CRYSPFLOW: PHYSICS-SUPERVISED FLOW MATCHING FOR CRYSTAL RELAXATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Crystal structure relaxation seeks stable, low-energy structures from unrelaxed configurations. First-principles methods achieve high accuracy through repeated energy and force evaluations, but incur substantial computational cost, while direct learning-based approaches largely rely on relaxed-structure supervision and underuse the local directional information encoded by the potential-energy surface (PES). We introduce CrySPFlow, a physics-supervised Flow Matching framework that combines geometric endpoint supervision with PES-derived force feedback during training. To make physical supervision reflect the model-induced relaxation dynamics, we propagate terminal force feedback through a short differentiable manifold rollout. We further introduce an endpoint-preserving gradient injection rule that incorporates physical descent information while preserving the geometric objective's first-order descent rate in raw-gradient space. The interatomic potential is used only during training, so inference remains purely flow-based without additional potential evaluations. On $SrTiO_3$, CrySPFlow reduces coordinate and bond-length MAEs by 40.6% and 48.5%, respectively, relative to the strongest baseline for each metric. On X–Mn–O, it reduces coordinate MAE by 9.26%, while experiments with adapted FlowMM, RG-VFM, and CrystalFlow backbones show consistent benefits of physics supervision across the tested architectures.

1 INTRODUCTION

Crystal structure relaxation is a fundamental step in computational materials science, transforming an initial configuration into a nearby stable structure by descending the underlying potential-energy surface (PES). Conventional first-principles relaxation repeatedly evaluates energies and forces from electronic-structure calculations and uses the resulting local information to update atomic positions and, when necessary, the unit cell (Kresse & Furthmüller, 1996; Bitzek et al., 2006). Machine-learned interatomic potentials such as MACE can substantially reduce the cost of these evaluations (Batatia et al., 2022), but relaxation with such potentials still requires an iterative optimization procedure for each new structure. Direct learning-based relaxation provides an alternative by amortizing this procedure into a predictor that maps an unrelaxed configuration toward its relaxed counterpart, as exemplified by DeepRelax (Yang et al., 2024). However, such approaches are primarily supervised by relaxed structures. Endpoint supervision specifies *where* a structure should arrive, but makes limited use of the local directional information encoded by the PES that determines *how* the structure should move during relaxation.

This distinction is important because crystal relaxation is governed by the local energy landscape. Atomic forces are the negative gradients of the potential energy and therefore provide local directions for reducing energy. Whereas a relaxed target constrains the final geometry, force information characterizes the local response of the PES around intermediate structures. Geometric and physical supervision are therefore complementary: the former constrains the destination of relaxation, while the latter provides physically grounded local update directions. The challenge is to exploit this information during learning without reintroducing iterative potential evaluations at inference time.

Flow Matching provides a natural framework for this purpose by representing structural transformation through a continuous learned vector field rather than a static input-to-output mapping. Recent methods such as FlowMM and CrystalFlow incorporate crystalline geometry and symmetry

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.

into flow-based modeling (Miller et al., 2024; Luo et al., 2025), while RG-VFM extends endpoint prediction to manifold-valued states (Zaghen et al., 2026). These approaches provide a geometric description of structural evolution, but do not explicitly constrain the learned flow using the underlying PES. Physics-aware flow models have begun to incorporate energetic information in other settings, including energy-informed guidance for molecular docking (GUO et al., 2025) and energy- and force-based alignment for molecular crystal generation (Subramanian et al., 2026). Supervised crystal relaxation introduces two additional challenges: physical feedback should account for the structure induced by the model's own evolution rather than an isolated prediction, and physical and geometric supervision may produce conflicting parameter updates. These considerations motivate propagating physical feedback through the learned flow while preserving the endpoint objective that defines the relaxation target.

To address these challenges, we introduce CrySPFlow, a physics-supervised Flow Matching framework that integrates PES-derived feedback into the learned relaxation dynamics while retaining geometric endpoint supervision. A frozen interatomic potential provides force information only during training. To make this feedback reflect the model-induced evolution, we propagate terminal force information through a short differentiable manifold rollout, allowing physical credit to reach preceding flow updates. We further introduce an endpoint-preserving physics injection rule that extracts a physical gradient component orthogonal to the geometric gradient and injects it within a controlled trust region. This preserves the first-order descent rate of the geometric objective while incorporating additional PES-informed descent information. CrySPFlow therefore transfers local physical information from the PES into the learned relaxation dynamics during training, while requiring no interatomic-potential evaluations at inference.

Experiments on $SrTiO_3$ and X–Mn–O demonstrate improved structural recovery under physics supervision. Controlled ablations isolate the effects of direct physical feedback, rollout-based credit propagation, and gradient routing, while experiments with adapted FlowMM, RG-VFM, and CrystalFlow backbones show that the proposed training principle is applicable beyond a single predictor architecture.

Our main contributions are: (1) we formulate crystal relaxation as a physics-supervised flow-learning problem, combining geometric endpoint supervision with local PES information without inference-time potential evaluations; (2) we introduce differentiable manifold unrolling to propagate terminal physical feedback through model-induced relaxation dynamics; and (3) we develop an endpoint-preserving gradient injection mechanism whose orthogonal corrections incorporate physical descent information while preserving the geometric objective's first-order descent rate in raw-gradient space. Together, these components form a modular training framework applicable to other flow-matching networks.

2 Related Work

Crystal relaxation and learned interatomic potentials. Conventional crystal relaxation combines repeated electronic-structure calculations (Kresse & Furthmüller, 1996) with force-driven optimization, such as FIRE (Bitzek et al., 2006). Neural-network potentials approximate potential-energy surfaces and their forces (Behler & Parrinello, 2007). SchNet and DimeNet learn geometric interactions through continuous-filter and directional message passing (Schütt et al., 2017; Gasteiger et al., 2020), while NequIP and MACE use equivariant representations for accurate interatomic potentials (Batzner et al., 2022; Batatia et al., 2022). M3GNet, CHGNet, and MACE-MP-0 extend learned potentials towards broad chemical coverage (Chen & Ong, 2022; Deng et al., 2023; Batatia et al., 2025). These models accelerate energy and force evaluation; relaxation with them still requires successive structural updates. DeepRelax instead learns a direct map from unrelaxed inputs to relaxed structures (Yang et al., 2024). Our framework combines endpoint prediction with physical feedback from a frozen potential, confining potential evaluations to training.

Flow and diffusion models for crystal structures. Neural ordinary differential equations provide a continuous-dynamics formulation for learned transformations (Chen et al., 2018). Flow Matching trains continuous flows through vector-field regression (Lipman et al., 2023), while Rectified Flow learns transport along straight interpolation paths (Liu et al., 2023). Riemannian Flow Matching extends this formulation to manifolds (Chen & Lipman, 2024). For crystals, CDVAE and Dif-

fCSP incorporate periodic structure into diffusion-based generation (Xie et al., 2022; Jiao et al., 2023). FlowMM and CrystalFlow model atomic coordinates and lattice geometry with continuous flows (Miller et al., 2024; Luo et al., 2025), and RG-VFM studies endpoint prediction on manifolds for material and protein design (Zaghen et al., 2026). These methods establish geometric formulations for structure generation. We focus on relaxation conditioned on an unrelaxed crystal, using paired endpoint supervision together with potential-derived force feedback.

Physical guidance and gradient conflict. ForceFM incorporates force-guided flow matching into molecular docking (GUO et al., 2025), while PackFlow uses potential energies and forces for reinforcement-learning alignment of molecular crystal generation (Subramanian et al., 2026). Combining geometric and physical objectives also connects to multi-objective learning (Sener & Koltun, 2018). PCGrad projects conflicting task gradients (Yu et al., 2020), and CAGrad regularizes optimization of the average objective using the worst local task improvement (Liu et al., 2021). CrySPFlow prioritizes endpoint supervision through orthogonal corrections that preserve its first-order descent rate in raw-gradient space, with a separate fallback for severe conflicts. Differentiable manifold unrolling propagates terminal physical feedback through earlier predictions, without inference-time potential queries.

3 PRELIMINARIES

3.1 FLOW MATCHING

Flow Matching (Lipman et al., 2023) learns a time-dependent vector field $v_\theta(t, x)$ that transports a source distribution p_0 to a target distribution p_1 through the ordinary differential equation $\mathrm{d}x_t/\mathrm{d}t = v_\theta(t, x_t)$.Given a pair (x_0, x_1) sampled from a coupling π of these distributions, a linear conditional path and its velocity are

$$x_t = (1-t)x_0 + tx_1, \qquad u_t = x_1 - x_0, \tag{1}$$

where $t \in [0, 1]$. The vector field is trained by conditional velocity regression:

$$\mathcal{L}_{\mathrm{FM}} = \mathbb{E}_{t\sim\mathcal{U}(0,1),\,(x_0,x_1)\sim\pi}\left[\|v_\theta(t, x_t) - u_t\|_2^2\right]. \tag{2}$$

At the population optimum, this objective recovers the marginal velocity field associated with the interpolating probability path.

