GHOST ATOMS: FLEXIBLE CRYSTAL GENERATION WITH VORONOI TESSELLATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Generative models for materials discovery fix the number of atoms per unit cell before generation, so the model never chooses how many atoms its crystal contains and cannot add or remove them. We restore this degree of freedom by padding every cell to a fixed number of atoms with fictitious *ghost atoms*, placed at well-spaced interstitial sites. We place them by iterative Voronoi maximin selection, so the padded training structures contain no unphysical atomic overlaps. Training the OMatG stochastic-interpolant model on an augmented MP-20 with Voronoi-placed ghosts improves the combined rate of stable or metastable, unique, and novel structures (SUN+MSUN) by 2.6 percentage points over a matched control that places ghosts uniformly at random, and by 6.0 points over the published OMatG model. The Voronoi-trained model learns to keep ghosts out of occupied space, placing far fewer near real atoms, so that converting a ghost into a real atom happens at a vacant site. Our best run reaches a 25.27% SUN+MSUN rate on LeMat-GenBench, a new state of the art.

1 INTRODUCTION

De novo generation of stable crystalline materials underpins the search for better battery cathodes, catalysts, and thermoelectrics (Curtarolo et al., 2013; Jain et al., 2013; Butler et al., 2018), and generative models have emerged as a promising route to accelerate this discovery (Xie et al., 2022; Jiao et al., 2023; Zeni et al., 2025; Okhotin et al., 2025; Hadži Veljković et al., 2026; Höllmer et al., 2025). Recent generative approaches involve diffusion models, flow matching, and stochastic interpolants that treat a crystal as a joint sample of its lattice, atomic species, and fractional coordinates (Xie et al., 2022; Jiao et al., 2023; Miller et al., 2024; Zeni et al., 2025; Höllmer et al., 2025), and are increasingly evaluated by the rate at which they produce structures that are simultaneously *stable*, *unique*, and *novel* (SUN; MSUN considers metastability instead of stability) (Betala et al., 2025). A shared feature of these models is that they operate on a *fixed* number of atoms per unit cell.

These fixed-cardinality generative models set the atom count N before generation begins, typically sampling it from the distribution of atom counts in the training set. The count is therefore imposed externally rather than chosen by the model, and the generative process has no mechanism to add or remove atoms. This matters because composition and atom count are not independent of structure. The thermodynamically preferred number of atoms in a unit cell depends on the lattice and species the model is simultaneously generating, so fixing N *a priori* prevents the model from exploring structures whose stability requires a different count. This constraint becomes even more limiting if one wishes to guide generation to produce materials with desirable properties (Zeni et al., 2025; Höllmer & Martiniani, 2026).

Contributions. To restore this degree of freedom, we introduce a new Voronoi-guided technique of padding crystals to a fixed N with removable *ghost atoms*. These ghost atoms, denoted by a reserved species (X), are stripped after generation, so that at inference each of the N slots is a decision the model makes to place a real atom or none at all (Figure 1). We place these ghosts at well-spaced interstitial sites found by iterative Voronoi tessellation, choosing at each step the site furthest from every existing atom. Because these are the largest voids in the cell, the padded training data never contains the overlapping or nonphysically close atoms that arbitrary placement would produce. This lets fixed-cardinality generative models adjust the number of atoms during generation, and we

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.

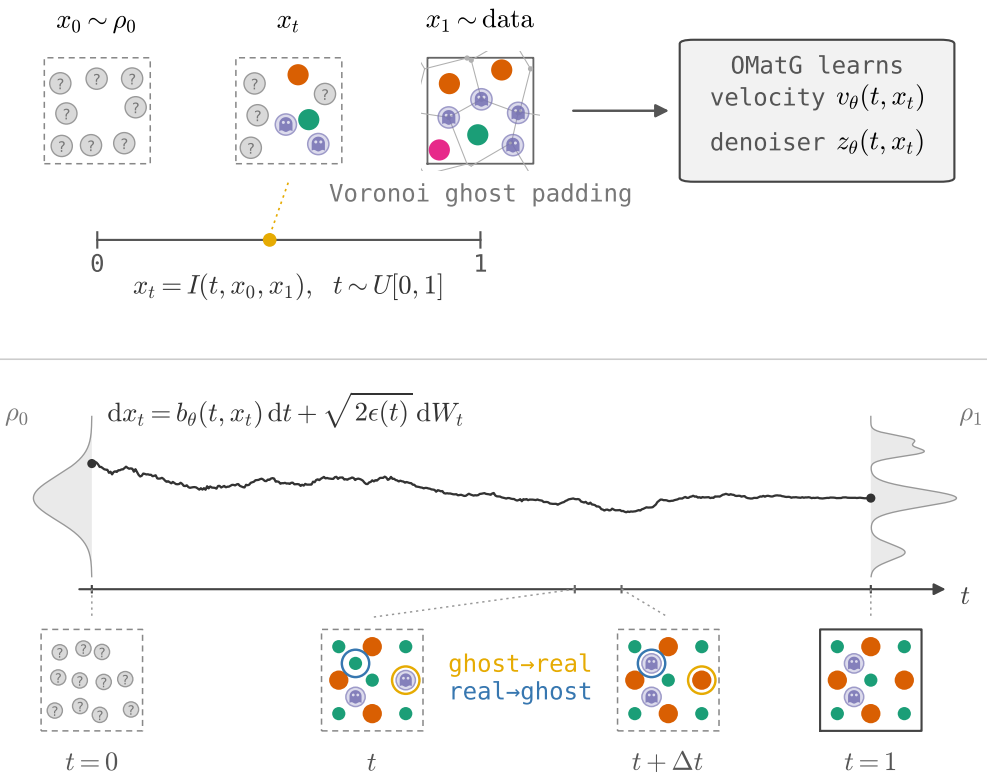


Figure 1: Overview of the method. **(top)** *Training.* Every training crystal is first padded to a fixed N by Voronoi ghost padding, so the target distribution $x_1 \sim \text{data}$ consists of cells with exactly N atoms whose padding sites lie in interstitial voids (Sections 3.4 and 3.5). A stochastic interpolant $x_t = I(t, x_0, x_1)$ (Section 3.2) with $t \sim U[0, 1]$ bridges the base distribution $x_0 \sim \rho_0$ to this padded target, and the network learns the velocity $v_\theta(t, x_t)$ and denoiser $z_\theta(t, x_t)$. **(bottom)** *Inference.* Generation integrates the SDE (Equation 3) from ρ_0 to ρ_1. Because the species field carries a ghost label, atoms may convert ghost→real or real→ghost along the trajectory (Section 3.3), which is the added degree of freedom: the model, not a prior over atom counts, decides how many real atoms the final cell contains. The diffusion coefficient $\epsilon(t)$ is an inference-time choice we retune post hoc (Section 3.7).

measure the benefit by their rate of producing stable or metastable, unique, and novel structures. We compare Voronoi padding with random padding and show that the Voronoi-trained model learns to keep its ghosts out of the space occupied by real atoms. We will also release our Voronoi-ghosted dataset, which other fixed-cardinality architectures can use with minimal changes to generate crystals with a variable number of atoms. Additionally, we find that the inference-time noise schedule of the stochastic interpolant has a significant impact on performance: decaying noise schedules improve generation quality while injecting more total noise. With these contributions, we set a new state of the art of **25.27%** combined SUN and MSUN rate on the LeMat-GenBench benchmark, improving on the previous best of 24.1% (Crystalite (Hadži Veljković et al., 2026)).

2 Related Work

Crystal generative models. CDVAE (Xie et al., 2022) combined a variational autoencoder with a score-based diffusion decoder for periodic crystals and introduced the MP-20 benchmark. DiffCSP (Jiao et al., 2023) applied equivariant joint diffusion to crystal structure prediction. FlowMM (Miller et al., 2024) brought Riemannian flow matching to the periodic setting, and MatterGen (Zeni et al., 2025) scaled diffusion-based generation with property-guided sampling. Crystalite (Hadži Veljković et al., 2026) achieved strong performance with a lightweight transformer.

OMatG (Höllmer et al., 2025) used stochastic interpolants for crystals, jointly modeling species, coordinates, and lattice. All of these methods operate on a fixed atom count per unit cell. We build our experiments on OMatG to exploit the flexibility of stochastic interpolants (Albergo et al., 2025), in particular the freedom to change the noise schedule after training (Section 3.7).

Variable cardinality. MiAD (Okhotin et al., 2025) independently addresses the fixed-cardinality constraint by introducing *mirage atoms* initialized uniformly at random within the unit cell. A key distinction between their design and ours is that MiAD masks the coordinate loss on mirage positions, so only the species labels are supervised; in our formulation, ghost coordinates are supervised targets, making their placement geometry a training signal. Because ghost atoms need no special treatment in the loss, our ghosted dataset is general-purpose: an existing fixed-cardinality architecture needs only minimal changes, such as an extra species label, to train on it. Our random-ghost control experiment (Section 4.2) isolates the effect of that geometry under supervised coordinates and is not a reimplementation of MiAD. In the sequence-modeling domain, Edit Flows (Havasi et al., 2025) address variable-length generation via discrete flow matching with insertion and deletion operations over continuous-time Markov chains, providing a complementary perspective on the variable-cardinality problem outside the materials setting.

