INFORMATION PRESENT, PATHWAY SHUT: A CONDITIONAL FLOW MODEL THAT NEVER USED ITS DIFFRACTION PATTERN

Anonymous authors
Paper under double-blind review

ABSTRACT

When conditioning fails to improve a conditional generative model, the null is usually read as a limit of the data or the task. We show that it can instead mean the model never used its condition, and we give inexpensive diagnostics that tell the two apart. Our case study is rigid-body flow matching for molecular crystal structure determination from powder X-ray diffraction (PXRD). The pattern enters through zero-initialised, $\tanh$-gated cross-attention. In every trained checkpoint the gate stayed near 5×10^{-5}. The pattern changed the predicted velocity by about one part in a million, classifier-free guidance was an identity, and our earlier guided-versus-unguided comparisons had been measuring prior-sampling noise. The gate's gradient was not vanishing. Forcing the gate fully open changed the loss identically for a crystal's own pattern and for another crystal's, because the attention weights behind the closed gate had never learned. An ideal-observer information-ceiling test shows that the pattern resolves 0.05 Å molecular displacements in all 200 held-out crystals tested, even at the lowest training count level. The flat conditioning is therefore a learning failure, not an information limit. We then test, under a pre-registered protocol with a deranged-pattern control, whether fine-tuning with the pathway forced open learns crystal-specific use of the pattern. It does not: even with the pathway open and the information present, the own-pattern advantage is confidently below a small effect (rank-biserial $r < 0.15$ on every head) and indistinguishable from the deranged-pattern control. That confident null is the full-corpus analysis; on the subset our representation defect provably cannot reach (Sohncke groups, $n = 1{,}116$) the pre-registered outcome is instead inconclusive, with the orientation bound at 0.152. We also identify an orientation target that is unlearnable in principle and a rigid-body representation defect that rebuilds improper symmetry copies with the wrong hand and, through a rigidity filter, leaves a corpus in which the two most common organic space groups are under-represented by close to an order of magnitude. We distil the diagnostics into a checklist for conditional generative models in the sciences.

1 INTRODUCTION

A conditional generator that ignores its condition produces the same flat result as a conditional generator whose condition carries too little information. The two failures call for opposite responses: the first for a change to training, the second for a change of task or data. In our experience the second reading is the default one, because it is the one a negative result invites. This paper is about a case where that default was wrong for two months, and about the cheap measurements that would have shown it on day one.

The case is structure determination of molecular crystals from powder X-ray diffraction (PXRD). The $3\mathrm{D} \to 1\mathrm{D}$ projection is lossy, and it is natural to suspect that the pattern underdetermines where rigid molecules sit in the cell. Riemannian flow matching (Lipman et al., 2023; Chen & Lipman, 2024) is an efficient generator for periodic systems (Miller et al., 2024; Jiao et al., 2023; Xie et al., 2022), and a rigid-body factorisation — freeze each conformer, generate the lattice, the fractional centroids and the per-molecule orientations on $\mathrm{SO}(3)$ — suits molecular crystals. We built such a model and

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.

conditioned it on the pattern through Flamingo-style (Alayrac et al., 2022) gated cross-attention. Guided and unguided sampling were statistically indistinguishable in every configuration we trained.

Our contributions are:

- **The pathway never opened** (§5). In all five trained cross-attention checkpoints the zero-initialised gate is $|\tanh g| \lesssim 10^{-4}$, guidance is an identity, and the gate's gradient is alive. Forcing the gate open reveals nothing crystal-specific behind it.
- **The information is present** (§6). An ideal-observer test puts a 0.05 Å centroid displacement two orders of magnitude above counting noise at the lowest training counts, for every crystal tested.
- **A pre-registered test of the remedy** (§7): a fine-tune with the gate forced open against an identical fine-tune on deranged patterns, with a teacher-forced endpoint, a correction family and outcome rules fixed and deposited before either run started. The outcome is a confident null: the opened pathway carries a pattern-generic signal and no crystal-specific one.
- **Two defects upstream of conditioning** (§4, §8): an orientation target with an analytic loss floor that the model sat on, and a rigid-body factorisation that rebuilds improper symmetry copies with the wrong hand and, through a rigidity filter, removes most centrosymmetric crystals from the corpus.

2 Related work

Generative crystal structure prediction. Diffusion and flow models generate the lattice and atom positions of periodic crystals jointly (Xie et al., 2022; Jiao et al., 2023; Miller et al., 2024), or an asymmetric unit plus symmetry operations (Levy et al., 2025). Rigid-body flows move whole building blocks: metal nodes and linkers (Kim et al., 2025), or molecules with given conformers over the lattice, torus centroids and $\mathrm{SO}(3)$ orientations (Zeng et al., 2026), the factorisation we use. Organic crystals have also been generated from molecular graphs (Vahediahmar et al., 2026; Subramanian et al., 2026). None of these conditions on diffraction.

Diffraction-conditioned structure determination. Generators conditioned on powder patterns include XtalNet (Lai et al., 2025), PXRDGen (Li et al., 2025), deCIFer (Johansen et al., 2026), PXRDnet (Guo et al., 2025), Crystalyze (Riesel et al., 2024), XRDiff (Segal et al., 2026) and RealPXRD-Solver (Li et al., 2026). Their benchmarks are inorganic or framework materials, and headline rates often assume side information: deCIFer's 94.5% match rate fixes composition and space group, against 5.0% from the pattern alone. Closest to our question, XRDiff initialises its FiLM conditioning to the identity and restricts training to polymorphs because composition alone could let a model "largely ignore" the pattern, and ED-CSP (Poloudenny et al., 2026), for electron diffraction, tests signal use directly by removing or swapping the diffraction input at test time.

Gated conditioning and ignored conditions. Zero-initialised gates make a new conditioning branch an exact identity at initialisation: Flamingo's $\tanh$ gates (Alayrac et al., 2022), ControlNet's zero convolutions (Zhang et al., 2023) and LLaMA-Adapter's gated attention (Zhang et al., 2024). In the reported cases the branch is taken up — Flamingo's gates grow quickly from zero in every layer, and ControlNet shows a "sudden convergence" to the condition, usually within 10k steps. A decoder that ignores its latent is a known stable equilibrium in VAEs (Bowman et al., 2016) and diffusion autoencoders (Wang et al., 2023). A flat guidance sweep alone does not show a condition is unused, since guidance matters mainly at intermediate noise (Kynkäänniemi et al., 2024); we therefore measure the gate and the field directly.