3.2 RIEMANNIAN FLOW MATCHING

Riemannian Flow Matching (Chen & Lipman, 2024) extends this construction to a manifold $(\mathcal{M}, g)$, where $v_\theta(t, x) \in T_x\mathcal{M}$ is a tangent vector and errors are measured using the Riemannian metric g. For endpoints connected by a unique minimizing geodesic, the conditional path and velocity take the form

$$x_t = \mathrm{Exp}_{x_0}\left(t\,\mathrm{Log}_{x_0}(x_1)\right), \qquad u_t = \frac{\mathrm{Log}_{x_t}(x_1)}{1-t}, \quad t < 1, \tag{3}$$

where Exp and Log denote the Riemannian exponential and logarithmic maps. The corresponding training objective is

$$\mathcal{L}_{\mathrm{RFM}} = \mathbb{E}_{t,\,(x_0,x_1)\sim\pi}\left[\|v_\theta(t, x_t) - u_t\|_{g_{x_t}}^2\right]. \tag{4}$$

This formulation accounts for the geometry of the state space through the interpolation path, tangent vector field and metric, providing the basis for modeling periodic atomic coordinates and lattice degrees of freedom.

4 METHODOLOGY

Our framework combines geometric endpoint supervision with force feedback from a frozen interatomic potential (Figure 1). An endpoint predictor defines the crystal relaxation flow; a short rollout propagates terminal force feedback through successive predictions; and a gradient projection controls its interaction with the geometric objective. The potential is used only during training.

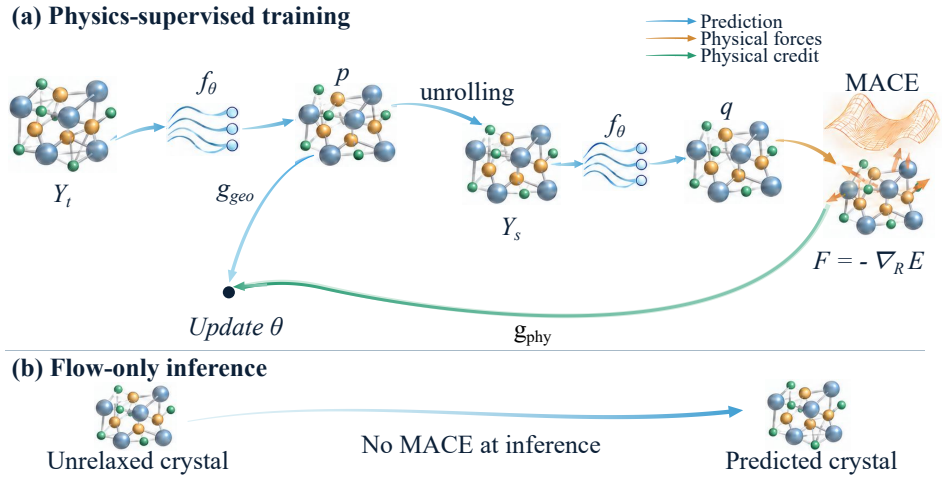


Figure 1: Training and inference framework. **(a) Physical supervision.** A direct endpoint prediction p advances the current state Y_t to Y_s, where the same network predicts the rollout endpoint q. A frozen MACE potential supplies atomic forces at the predicted structure with its cell held fixed. Detached forces propagate through the rollout and are combined with the geometric gradient. Blue, orange, and green arrows indicate prediction, forces, and physical credit, respectively. **(b) Inference.** The trained predictor relaxes the input structure without querying MACE.

4.1 Endpoint Flow on Crystal Geometry

For a fixed composition Z, we represent a periodic crystal by $Y = (q, S)$, where q contains fractional coordinates modulo periodic images and common translation, and $S = \log(AA^\top)$ represents the row-basis cell A.

$$Y \in \mathcal{M}_N = \mathcal{Q}_N \times \mathrm{Sym}(3), \qquad \mathcal{Q}_N = (\mathbb{T}^3)^N/\mathbb{T}^3, \qquad G = \exp S. \tag{5}$$

We write $\mathrm{Log}_q^G(\hat{q})$ for the lattice-metric displacement after periodic-image selection and removal of common translation. Appendix A gives its finite-search construction.

Given Y_t, the network predicts the relaxed endpoint $\hat{Y}_1 = f_\theta(t, Y_t, Z) = (\hat{q}_1, \hat{S}_1)$, defining

$$\frac{\mathrm{d}Y_t}{\mathrm{d}t} = \frac{1}{1-t}\begin{bmatrix}\mathrm{Log}_{q_t}^{G_t}(\hat{q}_1)\\ \hat{S}_1 - S_t\end{bmatrix}, \qquad G_t = \exp S_t. \tag{6}$$

Thus periodicity enters the predicted direction through the displacement operator. In Euclidean coordinates, this reduces to $(\hat{Y}_1 - Y_t)/(1-t)$.

Training states interpolate paired unrelaxed and relaxed structures using the target-cell metric and the log-cell chart (Appendix A.4). For target $(q_1^\star, A_1)$, let $G_1 = A_1 A_1^\top$, $S_1 = \log G_1$, and $\delta q_1 = \mathrm{Log}_{q_1^\star}^{G_1}(\hat{q}_1)$. The per-structure endpoint loss is

$$\ell_{\mathrm{geo}} = \frac{1}{2\sigma(t)^2}\left[\frac{\|\delta q_1 A_1\|_F^2}{N s_R^2} + \frac{\|\hat{S}_1 - S_1\|_F^2}{9 s_S^2}\right], \qquad \sigma(t) = \max(1-t, \sigma_{\min}), \tag{7}$$

where $s_R, s_S > 0$ set the coordinate and cell scales. Its batch aggregation defines $\mathcal{L}_{\mathrm{geo}}$ (Appendix A.4). Finite non-periodic patches use Cartesian coordinates.

4.2 Physical Supervision through Differentiable Unrolling

A predicted endpoint can change when the network is evaluated at a later state. We therefore query physical feedback at the endpoint of a short model rollout. Starting from the supervised state $Y_0 := Y_{t_0}$, set $t_k = t_0 + k(1-t_0)/K$ and $\alpha_k = 1/(K-k)$. With $(\hat{q}_{1,k}, \hat{S}_{1,k}) = f_\theta(t_k, Y_k, Z)$, the rollout is

$$\begin{aligned} q_{k+1} &= \mathrm{wrap}\big(q_k + \alpha_k \,\mathrm{Log}_{q_k}^{G_k}(\hat{q}_{1,k})\big),\\ S_{k+1} &= S_k + \alpha_k(\hat{S}_{1,k} - S_k), \qquad G_k = \exp S_k. \end{aligned} \tag{8}$$

We use the last network prediction as the final representative Y_K. For $K = 2$, the first prediction advances the state halfway through the remaining displacement and the second predicts the terminal structure. Appendix B defines the direct-to-unrolled discrepancy and proves exactness when successive endpoint predictions agree. This discrepancy is a diagnostic, not a training objective.

Let E_ψ be the frozen MACE potential and $F_\psi = -\nabla_R E_\psi$ its atomic forces. We recover the terminal cell in the input orientation and hold it fixed during differentiation:

$$\bar{A}_K = \mathrm{sg}[A(S_K; A_0)], \qquad R_K = q_K \bar{A}_K, \tag{9}$$

where sg denotes stop-gradient. This supplies coordinate-energy supervision without a direct stress objective. With $P_N = I_N - \mathbf{1}\mathbf{1}^\top/N$, the force surrogate is

$$\widetilde{\ell}_{\mathrm{phys}} = -\frac{1}{N} \left\langle \mathrm{sg}[P_N F_\psi(R_K, \bar{A}_K, Z)], P_N R_K \right\rangle_F . \tag{10}$$

Its gradient propagates the projected energy gradient through the active rollout Jacobian. Detaching the force avoids differentiating through the potential Hessian. Appendix C details the cell convention, gradient paths, and batch aggregation that defines $\mathcal{L}_{\mathrm{phys}}$.

