52.65±0.76	**0.0874**±0.0029

Stability, uniqueness, and novelty. Alongside oxidation-state fidelity, we evaluate the ability of each method to generate stable, unique, and novel crystals. Since *de novo* generation typically involves a trade-off between stability and novelty (Park & Walsh, 2026), we use the proportion of stable, unique, and novel (S.U.N.) structures as the primary evaluation metric. Following Zeni et al. (2025), generated structures are relaxed with the MatterSim machine-learning interatomic potential (Yang et al., 2024), and the predicted energies are used to compute energy above the convex hull, E_{hull}. A structure is considered stable if $E_{\mathrm{hull}} \leq 0.1$ eV/atom. We additionally report the mean E_{hull} and the root-mean-square deviation (RMSD) between the generated and relaxed structures. Further evaluation details are provided in Appendix E.1. OxiGen achieves the highest S.U.N. rate, lowest relaxation RMSD and second-lowest mean E_{hull} among the evaluated methods (Table 2). Relative to MatterGen, the increase in the S.U.N. rate corresponds to a higher proportion of stable structures, accompanied by a decrease in uniqueness and novelty (Appendix Table 9), consistent with OxiGen generating fewer crystals with chemically implausible compositions. We further evaluate OxiGen using LeMat-GenBench (Betala et al., 2025) and the proxy metrics proposed by Xie et al. (2022) (Appendix Tables 10 and 11). Overall, OxiGen substantially improves oxidation-state fidelity while maintaining strong performance on standard stability metrics.

Ablation study. We ablate the species representation and structured output layer to isolate the contribution of each component to OxiGen (Table 3). All ablation models use the same masked discrete diffusion process as OxiGen, differing only in the discrete state representation and whether the structured output layer is applied during training and sampling. We report the three oxidation-state metrics and the fraction of generated species assignments that satisfy charge neutrality, with the corresponding S.U.N. metrics reported in Appendix Table 12. The species representation alone provides only a modest improvement over the element representation, with the majority of generated assignments violating charge neutrality. Structured sampling enforces charge neutrality by construction and substantially improves all three oxidation-state metrics. These results suggest that explicitly representing oxidation states is insufficient on its own, and that enforcing global charge neutrality during sampling is needed to ensure chemically meaningful oxidation-state assignments. Training with the structured objective does not provide any consistent additional improvement while adding training overhead, and is therefore not used to train OxiGen.

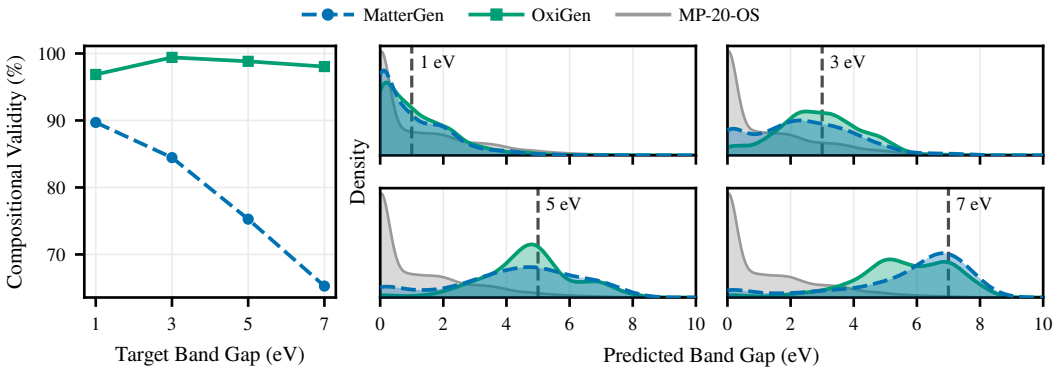


Figure 3: **Left**: Compositional validity at target band gaps of 1, 3, 5, and 7 eV. **Right**: Kernel density estimates of the predicted band-gap distributions for structures generated at each target. Results are shown for a single fine-tuning seed. Dashed vertical lines indicate the conditioning targets and the grey density shows the band-gap distribution for the MP-20-OS test set.

Conditional generation. Inverse design requires generating crystals with desired properties while maintaining generation quality. This becomes more challenging for targets that are weakly represented in the training data, as conditioning can push generation away from high-density regions of the learned distribution. Following Zeni et al. (2025), we fine-tune MatterGen and OxiGen on band-gap labels from MP-20-OS and generate 1,024 crystals conditioned on target band gaps of 1, 3, 5, and 7 eV using classifier-free guidance. These targets span the MP-20-OS band-gap distribution, which contains progressively fewer structures at higher band gaps (Figure 3). Band gaps are predicted from the relaxed structures using iComFormer, an SE(3)-invariant crystal graph transformer for property prediction (Yan et al., 2024). Further details on model fine-tuning and training the band-gap predictor are provided in Appendix D. As shown in Figure 3, both models shift the predicted band-gap distributions towards the conditioning targets. However, MatterGen's compositional validity decreases sharply as the target band gap increases, whereas OxiGen maintains consistently high validity across the full range. OxiGen also achieves a higher S.U.N. rate than MatterGen at all four conditioning targets (Appendix Table 14). These results demonstrate the value of structured sampling for inverse design, where conditioning on sparsely represented targets can otherwise lead to chemically invalid compositions.

6 Conclusions

In this work, we introduced OxiGen, a diffusion model for oxidation-state-aware crystal generation that explicitly represents oxidation states during generation and enforces charge neutrality using dynamic programming over a finite-state automaton. Across experiments, OxiGen more faithfully reproduces the oxidation-state distributions of synthesised materials, achieves the highest rate of stable, unique, and novel crystals among evaluated methods, and maintains high compositional validity during band-gap-conditioned generation. The proposed species representation and structured output layer require only minimal modification to the underlying diffusion model architecture, so we expect they could be readily incorporated into other crystal generative models. Overall, our results demonstrate how incorporating prior chemical knowledge through simple heuristics and constraints can guide generative models towards more chemically plausible materials.

Limitations. OxiGen inherits the limitations of formal integer oxidation states, which can be ambiguous or undefined for systems with substantial covalency, charge delocalisation, or metallic bonding (Karen, 2015; Walsh et al., 2018). As a result, OxiGen may mischaracterise or exclude valid materials for which a charge-neutral integer oxidation-state assignment is not chemically meaningful. OxiGen also relies on oxidation-state assignments inferred by our predictor, which may carry errors or biases. More broadly, charge neutrality is only one aspect of chemical validity: OxiGen does not enforce other constraints, such as electronegativity balance or consistency between an oxidation state and its local coordination and bonding environment. Future work could incorporate additional chemical constraints and explore more flexible inductive biases that retain the chemical information captured by oxidation states while moving beyond discrete formal assignments.

AI DISCLOSURE

In this work, we used generative AI tools to assist with the research process, including formalising the methodology, writing code, and implementing baselines. Generative AI tools were also used to help identify relevant literature, edit parts of the manuscript for readability, and create figures. The authors proposed the research question, developed the core methodology and experimental design and analysed the results. We have reviewed all AI-assisted work and take responsibility for the final content of this work, including text, claims or artefacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

We provide the information needed to reproduce our methodology and experiments. The diffusion model and structured inference procedure are described in Section 4, with further details in Appendix C. Dataset curation and preprocessing are described in Appendix B. Model architectures and hyperparameters are given in Appendices B.2, C.2 and D, and the evaluation procedure is described in Appendix E. Code and configuration files are provided as supplementary material.

REFERENCES

Dhruv Ahlawat, Vaibhav Mishra, Somaditya Singh, Mohd Zaki, Vaibhav Bihani, Hargun Singh Grover, Biswajit Mishra, Santiago Miret, Mausam, and N. M. Anoop Krishnan. A family of large language models for materials research with insights into model adaptability in continued pretraining. *Nature Machine Intelligence*, 8:435–448, 2026. doi: 10.1038/s42256-026-01199-8. URL https://www.nature.com/articles/s42256-026-01199-8.

Kareem Ahmed, Stefano Teso, Kai-Wei Chang, Guy Van den Broeck, and Antonio Vergari. Semantic Probabilistic Layers for Neuro-Symbolic Learning. In *Advances in Neural Information Processing Systems*, volume 35, pp. 29944–29959. Curran Associates, Inc., 2022. doi: 10.52202/068431-2171. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/c182ec594f38926b7fcb827635b9a8f4-Abstract-Conference.html.

Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured Denoising Diffusion Models in Discrete State-Spaces. In *Advances in Neural Information Processing Systems*, volume 34, pp. 17981–17993. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper/2021/hash/958c530554f78bcd8e97125b70e6973d-Abstract.html.

Siddharth Betala, Samuel P. Gleason, Ali Ramlaoui, Andy Xu, Georgia Channing, Daniel Levy, Clémentine Fourrier, Nikita Kazeev, Chaitanya K. Joshi, Sékou-Oumar Kaba, Félix Therrien, Alex Hernandez-Garcia, Rocío Mercado, N. M. Anoop Krishnan, and Alexandre Duval. LeMat-GenBench: A unified evaluation framework for crystal generative models, 2025. URL https://arxiv.org/abs/2512.04562.

Cyprien Bone, Matthew Walker, Kuangdai Leng, Luis M. Antunes, Ricardo Grau-Crespo, Amil Aligayev, Javier Dominguez, and Keith T. Butler. Discovery and recovery of crystalline materials with property-conditioned transformers, November 2025. URL https://arxiv.org/abs/2511.21299. arXiv:2511.21299 [cond-mat.mtrl-sci].