3 METHODS

3.1 CRYSTAL REPRESENTATION IN OMATG

Following OMatG (Höllmer et al., 2025), a unit cell with N atoms is the triplet $(\mathbf{A}, \mathbf{X}, \mathbf{L})$ of species $\mathbf{A} \in \mathcal{A}^N$, fractional coordinates $\mathbf{X} \in [0,1)^{N\times 3}$ on the three-dimensional torus, and lattice $\mathbf{L} \in \mathbb{R}^{3\times 3}$ whose rows are the lattice vectors; Cartesian positions are $\mathbf{XL}$, and the crystal is the periodic repetition of the cell. With ghost padding (Section 3.4), $N = 20$ for every structure and the species alphabet $\mathcal{A}$ contains the chemical elements plus the ghost species X. OMatG generates the continuous fields $\mathbf{X}$ and $\mathbf{L}$ with stochastic interpolants (Section 3.2; Liu et al., 2023; Lipman et al., 2023; Albergo et al., 2025) and the species $\mathbf{A}$ with discrete flow matching (Section 3.3), all on a shared time $t \in [0,1]$ and predicted by a single network that sees the current state of all three fields. The base distributions are uniform fractional coordinates, a lattice prior with log-normal lattice lengths fit to MP-20 and angles uniform on $[60^\circ, 120^\circ]$ (Miller et al., 2024), and fully masked species.

The network is the CSPNet encoder of DiffCSP (Jiao et al., 2023) as adapted in OMatG: a message-passing graph neural network (Duval et al., 2023) on the fully connected graph of the atoms in the cell, whose edge features are sinusoidal embeddings of periodic fractional-coordinate differences and whose lattice features are the inner products $\mathbf{LL}^\top$. Its outputs therefore respect the symmetries of the representation: permutation of atoms, rotation of the lattice, and translation of $\mathbf{X}$ modulo 1. It returns per-atom velocities and denoisers for $\mathbf{X}$ and a velocity for $\mathbf{L}$ (Section 3.2), and per-atom logits over $\mathcal{A}$ (Section 3.3).

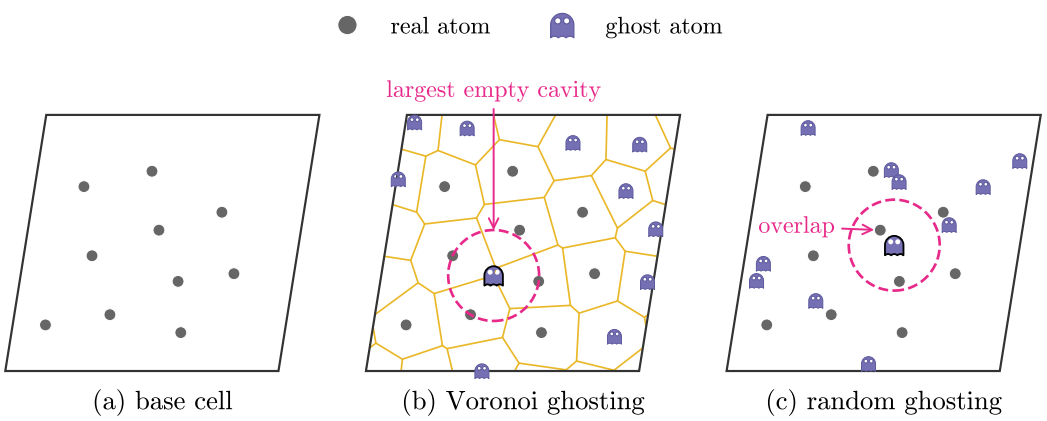
3.2 STOCHASTIC INTERPOLANTS

Stochastic interpolants (Albergo et al., 2025; 2024) provide a unified framework for continuous generative models. Given a base distribution ρ_0 and a target ρ_1, a stochastic interpolant, $I(t, x_0, x_1) = \alpha(t)x_0 + \beta(t)x_1 + \gamma(t)z$, is a function which bridges $x_0 \sim \rho_0$ with $x_1 \sim \rho_1$ where $z \sim \mathcal{N}(0, \mathbf{I})$. The time-dependent scheduling functions satisfy $\alpha(0) = \beta(1) = 1$ and $\alpha(1) = \beta(0) = \gamma(0) = \gamma(1) = 0$. $I(t, x_0, x_1)$ defines a path of densities $\rho_t \; \forall \; t \in [0,1]$ that recovers the base and target distributions at the endpoints. This path implies a velocity field, v, satisfying the continuity equation $\partial_t \rho_t + \nabla \cdot (\rho_t \, v) = 0$. A neural network, v_θ, can be trained to approximate v by minimizing a flow matching objective,

$$\mathcal{L}_v(\theta) = \mathbb{E}\left[\|v_\theta(t, x_t) - \dot{I}(t, x_0, x_1)\|^2\right]. \tag{1}$$

During inference we integrate $\dot{x}_t = v_\theta(t, x_t)$ from time $t = 0$ to $t = 1$, starting from initial condition x_0. A stochastic transport can be learned by parameterizing a denoiser network, z_θ, where the denoising loss is given by another objective,

$$\mathcal{L}_z(\theta) = \mathbb{E}\left[\|z_\theta(t, x_t) - z\|^2\right], \tag{2}$$



(a) base cell (b) Voronoi ghosting (c) random ghosting

Figure 2: Padding a partially filled unit cell to a fixed size of $N = 20$ atoms. **(a)** the base cell of real atoms; **(b)** iterative Voronoi maximin ghosting places each ghost one at a time at the largest remaining interstitial void, recomputing the tessellation after each insertion (Algorithm 1), so successive ghosts avoid both real atoms and previously placed ghosts; the tessellation of the final padded cell is overlaid; **(c)** uniformly random ghosting scatters them instead. In both padded panels the outlined marker is the *next* ghost each scheme would place, and the dashed circle is the same radius in both: the largest empty cavity available to the maximin rule. Under Voronoi placement that cavity is empty by construction, whereas the random draw puts a ghost in a cavity a real atom already occupies. Ghosts are stripped after generation.

which allows for the construction of a stochastic differential equation (SDE) for generating approximate data from ρ_1,

$$dx_t = b_\theta(t, x_t)\, dt + \sqrt{2\epsilon(t)}\, dW_t, \tag{3}$$

where $b_\theta(t, x_t) = v_\theta(t, x_t) - z_\theta(t, x_t)\epsilon(t)\gamma^{-1}(t)$ is a drift term, $\epsilon(t) \geq 0$ is a time-dependent diffusion coefficient, and W_t is a Wiener process. This injects stochasticity during sampling. The framework subsumes normalizing flows ($\epsilon = 0$) and score-based diffusion (Song et al., 2021) as special cases, and admits any noise schedule $\epsilon(t)$ at inference time without retraining—a property we exploit in Section 5.2. In OMatG the fractional coordinates use a periodic linear interpolant on the torus with a latent term $\gamma(t)z$ and are sampled with the SDE, while the lattice uses a linear interpolant without latent term ($\gamma = 0$) and is sampled with the ODE; only the coordinates therefore need a denoiser. Following OMatG and FlowMM (Höllmer et al., 2025; Miller et al., 2024), the learned velocity is also rescaled during sampling as $v_\theta \to (1 + st)\, v_\theta$ (velocity annealing), with $s = 6.33$ for the coordinates and $s = 1.07$ for the lattice (Table 5).

3.3 Discrete Flow Matching for Species

Species are generated with discrete flow matching (DFM) from a masking base distribution (Campbell et al., 2024). Every atom starts in a dedicated mask state, and the training interpolant reveals each atom's true species independently with probability t, leaving it masked otherwise. The network is trained with a cross-entropy loss to predict, for every atom, the distribution of its final species given the current state $(\mathbf{A}_t, \mathbf{X}_t, \mathbf{L}_t)$. Sampling simulates the corresponding continuous-time Markov chain in steps of size Δt: each masked atom is unmasked with probability $\Delta t\,(1 + \nu t)/(1 - t)$ to a species drawn from the predicted distribution, and each unmasked atom is re-masked with probability $\nu\,\Delta t$, where $\nu \geq 0$ sets the stochasticity; atoms still masked at the final step are unmasked. With $\nu = 0$ each atom's species is decided once. With $\nu > 0$ the model can revise its decisions, so an atom can pass from ghost to real, or from real to ghost, through the mask state (Figure 1). This is the mechanism by which a ghost-padded model adds or removes atoms during generation.