4.3 Endpoint-Preserving Physics Injection

Geometric supervision specifies the target structure, whereas force feedback describes the local energy landscape. Their gradients may conflict, connecting this setting to multi-objective optimization and gradient-conflict mitigation (Sener & Koltun, 2018; Yu et al., 2020; Liu et al., 2021). We give the geometric endpoint objective explicit priority when incorporating physical feedback. Let $g_{\mathrm{geo}} = \nabla_\theta \mathcal{L}_{\mathrm{geo}}$ and $g_{\mathrm{phys}} = \nabla_\theta \mathcal{L}_{\mathrm{phys}}$. For nonzero g_{geo}, we retain the physical component orthogonal to it in raw parameter-gradient space:

$$h = g_{\mathrm{phys}} - \frac{g_{\mathrm{geo}}^\top g_{\mathrm{phys}}}{\|g_{\mathrm{geo}}\|^2} g_{\mathrm{geo}}. \tag{11}$$

When $h \neq 0$, a correction with relative radius $\rho \geq 0$ gives

$$c^\star = \rho \|g_{\mathrm{geo}}\| \frac{h}{\|h\|}, \qquad g^\star = g_{\mathrm{geo}} + c^\star, \qquad g_{\mathrm{geo}}^\top g^\star = \|g_{\mathrm{geo}}\|^2. \tag{12}$$

This correction maximizes additional physical descent within the orthogonal trust region. For a raw step $\theta^+ = \theta - \eta g^\star$, it preserves the geometric objective's first-order descent rate. Theorem 3 in Appendix D states and proves this result.

For conflicting gradients, the radius is increased up to a prescribed cap to obtain $g_{\mathrm{phys}}^\top g^\star \geq 0$ (Appendix D.4). If this requires exceeding the cap, we use the nearest gradient in the physical half-space:

$$g^\star = g_{\mathrm{geo}} - \frac{g_{\mathrm{geo}}^\top g_{\mathrm{phys}}}{\|g_{\mathrm{phys}}\|^2} g_{\mathrm{phys}}. \tag{13}$$

This fallback preserves first-order non-increase of both objectives but can reduce the geometric descent rate. Appendices E–F give its derivation and zero-gradient cases. These guarantees concern raw gradients; optimizer transformations are treated separately in Appendix I.

Algorithm 1 combines geometric supervision with terminal physical feedback. DifferentiableRollout advances the structure along the network-predicted directions while retaining the active computation graph. DetachedForceSurrogate treats the terminal forces from frozen MACE as constants, allowing physical feedback to propagate through the complete active rollout without differentiating through MACE. EndpointPreservingInjection combines the geometric and physical gradients using the orthogonal correction in Eqs. 11 and 12, with adaptive-radius selection as described in Appendix D.4. When the required correction exceeds the radius cap, it applies the half-space projection in Eq. 13. Degenerate cases follow Appendix F. These first-order guarantees apply to raw gradients before clipping and optimizer transformations. At inference, our endpoint networks use a single forward evaluation for the main experiments.

Algorithm 1 CrySPFlow Training with Fully Unrolled Physical Feedback

Require: Paired dataset $\mathcal{D}$; endpoint network f_θ; frozen MACE potential E_ψ; rollout length K; physics interval K_{phys}; injection radii $0 \le \rho_0 \le \rho_{\max}$.

for each mini-batch $\mathcal{B}$ at update $m = 1, 2, \ldots$ **do**
  Sample training states $(t, Y_t, Y^\star)$ from $\mathcal{B}$.
  $\widehat{Y}_1 \leftarrow f_\theta(t, Y_t, Z)$
  $g_{\text{geo}} \leftarrow \nabla_\theta \mathcal{L}_{\text{geo}}(\widehat{Y}_1, Y^\star, t)$; $g \leftarrow g_{\text{geo}}$
  if $m \bmod K_{\text{phys}} = 0$ **then**
    $Y_K \leftarrow \textsc{DifferentiableRollout}(f_\theta, Y_t, t, \widehat{Y}_1, K)$
    $\widetilde{\mathcal{L}}_{\text{phys}} \leftarrow \textsc{DetachedForceSurrogate}(E_\psi, Y_K, \mathcal{B})$
    $g_{\text{phys}} \leftarrow \nabla_\theta \widetilde{\mathcal{L}}_{\text{phys}}$ ▷ Through the full active rollout
    $g \leftarrow \textsc{EndpointPreservingInjection}(g_{\text{geo}}, g_{\text{phys}}; \rho_0, \rho_{\max})$
  end if
  Update θ using g with the configured gradient clipping and optimizer.
end for
return θ

5 Experiments

We evaluated structural recovery on $SrTiO_3$ and X–Mn–O, examined how physical supervision and gradient routing affected prediction accuracy, and tested the same training principle across adapted backbones. The quantitative comparisons measure recovery quality, while the response diagnostics examine how force feedback changes the model's behaviour.

5.1 Experimental Setup

Data and protocol. We considered $SrTiO_3$ structure recovery with fixed atom identities and correspondences, together with X–Mn–O crystal relaxation. Reference relaxed coordinates provided geometric supervision and evaluation targets, including the supervised interpolation used to construct training states. Coordinate conditioning, lattice priors, and the surrounding environment for $SrTiO_3$ force queries were constructed from the unrelaxed input. A frozen MACE potential supplied force feedback on model-predicted structures. Its parameters remained fixed, but the feedback produced physical gradients for updating the structure predictor. Our main recovery results used a single forward evaluation at inference, without additional MACE queries.

Baselines and metrics. We compared DeepRelax (Yang et al., 2024), PaiNN (Schütt et al., 2021), EGNN (Satorras et al., 2021), and task-adapted CrystalFlow (Luo et al., 2025), FlowMM (Miller et al., 2024), and RG-VFM (Zaghen et al., 2026). The $SrTiO_3$ versions adapt these methods to coordinate recovery. Appendix K summarizes the training and evaluation protocols. We report coordinate and bond-length mean absolute errors (MAEs), raw and rotation-aligned lattice errors (Lat.-R/A), volume error (Vol.), relative volume error (Rel.V), and match rate (Match). Coordinate, bond, and lattice errors are in Å, volume error is in $Å^3$, and Rel. V and Match are percentages. For $SrTiO_3$, lattice and volume metrics are derived from affine fits of the recovered coordinates to an input-derived lattice prior; for X–Mn–O, they use the predicted and reference cells directly. These task-specific match rates are not compared across datasets. Appendix L provides the full metric definitions and matching settings. In Tables 1 and 2, errors are reported to six decimal places, and percentage-valued metrics to three decimal places. All results are point estimates. Lower errors and higher match rates are better; bold entries identify the best comparable values within each table or backbone.

5.2 Main Results

$SrTiO_3$. On $SrTiO_3$, CrySPFlow achieved the lowest errors across all six continuous metrics (Table 1). Coordinate MAE decreased from 0.013510 Å for RG-VFM to 0.008021 Å, while bond MAE decreased from 0.010509 Å for CrystalFlow to 0.005415 Å, corresponding to reductions of 40.6% and 48.5%, respectively. Lattice and volume errors were also substantially reduced. The match

Table 1: Structural recovery on $SrTiO_3$. Coordinate, bond, and lattice errors are reported in Å; volume errors in $Å^3$; relative volume errors and match rates in %.

Method	Coord.↓	Bond↓	Lat.-R↓	Lat.-A↓	Vol.↓	Rel.V↓	Match↑
DeepRelax	0.095223	0.095558	0.002675	0.002669	0.022462	0.409	64.346
PaiNN	0.027384	0.032047	0.002427	0.001987	0.055668	0.101	89.286
EGNN	0.076567	0.100955	0.005019	0.003825	0.104189	0.190	17.857
CrystalFlow	0.013776	0.010509	0.002239	0.001774	0.067118	0.122	67.857
FlowMM	0.015063	0.011510	0.002015	0.001671	0.077061	0.140	64.286
RG-VFM	0.013510	0.016411	0.001273	0.000750	0.017867	0.033	**97.621**
Ours	**0.008021**	**0.005415**	**0.000126**	**0.000103**	**0.003315**	**0.006**	96.429

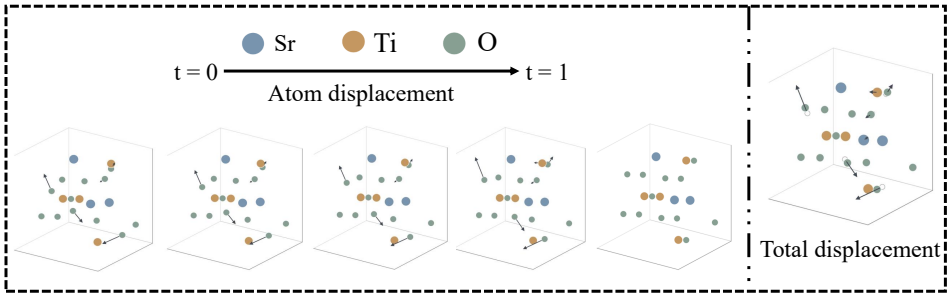


Figure 2: Auxiliary structural rollout for $SrTiO_3$.

rate reached 96.429%, slightly below RG-VFM's 97.621%, indicating that lower average structural errors do not necessarily translate to the highest threshold-based match rate. These results show that the benefit of physics supervision on $SrTiO_3$ is broad across coordinate, bond, lattice, and volume recovery, consistent with CrySPFlow's design of injecting PES-derived local descent information into model-induced relaxation updates while preserving geometric endpoint supervision.