What is new here. That a diffraction-conditioned generator might ignore its condition is not a new worry, and one test for it is already published: ED-CSP (Poloudenny et al., 2026) removes or swaps the diffraction input at test time, which is our deranged-pattern manipulation, and XRDiff (Segal et al., 2026) mitigates the concern by construction with identity-initialised FiLM and a polymorph-only training set. What we add is (i) reading the conditioning gate itself, a single scalar that settles the question before any sampling is run, (ii) measuring the velocity field's sensitivity to the condition in float64, where float32 inflates the ratio by 2–5$\times$ and can hide a dead pathway, (iii) an ideal-observer information ceiling that bounds what the condition could supply, which is what separates a shut pathway from an uninformative condition, and (iv) a pre-registered, controlled test of the remedy

rather than a post-hoc one. Items (i)–(iii) are diagnostics a reader can run on a trained checkpoint in minutes; the swap test alone cannot distinguish our two failure modes, because a model that ignores its condition and a condition that carries nothing both give a null swap.

3 SETUP

Model. A rigid-body flow over three manifolds: the lattice as a log-metric with the space-group family's constraints masked, fractional molecular centroids on the torus, and per-molecule orientations on $\mathrm{SO}(3)$ in the relative gauge of §4. Molecules are embedded by an E(n)-equivariant network over their fixed conformers and mixed by 4 attention layers (width 192, 5.69M parameters in total), with self-conditioning and per-structure optimal-transport coupling of the prior. A 1D residual convolutional encoder, initialised from a contrastively pretrained pattern–structure encoder, maps the 2,251-point pattern to a sequence of 141 tokens. Molecule tokens x attend to that sequence, and the result is added through a scalar gate initialised at zero,

$$x \leftarrow x + \tanh(g) \cdot \mathrm{MHA}\big(\mathrm{LN}(x),\, \mathrm{LN}(h),\, \mathrm{LN}(h)\big), \qquad g_0 = 0, \tag{1}$$

so the conditioned model is exactly the unconditioned backbone at initialisation. With probability 0.15 the condition is dropped during training, and sampling uses classifier-free guidance with weight w (Ho & Salimans, 2022); $w = 0$ is the unconditioned field.

Objective. Conditional flow matching (Lipman et al., 2023): for $t \sim \mathcal{U}(0, 1)$, x_t interpolates a prior draw and the data along the geodesics of each manifold, and each head regresses its velocity with a squared error, masked over padded molecules and weighted 1:8:1 (lattice : centroid : orientation). The prior is a log-normal cell around 10 Å^3 per atom (log-volume and shape deviations with standard deviation 0.3, family constraints masked), uniform centroids on the torus and uniform orientations. Each crystal's prior is drawn once and reused, and prior molecules are assigned to data molecules by a per-structure optimal-transport matching on torus distance. With probability 0.5 the model also receives its own detached terminal estimate (self-conditioning).

Training budget. All configurations use batch 8 — the memory ceiling for this backbone on a 24 GB card, since the molecule encoder's dense pairwise messages scale with the padded atom count — with AdamW, global gradient clipping, 500 warmup steps and cosine decay. The base models of Table 2 run 40,000 steps at learning rate 2×10^{-4}, best checkpoint at 37,500: about 5.0 epochs over the 59,727 training structures. The fine-tunes of §7 run 15,000 steps at 10^{-4}, about 2.0 epochs. These are small budgets by the standards of the gated-conditioning literature we compare against, and §9 states what that does and does not license.

Data. Molecular crystals from the Cambridge Structural Database (Groom et al., 2016). Of 246,829 candidate entries, 66,286 pass the rigid-body filters (§8 shows what the rigidity filter actually removes). They are split 59,727 train and 6,559 validation by CSD refcode family, so that all determinations of a compound fall on one side. CSD structures are licence-restricted; we release refcode manifests, configurations and code, not structures.

Conditioning simulation. Training renders each pattern under a fresh random degradation: isotropic strain, zero offset, Caglioti and Scherrer broadening (20–500 nm domains), preferred orientation, a Chebyshev background, an amorphous hump and Poisson counting noise (2×10^3–10^6 peak counts). Ranges are in Appendix A.

Evaluation. De-novo generation, $k = 10$ samples per crystal, 100 sampler steps and a rigid finisher; match@10 is a `StructureMatcher` (Ong et al., 2013) hit at stol $= 0.5$. Conditioning was assessed by paired, Holm-corrected (Holm, 1979) comparisons of graded metrics (R_{wp}, pattern cosine, cell-length error) between guided and unguided sampling of the same crystals.

4 THE PACKING HEADS WERE AT THEIR FLOORS

Our first model recovered 0 of 160 held-out structures. Installing the ground-truth unit cell into its generated packings, with the generated frame preserved, still gave $0/10$ matches, while installing the ground-truth packing was the only substitution that moved the RMS distance (Appendix B). A per-head decomposition on held-out data agrees (Table 1). Scored against the no-information floor that matters — the best constant predictor, not the zero predictor — no head was learning usefully:

the trained lattice loss sits 12.0% *above* its best constant and the centroid loss 2.6% above, and only the orientation head beats its floor, by 0.4%. The lattice head's 91.6% variance explained is against the zero predictor, a far weaker baseline, and almost all of it is the constant. We report both columns because the difference between them is the point: a head can look like it dominates the loss and still carry no information.

Table 1: Per-head loss against floors for the first model, held-out data (25 batches $\times$ 8 crystals).