Zhendong Cao, Xiaoshan Luo, Jian Lv, and Lei Wang. Space group informed transformer for crystalline materials generation. *Science Bulletin*, 70(21):3522–3533, November 2025. doi: 10.1016/j.scib.2025.09.035. URL https://www.sciencedirect.com/science/article/pii/S2095927325009752.

Michael Cardei, Jacob K. Christopher, Bhavya Kailkhura, Tom Hartvigsen, and Ferdinando Fioretto. Constrained discrete diffusion. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, pp. 12383–12416. Curran Associates, Inc., 2025. doi: 10.52202/085713-0415. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/124c1cc9f7750f512ed81cdbf2fd5bf5-Paper-Conference.pdf.

Rees Chang, Angela Pak, Alex Guerra, Ni Zhan, Nick Richardson, Elif Ertekin, and Ryan P. Adams. Space Group Equivariant Crystal Diffusion. In *Advances in Neural*

Information Processing Systems, volume 38, pp. 72772–72805. Curran Associates, Inc., 2025. doi: 10.52202/085713-2439. URL https://papers.nips.cc/paper_files/paper/2025/hash/697b2f31f99fb79f8a0a16e923b2471d-Abstract-Conference.html.

Anthony K. Cheetham and Ram Seshadri. Artificial Intelligence Driving Materials Discovery? Perspective on the Article: Scaling Deep Learning for Materials Discovery. *Chemistry of Materials*, 36(8):3490–3495, April 2024. doi: 10.1021/acs.chemmater.4c00643. URL https://doi.org/10.1021/acs.chemmater.4c00643.

Mouyang Cheng, Weiliang Luo, Hao Tang, Bowen Yu, Yongqiang Cheng, Weiwei Xie, Ju Li, Heather J. Kulik, and Mingda Li. Enhancing materials discovery with valence-constrained design in generative modeling. *Nature Computational Science*, August 2026. doi: 10.1038/s43588-026-01037-2. URL https://www.nature.com/articles/s43588-026-01037-2.

Jacob K. Christopher, Michael Cardei, Jinhao Liang, and Ferdinando Fioretto. Neuro-Symbolic Generative Diffusion Models for Physically Grounded, Robust, and Safe Generation. In *Proceedings of the International Conference on Neuro-symbolic Systems*, volume 288, pp. 188–213. PMLR, 28–30 May 2025. URL https://proceedings.mlr.press/v288/christopher25a.html.

Meihua Dang and Stefano Ermon. Constrained Decoding for Diffusion Language Models via Efficient Inference over Finite Automata, July 2026. URL https://arxiv.org/abs/2607.07026. arXiv:2607.07026 [cs.LG].

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, June 2019. doi: 10.21105/joss.01361. URL https://joss.theoj.org/papers/10.21105/joss.01361.

Paolo Dragone, Stefano Teso, and Andrea Passerini. Neuro-Symbolic Constraint Programming for Structured Prediction. In *Proceedings of the 15th International Workshop on Neural-Symbolic Learning and Reasoning*, volume 2986 of *CEUR Workshop Proceedings*, pp. 6–14. CEUR-WS.org, 2021. URL https://ceur-ws.org/Vol-2986/.

G.D. Forney. The Viterbi algorithm. *Proceedings of the IEEE*, 61(3):268–278, March 1973. doi: 10.1109/PROC.1973.9030. URL https://ieeexplore.ieee.org/document/1450960.

Nihang Fu, Jeffrey Hu, Ying Feng, Gregory Morrison, Hans-Conrad zur Loye, and Jianjun Hu. Composition Based Oxidation State Prediction of Materials Using Deep Learning Language Models. *Advanced Science*, 10(28):2301011, 2023. doi: 10.1002/advs.202301011. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/advs.202301011.

Johannes Gasteiger, Florian Becker, and Stephan Günnemann. GemNet: Universal Directional Graph Neural Networks for Molecules. In *Advances in Neural Information Processing Systems*, volume 34, pp. 6790–6802. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper/2021/hash/35cf8659cfcb13224cbd47863a34fc58-Abstract.html.

Nick Hoernle, Rafael Karampatsis, Vaishak Belle, and Yakov Kobi Gal. MultiplexNet: Towards Fully Satisfied Logical Constraints in Neural Networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36(5), pp. 5700–5709. Association for the Advancement of Artificial Intelligence, June 2022. doi: 10.1609/aaai.v36i5.20512. URL https://www.research.ed.ac.uk/en/publications/multiplexnet-towards-fully-satisfied-logical-constraints-in-neura/.

Philipp Höllmer, Thomas Egg, Maya Martirossyan, Eric Fuemmeler, Zeren Shui, Amit Gupta, Pawan Prakash, Adrian Roitberg, Mingjie Liu, George Karypis, Mark Transtrum, Richard Hennig, Ellad B. Tadmor, and Stefano Martiniani. Open Materials Generation with Stochastic Interpolants. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267, pp. 23417–23450. PMLR, July 2025. URL https://proceedings.mlr.press/v267/hollmer25a.html.

Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, July 2013. doi: 10.1063/1.4812323. URL https://doi.org/10.1063/1.4812323.

Rui Jiao, Wenbing Huang, Peijia Lin, Jiaqi Han, Pin Chen, Yutong Lu, and Yang Liu. Crystal Structure Prediction by Joint Equivariant Diffusion. In *Advances in Neural Information Processing Systems*, volume 36, pp. 17464–17497. Curran Associates, Inc., 2023. doi: 10.52202/075280-0767. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/38b787fc530d0b31825827e2cc306656-Abstract-Conference.html.

Rui Jiao, Wenbing Huang, Yu Liu, Deli Zhao, and Yang Liu. Space Group Constrained Crystal Generation. In *The Twelfth International Conference on Learning Representations*, pp. 6836–6853, May 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/1ba1b2133ac3d3dda0ce5cd67dab8456-Abstract-Conference.html.

Chaitanya K. Joshi, Xiang Fu, Yi-Lun Liao, Vahe Gharakhanyan, Benjamin Kurt Miller, Anuroop Sriram, and Zachary Ward Ulissi. All-atom Diffusion Transformers: Unified generative modelling of molecules and materials. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267, pp. 28393–28417. PMLR, July 2025. URL https://proceedings.mlr.press/v267/joshi25a.html.

Pavel Karen. Oxidation State, A Long-Standing Issue! *Angewandte Chemie International Edition*, 54(16):4716–4726, 2015. doi: 10.1002/anie.201407561. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201407561.

Nikita Kazeev, Wei Nong, Ignat Romanov, Ruiming Zhu, Andrey E. Ustyuzhanin, Shuya Yamazaki, and Kedar Hippalgaonkar. Wyckoff Transformer: Generation of Symmetric Crystals. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267, pp. 29495–29526. PMLR, July 2025. URL https://proceedings.mlr.press/v267/kazeev25a.html.

Daniel Levy, Siba Smarak Panigrahi, Sékou-Oumar Kaba, Qiang Zhu, Kin Long Kelvin Lee, Mikhail Galkin, Santiago Miret, and Siamak Ravanbakhsh. SymmCD: Symmetry-Preserving Crystal Generation with Diffusion Models. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=xnssGv9rpW.

Xiaoshan Luo, Zhenyu Wang, Qingchang Wang, Xuechen Shao, Jian Lv, Lei Wang, Yanchao Wang, and Yanming Ma. CrystalFlow: a flow-based generative model for crystalline materials. *Nature Communications*, 16(1):9267, October 2025. doi: 10.1038/s41467-025-64364-4. URL https://www.nature.com/articles/s41467-025-64364-4.

Maya M. Martirossyan, Hillary Pan, Philipp Höllmer, and Stefano Martiniani. Framework-constrained materials generation. In *AI for Accelerated Materials Design Workshop at ICLR 2026*, 2026. URL https://openreview.net/forum?id=Yk4FcrVcU2.

Kinga O. Mastej, Panyalak Detrattanawichai, Hyunsoo Park, Anthony Onwuli, Masahiro Negishi, and Aron Walsh. Chemical filters for ultra-high-throughput materials screening and generation, July 2026. URL https://arxiv.org/abs/2607.17910. arXiv:2607.17910 [cond-mat.mtrl-sci].

Amil Merchant, Simon Batzner, Samuel S. Schoenholz, Muratahan Aykol, Gowoon Cheon, and Ekin Dogus Cubuk. Scaling deep learning for materials discovery. *Nature*, 624(7990):80–85, December 2023. doi: 10.1038/s41586-023-06735-9. URL https://www.nature.com/articles/s41586-023-06735-9.

Mila AI4Science, Alex Hernandez-Garcia, Alexandre Duval, Alexandra Volokhova, Yoshua Bengio, Divya Sharma, Pierre Luc Carrier, Yasmine Benabed, Michał Koziarski, and Victor Schmidt. Crystal-GFN: sampling crystals with desirable properties and constraints, October 2023. URL https://arxiv.org/abs/2310.04925v1. arXiv:2310.04925v1 [cs.LG].

Benjamin Kurt Miller, Ricky T. Q. Chen, Anuroop Sriram, and Brandon M. Wood. FlowMM: Generating Materials with Riemannian Flow Matching. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 35664–35686. PMLR, July 2024. URL https://proceedings.mlr.press/v235/miller24a.html.

Andrey Okhotin, Maksim Nakhodnov, Nikita Kazeev, Mikhail Lazarev, Andrey E. Ustyuzhanin, and Dmitry Vetrov. MiAD: Mirage atom diffusion for de novo crystal generation, November 2025. URL https://arxiv.org/abs/2511.14426. arXiv:2511.14426 [cs.LG].

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, February 2013. doi: 10.1016/j.commatsci.2012.10.028. URL https://www.sciencedirect.com/science/article/pii/S0927025612006295.