Figure 2 visualizes an auxiliary endpoint-Euler rollout from the input towards the predicted structure. Colours identify Sr, Ti, and O, and arrows show coordinate displacements; the final panel summarizes the total displacement. The sequence illustrates the evolution of atomic positions during structural recovery.

X–Mn–O. On X–Mn–O, CrySPFlow achieved the best performance on six of the seven reported metrics (Table 2). Coordinate MAE decreased from 0.126785 Å for DeepRelax to 0.115041 Å, while absolute and relative volume errors were reduced to 3.645 $Å^3$ and 2.104%, respectively. In contrast, the improvement in bond MAE was marginal, and PaiNN remained stronger on rotation-aligned lattice error. These results are consistent with CrySPFlow's design of injecting PES-derived local descent information into model-induced relaxation updates while preserving geometric endpoint supervision. Taken together with the $SrTiO_3$ results, these findings suggest that physics supervision improves structural recovery across both datasets, while the magnitude and metric dependence of the gains vary across relaxation settings. Visual comparisons of structural recovery on $SrTiO_3$ and X–Mn–O are provided in Appendix M.

5.3 Physics Supervision and Gradient Routing

Gradient-routing configurations. We compared five configurations to separate the contribution of force feedback from the paths through which its gradients reach the predictor (Table 3). Configuration A used geometric supervision alone. Configuration B added direct physical supervision at the initial endpoint prediction. Configurations C–E retained this direct-physics base update and added different corrections derived from terminal force feedback. Configuration C (terminal local) retained only the gradient through the terminal prediction, with its input state detached from the preceding rollout. Configuration D (trajectory only) retained the additional gradient propagated through the preceding rollout state. Configuration E (terminal full) included both terminal-local and trajectory

Table 2: Structural recovery on X–Mn–O. Coordinate, bond, and lattice errors are reported in Å; volume errors in $Å^3$; relative volume errors and match rates in %.

Method	Coord.↓	Bond↓	Lat.-R↓	Lat.-A↓	Vol.↓	Rel.V↓	Match↑
DeepRelax	0.126785	0.139080	0.072110	0.159630	5.445080	3.949	84.540
PaiNN	0.153596	0.193883	0.110802	**0.105176**	5.922111	3.549	76.810
EGNN	0.168233	0.206787	0.121250	0.116595	7.006866	4.141	73.732
RG-VFM	0.279854	0.698664	0.280350	0.243800	7.191100	4.238	41.380
FlowMM	0.287954	0.358042	0.178870	0.217437	7.478900	6.213	64.298
CrystalFlow	0.217831	0.301958	0.100709	0.198322	9.224631	6.031	63.484
Ours	**0.115041**	**0.138924**	**0.069202**	0.117984	**3.645011**	**2.104**	**85.159**

Table 3: Physics supervision and gradient routing on $SrTiO_3$.

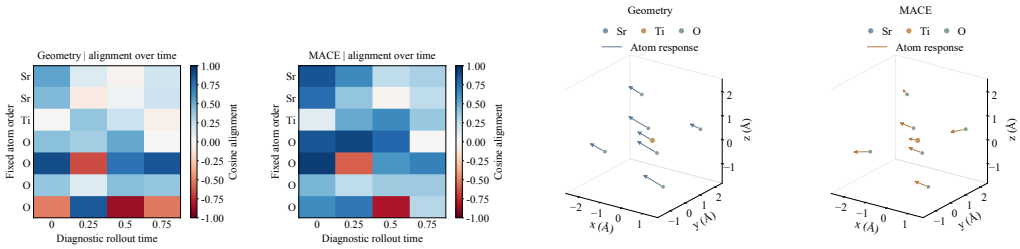
ID	Coord. MAE (Å)↓	Bond MAE (Å)↓
A	0.008844	0.006910
B	0.008663	0.005981
C	0.008462	0.005883
D	0.008262	0.005785
E	**0.008021**	**0.005415**

contributions. Thus, "trajectory only" describes the additional correction in D; it does not remove the direct physical supervision used in B. C–E shared the same terminal-query protocol, whereas B omitted the additional terminal query. Comparisons among C–E therefore isolate gradient routing more directly than comparisons against B.

Effect on recovery. Direct physical supervision reduced coordinate MAE from 0.008844 to 0.008663 Å and bond MAE from 0.006910 to 0.005981 Å (A→B). Each terminal-feedback correction further reduced both errors relative to B. Among these corrections, D outperformed C, and E achieved the lowest errors, 0.008021 Å for coordinates and 0.005415 Å for bonds. Relative to A, these represent reductions of 9.31% and 21.64%, respectively. The ordering indicates that, in this experiment, retaining the complete gradient through the rollout was more effective than either restricted route alone.

Response–force alignment and spatial responses. Figure 3 examines how MACE feedback changes the direction and spatial distribution of model-induced coordinate responses. At matched network weights and structural states, we measured first-order coordinate responses along the negative geometry-only and geometry-plus-MACE parameter gradients, using the geometry-only gradient norm as a common normalization factor. Panels (a) and (b) report the atom-wise cosine alignment between these responses and the corresponding MACE forces. Rows track the same seven atoms, and columns correspond to $t \in \{0, 0.25, 0.5, 0.75\}$. Blue indicates positive alignment, red indicates opposing directions, and values near zero indicate weak alignment. Adding MACE feedback shifted several entries towards stronger positive alignment. Mean alignment increased at all four sampled times, from 0.306 to 0.660 at $t = 0$ and from 0.153 to 0.325 at $t = 0.75$.

Panels (c) and (d) complement the heatmaps by visualizing coordinate responses in a second local example at $t = 0$. Both panels show the same Ti atom and six O atoms with identical coordinates, viewing angles, and axis limits. Blue and ochre arrows represent responses to geometry-only and geometry-plus-MACE updates, respectively; the panel labelled "MACE" denotes the latter. All arrows use a common display scale, preserving relative response magnitudes across atoms and methods. The geometry-only responses largely shared a common direction. Adding MACE feedback shortened the response of the central Ti atom and several O atoms, while increasing and redirecting the responses of the O atoms on the left and right of the displayed view. These changes show atom-dependent adjustments in both response direction and magnitude. The arrows depict first-order coordinate responses to parameter updates, whereas their alignment with atomic forces is quantified by the heatmaps. Together, these diagnostics show how physical feedback reshapes local coordinate responses and improves mean force alignment in the heatmap example.



(a) (b) (c) (d)

Figure 3: Effect of MACE feedback on response–force alignment and local coordinate responses.

Table 4: Physics guidance across adapted $SrTiO_3$ backbones.

	Baseline		+ MACE		+ MACE + Unrolling	
Backbone	Coord.↓	Bond↓	Coord.↓	Bond↓	Coord.↓	Bond↓
FlowMM-backbone	0.014260	0.006900	0.014245	0.006894	**0.014000**	**0.006823**
RG-VFM-backbone	0.015059	0.014110	0.014922	0.014091	**0.014814**	**0.013920**
CrystalFlow-backbone	0.013209	0.007654	0.013140	**0.007444**	**0.013137**	0.007448
Our backbone	0.008844	0.006910	0.008663	0.005981	**0.008021**	**0.005415**

5.4 Applicability Across Backbones

Comparison protocol. We tested whether the benefit of physical guidance extended beyond our predictor using adapted FlowMM, RG-VFM, and CrystalFlow implementations for $SrTiO_3$ recovery (Table 4). Within each external backbone, the baseline and two physical variants started from the same pretrained checkpoint with fresh optimizers and paired data order and sampling seeds.

For the three external backbones, the table compares the baseline, local MACE guidance (+ MACE), and MACE guidance with gradient propagation through the rollout (+ MACE + Unrolling). Both physical variants used the same forward sampling procedure and terminal force query. Local guidance propagated feedback only through the final prediction step; unrolling additionally propagated it through earlier sampling steps. Coordinate and bond MAEs are reported in Å, and bold values mark the minimum within each backbone. The external adapters used 20 Euler steps at inference. For our backbone, we reused the results of configurations A, B, and E from the main ablation, each evaluated with a single forward pass. These results were not obtained from a separate external-backbone fine-tuning experiment. All four implementations required no MACE queries at inference.