Algorithm 1 Voronoi maximin ghost padding

Require: cell with n atoms at $\{\mathbf{r}_i\}_{i=1}^{n}$, lattice $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$, target count N
Ensure: augmented cell with N atoms
1: **for** $k = 1$ **to** $N - n$ **do**
2: build $3\times3\times3$ supercell with jittered image translations
3: compute Voronoi tessellation of the supercell
4: $\mathcal{V} \leftarrow$ vertices lying in the central cell
5: $\mathcal{S} \leftarrow$ all 27 periodic images of the atoms placed so far
6: $\mathbf{r}_{\text{ghost}}^{(k)} \leftarrow \arg\max_{\mathbf{v}\in\mathcal{V}} \min_{\mathbf{r}\in\mathcal{S}} \|\mathbf{v} - \mathbf{r}\|$ (Eq. 4)
7: append ghost atom: $\mathbf{r}_{n+k} \leftarrow \mathbf{r}_{\text{ghost}}^{(k)}$, species X
8: **end for**

3.4 Ghost Atom Padding

A ghost atom is a placeholder carrying a reserved species label X that is disjoint from every real element. We augment each structure by adding ghost atoms until the cell contains a fixed number of atoms N. At inference time, after integration to $t = 1$, we strip ghost atoms to recover the physical crystal. For MP-20, whose cells contain at most 20 atoms, we pad to $N = 20$, so that every training structure and every generated structure occupies the same 20 slots.[1]

3.5 Voronoi Maximin Placement

We pad an n-atom crystal to the target count N by inserting the $N - n$ ghost atoms one at a time, placing each at the largest void currently available in the cell (Figure 2). We approximate the crystal's periodic boundary conditions by replicating the unit cell into a $3\times3\times3$ supercell and tessellating that rather than the unit cell alone. This is sufficient for our purpose: the vertex we select always lies in the central cell, and every candidate-to-atom distance is measured against all 27 images, so the void geometry around the placement site is captured faithfully.

We compute the Voronoi tessellation of the supercell using `freud` (Ramasubramani et al., 2020). Its vertices are the points locally equidistant from the surrounding atoms and therefore mark candidate interstitial sites; of these we keep the vertices $\mathcal{V}$ whose fractional coordinates fall within the central cell, up to a tolerance $\eta = 10^{-3}$. Writing $\mathcal{S}$ for the set of all 27 periodic images of the $n + k - 1$ atoms placed so far, we select the central-cell vertex that lies farthest from every existing atom, real or previously placed ghost,

$$\mathbf{r}_{\text{ghost}}^{(k)} = \arg\max_{\mathbf{v}\in\mathcal{V}} \min_{\mathbf{r}\in\mathcal{S}} \|\mathbf{v} - \mathbf{r}\|. \tag{4}$$

Geometrically, Equation 4 places each ghost at the center of the largest atom-free sphere centered in the central cell, as far as possible from the surrounding nuclei.[2]

After inserting the ghost atom we recompute the tessellation with the augmented point set, so that each newly placed ghost informs the placement of the next, and repeat until the cell holds N atoms (Algorithm 1; Appendix A illustrates the procedure). To resolve degeneracies which arise when tessellating highly symmetric unit cells we perturb each supercell image translation by an independent uniform jitter on $[-\delta/2, \delta/2]$ per lattice-vector component, with $\delta = 0.05$ in fractional coordinates (a maximum displacement of 0.025 along each lattice vector). This is large enough to remove degeneracies while leaving the interatomic-distance distribution essentially unchanged.

3.6 Model Training

We train the same OMatG model (Section 3.1), with a dedicated output channel for the ghost species, on the Voronoi-ghosted and randomly ghosted (Section 4.2) MP-20 datasets using the settings in Table 5, and sample with the SDE integrator for the coordinates (Equation 3, Appendix H).

[1] Internally, ghost atoms are stored with atomic number -1 and remapped to the fixed index 119, one above the heaviest element ($Z = 118$), at the model input, kept separate from the masking token at index 0.

[2] Voronoi vertices are precisely the circumcenters of a generalized triangulation in the working dimensionality.

3.7 Inference Noise Schedule

The diffusion coefficient $\epsilon(t)$ in Equation 3 sets how much noise the sampler injects into the fractional coordinates at time t. It enters only at sampling (Section 3.2), so the model is trained once and $\epsilon(t)$ is a post-training control. We use it to retune the schedule for the ghost-padded model instead of inheriting the bell-shaped schedule of OMatG (Section 5.2).

We summarize the total noise a schedule injects by its *quadratic variation* of the coordinate diffusion process, or noise budget (Øksendal, 2003):

$$V_\epsilon = \int_0^1 2\epsilon(t)\, dt. \tag{5}$$

This lets us compare schedules of different shape at a matched noise budget. We consider decaying, bell-shaped, and concave families (Table 6, Appendix J).

4 Experiments

Datasets. We train and evaluate our proposed ghost atoms method on Materials Project data (Jain et al., 2013). Specifically, we use the **MP-20** dataset consisting of inorganic crystal structures with at most 20 atoms in any given unit cell, an energy above the convex hull (Section 4.1) less than 0.08 eV/atom, and a formation energy less than 2.0 eV/atom (Xie et al., 2022). MP-20 is a standard benchmark for crystal generation (Xie et al., 2022; Jiao et al., 2023; Miller et al., 2024; Höllmer et al., 2025), which makes the effect of ghost atoms directly comparable with prior work.

4.1 De Novo Generation

We evaluate generated structures with LeMat-GenBench (Betala et al., 2025), stripping nonphysical ghost atoms first. Before any metric is computed, every structure passes a mandatory validity filter to ensure charge neutrality, a minimum interatomic-distance check, and physical-plausibility bounds on density and cell geometry (Appendix D). All rates we report are taken over this valid subset. **Stability** is assessed from the energy above the convex hull, e_{hull}, predicted by an ensemble of three machine-learning interatomic potentials (MLIPs: Orb-v3, MACE-MP, and UMA), each with its own reference hull, with the three values averaged (Rhodes et al., 2025; Batatia et al., 2022; 2025; Wood et al., 2025). The convex hull is built by plotting the formation energies of known compounds in a chemical system against composition and connecting the lowest-energy ones. A structure's e_{hull} is how far its energy lies above this hull, which is the energy it would release by decomposing into the compounds on the hull; structures on or below the hull ($e_{\mathrm{hull}} \leq 0$) are stable (Appendix D). Structures are relaxed with UMA (`uma-s-1p1`, `omat` task) before scoring. A structure is *stable* if $e_{\mathrm{hull}} \leq 0$ and *metastable* if $0 < e_{\mathrm{hull}} \leq 0.1$ eV/atom. **Uniqueness** and **Novelty** are determined by pairwise structure matching with pymatgen's `StructureMatcher` (Ong et al., 2013). Uniqueness counts each set of matching structures in the generated batch once, and novelty requires that a structure match no entry of the LeMat-Bulk reference dataset (Siron et al., 2025).

Combining these criteria, SUN (respectively MSUN) measures the fraction of structures that are simultaneously stable (respectively metastable), unique, and novel, applied as a hierarchical pipeline: stability, then uniqueness within the stable and metastable sets, then novelty within the unique structures. Because the stable and metastable sets are disjoint by construction, SUN and MSUN count non-overlapping structures, and we report their sum as a single combined rate.

4.2 Random Ghost Control

To isolate the effect of *how* ghosts are placed, we build a second augmented dataset nearly identical to the Voronoi-ghosted dataset except that ghost positions are drawn uniformly at random within the unit cell, subject to no geometric constraint. We then train and evaluate a model on this dataset under exactly the same architecture, optimizer, and inference settings used for the Voronoi-ghosted model. The two models therefore differ only in the ghost placement rule. As a result, any difference in their generated structures is attributable to placement geometry alone (Appendices B and C). Uniformly random placement is also the choice made by MiAD (Okhotin et al., 2025); see Section 2 for the comparison.

Table 1: De novo generation on MP-20, scored with LeMat-GenBench (%, over valid structures). SUN and MSUN are disjoint; their sum is the combined rate. For our two models we report the mean ± standard error over the same five sampling seeds; the best single Voronoi run reaches 25.27% and sets a new state of the art.

Model	Stable	Metastable	SUN	MSUN	SUN+MSUN
DiffCSP (Jiao et al., 2023)	2.3	29.8	0.1	8.5	8.6
MatterGen (Zeni et al., 2025)	2.0	33.4	0.2	15.0	15.2
MiAD (Okhotin et al., 2025)	6.4	**62.0**	1.0	16.6	17.6
OMatG (Höllmer et al., 2025)	11.6	49.8	1.0	18.0	19.0
Crystalite (Hadži Veljković et al., 2026)	**12.7**	51.6	**1.5**	22.6	24.1
OMatG + random ghosts (control)	7.6 ±0.2	46.4 ±0.6	0.66 ±0.04	21.8 ±0.4	22.4 ±0.5
OMatG + Voronoi ghosts (ours)	7.3 ±0.3	47.0 ±0.3	0.68 ±0.07	**24.3** ±0.1	**25.0** ±0.1

5 RESULTS

5.1 IMPACT OF GHOST ATOM PLACEMENT

Voronoi Performance. Table 1 compares our models with other models evaluated on LeMat-GenBench, including the un-ghosted OMatG submission and the current best-performing model, Crystalite (Hadži Veljković et al., 2026). Our Voronoi-ghosted model reaches $25.02 \pm 0.14\%$ combined SUN and MSUN (mean ± standard error over five seeds), and its best single run, 25.27%, sets a new state of the art, exceeding Crystalite's 24.1%.