Head	zero-predictor	best constant	trained	variance explained vs.	
				zero-predictor	best constant
lattice	1.2336	0.0928	0.1039	91.6%	−12.0%
centroid	0.1189	0.1160	0.1190	−0.1%	−2.6%
orientation	5.2002	5.0265	5.0063	3.7%	+0.4%

The orientation target was unlearnable in principle. For a uniformly distributed rotation, the expected squared geodesic residual of the zero predictor is

$$\mathbb{E}\|u_R\|^2 = \int_0^\pi \theta^2 \, \frac{1-\cos\theta}{\pi} \, \mathrm{d}\theta = \frac{\pi^2}{3} + 2 = 5.290, \tag{2}$$

and the model ended at ≈ 4.7. The absolute pose of copy m factorises as $R_m = \mathrm{rot}(g_m)\, R_{\text{asym}}$: the rotation of the generating space-group operation, which is learnable, times the pose of the asymmetric unit, which is a per-crystal free rotation that nothing in the packing or the pattern determines. Its dispersion across the corpus is $134.7°$, close to uniform, so the minimum-risk prediction was nothing. Referring orientations to the asymmetric unit removes that component. The orientation loss drops to 2.31, and held-out match@10 moves from $0/160$ to $5/160$ (Fisher exact two-sided $p = 0.061$; suggestive, not significant). Two further interventions — conditioning on the generating operation (not de-novo) and an $8\times$ centroid loss weight — gave $4/160$ and $7/160$, neither distinguishable from $5/160$ (Appendix B). The centroid head remained near its floor.

With the target repaired, the obvious remaining lever was the pattern: where the molecules sit is exactly what the diffraction intensities encode. Every paired conditioning gate we ran on the repaired models was flat.

5 The conditioning pathway never opened

Table 2 reads one scalar from each trained cross-attention checkpoint: the gate of Eq. 1.

Table 2: The trained cross-attention gate. Every configuration starts at $g = 0$.

Configuration	steps	$\tanh(g)$
relative gauge, $8\times$ centroid weight (base for §7)	37,500	-5.2×10^{-5}
relative gauge (best)	37,500	-9.2×10^{-5}
relative gauge (last)	40,000	-9.3×10^{-5}
+ space-group cosets	40,000	$+6.6 \times 10^{-5}$
absolute gauge	60,000	-4.2×10^{-5}

Consequence: guidance is an identity. On training batches the gated cross-attention output is a median 6.3×10^{-7} of the token-stream norm; ungated it would be $\approx 1.2\%$. Re-running the paired sampling evaluation with common random numbers — the same seed re-applied at the start of every guidance arm — on the base checkpoint (8 validation crystals, a short sampler) gives bit-identical per-crystal R_{wp}, cell-length and angle errors for $w = 0, 1, 2, 3$. Earlier gates drew each arm's prior independently, so every guided-versus-unguided comparison on these checkpoints compared prior draws with prior draws. That includes a Holm-significant cell-length "win" ($p = 3.5 \times 10^{-4}$) on the last configuration of Table 2, which had already failed to replicate after the gauge repair. Whatever produced it, it was not the pattern.

Not an optimiser failure. The gate is an ordinary AdamW parameter. Over six training batches its gradient after global clipping had median magnitude 4.3×10^{-4} — four orders of magnitude above

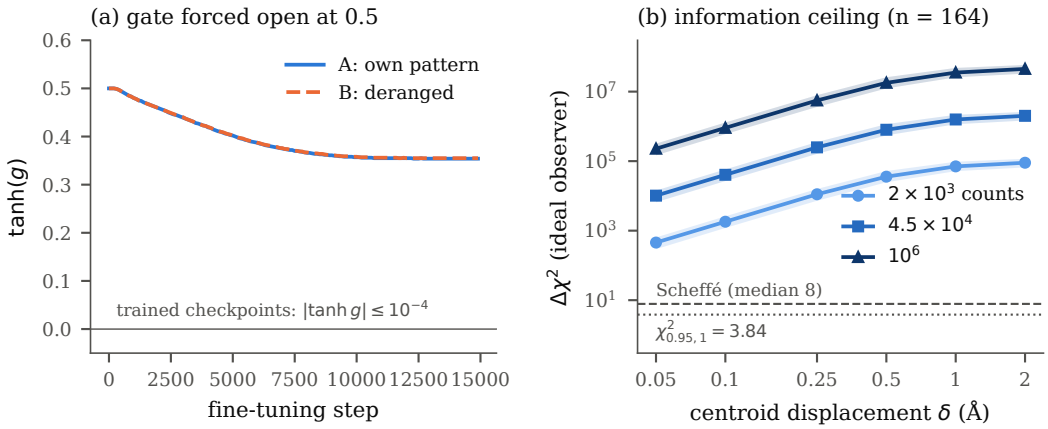


Figure 1: **(a)** The cross-attention gate during the two fine-tunes of §7, re-initialised at $\tanh g = 0.5$. The own-pattern run (A) and the deranged-pattern control (B) trace the same path, ending at 0.354 and 0.355; the grey band marks the trained checkpoints of Table 2. **(b)** Information ceiling (§6): ideal-observer $\Delta\chi^2$, scale profiled out, for symmetry-consistent centroid displacements. Median and interquartile range over (crystal, direction) pairs at three training count levels, against the single-parameter and Scheffé thresholds.

Adam's $\epsilon = 10^{-8}$ — with mixed signs, and the gate stayed within 10^{-4} of zero for 37,500 steps. The gradient reaching the attention weights is scaled by $\tanh(g)$ (Eq. 1), $\sim 5 \times 10^{-5}$ in total, and the encoder's is $\sim 10^{-5}$. The branch cannot learn while the gate is shut, and the gate receives no consistent push while the branch has learned nothing: a chicken-and-egg equilibrium at zero.

Nothing behind the gate. If the attention weights had learned something crystal-specific that the gate merely suppressed, forcing the gate open would help a crystal's own pattern more than another crystal's. Table 3 forces it open on the trained weights with every random draw shared between settings.

Table 3: Loss change when the trained gate is forced open, relative to its trained value: mean ± s.e. over 8 batches × 8 crystals of the 1,127-crystal pilot corpus, float64 on CPU; *own* vs a *deranged* pattern assignment.