Results. MACE guidance with unrolling reduced both coordinate and bond MAEs relative to the baseline for all four backbones. Coordinate error reductions ranged from 0.55% to 9.31%. Unrolling also lowered coordinate MAE relative to local guidance alone for all four backbones, although the improvement for CrystalFlow-backbone was small (0.013140 to 0.013137 Å). For bond MAE, CrystalFlow-backbone performed slightly better with local guidance alone (0.007444 versus 0.007448 Å). The results therefore support applicability across the tested adapted implementations, with the magnitude and metric dependence of the benefit varying by backbone.

6 Conclusion

We presented CrySPFlow, a physics-supervised flow-matching framework combining geometric endpoint supervision with frozen-potential force feedback through differentiable manifold unrolling and endpoint-preserving gradient injection. On $SrTiO_3$, coordinate and bond-length MAEs decreased by 40.6% and 48.5% relative to the strongest respective baselines. On X–Mn–O, coordinate MAE decreased by 9.26% relative to DeepRelax, and the match rate increased from 84.540% to 85.159%. Ablations supported full rollout gradients, while experiments demonstrated applicability across adapted flow backbones. Our main models require one forward pass and no potential queries at inference. Broader chemical evaluation and direct stress supervision remain future directions.

AI USE STATEMENT

Generative AI tools were used solely to improve the language and readability of the manuscript and to assist in checking author-written code for potential errors. They were not used to develop the research methodology, formulate theoretical claims or proofs, generate research data, or interpret experimental results. The authors reviewed the suggestions incorporated into the manuscript and code and take full responsibility for the final text, implementation, and scientific claims.

ETHICS STATEMENT

This work studies machine learning methods for crystal structure recovery and relaxation. It uses atomic structure data and computational experiments and does not involve human participants, personal data, or animal experiments. The intended application is to support computational materials research. Predicted structures and assessments based on learned interatomic potentials may be inaccurate outside the evaluated settings and should be independently validated before informing consequential materials design or experimental decisions.

REPRODUCIBILITY STATEMENT

The methodology section describes the proposed model, training objectives, and optimization procedure. Section 5.1 specifies the tasks, supervision protocol, baselines, and evaluation settings for $SrTiO_3$ and X–Mn–O. The appendices provide supporting mathematical derivations and proofs, together with their assumptions and applicability boundaries. Appendix K summarizes the data and experimental setup, while Appendix L defines the reported metrics and matching procedures. These descriptions support inspection and reproduction of the reported experiments.

REFERENCES

Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gabor Csanyi. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 11423–11436. Curran Associates, Inc., 2022. doi: 10.52202/068431-0830. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/4a36c3c51af11ed9f34615b81edb5bbc-Paper-Conference.pdf.

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://doi.org/10.1063/5.0297006.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural

networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13 (1):2453, 2022. doi: 10.1038/s41467-022-29939-5. URL https://doi.org/10.1038/s41467-022-29939-5.

Jörg Behler and Michele Parrinello. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.*, 98(14):146401, Apr 2007. doi: 10.1103/PhysRevLett.98.146401. URL https://link.aps.org/doi/10.1103/PhysRevLett.98.146401.

Erik Bitzek, Pekka Koskinen, Franz Gähler, Michael Moseler, and Peter Gumbsch. Structural Relaxation Made Simple. *Phys. Rev. Lett.*, 97(17):170201, Oct 2006. doi: 10.1103/PhysRevLett.97.170201. URL https://link.aps.org/doi/10.1103/PhysRevLett.97.170201.

Chi Chen and Shyue Ping Ong. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science*, 2(11):718–728, 2022. doi: 10.1038/s43588-022-00349-3. URL https://doi.org/10.1038/s43588-022-00349-3.

Ricky T. Q. Chen and Yaron Lipman. Flow Matching on General Geometries. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 47922–47945, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/d1f9936d3be6997ffffab692977eebe6-Paper-Conference.pdf.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural Ordinary Differential Equations. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf.

Bowen Deng, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Kevin Han, Christopher J. Bartel, and Gerbrand Ceder. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5(9):1031–1041, 2023. doi: 10.1038/s42256-023-00716-3. URL https://doi.org/10.1038/s42256-023-00716-3.

Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional Message Passing for Molecular Graphs. In *International Conference on Learning Representations (ICLR)*, 2020. URL https://openreview.net/forum?id=B1eWbxStPH.

HUANLEI GUO, Song LIU, and Bingyi Jing. ForceFM: Enhancing Protein-Ligand Predictions through Force-Guided Flow Matching. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 124268–124296. Curran Associates, Inc., 2025. doi: 10.52202/085713-4140. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/b3ee701cea4872c41356f32592a72289-Paper-Conference.pdf.

Rui Jiao, Wenbing Huang, Peijia Lin, Jiaqi Han, Pin Chen, Yutong Lu, and Yang Liu. Crystal Structure Prediction by Joint Equivariant Diffusion. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 17464–17497. Curran Associates, Inc., 2023. doi: 10.52202/075280-0767. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/38b787fc530d0b31825827e2cc306656-Paper-Conference.pdf.

G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169–11186, Oct 1996. doi: 10.1103/PhysRevB.54.11169. URL https://link.aps.org/doi/10.1103/PhysRevB.54.11169.

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*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-Averse Gradient Descent for Multi-task learning. In M. Ranzato, A. Beygelzimer, Y. Dauphin,

P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 18878–18890. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/9d27fdf2477ffbff837d73ef7ae23db9-Paper.pdf.

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*, 2023. URL https://openreview.net/forum?id=XVjTT1nw5z.

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

Benjamin Kurt Miller, Ricky T. Q. Chen, Anuroop Sriram, and Brandon M Wood. FlowMM: Generating Materials with Riemannian Flow Matching. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 35664–35686. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/miller24a.html.

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. URL https://doi.org/10.1016/j.commatsci.2012.10.028.

Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) Equivariant Graph Neural Networks. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9323–9332. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/satorras21a.html.

Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/303ed4c69846ab36c2904d3ba8573050-Paper.pdf.

Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9377–9388. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/schutt21a.html.

Ozan Sener and Vladlen Koltun. Multi-Task Learning as Multi-Objective Optimization. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/432aca3a1e345e339f35a30c8f65edce-Paper.pdf.

Akshay Subramanian, Elton Pan, Juno Nam, Maurice Weiler, Shuhui Qu, Cheol Woo Park, Tommi S. Jaakkola, Elsa Olivetti, and Rafael Gomez-Bombarelli. PackFlow: Generative Molecular Crystal Structure Prediction via Reinforcement Learning Alignment, 2026. URL https://arxiv.org/abs/2602.20140.

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. URL https://openreview.net/forum?id=03RLpj-tc_.

Ziduo Yang, Yi-Ming Zhao, Xian Wang, Xiaoqing Liu, Xiuying Zhang, Yifan Li, Qiujie Lv, Calvin Yu-Chian Chen, and Lei Shen. Scalable crystal structure relaxation using an iteration-free deep generative model with uncertainty quantification. *Nature Communications*, 15(1): 8148, 2024. doi: 10.1038/s41467-024-52378-3. URL https://doi.org/10.1038/s41467-024-52378-3.

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient Surgery for Multi-Task Learning. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 5824–5836. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/3fe78a8acf5fda99de95303940a2420c-Paper.pdf.

Olga Zaghen, Floor Eijkelboom, Alison Pouplin, Cong Liu, Max Welling, Jan-Willem van de Meent, and Erik Bekkers. Riemannian Variational Flow Matching for Material and Protein Design. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 34243–34286, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/39c5871aa13be86ab978cba7069cbcec-Paper-Conference.pdf.

A Crystal Quotient Geometry

This section defines the finite-search quotient geometry used in Sec. 4.1. We follow the row-vector convention

$$R = qA, \qquad G = AA^\top, \tag{14}$$

such that the physical squared length of a fractional displacement $u \in \mathbb{R}^3$ is

$$\|uA\|_2^2 = uGu^\top. \tag{15}$$

The cell state is represented in the log-Euclidean chart

$$S = \log G \in \mathrm{Sym}(3), \qquad G = \exp S. \tag{16}$$

The fractional coordinates lie on the translation quotient

$$\mathcal{Q}_N = (\mathbb{T}^3)^N / \mathbb{T}^3, \tag{17}$$

since a common periodic translation $q_i \mapsto q_i + c \pmod 1$ does not change the physical crystal.

A.1 Metric-Aware Minimum-Image Operator

For a skewed lattice, component-wise wrapping into $[-1/2, 1/2)^3$ does not necessarily minimize Cartesian displacement, because the physical metric depends on the off-diagonal entries of G. We therefore explicitly enumerate periodic images.

For search radius $r \geq 1$, define

$$\mathcal{N}_r = \{-r, \ldots, r\}^3. \tag{18}$$

For a fractional displacement u, the metric-aware minimum-image operator is

$$\mathcal{I}_G(u) = u + n_G^\star(u), \qquad n_G^\star(u) = \arg\min_{n \in \mathcal{N}_r} (u+n)G(u+n)^\top. \tag{19}$$

We use $r = 1$ by default, corresponding to 27 candidate images. Thus, Eq. 19 selects the shortest representative within the enumerated periodic neighborhood under the current cell metric.