The controlled comparison, however, is with the randomly ghosted model: it is identical in architecture, training, and inference, and differs only in where ghost atoms were placed during augmentation. Sampled with the same five seeds, so that each pair of runs starts from identical base samples, the control model reaches $22.41 \pm 0.48\%$ combined SUN and MSUN, so Voronoi placement accounts for an improvement of 2.6 ± 0.5 percentage points. The two models have identical validity (96.2%, $12{,}024$ of $12{,}500$ structures each) and stable rates that agree within their standard errors (7.3% versus 7.6%); the entire gain is in the metastable, unique, and novel fraction (MSUN, 24.3% versus 21.8%) rather than in raw stability. Both ghosted variants also exceed the best published OMatG result (19.0%). Appendix E shows the full e_{hull} distribution for the best model, and Appendix F reports crystal structure prediction.

Reproducing the Tessellation. Where ghosts are placed matters because their coordinate loss is not masked: they are supervised targets for the coordinate field rather than inert padding, and a model trained on physically implausible targets learns to reproduce them. We see this directly in where each model places its ghosts relative to real atoms.

Figure 3 shows the fraction of all ghosts lying within a distance d of the nearest real atom, as measured on the supercell over the $121{,}489$ Voronoi-model and $122{,}664$ random-model ghosts generated with the same five seeds. The two models are well separated at every d. Within 1 Å of a real atom the Voronoi model places 0.81–1.21% of its ghosts, against 23.3–24.3% for the randomly ghosted model (ranges over the five seeds). Within 0.5 Å the Voronoi model places a single ghost out of $121{,}489$, against 3.04% for the random model, and within 0.2 Å it places none at all, against 0.18%. No Voronoi ghost is coincident (within 0.05 Å) with a real atom in any of the five seeds; the randomly ghosted model produces three such overlaps.

We also verify that the Voronoi-trained model reproduces the geometry rather than memorizing positions by taking the generated structures across all 5 seeds, stripping the model's ghosts, and re-running the Voronoi maximin procedure of Section 3.5 on the remaining real-atom cell. Then, we compare the ghost–real distance distribution the model produced against the one the algorithm would have produced on the same real atoms. Over $11{,}593$ structures containing $121{,}489$ ghosts, the model's median $V^{1/3}$-normalized ghost–real distance is 0.357, against 0.359 for the re-tessellation—a difference of 0.6%. The agreement is stable across all five seeds (Table 4 in Appendix G).

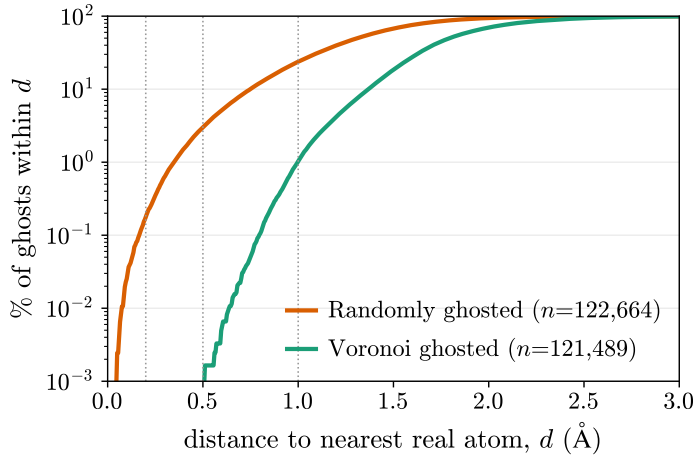


Figure 3: Fraction of generated ghosts within distance d of the nearest real atom, for both models over the same five seeds. Ghosts from the Voronoi-trained model sit far from real atoms, whereas nearly a quarter of the randomly ghosted model's ghosts lie within 1 Å of one.

5.2 Inference Noise Tunability

Introducing ghost atoms changes the geometry of the generative target, and the noise schedule inherited from the un-augmented OMatG model proves to be suboptimal. Because the schedule $\epsilon(t)$ is an inference-time quantity (Section 3.7), we are free to retune it without retraining.

At matched quadratic variation $V_\epsilon \approx 9.75$, the shape of the schedule still matters: holding the species stochasticity ν fixed and varying only the coordinate $\epsilon(t)$, we get a combined SUN+MSUN of 25.0% with linear decay, 23.8% with a bell schedule, 22.7% for an early-pulse polynomial, and 19.7% for a concave (late-loaded) schedule. At a lower budget ($V_\epsilon \approx 3.9$), bell and linear decay perform about equally; increasing the injected noise separates their SUN+MSUN rates and degrades the e_{hull} distribution of the non-decaying bell schedule (Figure 4).

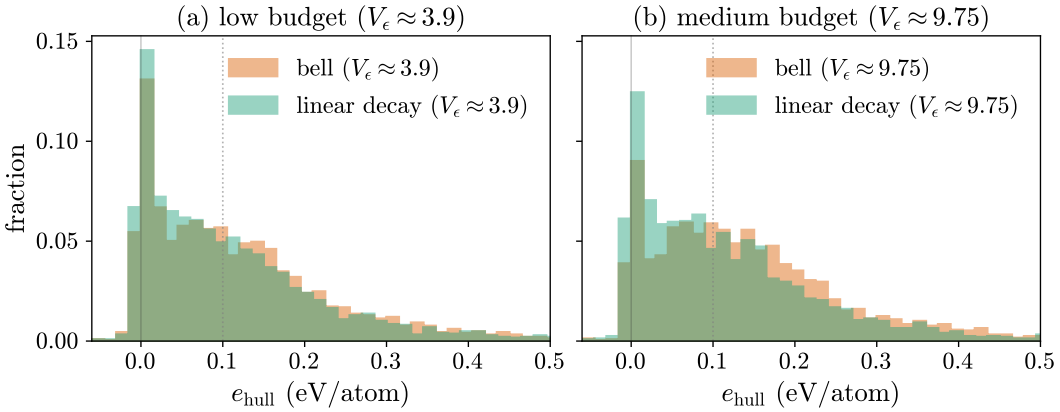


Figure 4: Energy-above-hull distributions at matched noise budget. (a) At low budget ($V_\epsilon \approx 3.9$), bell and linear-decay schedules produce similar distributions. (b) At medium budget ($V_\epsilon \approx 9.75$), the bell develops a high-e_{hull} shoulder while linear decay stays concentrated at low e_{hull}. Structures with $e_{\text{hull}} \leq 0$ are stable and those with $0 < e_{\text{hull}} \leq 0.1$ eV/atom are metastable.

The difference appears to lie in *when* the noise is injected. A decaying $\epsilon(t)$ front-loads noise while the structure is still forming and vanishes by $t = 1$; 75% of the linear decay budget is delivered

before $t = 0.5$. The bell schedule, by contrast, disperses noise nearly uniformly and holds $\sim 50\%$ of its budget after $t = 0.5$, when too little of the trajectory remains to correct it. Figure 4 shows the consequence: at low noise budget both schedules produce similar e_{hull} distributions, but at matched medium budget the bell develops a high-e_{hull} shoulder above 0.2 eV/atom, shifting mass away from the metastable band, while the linear-decay distribution stays concentrated at low e_{hull} (mean 0.146 vs. 0.123 eV/atom). The bell's mean e_{hull} climbs $+0.020$ eV/atom as budget rises from 3.9 to 9.75 and yet the decay's climbs only $+0.009$ eV/atom; the bell degrades roughly twice as fast with added noise.

6 Discussion

Our controlled experiment shows that the geometry of padding matters: we compare two identically configured models, aside from the ghost placement protocol in their training corpus, and find they differ by 2.6 percentage points in SUN + MSUN. The difference appears to be attributable to the model learning to respect the space occupied by real atoms (Section 5.1). Because Voronoi ghosts occupy genuine interstitial voids, the model's ghost-to-real decision operates at sites where an atom can plausibly sit, which is what increases the benefit of the added atom-count degree of freedom.

A second observation is that padding geometry and the inference noise schedule are separable controls that both must be set well. The benefit of physically grounded padding is realized under a decaying, high-budget noise schedule, which the stochastic-interpolant formulation lets us tune entirely after training (Section 5.2); with the schedule inherited from the un-augmented model, the ghost-padded model performs worse.