$\tanh(g)$	pattern	total	lattice	centroid
-0.95	own	-0.0002 ± 0.0039	-0.0015 ± 0.0005	0.0000
-0.95	deranged	-0.0003 ± 0.0042	-0.0015 ± 0.0005	0.0000
$+0.95$	own	$+0.0014 \pm 0.0041$	$+0.0018 \pm 0.0005$	0.0000
$+0.95$	deranged	$+0.0014 \pm 0.0044$	$+0.0018 \pm 0.0005$	0.0000

Own minus deranged is $+0.00006 \pm 0.00041$ at $\tanh g = +0.95$ and -0.00006 ± 0.00046 at -0.95. Fully opening the gate changes the loss by $\lesssim 0.1\%$, identically for both patterns, and there is no restoring force toward zero. Directly at the velocity field, with 32 training crystals in float64 at $t \in \{0.1, \ldots, 0.9\}$, the median relative change $\|v_{\text{own}} - v_{\text{null}}\|/\|v_{\text{own}}\|$ is 4×10^{-7}–3×10^{-6} for every head. (Float32 cannot resolve differences this small: sensitivity ratios read 2–5× larger.)

A masked second defect. Evaluation had also rendered the conditioning pattern at a clean operating point outside the training support in two coordinates (domain size 10^4 nm against 20–500 nm; noiseless against $\geq 2 \times 10^3$ counts). With the pathway shut this could not matter, and identical per-crystal scores under the clean and a train-matched operating point confirm that it did not. We fixed it anyway; it would matter to any model that reads the pattern.

6 THE PATTERN CONTAINS THE INFORMATION

A shut pathway explains the flat gates, but it does not show the pattern was worth reading. We measure an information ceiling. For 200 validation crystals (positions disjoint from every confirmatory crystal), we take the true lattice, orientations and profile parameters as known, displace the molecular centroids by δ in a random symmetry-consistent direction (8 directions per crystal and δ), and ask whether an ideal observer distinguishes the two patterns at the training count levels. The statistic is $\Delta\chi^2$ in count space with the scale fitted out, against the single-parameter 3.84 threshold ($\chi^2_{0.95}$, 1 d.o.f.) and a Scheffé-type simultaneous threshold $\chi^2_{0.95}(k_{\text{free}})$ over the crystal's free centroid subspace, which is the conservative bar: a displacement below it lies inside the joint 95% confidence region of all k_{free} centroid degrees of freedom and so is not resolvable by any search over centroid configurations. Across the 200 crystals k_{free} ranges from 2 to 17 (median 3), so the Scheffé threshold ranges from 5.99 to 27.59 (median 7.81). All detection fractions below are against *both* thresholds.

Table 4: Information ceiling, symmetry-consistent displacements, TYPICAL profile parameters. The three $\Delta\chi^2$ columns are the low, typical and high training count levels (2×10^3, 4.5×10^4, 10^6 peak counts). Detection is the fraction of (crystal, direction) pairs above both thresholds.

δ (Å)	noiseless R_{wp} [IQR]	median $\Delta\chi^2$ at counts		
		low	typical	high
0.05	0.042 [0.036, 0.049]	455	1.0×10^4	2.3×10^5
0.10	0.085	1.8×10^3	4.0×10^4	9.0×10^5
0.25	0.21	1.1×10^4	2.5×10^5	5.6×10^6
detected (both thresholds)		100%	100%	100%

A 0.05 Å displacement is detected for every crystal and direction, including at the training extremes of domain size and in every molecule-count bucket (2–16). A cruder observer that compares two noisy patterns by R_{wp} needs $\delta \approx 0.25$ Å at the lowest counts. This is a ceiling — the cell, orientations and profile are given — but the model's placement errors are orders of magnitude larger than 0.05 Å (a normalised `StructureMatcher` RMS near 1, where a match needs about 0.5). The pattern determines molecule placement far more finely than the model uses it, so the flat conditioning is a learning failure rather than an information limit.

7 A PRE-REGISTERED TEST: DOES AN OPENED PATHWAY LEARN?

Sections 5–6 predict that a model forced to route through an open pathway should learn crystal-specific use of the pattern. That prediction can fail: training might close the gate again, or learn only pattern-generic features. Because a post-hoc test of our own prediction would be weak evidence, we fixed the protocol in advance and deposited it in a public archive with restricted access before either model was trained (the identifier is withheld for review and available to the Area Chairs).

Models. Both runs fine-tune the first checkpoint of Table 2 for 15,000 steps with the gate re-initialised to $\tanh g = 0.5$, batch 8, learning rate 10^{-4} with warmup and cosine decay, and the same seed on the same device type, so they see identical batches and identical time, prior and condition-dropout draws. **A** sees each crystal's own pattern. **B**, the negative control, sees a random in-batch derangement: every crystal gets another crystal's pattern, so B cannot learn crystal-specific use. The final checkpoints are evaluated, not the lowest-loss ones.

Endpoint. Sampling-based metrics are noisy and indirect, so the primary endpoint is teacher-forced. For each of 1,600 held-out crystals and each head h, at identical x_t for $t \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, we compute the flow-matching loss with the crystal's own pattern and with a seeded wrong partner from a different compound family, and take the per-crystal gap $g_h = \text{mean}_t[\mathcal{L}_{\text{wrong}} - \mathcal{L}_{\text{own}}]$. Positive means the own pattern helps.

Decision rule. For model A, a two-sided Wilcoxon signed-rank test on g_h per head, Holm-corrected across the three heads. A head is SUPPORTED if Holm $p < 0.05$, the matched-pairs rank-biserial $r \geq 0.15$, and the median gap is positive. The same test on B is a validity check: if B is SUPPORTED on a head, the endpoint carries a crystal-specific signal that did not come from learned pattern use,

and A's result on that head is invalid. Outcomes were fixed in advance: **O1**, valid support on the centroid head; **O2**, support on the lattice or orientation head only; **O3**, a confident null, with every head's simultaneous bootstrap upper bound on r below 0.15; **O4**, otherwise. Because a wrong partner usually differs in space group and molecule count, O1 is described as *placement* use only if it survives a partner matched on both. A Sohncke-only reanalysis, unaffected by the defect of §8, is reported regardless of outcome.

Result: O3, a confident null. Both runs completed all $15{,}000$ steps with no skipped steps, and neither gate re-closed: $\tanh g = 0.354$ for A and 0.355 for B (Fig. 1a).