A.2 Anchor-Conditioned Translation-Quotient Log

Given a base configuration q and a target $\hat{q}$, define the raw fractional displacement

$$d_i = \hat{q}_i - q_i. \tag{20}$$

Let

$$\mathrm{cwrap}(x) = ((x + \tfrac{1}{2}) \bmod 1) - \tfrac{1}{2}. \tag{21}$$

For each valid anchor atom a, we first construct relative periodic displacements

$$r_i^{(a)} = \mathrm{cwrap}(d_i - d_a), \tag{22}$$

apply metric-aware image selection,

$$m_i^{(a)} = \mathcal{I}_G(r_i^{(a)}), \tag{23}$$

and restore the anchor branch,

$$w_i^{(a)} = d_a + m_i^{(a)}. \tag{24}$$

For a fixed anchor, the optimal common translation is obtained from

$$\min_{c \in \mathbb{R}^3} \frac{1}{N} \sum_{i=1}^{N} (w_i^{(a)} - c) G (w_i^{(a)} - c)^\top. \tag{25}$$

Proposition 1 (Optimal Common Translation)**.** *For $G \succ 0$, the unique minimizer of Eq. 25 is*

$$c_a^\star = \bar{w}^{(a)} = \frac{1}{N} \sum_{i=1}^{N} w_i^{(a)}. \tag{26}$$

Proof. Differentiating Eq. 25 with respect to c gives

$$-2 \left(\bar{w}^{(a)} - c \right) G. \tag{27}$$

Since $G \succ 0$, the unique stationary point is $c = \bar{w}^{(a)}$, and the Hessian is $2G \succ 0$. □

The translation-quotient residual is therefore

$$\xi_i^{(a)} = w_i^{(a)} - \bar{w}^{(a)}, \qquad \sum_{i=1}^{N} \xi_i^{(a)} = 0. \tag{28}$$

We score each anchor by its mean physical residual energy,

$$E_a = \frac{1}{N} \sum_{i=1}^{N} \xi_i^{(a)} G \left(\xi_i^{(a)} \right)^\top, \tag{29}$$

and select

$$a^\star = \arg\min_a E_a. \tag{30}$$

For compact notation, we denote the resulting implementation-level metric-aware quotient displacement operator by Log_q^G. It is a finite-search representative of the local quotient displacement rather than a globally exact logarithmic map over all $\mathbb{Z}^3$ images. Thus,

$$\mathrm{Log}_q^G(\hat{q}) = \left(\xi_1^{(a^\star)}, \ldots, \xi_N^{(a^\star)} \right). \tag{31}$$

This construction resolves the periodic-image ambiguity within the enumerated neighborhood together with the global translation gauge under the physical lattice metric. All subsequent uses of Log_q^G refer to this implementation-level operator.

A.3 Piecewise Differentiability

Equation 31 contains discrete minimizations over periodic images and anchors and is therefore not globally smooth. Away from image ties and anchor ties, however, the selected discrete branch remains constant in an open neighborhood. Within such a neighborhood, $\mathrm{Log}_q^G(\hat{q})$ reduces to ordinary smooth operations.

Consequently, the quotient operator is piecewise smooth and differentiable almost everywhere. Automatic differentiation therefore computes the derivative of the currently selected quotient branch. Our theoretical statements involving differentiation should be understood within such a fixed local branch.

A.4 Reference Path and Endpoint Objective

For an unrelaxed–relaxed pair $(q_0, A_0) \to (q_1^\star, A_1)$, define

$$S_0 = \log(A_0 A_0^\top), \qquad S_1 = \log(A_1 A_1^\top), \tag{32}$$

and

$$\xi_{01} = \mathrm{Log}_{q_0}^{G_1}(q_1^\star), \qquad G_1 = A_1 A_1^\top. \tag{33}$$

The supervised intermediate state is

$$q_t = \mathrm{wrap}(q_0 + t\xi_{01}), \qquad S_t = S_0 + t(S_1 - S_0). \tag{34}$$

Equation 34 is a reference interpolation within the selected local quotient branch and the log-Euclidean cell chart; we do not require it to be a globally shortest geodesic of a fully coupled crystal-state metric.

For completeness, the coordinate endpoint residual is

$$e_q = \mathrm{Log}_{q_1^\star}^{G_1}(\hat{q}_1), \qquad e_R = e_q A_1, \tag{35}$$

with

$$d_R^2 = \frac{1}{N}\|e_R\|_F^2, \qquad d_S^2 = \frac{1}{9}\|\hat{S}_1 - S_1\|_F^2. \tag{36}$$

Using

$$\sigma(t) = \max(1 - t, \sigma_{\min}), \tag{37}$$

the per-structure geometric objective is

$$\ell_{\mathrm{geo}} = \frac{1}{2}\left[\frac{d_R^2}{s_R^2\sigma(t)^2} + \frac{d_S^2}{s_S^2\sigma(t)^2}\right]. \tag{38}$$

For a minibatch, the implementation balances source and path examples before forming the geometric gradient. Let $\mathcal{B}_{\mathrm{src}}$ and $\mathcal{B}_{\mathrm{path}}$ denote the active source and path subsets and let

$$\mathcal{L}_c = \frac{1}{|\mathcal{B}_c|}\sum_{b\in\mathcal{B}_c} \ell_{\mathrm{geo}}^{(b)}, \qquad c \in \{\mathrm{src}, \mathrm{path}\}, \tag{39}$$

whenever the corresponding subset is nonempty. If $\mathcal{A}$ denotes the set of active branches in the current minibatch and $w_c > 0$ their prescribed weights, the batch geometric objective is

$$\mathcal{L}_{\mathrm{geo}} = \frac{\sum_{c\in\mathcal{A}} w_c \mathcal{L}_c}{\sum_{c\in\mathcal{A}} w_c}. \tag{40}$$

This active-weight normalization also covers minibatches in which only one branch is present.

B Endpoint-Consistent Manifold Unrolling

B.1 Remaining-Time-Normalized Discretization

Starting from the supervised state $Y_0 := Y_{t_0}$, we divide the remaining interval $[t_0, 1]$ into K uniform steps:

$$\Delta t = \frac{1 - t_0}{K}, \qquad t_k = t_0 + k\Delta t. \tag{41}$$

The relative fraction of the remaining endpoint displacement traversed at step k is

$$\alpha_k = \frac{\Delta t}{1 - t_k} = \frac{1}{K - k}. \tag{42}$$

Thus the update fraction is independent of the sampled starting time t_0.

At each step,

$$\hat{Y}_{1,k} = f_\theta(t_k, Y_k, Z) = (\hat{q}_{1,k}, \hat{S}_{1,k}), \tag{43}$$

and

$$\begin{aligned} q_{k+1} &= \mathrm{wrap}\big(q_k + \alpha_k \,\mathrm{Log}_{q_k}^{G_k}(\hat{q}_{1,k})\big), \\ S_{k+1} &= S_k + \alpha_k(\hat{S}_{1,k} - S_k), \qquad G_k = \exp S_k. \end{aligned} \tag{44}$$

B.2 Final-Step Representative Convention

At the final step,

$$\alpha_{K-1} = 1. \tag{45}$$

On the quotient manifold, applying the quotient displacement with unit step reaches a representative physically equivalent to the predicted endpoint. To avoid an unnecessary gauge-equivalent tensor transformation, the implementation directly uses the network endpoint representative at the final step:

$$q_K = \hat{q}_{1,K-1}, \qquad S_K = \hat{S}_{1,K-1}. \tag{46}$$

An immediate consequence is exact tensor-level compatibility for $K = 1$:

$$Y_K = f_\theta(t_0, Y_{t_0}, Z). \tag{47}$$

B.3 Direct-to-Unrolled Discrepancy

The direct endpoint and the two-step terminal endpoint are

$$\hat{Y}_1^{(0)} = f_\theta(t_0, Y_{t_0}, Z), \qquad \hat{Y}_1^{(1)} = Y_2. \tag{48}$$

To specify the discrepancy without assuming a globally exact quotient distance, use the target cell A_1 as a common diagnostic metric, $G_1 = A_1 A_1^\top$, and define

$$\Delta^2_{\mathrm{roll}} = \frac{\| \operatorname{Log}^{G_1}_{\hat{q}_1^{(0)}}(\hat{q}_1^{(1)}) A_1 \|_F^2}{N s_R^2} + \frac{\| \hat{S}_1^{(1)} - \hat{S}_1^{(0)} \|_F^2}{9 s_S^2}. \tag{49}$$

This is a diagnostic on paired training examples using the finite-search operator from Appendix A; it is not asserted to be a globally exact geodesic distance. It measures disagreement between endpoint predictions, not which endpoint has lower energy. We do not optimize this quantity or use it as a structural match criterion. The following theorem gives a sufficient condition for it to vanish.