Limitations. The Voronoi tessellation approximates periodicity with a $3 \times 3 \times 3$ supercell rather than exact periodic boundary conditions, and stability is assessed with an MLIP ensemble rather than DFT, which is faster, but MLIP errors can flip the stability label of structures close to the convex hull (Riebesell et al., 2025). The SUN + MSUN rate is a useful benchmark, but practical impact will require guiding generation toward desirable properties, for example with inference-time reinforcement learning as in OMatG-IRL (Höllmer & Martiniani, 2026). Finally, our study of noise schedules covers a single model and dataset; whether decaying schedules help other stochastic-interpolant models, including outside materials, remains to be tested.

Future work. In selecting the Voronoi vertices, the maximin rule is purely geometric; weighting or nudging the selected vertex by neighboring charges toward a Coulomb-optimal interstitial site is a natural extension, as well as incorporating other inductive biases. Noise-schedule tuning was not the focus of this work, and our schedule experiments were empirical; a theoretical treatment or a broader sweep may find further gains. The learned ghost-to-real degree of freedom also invites a direct test: take a Voronoi-ghosted real crystal, turn one generated ghost into a real atom, relax the structure, and measure the geometric cost. Over many such trials, with an unbiased choice of which ghost to convert, we expect this cost to be lower than for other padding schemes. Finally, application to other fixed-cardinality models such as diffusion and flow-matching models, which face the same padding choice, would help to establish the generality of this approach.

7 Conclusion

We introduced the technique of *ghost atoms* placed by iterative Voronoi maximin selection, a padding scheme that lets a fixed-cardinality generative model for crystal discovery choose its own atom count while keeping every padding position in physically plausible space. On MP-20 with OMatG, Voronoi-placed ghosts reach $25.02 \pm 0.14\%$ combined SUN and MSUN (mean $\pm$ standard error over five seeds), with a best run of 25.27% that sets a new state of the art, and they improve on an identically configured randomly ghosted control by 2.6 percentage points. The model reproduces the geometric placement rule rather than memorizing padding, so its decision to introduce or remove an atom is made at a physically realistic location—the degree of freedom the method was designed to provide.

AI USE STATEMENT

In this work, we used generative AI tools to implement methods, provide feedback on research methodology and experiments, support qualitative and thematic data analysis, and interpret results. We have not used generative AI tools to generate synthetic data sets (we use only public data), help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, propose or refine hypotheses, assist with translation, or clean and reformat datasets. Additionally, we used generative AI tools to help write code, find and check references, and draft and edit parts of the text, all of which the authors reviewed and revised. We have reviewed all AI-assisted work: we treated AI-generated code and critiques as starting points, multiple authors reviewed, edited, and finalized every implementation and decision, and the authors verified every AI-assisted analysis and interpretation. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work develops generative models for inorganic crystal structures, trained on the public MP-20 dataset; it involves no human subjects or personal data. Materials discovery has many possible downstream applications and implications, but we do not believe any of them need to be highlighted for this work. Some authors are affiliated with a company that develops AI methods for materials discovery, we disclose this as a potential conflict of interest.

REPRODUCIBILITY STATEMENT

Section 3 specifies the crystal representation, the generative processes, and the ghost-padding procedure (Algorithm 1). Appendix I lists the full training and sampling configuration (Table 5), and Appendix J defines every inference noise schedule we use (Table 6). We evaluate with the public LeMat-GenBench benchmark in its `comprehensive_multi_mlip_hull` configuration (Section 4.1, Appendix D) and report every rate for our models as a mean and standard error over five sampling seeds (Table 1). Our models are built on the open-source OMatG code (Höllmer et al., 2025), and we will release the Voronoi-ghosted MP-20 dataset used for training.

REFERENCES

Michael S. Albergo, Mark Goldstein, Nicholas M. Boffi, Rajesh Ranganath, and Eric Vanden-Eijnden. Stochastic interpolants with data-dependent couplings. In *International Conference on Machine Learning (ICML)*, volume 235 of *Proceedings of Machine Learning Research*, pp. 921–937, 2024. URL https://proceedings.mlr.press/v235/albergo24a.html.

Michael S. Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025. URL https://jmlr.org/papers/v26/23-1605.html.

Ilyes Batatia, Dávid Péter Kovács, Gregor N. C. Simm, Christoph Ortner, and Gábor Csányi. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. URL https://arxiv.org/abs/2206.07697.

Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M. Elena, Dávid P. Kovács, Janosh Riebesell, Xavier R. Advincula, Mark Asta, Matthew Avaylon, William J. Baldwin, Fabian Berger, Noam Bernstein, Arghya Bhowmik, Filippo Bigi, Samuel M. Blau, Vlad Cărare, Michele Ceriotti, Sanggyu Chong, James P. Darby, Sandip De, Flaviano Della Pia, Volker L. Deringer, Rokas Elijošius, Zakariya El-Machachi, Edvin Fako, Fabio Falcioni, Andrea C. Ferrari, John L. A. Gardner, Mikołaj J. Gawkowski, Annalena Genreith-Schriever, Janine George, Rhys E. A. Goodall, Jonas Grandel, Clare P. Grey, Petr Grigorev, Shuang Han, Will Handley, Hendrik H. Heenen, Kersti Hermansson, Cheuk Hin Ho, Stephan Hofmann, Christian Holm, Jad Jaafar, Konstantin S. Jakob, Hyunwook Jung, Venkat Kapil, Aaron D. Kaplan, Nima Karimitari, James R. Kermode, Panagiotis Kourtis, Namu Kroupa, Jolla Kullgren, Matthew C. Kuner, Domantas Kuryla, Guoda

Liepuoniute, Chen Lin, Johannes T. Margraf, Ioan-Bogdan Magdău, Angelos Michaelides, J. Harry Moore, Aakash A. Naik, Samuel P. Niblett, Sam Walton Norwood, Niamh O'Neill, Christoph Ortner, Kristin A. Persson, Karsten Reuter, Andrew S. Rosen, Louise A. M. Rosset, Lars L. Schaaf, Christoph Schran, Benjamin X. Shi, Eric Sivonxay, Tamás K. Stenczel, Christopher Sutton, Viktor Svahn, Thomas D. Swinburne, Jules Tilly, Cas van der Oord, Santiago Vargas, Eszter Varga-Umbrich, Tejs Vegge, Martin Vondrák, Yangshuai Wang, William C. Witt, Thomas Wolf, Fabian Zills, and Gábor Csányi. A foundation model for atomistic materials chemistry. *The Journal of Chemical Physics*, 163(18):184110, 2025. doi: 10.1063/5.0297006. URL https://arxiv.org/abs/2401.00096.

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. *arXiv preprint arXiv:2512.04562*, 2025. URL https://arxiv.org/abs/2512.04562.

Keith T. Butler, Daniel W. Davies, Hugh Cartwright, Olexandr Isayev, and Aron Walsh. Machine learning for molecular and materials science. *Nature*, 559(7715):547–555, 2018. doi: 10.1038/s41586-018-0337-2. URL https://www.nature.com/articles/s41586-018-0337-2.

Andrew Campbell, Jason Yim, Regina Barzilay, Tom Rainforth, and Tommi Jaakkola. Generative flows on discrete state-spaces: Enabling multimodal flows with applications to protein co-design. In *International Conference on Machine Learning (ICML)*, volume 235 of *Proceedings of Machine Learning Research*, pp. 5453–5512, 2024. URL https://arxiv.org/abs/2402.04997.

Stefano Curtarolo, Gus L. W. Hart, Marco Buongiorno Nardelli, Natalio Mingo, Stefano Sanvito, and Ohad Levy. The high-throughput highway to computational materials design. *Nature Materials*, 12 (3):191–201, 2013. doi: 10.1038/nmat3568. URL https://www.nature.com/articles/nmat3568.

Alexandre Duval, Simon V. Mathis, Chaitanya K. Joshi, Victor Schmidt, Santiago Miret, Fragkiskos D. Malliaros, Taco Cohen, Pietro Liò, Yoshua Bengio, and Michael Bronstein. A hitchhiker's guide to geometric GNNs for 3D atomic systems. *arXiv preprint arXiv:2312.07511*, 2023. URL https://arxiv.org/abs/2312.07511.

Tin Hadži Veljković, Joshua Rosenthal, Ivor Lončarić, and Jan-Willem van de Meent. Crystalite: A lightweight transformer for efficient crystal modeling. *arXiv preprint arXiv:2604.02270*, 2026. URL https://arxiv.org/abs/2604.02270.

Marton Havasi, Brian Karrer, Itai Gat, and Ricky T. Q. Chen. Edit Flows: Variable length discrete flow matching with sequence-level edit operations. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/cb43f46154e750746602faaffd65fbbb-Abstract-Conference.html.

Philipp Höllmer and Stefano Martiniani. Open materials generation with inference-time reinforcement learning. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=xfHppnGXaH.

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 *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=gHGrzxFujU.