Hyunsoo Park and Aron Walsh. Guiding generative models to uncover diverse and novel crystals via reinforcement learning. *Nature Machine Intelligence*, 8(7):1087–1099, July 2026. doi: 10.1038/s42256-026-01262-4. URL https://www.nature.com/articles/s42256-026-01262-4.

Hyunsoo Park, Zhenzhu Li, and Aron Walsh. Has generative artificial intelligence solved inverse materials design? *Matter*, 7(7):2355–2367, July 2024. doi: 10.1016/j.matt.2024.05.017. URL https://www.cell.com/matter/abstract/S2590-2385(24)00242-X.

Hyunsoo Park, Anthony Onwuli, and Aron Walsh. Exploration of crystal chemical space using text-guided generative artificial intelligence. *Nature Communications*, 16(1):4379, 2025. doi: 10.1038/s41467-025-59636-y. URL https://www.nature.com/articles/s41467-025-59636-y.

Hyunsoo Park, Kinga O. Mastej, Panyalak Detrattanawichai, Ryan Nduma, and Aron Walsh. Closing the synthesis gap in computational materials design. *Nature Synthesis*, 5(5):651–659, May 2026a. doi: 10.1038/s44160-026-01027-2. URL https://www.nature.com/articles/s44160-026-01027-2.

Junkil Park, Junyoung Choi, and Yousung Jung. Generative modelling of inorganic materials with explicit electronic structure. *Nature Communications*, 17(1):7070, June 2026b. ISSN 2041-1723. doi: 10.1038/s41467-026-73985-2. URL https://www.nature.com/articles/s41467-026-73985-2.

L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, February 1989. doi: 10.1109/5.18626. URL https://ieeexplore.ieee.org/document/18626.

Subham Sekhar Sahoo, Marianne Arriola, Yair Schiff, Aaron Gokaslan, Edgar Marroquin, Justin T. Chiu, Alexander Rush, and Volodymyr Kuleshov. Simple and Effective Masked Diffusion Language Models. In *Advances in Neural Information Processing Systems*, volume 37, pp. 130136–130184. Curran Associates, Inc., 2024. doi: 10.52202/079017-4135. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/eb0b13cc515724ab8015bc978fdde0ad-Abstract-Conference.html.

Kiyoung Seong, Sungsoo Ahn, Sehui Han, and Changyoung Park. Multimodal crystal flow: Any-to-any modality generation for unified crystal modeling, February 2026. URL https://arxiv.org/abs/2602.20210. arXiv:2602.20210 [cs.LG].

Tarun Suresh, Debangshu Banerjee, Shubham Ugare, Sasa Misailovic, and Gagandeep Singh. DINGO: Constrained Inference for Diffusion LLMs. In *Advances in Neural Information Processing Systems*, volume 38, pp. 160518–160551. Curran Associates, Inc., 2025. doi: 10.52202/085713-5362. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/eb17a2030d1bd4a1bd29531bcd626705-Abstract-Conference.html.

Federico Tomasi, Dmitrii Moor, Alice Wang, and Mounia Lalmas. Primal-Dual Guided Decoding for Constrained Discrete Diffusion, May 2026. URL https://arxiv.org/abs/2605.09749. arXiv:2605.09749 [cs.AI].

Emile van Krieken, Pasquale Minervini, Edoardo Maria Ponti, and Antonio Vergari. Neurosymbolic Diffusion Models. In *Advances in Neural Information Processing Systems*, volume 38, pp. 135889–135931. Curran Associates, Inc., 2025. doi: 10.52202/085713-4535. URL https://papers.nips.cc/paper_files/paper/2025/hash/c60bd92a01804b7df0540ed7ca2f7c05-Abstract-Conference.html.

Tin Hadži Veljković, Joshua Rosenthal, Ivor Lončarić, and Jan-Willem van de Meent. Crystalite: A Lightweight Transformer for Efficient Crystal Modeling, July 2026. URL https://arxiv.org/abs/2604.02270. arXiv:2604.02270 [cs.LG].

Aron Walsh, Alexey A. Sokol, John Buckeridge, David O. Scanlon, and C. Richard A. Catlow. Oxidation states and ionicity. *Nature Materials*, 17(11):958–964, November 2018. doi: 10.1038/s41563-018-0165-7. URL https://www.nature.com/articles/s41563-018-0165-7.

Tian Xie, Xiang Fu, Octavian-Eugen Ganea, Regina Barzilay, and Tommi S. Jaakkola. Crystal diffusion variational autoencoder for periodic material generation. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=03RLpj-tc_.

Andy Xu, Rohan Desai, Larry Wang, Ethan Ritz, and Gabriel Hope. PLaID++: A Preference Aligned Language Model for Targeted Inorganic Materials Design. In *Proceedings of the 43rd International Conference on Machine Learning*, volume 306. PMLR, July 2026. URL https://arxiv.org/abs/2509.07150.

Keqiang Yan, Cong Fu, Xiaofeng Qian, Xiaoning Qian, and Shuiwang Ji. Complete and Efficient Graph Transformers for Crystal Material Property Prediction. In *International Conference on Learning Representations*, pp. 2564–2590, May 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/0ab51646ca369140c3c3ece011b66587-Abstract-Conference.html.

Han Yang, Chenxi Hu, Yichi Zhou, Xixian Liu, Yu Shi, Jielan Li, Guanzhi Li, Zekun Chen, Shuizhou Chen, Claudio Zeni, Matthew Horton, Robert Pinsler, Andrew Fowler, Daniel Zügner, Tian Xie, Jake Smith, Lixin Sun, Qian Wang, Lingyu Kong, Chang Liu, Hongxia Hao, and Ziheng Lu. MatterSim: A Deep Learning Atomistic Model Across Elements, Temperatures and Pressures, May 2024. URL https://arxiv.org/abs/2405.04967. arXiv:2405.04967 [cond-mat.mtrl-sci].

Zhenpeng Yao, Yanwei Lum, Andrew Johnston, Luis Martin Mejia-Mendoza, Xin Zhou, Yonggang Wen, Alán Aspuru-Guzik, Edward H. Sargent, and Zhi Wei Seh. Machine learning for a sustainable energy future. *Nature Reviews Materials*, 8(3):202–215, March 2023. doi: 10.1038/s41578-022-00490-5. URL https://www.nature.com/articles/s41578-022-00490-5.

Xiaohan Yi, Guikun Xu, Zhong Zhang, Liu Liu, Yatao Bian, Xi Xiao, and Peilin Zhao. CrystalDiT: Simple Diffusion Transformers for Crystal Generation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(2):1462–1470, March 2026. ISSN 2374-3468. doi: 10.1609/aaai.v40i2.37121. URL https://ojs.aaai.org/index.php/AAAI/article/view/37121.

D. Zagorac, H. Müller, S. Ruehl, J. Zagorac, and S. Rehme. Recent developments in the Inorganic Crystal Structure Database: theoretical crystal structure data and related features. *Journal of Applied Crystallography*, 52(5):918–925, October 2019. doi: 10.1107/S160057671900997X. URL https://journals.iucr.org/j/issues/2019/05/00/in5024/.

Claudio Zeni, Robert Pinsler, Daniel Zügner, Andrew Fowler, Matthew Horton, Xiang Fu, Zilong Wang, Aliaksandra Shysheya, Jonathan Crabbé, Shoko Ueda, Roberto Sordillo, Lixin Sun, Jake Smith, Bichlien Nguyen, Hannes Schulz, Sarah Lewis, Chin-Wei Huang, Ziheng Lu, Yichi Zhou, Han Yang, Hongxia Hao, Jielan Li, Chunlei Yang, Wenjie Li, Ryota Tomioka, and Tian Xie. A generative model for inorganic materials design. *Nature*, 639(8055):624–632, March 2025. doi: 10.1038/s41586-025-08628-5. URL https://www.nature.com/articles/s41586-025-08628-5.

A Structured Output Layer

Proof of Proposition 1. We show by induction on i that $V_i(c)$ is the maximum unnormalised weight of any assignment to the first i sites with cumulative charge c. The claim holds for $i = 0$ since $V_0(c) = \mathbb{1}[c = 0]$. Assuming it holds for $i - 1$, any assignment to the first i sites with cumulative charge c can be decomposed into an assignment to the first $i - 1$ sites with cumulative charge $c - z(s_{0,i})$ and a choice of $s_{0,i}$. Hence,

$$V_i(c) = \max_{s_{0,i} \in \mathcal{S}} p_{\theta,i}(s_{0,i} \mid \boldsymbol{M}_t, t)\, V_{i-1}(c - z(s_{0,i}))\,, \tag{16}$$

which is therefore the maximum unnormalised weight among all such assignments.

It follows that $V_N(0)$ is the maximum unnormalised weight over all charge-neutral assignments. Since the normalising constant $Z_\theta(\boldsymbol{M}_t, t)$ does not depend on the assignment, maximising the unnormalised weight is equivalent to maximising $\widetilde{p}_\theta(\boldsymbol{S}_0 \mid \boldsymbol{M}_t, t)$ over its support. Backtracking through the stored maximisers therefore recovers the exact MAP charge-neutral assignment. □

Proof of Proposition 2. Let $\boldsymbol{S}_0 = (s_{0,1}, \ldots, s_{0,N})$ be a charge-neutral species assignment, with cumulative charges $c_i = \sum_{j=1}^{i} z(s_{0,j})$ and $c_0 = 0$. Under the sequential sampling rule in Eq. (15), its probability is

$$\prod_{i=1}^{N} \frac{p_{\theta,i}(s_{0,i} \mid \boldsymbol{M}_t, t)\, B_i(c_i)}{B_{i-1}(c_{i-1})} = \frac{\prod_{i=1}^{N} p_{\theta,i}(s_{0,i} \mid \boldsymbol{M}_t, t)\, B_N(c_N)}{B_0(c_0)} \tag{17}$$

$$= \frac{\prod_{i=1}^{N} p_{\theta,i}(s_{0,i} \mid \boldsymbol{M}_t, t)}{Z_\theta(\boldsymbol{M}_t, t)}, \tag{18}$$

where the backward-weight factors telescope. Since $\boldsymbol{S}_0$ is charge neutral, $c_N = 0$, so $B_N(c_N) = B_N(0) = 1$, while $c_0 = 0$ and $B_0(0) = Z_\theta(\boldsymbol{M}_t, t)$.