B.4 Endpoint-Consistent Exactness

Theorem 1 (Exactness under Endpoint Consistency)**.** *Consider the K-step rollout in Eq. 8. Suppose the visited image/anchor branch remains fixed and*

$$f_\theta(t_k, Y_k, Z) = Y^\star, \qquad k = 0, \dots, K-1.$$

Let

$$\xi_k = \operatorname{Log}^{G_k}_{q_k}(q^\star), \qquad r_k^S = S^\star - S_k.$$

Then, for $k = 0, \dots, K-2$,

$$\xi_{k+1} = \frac{K-k-1}{K-k}\xi_k, \qquad r_{k+1}^S = \frac{K-k-1}{K-k} r_k^S.$$

Under the final-step endpoint-representative convention, $Y_K = Y^\star$. In particular, for $K = 2$,

$$\hat{Y}_1^{(1)} = \hat{Y}_1^{(0)}, \qquad \Delta_{\mathrm{roll}} = 0.$$

Proof. Assume that all visited coordinate pairs $(q_k, q^\star)$ remain in the same image/anchor branch of the quotient operator and that

$$f_\theta(t_k, Y_k, Z) = Y^\star = (q^\star, S^\star) \tag{50}$$

for every k.

For $k = 0, \dots, K-2$, define the remaining coordinate displacement

$$\xi_k = \operatorname{Log}^{G_k}_{q_k}(q^\star). \tag{51}$$

Within a fixed image/anchor branch, all discrete periodic shifts and the selected anchor are constant. Conditioned on that branch, the quotient displacement is affine in the fractional coordinates. Although G_k may change with the cell state, it only affects which discrete branch is selected; once the

branch is fixed, it does not alter the affine residual expression. Using the common endpoint $q^\star$ and the update rule therefore gives

$$\xi_{k+1} = (1-\alpha_k)\xi_k, \qquad k = 0, \dots, K-2. \tag{52}$$

Similarly, with

$$r_k^S = S^\star - S_k, \tag{53}$$

we have

$$r_{k+1}^S = (1-\alpha_k) r_k^S, \qquad k = 0, \dots, K-2. \tag{54}$$

Since

$$1 - \alpha_k = \frac{K-k-1}{K-k}, \tag{55}$$

the pre-final residuals contract by the stated remaining-time factor.

At the final step, the implementation uses the endpoint representative convention in Eq. 46. Because the endpoint map again returns the same $Y^\star$, it directly sets

$$q_K = q^\star, \qquad S_K = S^\star, \tag{56}$$

and hence $Y_K = Y^\star$ exactly. For $K = 2$, the second endpoint prediction therefore equals the direct endpoint prediction,

$$\hat{Y}_1^{(1)} = \hat{Y}_1^{(0)}, \tag{57}$$

which implies $\Delta_{\mathrm{roll}} = 0$. □

B.5 Rollout Jacobian

Write each unrolling step abstractly as

$$Y_{k+1} = \Phi_k(Y_k, \theta). \tag{58}$$

Because the starting state Y_0 is constructed from supervised data,

$$D_\theta Y_0 = 0. \tag{59}$$

The chain rule gives

$$D_\theta Y_{k+1} = D_Y \Phi_k \, D_\theta Y_k + D_\theta \Phi_k. \tag{60}$$

Proposition 2 (Rollout Jacobian Expansion)**.** *The terminal-state Jacobian satisfies*

$$D_\theta Y_K = \sum_{j=0}^{K-1} \left(D_Y \Phi_{K-1} \cdots D_Y \Phi_{j+1} \right) D_\theta \Phi_j, \tag{61}$$

where an empty product is the identity map.

Proof. The result follows by repeatedly substituting the one-step chain-rule recursion. For $K = 2$,

$$D_\theta Y_2 = D_Y \Phi_1 D_\theta \Phi_0 + D_\theta \Phi_1. \tag{62}$$

The general expression follows by induction. □

Consequently, for any differentiable terminal functional $\mathcal{P}(Y_K)$,

$$\nabla_\theta \mathcal{P} = (D_\theta Y_K)^\top \nabla_{Y_K} \mathcal{P}. \tag{63}$$

Physics evaluated at the rollout terminal state therefore propagates to all earlier endpoint evaluations, rather than only to the final network call.

We emphasize that the training rollout consists of a teacher-forced reference prefix up to t_0 followed by a model-unrolled suffix from t_0 to 1. It is not a free-running model trajectory starting from $t = 0$.

C Frozen-Potential Physical Supervision

C.1 Reference-Cell Gauge

The Gram matrix $G = AA^\top$ determines the lattice only up to a right orthogonal transformation. To recover a lattice representative consistent with the input-cell orientation, define

$$G_0 = A_0 A_0^\top, \qquad L_0 = \mathrm{chol}(G_0), \tag{64}$$

and

$$Q_0 = L_0^{-1} A_0. \tag{65}$$

Since

$$Q_0 Q_0^\top = L_0^{-1} G_0 L_0^{-\top} = I, \tag{66}$$

Q_0 is orthogonal.

For a predicted log metric S, we use

$$A(S; A_0) = \mathrm{chol}(\exp S)\, Q_0. \tag{67}$$

By construction,

$$A(S; A_0) A(S; A_0)^\top = \exp S. \tag{68}$$

C.2 Fixed-Final-Cell Physics

Before querying the frozen potential, we stop gradients through the final predicted cell:

$$\bar{A}_K = A(\mathrm{sg}[S_K]; \mathrm{sg}[A_0]), \qquad \frac{\partial \bar{A}_K}{\partial \theta} = 0. \tag{69}$$

The live Cartesian coordinates are

$$R_K(\theta) = q_K(\theta) \bar{A}_K. \tag{70}$$

Hence the physical branch supplies a fixed-cell coordinate-energy gradient. It does not directly supervise lattice stress.

Holding the final cell fixed removes the direct derivative through $R_K = q_K \bar{A}_K$. It does not, by itself, specify whether intermediate cell states remain live in the rollout. If they remain live, they can affect subsequent coordinate predictions; if they are detached, that path is absent. The rollout Jacobian in Appendix B.5 is always the Jacobian of the active computational graph, including its stop-gradient operations. No stress or cell-mediated gradient is implied solely by the fixed-final-cell force query.

C.3 Detached-Force VJP

Let the frozen interatomic potential be

$$E_\psi(R, A, Z), \tag{71}$$

with force

$$F_\psi(R, A, Z) = -\nabla_R E_\psi(R, A, Z). \tag{72}$$

Define the translation projector

$$P_N = I_N - \frac{1}{N} \mathbf{1}\mathbf{1}^\top, \tag{73}$$

for which

$$P_N^\top = P_N, \qquad P_N^2 = P_N. \tag{74}$$

Because atoms are stored along the row dimension, $R, F \in \mathbb{R}^{N\times 3}$ and P_N acts from the left. Component-wise,

$$(P_N R)_{ia} = R_{ia} - \frac{1}{N} \sum_{j=1}^{N} R_{ja}, \tag{75}$$

and analogously for the force matrix. Equivalently, after vectorization the projector is $P_N \otimes I_3$.

The per-structure detached-force surrogate is

$$\widetilde{\ell}_{\mathrm{phys}} = -\frac{1}{N}\left\langle \mathrm{sg}\left[P_N F_\psi(R_K(\theta), \bar{A}_K, Z)\right], P_N R_K(\theta)\right\rangle_F . \tag{76}$$

For a minibatch of B structures, with structure-dependent atom counts N_b, the physical objective used to form the raw parameter gradient is

$$\mathcal{L}_{\mathrm{phys}} = \frac{1}{B}\sum_{b=1}^{B} \widetilde{\ell}^{(b)}_{\mathrm{phys}}, \qquad g_{\mathrm{phys}} = \nabla_\theta \mathcal{L}_{\mathrm{phys}}. \tag{77}$$

The normalization $1/N_b$ is already included inside each $\widetilde{\ell}^{(b)}_{\mathrm{phys}}$.