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 (NeurIPS)*, 2023. URL https://arxiv.org/abs/2309.04475.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations (ICLR)*, 2023. URL https://arxiv.org/abs/2210.02747.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations (ICLR)*, 2023. URL https://arxiv.org/abs/2209.03003.

Benjamin Kurt Miller, Ricky T. Q. Chen, Anuroop Sriram, and Brandon M. Wood. FlowMM: Generating materials with Riemannian flow matching. In *International Conference on Machine Learning (ICML)*, 2024. URL https://arxiv.org/abs/2406.04713.

Andrey Okhotin, Maksim Nakhodnov, Nikita Kazeev, Mikhail Lazarev, Andrey Ustyuzhanin, and Dmitry Vetrov. MiAD: Mirage atom diffusion for de novo crystal generation. *arXiv preprint arXiv:2511.14426*, 2025. URL https://arxiv.org/abs/2511.14426. v2, 28 May 2026.

Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer, 6th edition, 2003. doi: 10.1007/978-3-642-14394-6. URL https://link.springer.com/book/10.1007/978-3-642-14394-6.

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.

Vyas Ramasubramani, Bradley D. Dice, Eric S. Harper, Matthew P. Spellings, Joshua A. Anderson, and Sharon C. Glotzer. freud: A software suite for high throughput analysis of particle simulation data. *Computer Physics Communications*, 254:107275, 2020. doi: 10.1016/j.cpc.2020.107275. URL https://arxiv.org/abs/1906.06317.

Benjamin Rhodes, Sander Vandenhaute, Vaidotas Šimkus, James Gin, Jonathan Godwin, Tim Duignan, and Mark Neumann. Orb-v3: atomistic simulation at scale. *arXiv preprint arXiv:2504.06231*, 2025. URL https://arxiv.org/abs/2504.06231.

Janosh Riebesell, Rhys E. A. Goodall, Philipp Benner, Yuan Chiang, Bowen Deng, Gerbrand Ceder, Mark Asta, Alpha A. Lee, Anubhav Jain, and Kristin A. Persson. A framework to evaluate machine learning crystal stability predictions. *Nature Machine Intelligence*, 7(6):836–847, 2025. doi: 10.1038/s42256-025-01055-1. URL https://www.nature.com/articles/s42256-025-01055-1.

Martin Siron, Inel Djafar, Ali Ramlaoui, Etienne du Fayette, Amandine Rossello, Edvin Fako, Matthew McDermott, Felix Therrien, Luis Barroso-Luque, Flaviu Cipcigan, Philippe Schwaller, Thomas Wolf, and Alexandre Duval. LeMat-Bulk: Aggregating, and de-duplicating quantum chemistry materials databases. *arXiv preprint arXiv:2511.05178*, 2025. URL https://arxiv.org/abs/2511.05178.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations (ICLR)*, 2021. URL https://arxiv.org/pdf/2011.13456.

Brandon M. Wood, Misko Dzamba, Xiang Fu, Meng Gao, Muhammed Shuaibi, Luis Barroso-Luque, Kareem Abdelmaqsoud, Vahe Gharakhanyan, John R. Kitchin, Daniel S. Levine, Kyle Michel, Anuroop Sriram, Taco Cohen, Abhishek Das, Ammar Rizvi, Sushree Jagriti Sahoo, Zachary W. Ulissi, and C. Lawrence Zitnick. UMA: A family of universal models for atoms. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. URL https://arxiv.org/pdf/2506.23971.

Tian Xie, Xiang Fu, Octavian-Eugen Ganea, Regina Barzilay, and Tommi Jaakkola. Crystal diffusion variational autoencoder for periodic material generation. In *International Conference on Learning Representations (ICLR)*, 2022. URL https://arxiv.org/abs/2110.06197.

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, 2025. doi: 10.1038/s41586-025-08628-5. URL https://www.nature.com/articles/s41586-025-08628-5.

A Voronoi Maximin Tessellation

Figure 5 illustrates the iterative ghost placement procedure. Starting from the base unit cell (a), a 3×3 periodic supercell is constructed and Voronoi-tessellated (b). At each step, a ghost atom is placed at the Voronoi vertex farthest from any existing atom (the largest cavity), the tessellation is recomputed, and the process repeats until the target cardinality $N{=}20$ is reached (c–f). By construction, ghosts occupy the largest interstitial voids, maximizing the minimum ghost–real separation.

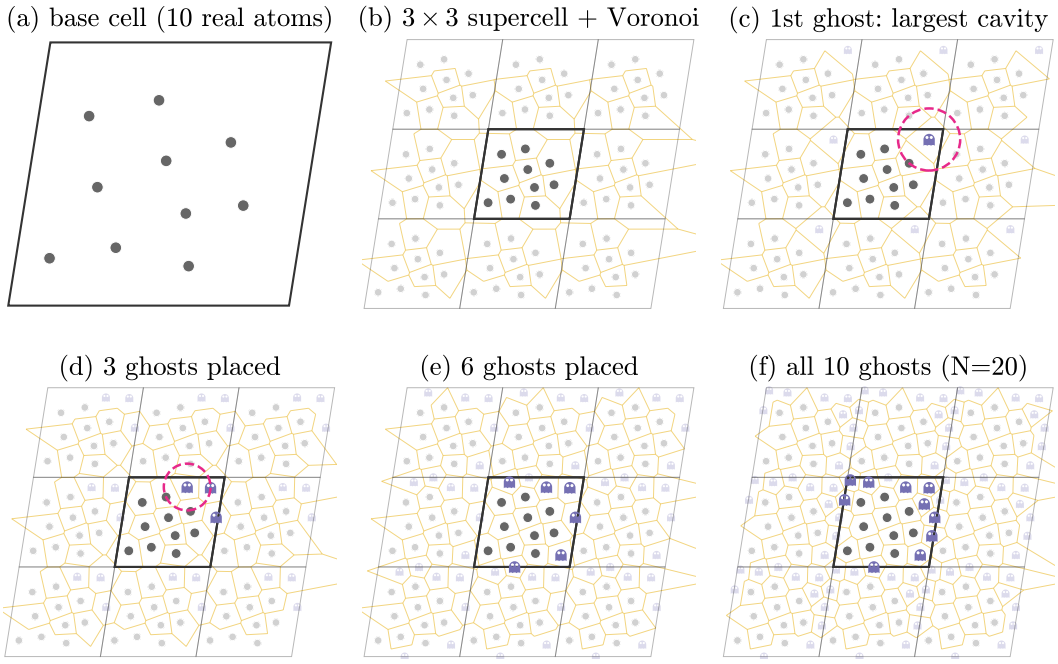


Figure 5: Voronoi maximin ghost placement on a 2D demo cell ($N_{\text{real}}{=}10$, $N{=}20$). (a) Unit cell with real atoms. (b) $3{\times}3$ supercell with Voronoi tessellation. (c) First ghost placed at the largest cavity (dashed circle). (d) Three ghosts. (e) Six ghosts. (f) All ten ghosts.

B Training Data Geometry

We validate that Voronoi ghost placement maintains physically reasonable interatomic separations by comparing it to a random baseline. We take the first $5{,}000$ structures of the Voronoi-ghosted training set, strip their ghosts, and re-place them uniformly at random in fractional coordinates. Figure 6 reports $V^{1/3}$-normalized nearest-neighbor distances over the $4{,}515$ of these structures that contain ghosts and at least two real atoms.

Voronoi placement (top row) produces ghost–real distances with mean 0.366 and minimum 0.244, confirming that ghosts sit in interstitial cavities well separated from real atoms. Random placement (bottom row) yields a lower mean (0.263) and minimum near zero (0.008), with a substantial tail of sub-physical contacts. The worst contact per crystal (panels c, f) shows the difference most clearly: Voronoi concentrates at ∼0.30 with no structure below 0.13, while random placement produces worst contacts near 0.10 with many below 0.05—distances at which atoms would overlap.