Thus the sequential procedure assigns probability $\widetilde{p}_\theta(\boldsymbol{S}_0 \mid \boldsymbol{M}_t, t)$ to every charge-neutral species assignment. Any non-neutral assignment has probability zero because $B_N(c_N) = 0$ whenever $c_N \neq 0$. Therefore the procedure samples exactly from the structured distribution. □

B Oxidation-State Prediction Details

Oxidation-state prediction assigns an oxidation state to each atomic site of a crystal with known composition and structure. Given a crystal $\boldsymbol{M} = (\boldsymbol{A}, \boldsymbol{F}, \boldsymbol{L})$ with element identities $\boldsymbol{A} = (a_1, \ldots, a_N) \in \mathcal{A}^N$, we predict an oxidation-state assignment $\boldsymbol{Z} = (z_1, \ldots, z_N) \in \mathcal{Z}^N$, where $\mathcal{Z}$ is a vocabulary of integer oxidation states. A predictor defines a factorised distribution

$$p_\phi(\boldsymbol{Z} \mid \boldsymbol{M}) = \prod_{i=1}^{N} p_{\phi,i}(z_i \mid \boldsymbol{M}), \tag{19}$$

where each site is predicted over the full oxidation-state vocabulary $\mathcal{Z}$. The predictor is trained using the categorical cross-entropy loss,

$$\mathcal{L}_{\mathrm{CE}}(\phi) = -\frac{1}{N} \sum_{i=1}^{N} \log p_{\phi,i}(z_i^\star \mid \boldsymbol{M}), \tag{20}$$

where $z_i^\star$ denotes the ground-truth oxidation state at site i. At inference, unconstrained predictions are obtained independently at each site by argmax decoding,

$$\hat{z}_i = \arg\max_{z \in \mathcal{Z}} p_{\phi,i}(z \mid \boldsymbol{M}). \tag{21}$$

Argmax decoding does not enforce that the predicted oxidation state is chemically admissible for the corresponding element or that the complete assignment is charge neutral.

B.1 Vocabulary and Dataset Construction

We define the oxidation-state vocabulary using the `SMACT` v4.0.0 ICSD-24 oxidation-state filter with `consensus=6`, `commonality=low`, and `include_zero=False` (Davies et al., 2019). We use a stricter consensus threshold than the `SMACT` default of `consensus=3` to exclude extremely unusual oxidation states from the vocabulary. The admissible element–oxidation-state pairs cover 90 elements with oxidation states drawn from

$$\mathcal{Z} = \{-5, -4, -3, -2, -1, +1, +2, +3, +4, +5, +6, +7, +8\}, \tag{22}$$

with an admissible subset $\mathcal{Z}(a) \subseteq \mathcal{Z}$ for each element a.

We construct an oxidation-state-labelled crystal structure dataset from the Inorganic Crystal Structure Database (ICSD) (Zagorac et al., 2019). We apply the following curation procedure:

1. Retain only structures for which every atomic site has an integer oxidation-state label, the complete assignment is charge neutral, and each site label belongs to the corresponding element-specific admissible set $\mathcal{Z}(a)$.
2. Remove duplicate structures and structures with more than 200 atoms.
3. Convert all retained structures to primitive Niggli-reduced cells.
4. Define the set of elements with known mixed-valence chemistry, $\mathcal{A}_{\mathrm{mv}}$, using `SMACT`'s `MIXED_VALENCE_ELEMENTS` set, and exclude structures in which any element outside this set is assigned more than one oxidation state.

We then group structures by reduced composition and construct composition-disjoint training, validation, and test splits such that no reduced composition occurs in more than one split. The resulting dataset contains 22,879 training, 2,831 validation, and 2,782 test structures.

B.2 Model Architectures

BERTOS. BERTOS is a composition-only oxidation-state predictor based on a BERT transformer (Fu et al., 2023). A composition is expanded into an electronegativity-ordered sequence of element tokens, which is processed by the transformer to produce contextual representations for each site. A shared classifier predicts an oxidation state for each token. The encoder is first pre-trained using masked-language modelling over element tokens and is then fine-tuned for oxidation-state prediction.

CrystaliteOS. Our predictor, which we refer to as CrystaliteOS, is adapted from the Crystalite transformer (Veljković et al., 2026). We retain the geometry-enhanced transformer backbone, remove the diffusion-time conditioning and prediction heads, and replace them with a linear per-site classifier over the oxidation-state vocabulary $\mathcal{Z}$. Unlike BERTOS, the predictor takes the complete crystal structure $\boldsymbol{M} = (\boldsymbol{A}, \boldsymbol{F}, \boldsymbol{L})$ as input.

Architecture and training settings are given in Table 5. For each decoding method, we select the checkpoint with the highest validation crystal accuracy under that same decoding method. We therefore report argmax and Viterbi results using their respective best validation checkpoints.

B.3 Structured Inference

For CrystaliteOS, we apply the structured Viterbi decoding of Section 4.3 to oxidation-state prediction. Given the factorised predictive distribution

$$p_\phi(\boldsymbol{Z} \mid \boldsymbol{M}) = \prod_{i=1}^{N} p_{\phi,i}(z_i \mid \boldsymbol{M}), \tag{23}$$

we define the structured distribution

$$\widetilde{p}_\phi(\boldsymbol{Z} \mid \boldsymbol{M}) \propto \mathbb{1}\left[\sum_{i=1}^{N} z_i = 0\right] \prod_{i=1}^{N} \mathbb{1}[z_i \in \mathcal{Z}(a_i)]\, p_{\phi,i}(z_i \mid \boldsymbol{M}), \tag{24}$$

whose support is restricted to charge-neutral assignments that use an admissible oxidation state for each element. The same dynamic program described in Section 4.3 can be used for exact inference under this distribution.

For CrystaliteOS, we additionally restrict mixed valence to elements in $\mathcal{A}_{\mathrm{mv}}$. For each element $a \notin \mathcal{A}_{\mathrm{mv}}$, all sites occupied by a are constrained to have the same oxidation state. This restriction can be incorporated exactly into the dynamic program by grouping sites of the same element. Let

$$I_a = \{i : a_i = a\} \tag{25}$$

denote the set of sites occupied by element a, and let $n_a = |I_a|$ denote its multiplicity. For an admissible oxidation state $z \in \mathcal{Z}(a)$, assigning oxidation state z to every site in I_a has unnormalised weight

$$w_a(z) = \prod_{i \in I_a} p_{\phi,i}(z \mid \boldsymbol{M}), \tag{26}$$

and contributes $n_a z$ to the cumulative charge. For each $a \notin \mathcal{A}_{\mathrm{mv}}$, the n_a site-level transitions can therefore be replaced by a single transition with weight $w_a(z)$ and charge increment $n_a z$. This preserves the unnormalised weight of every assignment satisfying the single-valence constraint, so the forward–backward and Viterbi recursions apply unchanged over the restricted feasible set. Elements in $\mathcal{A}_{\mathrm{mv}}$ are not grouped and instead retain one dynamic-programming step per site, allowing different sites of the same element to adopt different oxidation states. We use Viterbi decoding to return the MAP assignment satisfying element-wise admissibility, charge neutrality, and the mixed-valence constraint.

B.4 MP-20 OXIDATION-STATE LABELLING

We use CrystaliteOS with charge-neutral Viterbi decoding to annotate MP-20 and apply the following curation procedure:

1. Convert all structures to Niggli-reduced cells.
2. Remove structures containing elements outside the oxidation-state vocabulary.
3. Assign oxidation state zero to every site for structures in which all elements are classified as metals by `pymatgen`, following the conventional treatment of metallic alloys.
4. Remove non-metallic elemental structures. Oxidation state zero is typically only assigned to non-metals in elemental structures, which constitute 0.2% of the dataset; excluding them avoids expanding the species vocabulary to accommodate these cases.
5. Assign oxidation states to all remaining structures using CrystaliteOS.
6. Remove structures with no feasible charge-neutral assignment, or with a chemically implausible predicted assignment in which the same element adopts more than two distinct oxidation states or both positive and negative oxidation states.

Table 4 summarises the filtering pipeline. Starting from 45,229 MP-20 structures, the final MP-20-OS dataset contains 42,576 structures, comprising 25,536 training, 8,495 validation, and 8,545 test structures.

Table 4: MP-20-OS filtering. Rows under *removed* report structures excluded at each filtering stage.

Stage	Train	Val.	Test	Total
MP-20	27,136	9,047	9,046	45,229
Removed				
Outside vocabulary	289	90	83	462
Elemental non-metal	55	20	10	85
No feasible charge-neutral assignment	1,003	359	327	1,689
Chemically implausible assignment	253	83	81	417
Alloys	7,926	2,691	2,704	13,321
Labelled non-alloys	17,610	5,804	5,841	29,255
Total	25,536	8,495	8,545	42,576

Table 5: Architecture and training settings for BERTOS and CrystaliteOS.