The gate did, however, move *down* from its 0.5 initialisation, and by the same amount in both arms. The gate sits in the single AdamW parameter group (weight decay 0.01), but decoupled decay under this learning-rate schedule would shrink $\tanh g$ only from 0.500 to 0.497 over $15{,}000$ steps — 0.6% of the observed 29% drop. The descent is therefore gradient-driven, and it is indistinguishable between a run that sees informative patterns and a run that sees scrambled ones. Whatever the gate's gradient is tracking, it is not the content of the condition. We read this as the same equilibrium seen from the other side: with nothing crystal-specific behind the pathway, the objective prefers a smaller contribution from it, regardless of which pattern is supplied. No head of A is supported, and every simultaneous upper bound on r lies below 0.15 (Table 5). B is null in the same way, and the per-crystal difference between A's and B's gaps has $|r| \le 0.011$ ($p \ge 0.70$) on every head. The median gaps are 10^{-9}–10^{-7}, against losses of order one. Against a partner matched on space group and molecule count ($n = 1{,}581$) the outcome is again O3. Restricted to Sohncke-group crystals ($n = 1{,}116$), which the defect of §8 cannot affect, it is O4: no head is supported, but the orientation head's upper bound is 0.152, just above the floor.

Table 5: Pre-registered open-gate test on $n = 1{,}600$ held-out crystals: gap $\mathcal{L}_{\text{wrong}} - \mathcal{L}_{\text{own}}$ averaged over t; matched-pairs rank-biserial r with its simultaneous (98.3%) bootstrap interval. NULL: no support, upper bound below 0.15.

Model	head	median gap	r [simultaneous CI]	Holm p	verdict
A (own)	centroid	$+1.0 \times 10^{-9}$	$+0.014$ $[-0.058, 0.083]$	0.88	NULL
A (own)	lattice	-2.9×10^{-8}	-0.022 $[-0.092, 0.046]$	0.88	NULL
A (own)	orientation	$+1.2 \times 10^{-7}$	$+0.053$ $[-0.015, 0.122]$	0.19	NULL
B (deranged)	centroid	$+9.1 \times 10^{-8}$	$+0.013$ $[-0.055, 0.084]$	1.00	NULL
B (deranged)	lattice	-3.2×10^{-7}	-0.001 $[-0.071, 0.070]$	1.00	NULL
B (deranged)	orientation	-1.9×10^{-6}	-0.007 $[-0.074, 0.061]$	1.00	NULL

What the open pathway does instead (exploratory). The pathway is not inert: opening it changes the predicted velocity by 0.13–1.4% relative to the null condition, against $\sim 10^{-6}$ with the trained gate. But in A at most 0.32% of that change depends on which crystal's pattern is supplied (median $\|v_{\text{own}} - v_{\text{wrong}}\| / \|v_{\text{own}} - v_{\text{null}}\|$ over heads and t), and against no pattern at all the own pattern moves the loss by $|r| \le 0.10$ everywhere, in A and B alike. B's pathway is more pattern-dependent by this measure (3.5–13%) with an equally null loss gap; we cannot explain that difference. Training learned a pattern-generic contribution and no crystal-specific use.

7.1 Exploratory checks around the pre-registered result

The protocol above is one architecture, one initialisation and one seed, with no positive control showing that this gate and this test *can* return support. The checks below were designed after the outcome of record was known; they use the same endpoint, crystals and decision rule, but they are exploratory and cannot alter O3.

A synthetic positive control (Fig. 2). We built a toy with the paper's gate form $\text{tok} + \tanh(g)\,\delta$, its optimiser recipe and its endpoint: flow matching on a target of 8 informative dimensions plus D_n uninformed ones, conditioned on a 64-dimensional linear view of the informative part at signal-to-noise s, over $s \in \{0, 0.3, 1, 3, 10\}$, $N \in \{512, 4096, 32768\}$ training samples, $D_n \in \{0, 16, 128\}$, three seeds, from a zero and from an open gate, with deranged-training twins (282 runs, CPU). Three things hold. The test returns SUPPORT whenever it should ($r = 1.00$ on every informative, open-gate cell) and never for a deranged twin ($|r| \le 0.10$ over twelve). A zero-initialised gate is *not* generally

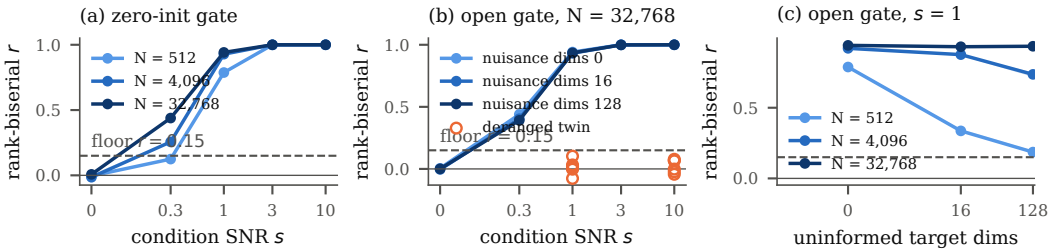


Figure 2: Synthetic sweep with the paper's gate, recipe and test (median, IQR over seeds): r against the condition's signal-to-noise s from **(a)** a zero-initialised and **(b)** an open gate (hollow: deranged-training twins); **(c)** against uninformed target dimensions.

stuck: it learns in 97 of 135 cells, with the boundary in (s, N) — never at $s = 0$, at $s = 0.3$ only with $N \geq 4{,}096$, and almost always at $s \geq 1$ (Fig. 2a). Even with no information, Adam walks a zero-initialised gate to a median $|\tanh g| = 0.46$ within 6,000 steps; the real gate stayed within 10^{-4} of zero for 37,500, so it was held, not left alone. The equilibrium of §5 is therefore a property of this model and data, not of zero-initialised gates in general: the real model's effective per-crystal signal sits below the toy's $s = 0.3$ at $N = 512$, or something the toy lacks holds its gate. The sign consistency of the gate's gradient at initialisation does *not* predict escape (AUC 0.68). Last, the target-variance hypothesis of §9 holds as a small-data effect: with the gate open at $s = 1$, D_n from 0 to 128 lowers the median r from 0.79 to 0.19 at $N = 512$ and from 0.92 to 0.74 at $N = 4{,}096$, but not at $N = 32{,}768$ (0.94; Fig. 2c).