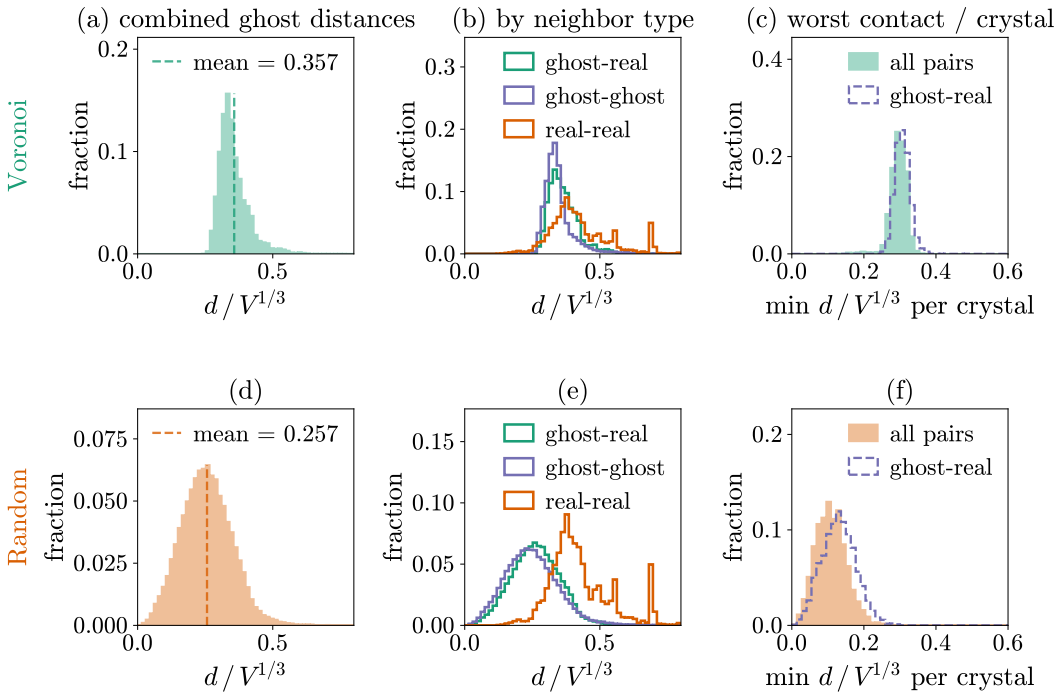


Figure 6: Nearest-neighbor distance distributions ($d/V^{1/3}$) in the training set. **Top:** Voronoi ghost placement. **Bottom:** Random placement on the same structures. (a, d) Combined ghost distances. (b, e) Breakdown by neighbor type. (c, f) Worst contact per crystal.

Table 2: Distance statistics ($d/V^{1/3}$): Voronoi vs. random ghost placement.

Placement	Pair type	Mean	Median	Min	Std
Voronoi	Ghost–real	0.366	0.355	0.244	0.061
	Ghost–ghost	0.349	0.335	0.241	0.068
	Worst/crystal	0.295	0.297	0.131	0.025
Random	Ghost–real	0.263	0.262	0.008	0.092
	Ghost–ghost	0.250	0.243	0.009	0.103
	Worst/crystal	0.109	0.107	0.008	0.042

C Generated Structure Geometry

Figure 7 repeats the analysis of Appendix B on generated structures before ghost stripping: the five Voronoi-model seeds (11,547 structures, 120,615 ghosts) and the randomly ghosted control sampled with the same five seeds and linear-decay schedule (11,586 structures, 122,189 ghosts). As in the training set, structures with no ghosts or fewer than two real atoms are excluded. Each model reproduces the ghost geometry of its own training data: the mean ghost–real distance is 0.365 for the Voronoi model and 0.265 for the random model (training: 0.366 and 0.263), and the mean

ghost–ghost distance is 0.348 and 0.249 (training: 0.349 and 0.250). The mean worst contact per crystal is 0.284 for the Voronoi model and 0.108 for the random model. The low tail of the Voronoi model's all-pairs worst contact comes from real–real pairs; its closest ghost–real contact in any structure is 0.052.

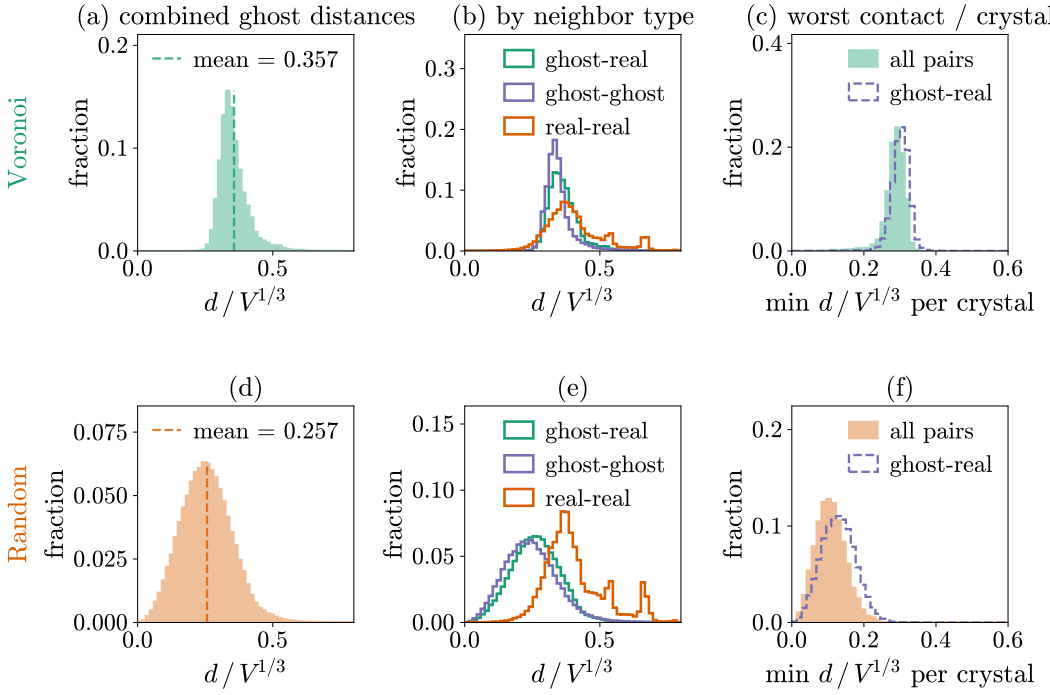


Figure 7: Nearest-neighbor distance distributions ($d/V^{1/3}$) in generated structures, ghosts included. **Top:** Voronoi-trained model, five seeds pooled. **Bottom:** randomly ghosted control, same five seeds and sampler settings. Panels as in Figure 6; unlike the training data, real–real distances differ between rows because each model generates its own real atoms.

D Evaluation Details

We evaluate generated structures with LeMat-GenBench (Betala et al., 2025) using its `comprehensive_multi_mlip_hull` configuration. After ghost atoms are stripped, every structure goes through a mandatory validity filter, and only structures that pass all three checks below are scored; every rate in Table 1 is a fraction of these valid structures.

Charge neutrality. A structure passes if it is metallic or if its atoms can be assigned oxidation states that sum to zero. It counts as metallic if a composition-based metallicity score exceeds 0.7 or if all of its bond-valence sums vanish. Otherwise, oxidation states are assigned by pymatgen's bond-valence analysis (Ong et al., 2013), and the structure passes if they sum to within 0.1 of zero. If bond-valence analysis cannot assign oxidation states, the benchmark searches for a charge-balanced assignment among the common oxidation states of each element (a reference table distributed with LeMat-GenBench) and, failing that, for any charge-balanced assignment in which more electronegative elements take lower oxidation states.

Minimum interatomic distance. A structure passes if every pair of atoms i and j, taken at their nearest periodic images, is at least $0.5\,(0.7\,\text{Å} + r_i + r_j)$ apart, where r_i is the atomic radius of element i tabulated in pymatgen.

Physical plausibility. A structure must pass five checks. Its mass density must lie between 0.01 and 25 g cm^{-3}, and its number density between 10^{-5} and 0.5 atoms per Å^3. Its unit cell must have a

volume above 1 Å^3, lattice lengths between 1 and 100 Å, and angles strictly between $0°$ and $180°$. Finally, pymatgen must be able to write the structure to a CIF file and assign it a space group.

Stability and the convex hull. For the elements in a structure, the formation energies of the reference compounds are plotted against composition, and the convex hull connects the lowest-energy ones: at every composition it gives the lowest energy that a single compound, or a combination of compounds, can reach. A structure's energy above the hull, e_{hull}, is how far its formation energy lies above the hull at its composition, which equals the energy it would release by decomposing into the most stable compounds. Structures on or below the hull ($e_{\text{hull}} \leq 0$) are thermodynamically stable at zero temperature. After relaxation with UMA, each of Orb-v3, MACE-MP, and UMA computes e_{hull} against its own reference hull, and the benchmark averages the three values. A structure is stable if this average is at most 0 and metastable if it lies in $(0, 0.1]$ eV/atom.

E Energy Above Hull Distribution

Figure 8 shows the per-structure energy above hull (e_{hull}) for the Voronoi-ghosted model with the linear-decay schedule, pooled across all five seeds (n=14,987). The distribution peaks near zero with a long right tail. Of valid structures, 6.9% are thermodynamically stable ($e_{\text{hull}} \leq 0$) and 53.0% have $e_{\text{hull}} \leq 0.1$ eV/atom (stable or metastable), with mean 0.123 and median 0.091 eV/atom.