(a) Architecture

Setting	**BERTOS**	**CrystaliteOS**
Hidden dimension	120	128
Transformer layers	12	12
Attention heads	4	4
MLP hidden dimension	512	256
Dropout	0.1	0.1
Attention dropout	0.1	0.1
Coordinate embedding	–	RFF, 32 freqs.
GEM distance bias	–	Enabled
GEM edge-aware bias	–	Enabled
Edge-bias freqs. / hidden / RBF / max	–	12 / 256 / 32 / 4
GEM sharing	–	Enabled
PBC search radius R	–	1

(b) Training

Setting	**BERTOS**	**CrystaliteOS**
Optimiser	AdamW	AdamW
Learning rate	10^{-3}	10^{-3}
Weight decay	0	0
Batch size	256	64
Training epochs	150	150
LR schedule	Linear decay	Cosine decay
Warmup	5% of steps	1000 steps
Gradient clipping	1.0	1.0
Precision	FP32	FP32
Checkpoint selection	Validation crystal accuracy	Validation crystal accuracy
MLM epochs	5 epochs	–
MLM mask probability	0.15	–
MLM learning rate	10^{-4}	–
MLM LR schedule	Constant	–
MLM weight decay	0.01	–

C OxiGen Details

OxiGen adapts the MatterGen crystal diffusion model (Zeni et al., 2025). A crystal $\boldsymbol{M} = (\boldsymbol{S}, \boldsymbol{F}, \boldsymbol{L})$ is generated by jointly denoising its categorical species assignment $\boldsymbol{S}$, fractional coordinates $\boldsymbol{F}$, and lattice matrix $\boldsymbol{L}$. The forward diffusion process factorises across the three components,

$$q(\boldsymbol{M}_t \mid \boldsymbol{M}_r) = q(\boldsymbol{S}_t \mid \boldsymbol{S}_r)\, q(\boldsymbol{F}_t \mid \boldsymbol{F}_r)\, q(\boldsymbol{L}_t \mid \boldsymbol{L}_r) \qquad 0 \le r < t \le 1. \tag{27}$$

We retain MatterGen's coordinate diffusion, lattice diffusion and network architecture. We briefly summarise these components below and describe our modified species diffusion in Appendix C.1.

Coordinate diffusion. Coordinate diffusion is performed on fractional coordinates, which naturally live on the flat torus $\mathbb{T}^3 = \mathbb{R}^3/\mathbb{Z}^3$. The diffusion process uses variance-exploding diffusion with wrapped-normal noise. For a fractional coordinate $\boldsymbol{f}$, the noising process is defined as

$$q(\boldsymbol{f}_t \mid \boldsymbol{f}_0) = \mathcal{N}_{\mathrm{W}}\big(\boldsymbol{f}_t; \boldsymbol{f}_0, \sigma_t^2(N)\boldsymbol{I}\big)\,, \tag{28}$$

where $\mathcal{N}_{\mathrm{W}}$ denotes a Gaussian distribution wrapped onto $[0,1)^3$. As σ_t increases, this distribution approaches the uniform distribution on the unit torus. To approximately remove the dependence of the corresponding Cartesian noise scale on unit-cell size, the fractional-coordinate standard deviation is rescaled according to

$$\sigma_t(N) = \frac{\sigma_t}{N^{1/3}}, \tag{29}$$

where N is the number of atomic sites in the unit cell. Since Cartesian and fractional coordinates satisfy $\boldsymbol{X} = \boldsymbol{L}\boldsymbol{F}$, the network predicts coordinate scores in Cartesian space and then transforms these to fractional-coordinate scores.

Lattice diffusion. The arbitrary global rotation of the unit cell is removed using the polar decomposition

$$\boldsymbol{L} = \boldsymbol{U}\tilde{\boldsymbol{L}}, \tag{30}$$

where $\boldsymbol{U}$ is orthogonal and $\tilde{\boldsymbol{L}}$ is symmetric positive definite. Diffusion is then performed on $\tilde{\boldsymbol{L}}$ and restricted to the space of symmetric matrices. A variance-preserving diffusion process is used to avoid unrealistically large noisy unit cells. The noising process is defined as

$$q(\tilde{\boldsymbol{L}}_t \mid \tilde{\boldsymbol{L}}_0) = \mathcal{N}\Big(\sqrt{\bar{\alpha}_t}\tilde{\boldsymbol{L}}_0 + (1-\sqrt{\bar{\alpha}_t})\mu(N)\boldsymbol{I},\ (1-\bar{\alpha}_t)\sigma_L^2(N)\boldsymbol{I}\Big), \tag{31}$$

where $\bar{\alpha}_t$ is the variance-preserving signal schedule. The prior is centred on an approximately cubic lattice, and $\mu(N)$ and $\sigma_L(N)$ are scaled with the number of atoms so that the typical atomic density and signal-to-noise ratio remain approximately independent of system size.

Network architecture. The reverse process is parameterised by an SE(3)-equivariant GemNet-dT message-passing network (Gasteiger et al., 2021), which processes the complete noisy crystal $\boldsymbol{M}_t$ together with the diffusion time. The network jointly predicts Cartesian coordinate scores, symmetric lattice scores, and logits for the clean species assignments. Thus, although the three forward corruption processes factorise, their reverse predictions are coupled through a single network conditioned on the full noisy crystal.

C.1 Species Diffusion

MatterGen uses an absorbing-state D3PM (Austin et al., 2021) discrete diffusion process over atomic identities $\boldsymbol{A} \in \mathcal{A}^N$. We instead characterise the categorical state at each site as a species $s = (a, z)$ consisting of an element a and oxidation state z, and replace the discrete-time D3PM with the continuous-time absorbing-state diffusion process of MDLM (Sahoo et al., 2024).

Species vocabulary. We use the species vocabulary defined in Appendix B.1 and additionally include $z = 0$ for elements classified as metals by `pymatgen`. The resulting vocabulary contains 428 species over 90 elements and a mask token, giving a total model vocabulary of 429 states. MatterGen initialises each atomic node with an element embedding e_{element}; OxiGen replaces this with

$$e = e_{\text{element}} + \omega_{\text{os}} e_{\text{os}}, \tag{32}$$

where e_{os} is a learned oxidation-state embedding of the same dimension and ω_{os} is a learned global scalar initialised at one. The categorical output remains a single linear head, now defined over the joint element–oxidation-state species vocabulary.

Forward diffusion. Let m denote the absorbing mask token and let $\alpha_t \in [0, 1]$ denote the probability that a site remains unmasked at time t, with α_t monotonically decreasing from $\alpha_0 = 1$ to $\alpha_1 = 0$ in the idealised continuous-time process. Given a clean species assignment $\boldsymbol{S}_0 = (s_{0,1}, \ldots, s_{0,N})$, sites are masked independently according to

$$q(\boldsymbol{S}_t \mid \boldsymbol{S}_0) = \prod_{i=1}^{N} \left[\alpha_t\, \mathbb{1}[s_{t,i} = s_{0,i}] + (1-\alpha_t)\, \mathbb{1}[s_{t,i} = m]\right]. \tag{33}$$

For $0 \le r < t \le 1$, the corresponding single-site posterior is

$$q(s_{r,i} \mid s_{t,i}, s_{0,i}) = \begin{cases} \mathbb{1}[s_{r,i} = s_{t,i}], & s_{t,i} \neq m, \\ \dfrac{1-\alpha_r}{1-\alpha_t}\mathbb{1}[s_{r,i} = m] + \dfrac{\alpha_r - \alpha_t}{1-\alpha_t}\mathbb{1}[s_{r,i} = s_{0,i}], & s_{t,i} = m. \end{cases} \tag{34}$$

Denoising objective. The denoiser predicts $p_{\theta,i}(s_{0,i} \mid \boldsymbol{M}_t, t)$ from the complete noisy crystal $\boldsymbol{M}_t = (\boldsymbol{S}_t, \boldsymbol{F}_t, \boldsymbol{L}_t)$. Following Sahoo et al. (2024), the denoiser assigns zero probability to the mask token and deterministically copies the species at sites that are already unmasked. Consequently, only masked sites contribute to the denoising objective. For the factorised clean-state prediction

$$p_\theta(\boldsymbol{S}_0 \mid \boldsymbol{M}_t, t) = \prod_{i=1}^{N} p_{\theta,i}(s_{0,i} \mid \boldsymbol{M}_t, t), \tag{35}$$

the continuous-time MDLM NELBO for the species component is

$$\mathcal{L}_{\text{MDLM}} = \mathbb{E}_{\boldsymbol{M}_0} \int_0^1 \mathbb{E}_{\boldsymbol{M}_t|\boldsymbol{M}_0} \left[\lambda(t) \sum_{i=1}^{N} \mathbb{1}[s_{t,i} = m] \, (-\log p_{\theta,i}(s_{0,i} \mid \boldsymbol{M}_t, t)) \right] dt, \tag{36}$$

where

$$\lambda(t) = -\frac{\alpha_t'}{1-\alpha_t}. \tag{37}$$

In implementation, we use the numerically stabilised schedule $\alpha_t = 1 - (1-\epsilon)t$ with $\epsilon = 10^{-3}$, which approximates the fully masked terminal condition $\alpha_1 = 0$ of the idealised continuous-time process.

Structured training objective. In the ablation study, we additionally train OxiGen using the charge-neutral clean-state distribution $\widetilde{p}_\theta(\boldsymbol{S}_0 \mid \boldsymbol{M}_t, t)$ defined in Eq. (6). A general derivation of the continuous-time masked-diffusion NELBO for arbitrary non-factorised clean-state distributions is given by van Krieken et al. (2025). Their Theorem C.1 shows that, under zero masking probability, carry-over unmasking, and a proper fully masked terminal prior, the continuous-time objective depends only on the one-site marginals of the clean-state distribution.