Seeds, a from-scratch open gate, and an un-gated architecture (Table 6). The bootstrap of Table 5 resamples crystals, not training runs. Repeating the A/B pair with seeds 1–3 gives $r \in [-0.039, +0.014]$ on the centroid head and $[-0.029, +0.053]$ on orientation, with every upper bound below 0.15 in all four seeds (record included): the record was not a lucky draw. To answer "the fine-tuned attention had never learned" and "train it longer", a pair trained from scratch with the base recipe of Table 2 (40,000 steps) and the gate open from step 0 is also O3 on both corpora, and its gate *re-closed*: from 0.5 it fell monotonically to $\tanh g = 0.093$ (A) and 0.127 (B) while both reached the base model's training loss. Longer training does not open the pathway; it shuts it. And an un-gated pair with the original additive injection — a single pooled vector added to every token from step 0, where the zero-initialisation equilibrium cannot exist — reaches the base model's loss in 40,000 steps and is O3 on both corpora, its deranged twin null in the same way and $|r| \leq 0.05$ between them. The null does not depend on the gate. Across four seeds, a gate open from the first step, and an architecture with no gate at all, the endpoint returns the same confident null: whatever keeps this model from using its pattern, it is not the seed, the gate, or the training budget.

Table 6: Exploratory pairs, crystals and endpoint of Table 5: own-pattern model's r per head with its simultaneous upper bound (ub; all $< 0.15 \Rightarrow$ O3); outcome full / Sohncke.

pair	$\tanh g$ A / B	centroid r (ub)	lattice r (ub)	orientation r (ub)	outcome
record (seed 0)	0.354 / 0.355	+0.014 (0.08)	−0.022 (0.05)	+0.053 (0.12)	O3 / O4
seed 1	0.326 / 0.324	−0.002 (0.07)	−0.015 (0.06)	+0.034 (0.10)	O3 / O3
seed 2	0.361 / 0.360	+0.006 (0.07)	+0.007 (0.08)	+0.002 (0.07)	O3 / O3
seed 3	0.335 / 0.335	−0.039 (0.03)	−0.009 (0.06)	−0.029 (0.04)	O3 / O3
scratch, gate open, 40k	0.093 / 0.127	−0.021 (0.05)	−0.023 (0.05)	+0.025 (0.09)	O3 / O3
additive, un-gated, 40k	–	0.000 (0.07)	+0.048 (0.12)	+0.015 (0.08)	O3 / O3

8 A REPRESENTATION DEFECT: IMPROPER COPIES TAKE THE WRONG HAND

The rigid-body factorisation stores one reference conformer per species and aligns every copy to it with a proper rotation. A copy related to the reference by an improper operation — inversion, mirror or glide — of a conformer that is not achiral cannot be superimposed that way. If the best proper-rotation residual still passes the rigidity tolerance, the copy is stored and rebuilt with the

reference's hand. Among the 200 crystals of §6, the rebuilt structure has lower symmetry than its label for $22/39$ in $P2_1/c$ and $9/11$ in $P\bar{1}$, and for none of the 127 in the Sohncke groups $P2_12_12_1$ and $P2_1$.

The same mechanism shaped the corpus. Where a conformer is genuinely chiral, the mirror-image copy fails the tolerance and the crystal is rejected as "non-rigid". In random samples of 8,000 kept and 8,000 rejected entries, the keep rate is 72.8–81.4% for the Sohncke groups $P2_1$, $C2$ and $P2_12_12_1$ and 10.4–17.7% for groups with improper operations ($P\bar{1}$, $P2_1/c$, $C2/c$, $Pbca$, Cc, $Pna2_1$). A sevenfold drop between the two classes is not what conformational flexibility looks like. The two most common organic space groups are under-represented by close to an order of magnitude. For the centrosymmetric crystals that remain, the flow target and the conditioning pattern describe different structures, and every match rate in this paper is scored against the rebuilt ground truth. The fix — per-copy handedness, or improper alignments with a reflected conformer — requires rebuilding the corpus and retraining, and is outside this paper.

9 Discussion

A checklist. Each item below would have caught a defect in this paper at negligible cost, and none is specific to crystals.

1. *Log every conditioning gate and assert it moved* before evaluating conditioning.
2. *Evaluate paired conditioning arms with common random numbers.* A dead pathway then shows up as identical arms instead of noise.
3. *Measure the field's sensitivity to the condition*, $\|v_{\text{cond}} - v_{\text{null}}\|/\|v_{\text{null}}\|$, in float64, before pre-registering any effect of conditioning.
4. *Bound the information* with an ideal observer before reading a null as an information limit.
5. *Check converged losses against analytic floors*, which separates "needs more data" from "asks for something that does not exist".
6. *Diff the evaluation-time condition pipeline against training's*, field by field.

What this paper does not show. It does not demonstrate competitive PXRD structure determination: our best de-novo figure is 7 of 160 molecular crystals, scored against partly wrong-handed ground truth. Rates from the inorganic generators of §2 are not comparable, differing in chemistry and often in the side information supplied. It does not show that zero-initialised gates generally fail to open; §7.1 shows they do not, and that ours was held shut by something specific to this model and data. The open-gate null is equally specific: this objective, from this initialisation, over 15,000 steps, did not learn crystal-specific use, and neither did 40,000 steps from scratch with the gate open from step 0, which re-closed it (§7.1). Because the protocol runs guided sampling only after a positive teacher-forced result, it makes no direct claim about generation.

What the null leaves. The closed-gate equilibrium explains why the gate stayed shut, not why an open pathway stays pattern-generic when the pattern determines placement to 0.05 Å. Two hypotheses: at the noise levels that dominate the objective, the per-crystal gain is small next to the flow-matching target's variance, which §7.1 shows can push r below the floor when data are scarce; and for centrosymmetric crystals the pattern and the target disagree (§8), which the Sohncke-only reanalysis only partly controls.

A confound we cannot remove. The head the pattern should drive is the centroid head, and that head never beat its no-information floor (Table 1), before conditioning, after the gauge repair, or under an $8\times$ loss weight. A conditioning benefit is hard to detect on a head that is not learning from *any* input. This does not explain the null on the lattice and orientation heads, which do move, nor why the open pathway is pattern-generic rather than merely weak; but the centroid result is an upper bound on what this endpoint could detect, not a clean test of pattern use. Establishing a head that learns unconditionally before asking whether the condition improves it is the ordering we would use again.