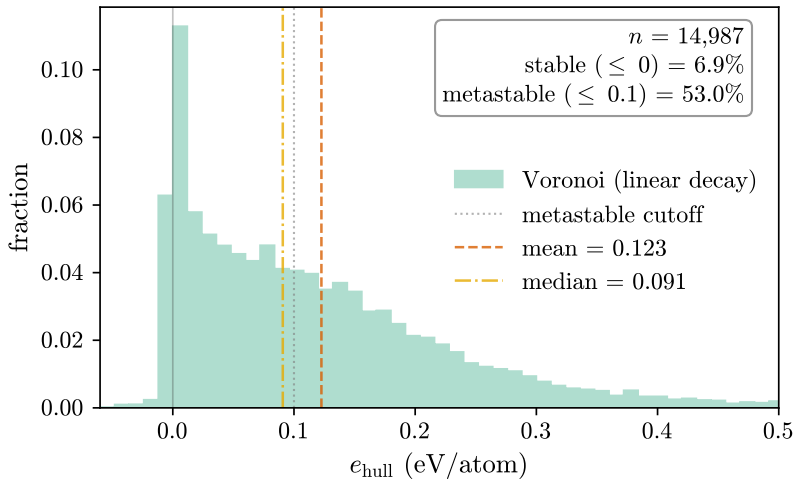


Figure 8: Energy above hull distribution for the Voronoi-ghosted model (linear-decay schedule, five seeds pooled). Dotted line marks the metastable cutoff at 0.1 eV/atom.

F Crystal Structure Prediction

We evaluate crystal structure prediction (CSP) by generating structures conditioned on a target composition and measuring how closely they reconstruct known crystals. Table 3 reports match rates and root-mean-square errors under three settings: over all generated structures, over the validity-filtered subset, and including ghost atoms in the distance calculation (before stripping). The match rate of 69.1% over all structures indicates the model can reconstruct known crystals from composition alone; the slight drop to 62.7% under validity filtering reflects that some matched structures fail the physical plausibility checks. The ghost-inclusive column measures reconstruction quality before ghost stripping, where the additional atoms increase both RMSE and cRMSE, confirming that ghost positions are approximate interstitial estimates rather than exact reconstructions.

Table 3: CSP match rate and reconstruction error.

	All structures	Valid only	Ghost-inclusive
Match rate	0.691	0.627	0.552
RMSE (Å)	0.082	0.080	0.179
cRMSE (Å)	0.211	0.237	0.323

G RETESSELLATION SELF-CONSISTENCY

To test whether generated ghost positions are consistent with the Voronoi placement rule, we strip ghosts from generated crystals, re-apply Voronoi maximin tessellation, and compare the resulting ghost–real distances. Figure 9 shows the $V^{1/3}$-normalized ghost–real distance distributions for model-generated ghosts versus re-tessellated ghosts, pooled across all five seeds (11,593 structures). The distributions are nearly identical (median 0.357 vs 0.359, a 0.6% difference), suggesting that the model has learned the Voronoi placement geometry.

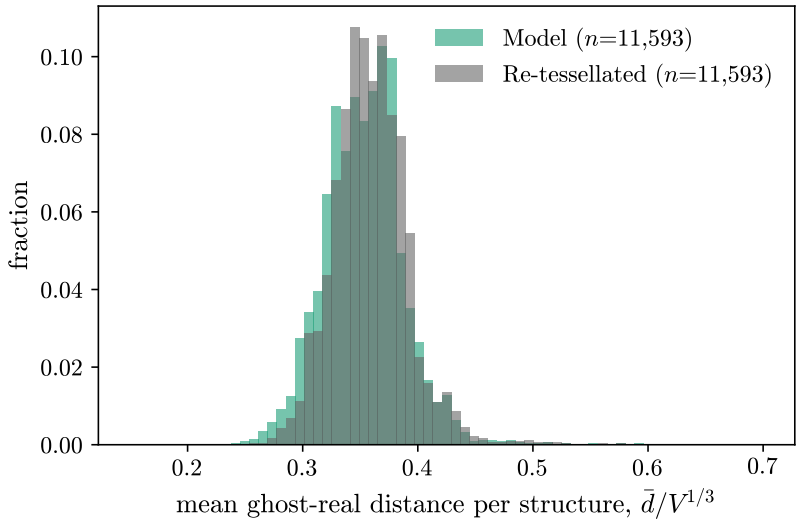


Figure 9: Ghost–real distance distributions: model-generated vs. re-tessellated. Pooled over five seeds (n=11,593).

Table 4: Self-consistency, $V^{1/3}$-normalized median ghost–real distance from the model's own ghosts vs. Voronoi re-tessellation of the same real-atom cells. Structures are the generated cells that contain at least one ghost and one real atom: of the 12,500 generated cells (2,500 per seed), 904 have all 20 slots real and 3 have none.

Seed	Structures	Ghosts	Model	Re-tessellated	Diff
1	2,286	23,711	0.357	0.358	+0.4%
2	2,320	24,381	0.356	0.359	+0.7%
3	2,323	24,634	0.357	0.360	+0.7%
4	2,305	24,132	0.357	0.359	+0.6%
5	2,359	24,631	0.357	0.358	+0.5%
All	11,593	121,489	0.357	0.359	+0.6%

H Inference Trajectory

Figure 10 shows five frames from a single sampling trajectory, illustrating how species identity resolves during the generative process. At t=0, all 20 atoms are masked and at uniformly random positions. By t=0.5, the model has resolved most real atom identities (O, V) and begun assigning ghost labels. At t=1, the final crystal contains 10 real atoms (ZrV_2O_7) and 10 ghost atoms occupying interstitial positions.

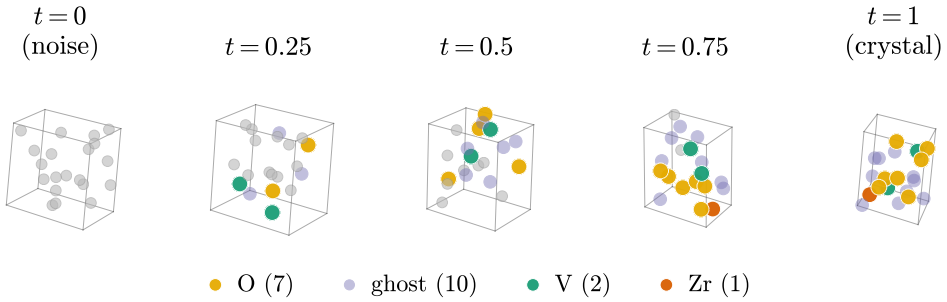


Figure 10: Inference trajectory, 3D snapshots at five timesteps showing species resolution during sampling. Real atoms are colored by element; ghost atoms (purple, transparent) occupy interstitial positions.

I Training Configuration

Table 5: Training and model configuration.

Setting	**Value**
Encoder	CSPNet (12.4M params) on the fully connected graph: hidden dim 512, time-embedding dim 256, 6 message-passing layers, ghost-species channel
Species interpolant	discrete flow matching with masking, $\nu = 0.189$
Coordinate interpolant	periodic linear, $\gamma(t) = \sqrt{a\,t(1-t)}$ with $a = 0.018$; SDE sampling
Lattice interpolant	linear, $\gamma = 0$; ODE sampling
Lattice prior	log-normal lengths (FlowMM fit to MP-20), angles uniform on $[60°, 120°]$
Optimizer	AdamW, lr 1.97×10^{-4}, weight decay 3.1×10^{-4}
LR schedule	cosine annealing ($T_{\max}$=2000, $\eta_{\min}$=10^{-7})
Batch size	256
Training	3000 epochs; best-validation checkpoint (epoch 1963)
Gradient clipping	0.5 (by value); precision 32-bit
Velocity annealing s (coord / lattice)	6.33 / 1.07
Integration steps (sampling)	710 on $t \in [10^{-3}, 1 - 10^{-3}]$

J INFERENCE-TIME NOISE SCHEDULES

Table 6: Inference noise schedules $\epsilon(t)$ considered. Here $c > 0$ is a scale factor; the bell schedule has center μ and width w; the exponential schedule has decay rate λ; the early-pulse schedule peaks at t_0 with half-width δ; and the concave schedule has exponent p. In every case the scale c is fixed by holding the quadratic variation, or injected noise budget, $V_\epsilon = \int_0^1 2\epsilon(t)\,dt$ (Equation 5) constant across schedules. In practice we take $t \in [10^{-3}, 1 - 10^{-3}]$, which keeps the inverse-time schedules finite.

Schedule	$\boldsymbol{\epsilon(t)}$
bell (inherited)	$\dfrac{c}{\left(1 + e^{-(t-\mu)/w}\right)\left(1 + e^{-(1-\mu-t)/w}\right)}$
linear decay	$c\,(1 - t)$
exponential	$c\,e^{-\lambda t}$
inverse-time	$c\,\dfrac{1-t}{t}$
square-root inverse-time	$c\,\dfrac{1-t}{\sqrt{t}}$
early-pulse polynomial	$\max\left(0,\ c\left[1 - \left(\dfrac{t - t_0}{\delta}\right)^2\right]\right)$
concave (late-loaded)	$c\,(1 - t^{\,p})$