Consider a reverse interval $[r, t]$ of width $\Delta t = t - r$. For a site that is masked at time t, Eq. (34) gives the probability of revealing its clean species during this interval as

$$\frac{\alpha_r - \alpha_t}{1-\alpha_t} = -\frac{\alpha_t'}{1-\alpha_t}\Delta t + O(\Delta t^2) = \lambda(t)\Delta t + O(\Delta t^2). \tag{38}$$

Thus, revealing one site has probability $O(\Delta t)$, whereas revealing two or more sites in the same interval has probability $O(\Delta t^2)$ or smaller. Summed over $O(\Delta t^{-1})$ reverse intervals, the contribution of multi-site reveal events vanishes as $\Delta t \to 0$. Consequently, only single-site reveal events contribute in the continuous-time limit, and the NELBO depends on the structured distribution only through its one-site marginals $\widetilde{p}_{\theta,i}(s_{0,i} \mid \boldsymbol{M}_t, t)$.

Replacing the factorised clean-state probabilities in Eq. (36) by these exact structured marginals gives

$$\mathcal{L}_{\text{neutral}} = \mathbb{E}_{\boldsymbol{M}_0} \int_0^1 \mathbb{E}_{\boldsymbol{M}_t|\boldsymbol{M}_0} \left[\lambda(t) \sum_{i=1}^{N} \mathbb{1}[s_{t,i} = m] \, (-\log \widetilde{p}_{\theta,i}(s_{0,i} \mid \boldsymbol{M}_t, t)) \right] dt. \tag{39}$$

The marginals $\widetilde{p}_{\theta,i}$ are computed exactly using the forward–backward procedure in Section 4.3. The structured objective therefore has the same site-wise form as the standard MDLM objective, but evaluates each masked-site loss under the charge-neutral clean-state distribution. We use this objective only for the ablation study; the default OxiGen model is trained with Eq. (36) and applies the structured output layer only during sampling.

C.2 Training and sampling

For the coordinates and lattice, we follow the training and sampling procedures of MatterGen. Species are generated using the charge-neutral sampling procedure described in Section 4.3. Following MatterGen, generation excludes the noble gases, Tc, Pm, and all elements with atomic number $Z \geq 84$. We retain MatterGen's network, coordinate-diffusion, and lattice-diffusion training and sampling hyperparameters without additional tuning. The architecture, optimisation, diffusion, and sampling settings are summarised in Table 7.

C.3 Computational Overhead

Structured inference requires a single backward dynamic-programming pass to draw a charge-neutral sample at each diffusion step using Eq. (15), while structured training requires one forward and one backward dynamic-programming pass to compute the marginals of the structured distribution using Eq. (13). Each pass has time complexity $O(N^2|\mathcal{S}|(z_{\max} - z_{\min}))$, where $z_{\max}$ and $z_{\min}$ are the maximum and minimum oxidation states in the species vocabulary.

Table 6 reports the overhead of the structured output layer relative to the same model without it. Training time is measured per epoch on MP-20-OS, and sampling time per batch of 128 structures

over a full 1000-step reverse-diffusion trajectory, both on a single NVIDIA GH200 120GB with batch size 128. Training OxiGen for 900 epochs on four GH200 GPUs with batch size 128 per GPU (512 in total; Table 7) takes approximately 3.5 hours.

Table 6: Computational overhead of the structured output layer, reported as mean ± std over five epochs or five batches. Percentage changes are relative to the model without the structured output layer.

Structured Output Layer	Training Time (s/epoch) ↓	Sampling Time (s/batch) ↓
✗	$78.797_{\pm 0.101}$	$309.527_{\pm 4.923}$
✓	$98.906_{\pm 0.057}$	$329.716_{\pm 4.809}$
Change (%)	$+25.52$	$+6.52$

D Band-Gap Conditional Generation

Conditional generation. We fine-tune MatterGen and OxiGen for band-gap conditional generation following Zeni et al. (2025). Both models are fine-tuned on MP-20-OS using the DFT band-gap labels provided with MP-20. Band-gap values are standardised and encoded using a fixed sinusoidal embedding, which is incorporated before each GemNet-dT message-passing layer through an adapter consisting of a two-layer MLP and a zero-initialised linear mix-in layer. Fine-tuning updates both the pre-trained denoiser and the conditioning parameters. The models are trained with conditional dropout and sampled using classifier-free guidance, applied to both the continuous coordinate and lattice scores and the discrete atomic-identity or species predictions. Fine-tuning and generation settings are summarised in Table 8.

Band-gap prediction. We evaluate conditional generation using iComFormer (Yan et al., 2024), an SE(3)-invariant crystal transformer for property prediction. Following Yan et al. (2024), we train the model on the Materials Project 2018.6.1 band-gap benchmark using the authors' provided data split, code, and hyperparameters. Our trained model achieves a test mean absolute error of $0.193\ \mathrm{eV}$, matching the value reported by Yan et al. (2024). For generated crystals, band gaps are predicted from the MatterSim-relaxed structures rather than the unrelaxed model outputs.

E Evaluation Procedure

E.1 Stability, uniqueness, and novelty

We follow the released MatterGen evaluation procedure (Zeni et al., 2025). Evaluation excludes structures containing noble gases, Tc, Pm, and all elements with atomic number $Z \geq 84$ for all methods. Generated structures are first relaxed with MatterSim v1.0.0-1M (Yang et al., 2024), and stability, uniqueness, novelty, and relaxation metrics are computed from the relaxed structures. Energy above the convex hull, E_{hull}, is computed using `pymatgen` (Ong et al., 2013) phase diagrams with the Alex-MP reference dataset and Materials Project 2020 compatibility corrections provided by Zeni et al. (2025). A structure is considered stable if $E_{\mathrm{hull}} \leq 0.1\ \mathrm{eV/atom}$. Uniqueness and novelty are computed via the `StructureMatcher` from `pymatgen` with `ltol` $= 0.2$, `stol` $= 0.3$, and `angle_tol` $= 5^\circ$. Uniqueness is evaluated within each generated set, and a structure is considered novel if it does not match any entry in the Alex-MP reference dataset. Oxidation-state labels are removed before structure matching. The relaxation RMSD is computed by matching each generated structure to its relaxed counterpart using `StructureMatcher` with `ltol` $= 0.5$, `stol` $= 0.5$, and `angle_tol` $= 10^\circ$. A structure is counted as S.U.N. if it is simultaneously stable, unique, and novel. Stability and S.U.N. rates are reported relative to the total number of generated structures. Uniqueness, novelty, mean E_{hull}, and relaxation RMSD are computed over structures that were successfully relaxed.

Table 7: Architecture, training, and sampling settings for OxiGen.

Setting	Value
Hidden / atom / edge dimension	512 / 512 / 512
Interaction blocks	4
Cutoff	7.0 Å
Maximum neighbours	50
Optimiser	Adam
Learning rate	10^{-4}
LR schedule	ReduceLROnPlateau
LR factor / patience / minimum	0.6 / 100 epochs / 10^{-6}
Batch size	512
Training epochs	900
Precision	FP32
Gradient clipping	0.5, value-based
Checkpoint selection	Last epoch
EMA	None
Cell / position / species loss weights	1.0 / 0.1 / 0.01
Mask schedule	$\alpha_t = 1 - (1 - \epsilon)t$
ϵ / training-time range	10^{-3} / $t \sim \mathcal{U}(10^{-3}, 1)$
Reverse steps	1000
Reverse-time grid	Uniform in t
Position/cell corrector	1 Langevin step
Species corrector	None

Table 8: Band-gap conditional fine-tuning and sampling settings for OxiGen and MatterGen.

Setting	Value
Property embedding dimension	512
Null condition	Learned embedding
Fine-tuning mode	Full-model
Initialisation	Unconditional last epoch
Optimiser	Adam
Learning rate	5×10^{-6}
LR schedule	ReduceLROnPlateau
LR factor / patience / minimum	0.6 / 100 epochs / 10^{-6}
Batch size	128
Epochs	200
Conditional dropout	0.2
Gradient clipping / precision	0.5 / FP32
Guidance strength γ	2
Reverse steps	1000

E.2 Oxidation-state metrics

We evaluate compositional validity, the oxidation-state frequency score, and the oxidation-state distribution distance using `SMACT` v4.0.0 (Davies et al., 2019). We exclude elemental structures and structures for which all elements are classified as metals by `pymatgen`. Oxidation-state frequencies are taken from the ICSD-24 species-occurrence table with `consensus=3` and `include_zero=False`. We compute compositional validity using `smact_validity` with `consensus=3`, `commonality=low`, `include_zero=False`, `mixed_valence=False`, and `include_alloys=False`.

For an element assignment $\boldsymbol{A} = (a_1, \ldots, a_N)$, let $\mathcal{N}(\boldsymbol{A})$ denote the set of valid charge-neutral oxidation-state assignments enumerated by `SMACT`. Because mixed valence is not permitted, each element is assigned a single oxidation state within any $\boldsymbol{S} \in \mathcal{N}(\boldsymbol{A})$. We write $\mathrm{uniq}(\boldsymbol{S})$ for the set

of distinct species (a, z) appearing in $\boldsymbol{S}$, and let $f_a(z) \in [0, 1]$ denote the normalised ICSD-24 occurrence frequency of oxidation state z for element a.

For each composition, we define the oxidation-state frequency score as the highest geometric mean species frequency over all valid assignments,

$$f_{\mathrm{OS}}(\boldsymbol{A}) = 100 \max_{\boldsymbol{S} \in \mathcal{N}(\boldsymbol{A})} \left(\prod_{(a,z) \in \mathrm{uniq}(\boldsymbol{S})} f_a(z) \right)^{1/|\,\mathrm{uniq}(\boldsymbol{S})|}. \tag{40}$$

Each distinct species contributes once, regardless of its multiplicity in the composition. We report the mean of $f_{\mathrm{OS}}(\boldsymbol{A})$ over structures with at least one valid assignment.