Limitations. Simulated patterns only; one training run per configuration except the four-seed open-gate pair; the sweep of Table 3 and the velocity measurement use small samples (64, 32 and 8 crystals); the information ceiling assumes the cell, orientations and profile are known; and the representation defect of §8 affects the corpus composition and the ground truth of every model here.

REPRODUCIBILITY STATEMENT

The pre-registration, its analysis script and the training and evaluation scripts were deposited before the fine-tunes of §7 started, as the version-controlled files at a named commit with per-file checksums. The analysis of record refuses inputs that are not the declared checkpoints: swapped arms, checkpoints before the final step, runs with more than 1% skipped steps, or runs whose recorded code fingerprint or settings differ. Both fine-tunes ran concurrently on the two RTX 4090 GPUs of one machine, about 40 minutes each, under PyTorch 2.11 with CUDA 12.8.

Because a rented machine receives a source tarball rather than a git checkout, every checkpoint records a *code fingerprint*: a SHA-256 over the line-ending-normalised contents of every Python file in the model and backbone packages. The two fine-tunes carry identical fingerprints, and the analysis of record recomputes and compares them, so a checkpoint cannot be silently swapped for one built by different code. Both runs used seed 0, batch 8, learning rate 10^{-4} with 500 warmup steps and cosine decay, $15{,}000$ steps, and completed with zero skipped steps; the primary endpoint was evaluated in float64 on validation positions $[160, 1760)$. The unit-test suite (102 tests at the time of deposit) covers the derangement, the fingerprint, the decision rule and the refusal paths.

The supplementary bundle contains the model and training code, the pre-registration and its analysis script, the configurations, the exact commands, and the CSD refcode manifests that define the corpus and the splits. CSD structures are licence-restricted and are not redistributed; any licence holder can regenerate the corpus and the simulated patterns from the manifests and the build script.

ETHICS STATEMENT

This work uses licensed crystallographic data under the terms of that licence and redistributes no structures. We are not aware of ethical concerns arising from the methods or results.

AI USE STATEMENT

We used a generative AI coding agent (Fable 5.1, Claude, Anthropic) throughout this work. Per the ICLR 2027 policy we disclose its use by category, in the terms of the submission form.

To aid or polish writing. The agent edited every section of this paper for clarity, length and consistency of notation, including the abstract, the captions and the statements, working from the author's instructions and the committed run outputs.

For retrieval and discovery. The agent searched for and summarised the related work in §2, including the comparison to prior PXRD-conditioned generators. Every claim it produced about a cited paper was then checked against that paper's text in a separate pass: of 96 claims, 73 were confirmed, 6 were corrected and 17 that could not be verified were removed rather than softened.

Not used for generating datasets (the diffraction patterns are physics simulations of deposited structures, §3) or for proving mathematical claims. The author directed the work, reviewed and accepted every design decision, and is responsible for every claim.

REFERENCES

Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katie Millican, Malcolm Reynolds, Roman Ring, Eliza Rutherford, Serkan Cabi, Tengda Han, Zhitao Gong, Sina Samangooei, Marianne Monteiro, Jacob Menick, Sebastian Borgeaud, Andrew Brock, Aida Nematzadeh, Sahand Sharifzadeh, Mikolaj Binkowski, Ricardo Barreira, Oriol Vinyals, Andrew Zisserman, and Karen Simonyan. Flamingo: a visual language model for few-shot learning. In *Advances in Neural Information Processing Systems*, volume 35, pp. 23716–23736, 2022. arXiv:2204.14198.

Samuel R. Bowman, Luke Vilnis, Oriol Vinyals, Andrew Dai, Rafal Jozefowicz, and Samy Bengio. Generating sentences from a continuous space. In *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*, pp. 10–21, 2016. doi: 10.18653/v1/K16-1002.

Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *International Conference on Learning Representations (ICLR)*, 2024. arXiv:2302.03660.

Colin R. Groom, Ian J. Bruno, Matthew P. Lightfoot, and Suzanna C. Ward. The Cambridge Structural Database. *Acta Crystallographica Section B*, 72:171–179, 2016. doi: 10.1107/S2052520616003954.

Gabe Guo, Tristan Luca Saidi, Maxwell W. Terban, Michele Valsecchi, Simon J. L. Billinge, and Hod Lipson. Ab initio structure solutions from nanocrystalline powder diffraction data via diffusion models. *Nature Materials*, 24:1726–1734, 2025. doi: 10.1038/s41563-025-02220-y. arXiv:2406.10796. Model name: PXRDnet. An Author Correction was published 7 Jul 2025.

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

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70, 1979.

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, 2023.

Frederik Lizak Johansen, Ulrik Friis-Jensen, Erik B. Dam, Kirsten M. Ø. Jensen, Rocío Mercado, and Raghavendra Selvan. deCIFer: Crystal structure prediction from powder diffraction data using autoregressive language models. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. URL https://openreview.net/forum?id=LftFQ35l47. arXiv:2502.02189 (v4 = accepted version).

Nayoung Kim, Seongsu Kim, Minsu Kim, Jinkyoo Park, and Sungsoo Ahn. MOFFlow: Flow matching for structure prediction of metal-organic frameworks. In *International Conference on Learning Representations (ICLR)*, 2025. arXiv:2410.17270.

Tuomas Kynkäänniemi, Miika Aittala, Tero Karras, Samuli Laine, Timo Aila, and Jaakko Lehtinen. Applying guidance in a limited interval improves sample and distribution quality in diffusion models. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://openreview.net/forum?id=nAIhvNy15T. arXiv:2404.07724. Page range not verified.

Qingsi Lai, Fanjie Xu, Lin Yao, Zhifeng Gao, Siyuan Liu, Hongshuai Wang, Shuqi Lu, Di He, Liwei Wang, Linfeng Zhang, Cheng Wang, and Guolin Ke. End-to-end crystal structure prediction from powder X-ray diffraction. *Advanced Science*, 12(8):2410722, 2025. doi: 10.1002/advs.202410722. arXiv:2401.03862. Model name: XtalNet.

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 *International Conference on Learning Representations (ICLR)*, 2025. URL https://openreview.net/forum?id=xnssGv9rpW. arXiv:2502.03638.