For Figure 2, we additionally consider the minimum species frequency within each valid assignment. We define

$$f_{\mathrm{OS}}^{\min}(\boldsymbol{A}) = \max_{\boldsymbol{S} \in \mathcal{N}(\boldsymbol{A})} \min_{(a,z) \in \mathrm{uniq}(\boldsymbol{S})} f_a(z), \tag{41}$$

which is the highest minimum species frequency achievable by any valid assignment for the composition. At a threshold $\tau \in [0, 1]$, a composition is therefore counted as valid if it admits at least one assignment in which every constituent species has frequency at least τ, equivalently if $f_{\mathrm{OS}}^{\min}(\boldsymbol{A}) \geq \tau$. We report the fraction of evaluated structures satisfying this condition.

To compare oxidation-state distributions, we account for compositions that admit multiple valid assignments by assigning each $\boldsymbol{S} \in \mathcal{N}(\boldsymbol{A})$ a weight proportional to the product of its species frequencies,

$$q_{\boldsymbol{A}}(\boldsymbol{S}) = \frac{\displaystyle\prod_{(a,z) \in \mathrm{uniq}(\boldsymbol{S})} f_a(z)}{\displaystyle\sum_{\boldsymbol{S}' \in \mathcal{N}(\boldsymbol{A})} \prod_{(a,z) \in \mathrm{uniq}(\boldsymbol{S}')} f_a(z)}. \tag{42}$$

These weights define a distribution over the valid assignments of each composition. For a dataset $\mathcal{D}$, the oxidation-state distribution for element a is then obtained by marginalising over these assignments,

$$p_{\mathcal{D}}(z \mid a) \propto \sum_{\boldsymbol{A} \in \mathcal{D}} \sum_{\boldsymbol{S} \in \mathcal{N}(\boldsymbol{A})} q_{\boldsymbol{A}}(\boldsymbol{S}) \mathbb{1}[(a, z) \in \mathrm{uniq}(\boldsymbol{S})], \tag{43}$$

with the distribution normalised over z separately for each element.

Finally, we compare the generated and reference oxidation-state distributions using the Jensen–Shannon distance, macro-averaged over the set $\mathcal{A}'$ of elements present in both datasets,

$$d_{\mathrm{OS}} = \frac{1}{|\mathcal{A}'|} \sum_{a \in \mathcal{A}'} \sqrt{\mathrm{JSD}\left(p_{\mathrm{gen}}(\cdot \mid a) \,\|\, p_{\mathrm{ref}}(\cdot \mid a)\right)}. \tag{44}$$

We use the MP-20-OS test split as the reference distribution. Compositions with no valid assignment are excluded from the oxidation-state frequency and distribution metrics.

F Additional Results

F.1 De Novo Generation

Table 9 reports the stability, uniqueness, novelty, and relaxation metrics for the *de novo* generation experiments in Section 5.2. The evaluated methods show the expected trade-off between stability and novelty. Crystalite achieves the highest stable fraction but the lowest novelty among the compared methods, while MatterGen achieves the highest novelty. OxiGen achieves a strong balance across these metrics, resulting in the highest S.U.N. rate, together with the lowest relaxation RMSD.

Proxy metrics. Table 10 reports the proxy metrics proposed by Xie et al. (2022). All metrics are computed on the unrelaxed structures. Validity and coverage are reported over all generated structures, while the distribution metrics are computed over the subset of structures that are both compositionally and structurally valid. Compositional validity here is evaluated over all structures, including alloys and elemental structures, with `include_alloys=True`, and therefore differs

Table 9: Additional *de novo* generation results, reported as mean ± std over five independently seeded training runs, with 10,240 generated structures per seed. Best values are shown in **bold** and second-best values are underlined.

Method	Stable (%) ↑	Unique (%) ↑	Novel (%) ↑	S.U.N. (%) ↑	E_{hull} (eV/atom) ↓	RMSD (Å) ↓
DiffCSP	43.34±1.29	95.80±0.58	71.58±1.57	18.24±1.38	0.2007±0.0101	0.3555±0.0427
MatterGen	46.11±3.17	97.74±0.49	**75.32**±2.72	24.54±0.91	0.1799±0.0085	0.1747±0.0203
FlowMM	36.04±0.83	**97.99**±0.21	74.07±1.01	13.85±0.39	0.2481±0.0104	0.6704±0.0347
OMatG	49.12±0.39	95.12±0.33	65.42±0.87	19.03±0.78	0.1831±0.0030	0.5244±0.0278
Crystalite	**67.17**±2.18	93.03±0.91	52.25±2.61	22.83±0.76	**0.1437**±0.0083	0.2496±0.0198
OxiGen	52.59±2.25	96.45±0.64	70.68±2.31	**26.38**±1.01	0.1575±0.0080	**0.1679**±0.0117

Table 10: Proxy metrics for *de novo* generation, reported as mean ± std over five independently seeded training runs, with 10,240 generated structures per seed. Best values are shown in **bold** and second-best values are underlined.

Method	Validity		Coverage		Distribution	
	Comp. (%) ↑	Struct. (%) ↑	Cov-R (%) ↑	Cov-P (%) ↑	wdist ρ ↓	wdist N_{el} ↓
DiffCSP	87.29±1.97	99.89±0.05	99.58±0.05	99.70±0.10	0.2514±0.1340	0.5052±0.0557
MatterGen	88.71±0.64	99.97±0.04	**99.71**±0.08	99.69±0.07	**0.1416**±0.0576	0.2398±0.0791
FlowMM	87.58±0.17	91.80±0.86	99.27±0.11	98.83±0.18	1.3027±0.4222	0.1200±0.0125
OMatG	89.06±0.37	99.11±0.19	99.59±0.09	99.83±0.03	0.2672±0.0152	**0.0549**±0.0097
Crystalite	89.12±0.41	99.68±0.06	99.67±0.09	**99.87**±0.04	0.1541±0.0266	0.1151±0.0190
OxiGen	**95.24**±0.44	**99.98**±0.01	99.60±0.08	99.80±0.01	0.2045±0.0869	0.3137±0.0340

from the results reported in Table 2. OxiGen is comparable to MatterGen on structural validity and coverage, and slightly worse on element-count distribution distance.

LeMat-GenBench. Table 11 reports results under the standardised LeMat-GenBench evaluation procedure (Betala et al., 2025). We evaluate 2,500 structures generated by the OxiGen seed with the highest S.U.N. rate under our primary evaluation, pre-relaxed using MatterSim. We compare against methods trained on MP-20 on the public LeMat-GenBench leaderboard, with results taken directly from the leaderboard; OxiGen is instead trained on MP-20-OS, a filtered subset of MP-20 (Appendix B.4). The baselines evaluated on pre-relaxed structures are WyFormer and WyFormer-DFT (Kazeev et al., 2025), PLaID++ (Xu et al., 2026), OMatG-FC (Martirossyan et al., 2026), MatterGen (Zeni et al., 2025), MiAD (Okhotin et al., 2025), OMatG (Höllmer et al., 2025), MCFlow (Seong et al., 2026), and Crystalite (Veljković et al., 2026). The remaining baselines were evaluated on unrelaxed structures: Crystal-GFN (Mila AI4Science et al., 2023), LLaMat2 and LLaMat3 (Ahlawat et al., 2026), ADiT (Joshi et al., 2025), SymmCD (Levy et al., 2025), CrystalFormer (Cao et al., 2025), CrystaLLM-π (Bone et al., 2025), DiffCSP++ (Jiao et al., 2024), DiffCSP (Jiao et al., 2023), Chemeleon1 (Park et al., 2025), and Chemeleon2 (Park & Walsh, 2026). LeMat-GenBench reports the fractions of stable ($E_{\text{hull}} \leq 0$) and metastable ($E_{\text{hull}} \leq 0.1$ eV/atom) structures; S.U.N. denotes the fraction of generated structures that are stable, unique, and novel, whereas M.S.U.N. uses metastability in place of stability. These definitions differ from our primary evaluation, which uses $E_{\text{hull}} \leq 0.1$ eV/atom as the stability threshold. Under the LeMat-GenBench procedure, OxiGen achieves the highest validity, uniqueness, and M.S.U.N. rate among all evaluated methods, and is competitive across the other metrics.

F.2 Ablation Study

In this section we provide additional results for the ablations in Section 5.2.

Stability, uniqueness, and novelty. Table 12 reports the stability, uniqueness, novelty, and relaxation metrics. Relative to the element-representation baseline, OxiGen slightly increases both the

stability and S.U.N. rate. Comparing the species models, structured sampling increases stability and slightly increases the S.U.N. rate, while modestly reducing uniqueness and novelty. Structured training provides no further improvement.

Oxidation-state agreement. Table 13 reports the oxidation-state agreement between OxiGen and CrystaliteOS. We label each unrelaxed generated structure using CrystaliteOS, excluding metallic alloys and elemental structures. Site agreement is the fraction of atomic sites for which the generated oxidation state matches the CrystaliteOS prediction and crystal agreement is the fraction of generated crystals for which the two assignments agree at every site. These metrics are not defined for the element representation because it does not explicitly generate oxidation states. Structured sampling substantially improves the agreement between the generated oxidation states and the CrystaliteOS assignments. Structured training again provides no additional improvement. These results suggest that structured sampling not only guarantees charge neutrality, but improves the proportion of generated oxidation states that agree with those inferred from the generated structure.