Qi Li, Rui Jiao, Liming Wu, Tiannian Zhu, Wenbing Huang, Shifeng Jin, Yang Liu, Hongming Weng, and Xiaolong Chen. Powder diffraction crystal structure determination using generative models. *Nature Communications*, 16:7428, 2025. doi: 10.1038/s41467-025-62708-8. arXiv:2409.04727. Model name: PXRDGen.

Qi Li, Mingyu Guo, Rui Jiao, Jing Gao, Fanjie Xu, Haonan Xue, Weixiong Zhang, Wenbing Huang, Junchi Yan, Linfeng Zhang, Cheng Wang, Zhuang Yan, Guolin Ke, Weinan E, Zhiyong Tang, Shifeng Jin, and Lin Yao. Experimental powder X-ray diffraction crystal structure determination with RealPXRD-Solver. *arXiv preprint arXiv:2603.00965*, 2026.

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. arXiv:2210.02747.

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. arXiv:2406.04713.

Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L. Chevrier, Kristin A. Persson, and Gerbrand Ceder. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013. doi: 10.1016/j.commatsci.2012.10.028.

Germain Poloudenny, Arnaud Demortière, and Yaël Frégier. ED-CSP: Crystal structure prediction from electron diffraction. *arXiv preprint arXiv:2608.06448*, 2026. v3, 27 Aug 2026. Electron diffraction, not PXRD.

Eric A. Riesel, Tsach Mackey, Hamed Nilforoshan, Minkai Xu, Catherine K. Badding, Alison B. Altman, Jure Leskovec, and Danna E. Freedman. Crystal structure determination from powder diffraction patterns with generative machine learning. *Journal of the American Chemical Society*, 146(44):30340–30348, 2024. doi: 10.1021/jacs.4c10244. Model name: Crystalyze (as named by citing works; full text not accessed).

Nofit Segal, Mingda Li, Benjamin Kurt Miller, and Rafael Gómez-Bombarelli. XRDiff: Crystal structure prediction from powder X-ray diffraction data using diffusion models. *arXiv preprint arXiv:2606.14003*, 2026.

Akshay Subramanian, Elton Pan, Juno Nam, Maurice Weiler, Shuhui Qu, Cheol Woo Park, Tommi S. Jaakkola, Elsa Olivetti, and Rafael Gómez-Bombarelli. PackFlow: Generative molecular crystal structure prediction via reinforcement learning alignment. *arXiv preprint arXiv:2602.20140*, 2026.

Mohammadmahdi Vahediahmar, Matthew A. McDonald, and Feng Liu. OrgFlow: Generative modeling of organic crystal structures from molecular graphs. *arXiv preprint arXiv:2602.20195*, 2026.

Yingheng Wang, Yair Schiff, Aaron Gokaslan, Weishen Pan, Fei Wang, Christopher De Sa, and Volodymyr Kuleshov. InfoDiffusion: Representation learning using information maximizing diffusion models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 36336–36354, 2023.

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*, 2022.

Cheng Zeng, Harry W. Sullivan, Thomas Egg, Maya M. Martirossyan, Philipp Höllmer, Jirui Jin, Adrian Roitberg, Richard G. Hennig, Stefano Martiniani, Ellad B. Tadmor, and Mingjie Liu. MolCrystalFlow: Molecular crystal structure prediction via flow matching. *arXiv preprint arXiv:2602.16020*, 2026. Author order follows the v3 PDF; the arXiv metadata lists Hennig before Roitberg.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3836–3847, 2023. arXiv:2302.05543.

Renrui Zhang, Jiaming Han, Chris Liu, Aojun Zhou, Pan Lu, Yu Qiao, Hongsheng Li, and Peng Gao. LLaMA-Adapter: Efficient fine-tuning of large language models with zero-initialized attention. In *International Conference on Learning Representations (ICLR)*, 2024. URL https://openreview.net/forum?id=d4UiXAHN2W. arXiv:2303.16199. The ICLR PDF lists 8 authors; older arXiv listings list 10.

A Training-time degradation ranges

B Oracle substitution and the match ladder

Table 7: Per-pattern degradation drawn during training. The historical clean evaluation operating point is shown for comparison; it lies outside the training support in domain size and counts.

Parameter	training range	clean operating point
isotropic strain	$[-0.005, 0.005]$	0
zero offset ($^\circ 2\theta$)	$[-0.10, 0.10]$	0
Caglioti U, V, W (deg^2)	$[0, 0.02]$, $[-0.02, 0.005]$, $[0.001, 0.03]$	$0.002, -0.001, 0.004$
domain size (nm)	$[20, 500]$, log-uniform	$\mathbf{10^4}$
pseudo-Voigt η	$[0.1, 0.9]$	0.5
March–Dollase r	$[0.7, 1.3]$ with prob. 0.5, else 1	1
background scale	$[0, 0.15]$ with prob. 0.8, else 0	0
amorphous hump scale	$[0, 0.10]$	0
peak counts	$[2 \times 10^3, 10^6]$, log-uniform	**noiseless**

Table 8: Frame-preserving oracle substitution on the first model ($n = 10$ validation crystals, $w = 0$, medians). `gtsanity` confirms the machinery is exact.

Variant	R_{wp}	RMS-to-GT	len err	ang err	match
baseline	0.999	1.080	0.218	8.0°	0/10
family-masked cell	0.999	1.105	0.211	0.0°	0/10
ground-truth cell	0.995	1.212	0.0	0.0°	0/10
ground-truth packing	0.974	0.685	0.218	8.0°	0/10
`gtsanity`	0.000	0.000	0.0	0.0°	10/10

Table 9: De-novo match@10 at $w = 0$, $n = 160$, with Wilson 95% intervals, scored against rebuilt ground truth (§8). Only the gauge step is a mechanism change.

Configuration	match@10	Wilson 95% CI	orientation loss
absolute gauge	0.000 (0/160)	[0.000, 0.023]	≈ 4.7 (floor)
relative gauge	0.031 (5/160)	[0.013, 0.071]	2.31
+ space-group cosets	0.025 (4/160)	[0.010, 0.063]	—
+ 8× centroid weight	0.044 (7/160)	[0.021, 0.088]	—