Table 11: *De novo* generation results under the LeMat-GenBench evaluation procedure. OxiGen results are computed over 2,500 generated structures; results for all other models trained only on MP-20 are taken from the public LeMat-GenBench leaderboard. Rows are grouped according to whether the submitted structures are pre-relaxed. Best and second-best values within the pre-relaxed group are shown in **bold** and underlined, respectively.

Model	Valid (%) ↑	Unique (%) ↑	Novel (%) ↑	Stable (%) ↑	Metastable (%) ↑	S.U.N. (%) ↑	M.S.U.N. (%) ↑	E_{hull} (eV/atom) ↓
Pre-Relaxed Structures								
WyFormer	93.4	93.0	66.4	0.5	15.7	0.1	1.9	0.4988
PLaID++	96.0	77.8	24.2	12.4	60.7	1.0	7.6	0.0854
WyFormer-DFT	95.2	95.0	66.4	3.7	24.8	0.4	7.8	0.2708
OMatG-FC	97.2	92.8	28.9	**18.4**	57.9	**1.7**	12.0	**0.0694**
MatterGen	95.7	95.1	**70.5**	2.0	33.4	0.2	15.0	0.1834
MiAD	96.2	94.3	40.2	6.4	**62.0**	1.0	16.6	0.0804
OMatG	96.4	95.2	51.2	11.6	49.8	1.0	18.0	0.0956
MCFlow	97.2	96.3	52.2	11.9	49.3	0.7	18.9	0.0987
Crystalite	97.2	95.8	53.2	12.7	51.6	1.5	22.6	0.0905
OxiGen	**99.6**	**98.5**	68.8	5.0	49.5	0.6	**26.7**	0.1203
Non-Pre-Relaxed Structures								
Crystal-GFN	51.7	51.7	51.7	0.0	0.0	0.0	0.0	2.0858
LLaMat3	15.4	15.2	10.5	0.1	2.1	0.0	0.2	1.7067
ADiT	90.6	87.8	26.0	0.4	36.5	0.0	1.0	0.3333
LLaMat2	84.4	81.4	30.0	0.7	34.7	0.1	2.1	0.4395
SymmCD	73.4	73.0	47.0	1.4	18.6	0.1	2.4	0.8761
CrystalFormer	69.9	69.4	31.8	1.4	28.8	0.0	3.1	0.7039
CrystaLLM-π	86.8	84.9	24.9	2.9	49.0	0.3	3.6	0.3205
DiffCSP++	95.3	95.1	62.0	1.0	26.4	0.2	5.0	0.4093
DiffCSP	95.7	94.8	66.2	2.3	29.8	0.1	8.5	0.2747
Chemeleon1	95.4	94.8	64.0	3.6	34.4	0.3	11.3	0.2055
Chemeleon2	95.2	88.1	71.6	0.0	39.8	0.0	21.2	0.1557

F.3 Band-Gap Conditional Generation

Table 14 reports additional results for the band-gap-conditioned generation experiments in Section 5.2. The band-gap RMSE measures the root-mean-square error between the conditioning target and the iComFormer-predicted band gap of the relaxed generated structures. OxiGen achieves a higher S.U.N. rate than MatterGen at all four conditioning targets, due to a substantially higher stability rate despite a lower novelty rate. Band-gap RMSE is comparable at 1 eV and better for OxiGen at the 3, 5, and 7 eV targets.

Table 12: Ablation results for stability, uniqueness, novelty, and relaxation quality. Results are reported as mean ± std over five independently seeded training runs, with 10,240 generated structures per seed. The method names indicate whether structured training / structured sampling are used, respectively. Best values are shown in **bold** and second-best values are underlined; the shaded row is OxiGen.

Method	Stable (%) ↑	Unique (%) ↑	Novel (%) ↑	S.U.N. (%) ↑	E_{hull} (eV/atom) ↓	RMSD (Å) ↓
Element ✗/✗	47.47±2.65	**97.48**±0.36	**74.68**±2.03	25.21±0.60	0.1709±0.0103	0.1728±0.0098
Species ✗/✗	49.72±2.53	97.28±0.35	73.20±1.80	26.08±0.92	0.1646±0.0035	**0.1616**±0.0102
Species ✗/✓	**52.59**±2.25	96.45±0.64	70.68±2.31	**26.38**±1.01	**0.1575**±0.0080	0.1679±0.0117
Species ✓/✓	50.35±2.39	96.80±0.70	72.04±2.03	25.62±0.69	0.1696±0.0075	0.1853±0.0175

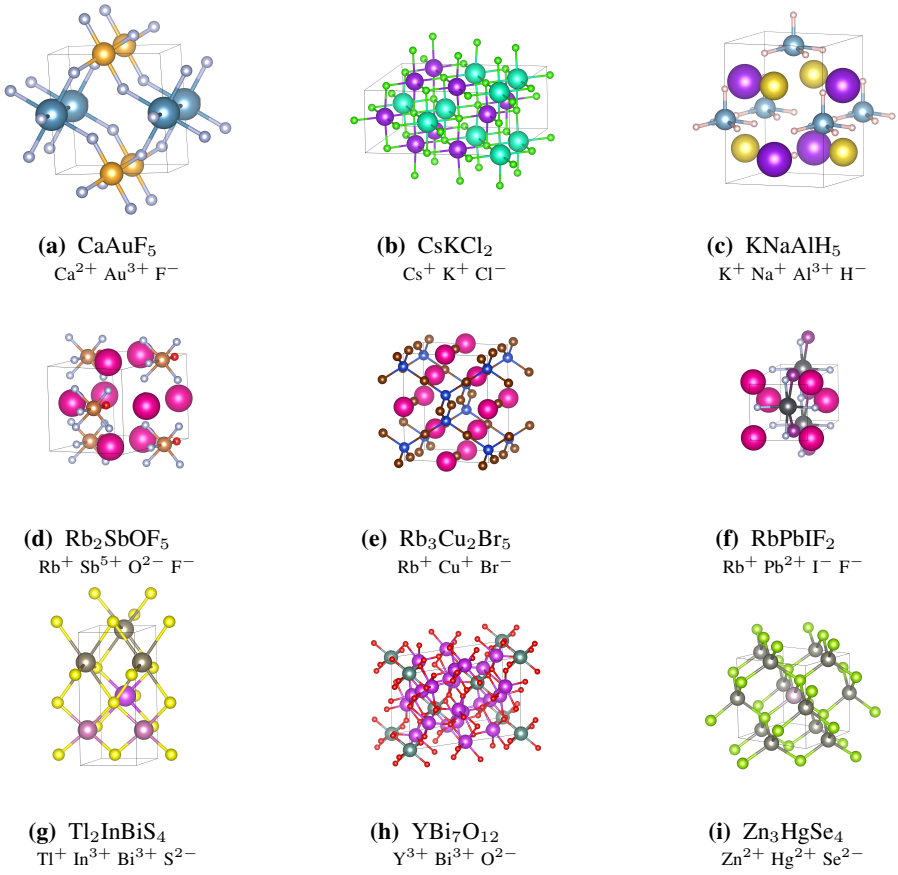
Table 13: Ablation results for charge neutrality and oxidation-state agreement. Results are reported as mean ± std over five independently seeded training runs, with 10,240 generated structures per seed. The method names indicate whether structured training / structured sampling are used, respectively. Best values are shown in **bold** and second-best values are underlined; the shaded row is OxiGen.

Method	Charge Neutrality (%) ↑	Site Agreement (%) ↑	Crystal Agreement (%) ↑
Element ✗/✗	–	–	–
Species ✗/✗	39.08±2.60	74.45±1.84	33.57±2.29
Species ✗/✓	**100.00**±0.00	**90.69**±0.81	**72.47**±1.40
Species ✓/✓	**100.00**±0.00	90.45±0.39	71.49±1.32

Table 14: Band-gap-conditioned generation results for target band gaps of 1, 3, 5, and 7 eV. Results are shown for a single fine-tuning seed, with 1,024 generated structures at each target.

Method	Target (eV)	RMSE (eV) ↓	Stable (%) ↑	Unique (%) ↑	Novel (%) ↑	S.U.N. (%) ↑	E_{hull} (eV/atom) ↓	RMSD (Å) ↓
MatterGen	1	1.116	49.32	99.02	82.42	33.59	0.1847	0.1961
	3	1.724	45.90	98.93	83.98	31.05	0.1530	0.2684
	5	2.185	49.51	94.82	82.32	31.54	0.1513	0.3016
	7	2.622	55.37	86.82	84.18	35.16	0.1456	0.3338
OxiGen	1	1.138	60.55	96.48	73.63	35.84	0.1489	0.1813
	3	1.322	71.48	96.29	71.97	43.55	0.0877	0.1895
	5	1.486	72.75	91.11	70.41	42.97	0.0816	0.2303
	7	2.125	68.95	87.79	75.98	41.89	0.0878	0.2580

F.4 Example Crystals



(a) $CaAuF_5$ — Ca^{2+} Au^{3+} F^{-}

(b) $CsKCl_2$ — Cs^{+} K^{+} Cl^{-}

(c) $KNaAlH_5$ — K^{+} Na^{+} Al^{3+} H^{-}

(d) Rb_2SbOF_5 — Rb^{+} Sb^{5+} O^{2-} F^{-}

(e) $Rb_3Cu_2Br_5$ — Rb^{+} Cu^{+} Br^{-}

(f) $RbPbIF_2$ — Rb^{+} Pb^{2+} I^{-} F^{-}

(g) Tl_2InBiS_4 — Tl^{+} In^{3+} Bi^{3+} S^{2-}

(h) YBi_7O_{12} — Y^{3+} Bi^{3+} O^{2-}

(i) Zn_3HgSe_4 — Zn^{2+} Hg^{2+} Se^{2-}

Figure 4: Examples of stable, unique, and novel crystals generated by OxiGen.