Theorem 2 (Detached-Force VJP Equivalence)**.** *At the current parameter value, under a fixed final cell and a fixed neighbor-graph branch,*

$$\nabla_\theta \widetilde{\ell}_{\mathrm{phys}} = \frac{1}{N} J^\top_{R_K,\theta} P_N \nabla_R E_\psi. \tag{78}$$

If the potential is exactly invariant to global translation, then $P_N \nabla_R E_\psi = \nabla_R E_\psi$ and

$$\nabla_\theta \widetilde{\ell}_{\mathrm{phys}} = \nabla_\theta \left[\frac{1}{N} E_\psi(R_K(\theta), \bar{A}_K, Z)\right]_{\bar{A}_K \text{ fixed}} . \tag{79}$$

Proof. Since the force is detached,

$$\nabla_\theta \widetilde{\ell}_{\mathrm{phys}} = -\frac{1}{N} J^\top_{P_N R_K,\theta} P_N F_\psi \tag{80}$$

$$= \frac{1}{N} J^\top_{P_N R_K,\theta} P_N \nabla_R E_\psi. \tag{81}$$

Using $J_{P_N R_K,\theta} = P_N J_{R_K,\theta}$ and $P_N^\top P_N = P_N$ yields Eq. 78.

If

$$E_\psi(R + \mathbf{1}c^\top, A, Z) = E_\psi(R, A, Z) \tag{82}$$

for every c, differentiation with respect to c gives

$$\mathbf{1}^\top \nabla_R E_\psi = 0, \tag{83}$$

and hence $P_N \nabla_R E_\psi = \nabla_R E_\psi$. □

Importantly, differentiating through the force itself would introduce

$$\frac{\partial F_\psi}{\partial R} = -\nabla^2_R E_\psi. \tag{84}$$

The stop-gradient in Eq. 76 removes this second-order path, so the physical update requires only a vector–Jacobian product with the energy gradient.

D Endpoint-Preserving Physics Injection

All gradient geometry in this section is defined in the Euclidean space of flattened trainable parameters. Let

$$g_{\mathrm{geo}} = \nabla_\theta \mathcal{L}_{\mathrm{geo}}, \qquad g_{\mathrm{phys}} = \nabla_\theta \mathcal{L}_{\mathrm{phys}}. \tag{85}$$

D.1 Orthogonal Physical Component

For $g_{\mathrm{geo}} \neq 0$, define

$$h = g_{\mathrm{phys}} - \frac{g^\top_{\mathrm{geo}} g_{\mathrm{phys}}}{\|g_{\mathrm{geo}}\|^2} g_{\mathrm{geo}}, \tag{86}$$

so that

$$g^\top_{\mathrm{geo}} h = 0. \tag{87}$$

For $g_{\mathrm{phys}} \neq 0$, define the angle $\phi \in [0, \pi]$ by

$$\cos\phi = \frac{g^\top_{\mathrm{geo}} g_{\mathrm{phys}}}{\|g_{\mathrm{geo}}\|\|g_{\mathrm{phys}}\|}. \tag{88}$$

Proposition 3 (Novel Physics Fraction). *If $g_{\text{phys}} \neq 0$, then*

$$\frac{\|h\|}{\|g_{\text{phys}}\|} = \sin\phi. \tag{89}$$

Proof. The parallel and orthogonal components of g_{phys} form an orthogonal decomposition. The parallel component has norm $\|g_{\text{phys}}\||\cos\phi|$, and the result follows from the Pythagorean identity. □

D.2 Optimal Physics Injection

Theorem 3 (Optimal Endpoint-Preserving Physics Injection). *Assume $g_{\text{geo}} \neq 0$, $h \neq 0$, and $\rho \geq 0$. Consider*

$$\max_c \; g_{\text{phys}}^\top c \quad \text{s.t.} \quad g_{\text{geo}}^\top c = 0, \qquad \|c\| \leq \rho\|g_{\text{geo}}\|. \tag{90}$$

Its unique optimum is

$$c^\star = \rho\|g_{\text{geo}}\|\frac{h}{\|h\|}, \qquad g^\star = g_{\text{geo}} + c^\star. \tag{91}$$

Moreover,

$$g_{\text{geo}}^\top g^\star = \|g_{\text{geo}}\|^2. \tag{92}$$

Hence, for the raw-gradient update $\theta^+ = \theta - \eta g^\star$,

$$\mathcal{L}_{\text{geo}}(\theta - \eta g^\star) = \mathcal{L}_{\text{geo}}(\theta) - \eta\|g_{\text{geo}}\|^2 + \mathcal{O}(\eta^2), \tag{93}$$

which has the same first-order endpoint descent rate as using g_{geo} alone. At the same time, $c^\star$ maximizes the additional first-order physical descent within the prescribed trust region.

Proof. Every feasible correction c satisfies $g_{\text{geo}}^\top c = 0$. Decomposing

$$g_{\text{phys}} = g_{\text{phys}}^{\parallel} + h \tag{94}$$

with $g_{\text{phys}}^{\parallel} \parallel g_{\text{geo}}$ therefore gives

$$g_{\text{phys}}^\top c = h^\top c. \tag{95}$$

By Cauchy–Schwarz,

$$h^\top c \leq \|h\|\,\|c\| \leq \rho\|g_{\text{geo}}\|\|h\|. \tag{96}$$

For $h \neq 0$, equality holds if and only if c is aligned with h and lies on the trust-region boundary. Hence the unique optimum is

$$c^\star = \rho\|g_{\text{geo}}\|\frac{h}{\|h\|}. \tag{97}$$

Since $g_{\text{geo}}^\top c^\star = 0$,

$$g_{\text{geo}}^\top(g_{\text{geo}} + c^\star) = \|g_{\text{geo}}\|^2. \tag{98}$$

Therefore

$$\mathcal{L}_{\text{geo}}(\theta - \eta g^\star) = \mathcal{L}_{\text{geo}}(\theta) - \eta g_{\text{geo}}^\top g^\star + \mathcal{O}(\eta^2) \tag{99}$$

$$= \mathcal{L}_{\text{geo}}(\theta) - \eta\|g_{\text{geo}}\|^2 + \mathcal{O}(\eta^2). \tag{100}$$

Thus the raw fused gradient preserves exactly the same first-order endpoint descent rate as g_{geo} alone while maximizing the additional first-order physical descent inside the feasible trust region. □

D.3 Additional Identities

The orthogonality of the correction yields an exact update-norm identity.

Proposition 4 (Pythagorean Update Identity)**.** *In the orthogonal trust-region branch, let ρ denote the actual radius used by the update. Then*

$$\|g^\star\| = \|g_{\mathrm{geo}}\|\sqrt{1+\rho^2}. \tag{101}$$

Proof. Since $g_{\mathrm{geo}} \perp c^\star$ and $\|c^\star\| = \rho\|g_{\mathrm{geo}}\|$,

$$\|g^\star\|^2 = \|g_{\mathrm{geo}}\|^2 + \|c^\star\|^2 = (1+\rho^2)\|g_{\mathrm{geo}}\|^2. \tag{102}$$

□

The corresponding physical margin is also available in closed form.

Proposition 5 (Physics Margin)**.** *In the orthogonal branch,*

$$\frac{g_{\mathrm{phys}}^\top g^\star}{\|g_{\mathrm{phys}}\|\|g_{\mathrm{geo}}\|} = \cos\phi + \rho\sin\phi. \tag{103}$$

Proof. Using $g^\star = g_{\mathrm{geo}} + c^\star$, the first term contributes $\|g_{\mathrm{phys}}\|\|g_{\mathrm{geo}}\|\cos\phi$. Moreover,

$$g_{\mathrm{phys}}^\top h = \|h\|^2, \tag{104}$$

so

$$g_{\mathrm{phys}}^\top c^\star = \rho\|g_{\mathrm{geo}}\|\|h\| = \rho\|g_{\mathrm{geo}}\|\|g_{\mathrm{phys}}\|\sin\phi. \tag{105}$$

Combining the two terms proves Eq. 103. □

D.4 Adaptive Trust Radius

Assume $g_{\mathrm{geo}} \neq 0$, $g_{\mathrm{phys}} \neq 0$, and $0 \leq \rho_0 \leq \rho_{\max}$. For the raw update $\theta^+ = \theta - \eta g^\star$ with $\eta > 0$, first-order non-increase of the physical objective requires

$$g_{\mathrm{phys}}^\top g^\star \geq 0, \tag{106}$$

because its first-order change is $-\eta g_{\mathrm{phys}}^\top g^\star$. The physical-margin identity in the preceding subsection gives the normalized margin as $\cos\phi + \rho\sin\phi$. The minimum required radius is therefore

$$\rho_{\mathrm{req}} := \begin{cases} 0, & \cos\phi \geq 0, \\ -\dfrac{\cos\phi}{\sin\phi}, & \cos\phi < 0,\ \sin\phi > 0, \\ +\infty, & \cos\phi < 0,\ \sin\phi = 0. \end{cases} \tag{107}$$

The first case requires no correction for physical non-increase. The second gives the smallest feasible orthogonal correction. In the antiparallel case, $h = 0$, and no orthogonal correction can improve the physical margin.

For finite ρ_{req}, define the capped radius

$$\rho_{\mathrm{used}} = \min\{\max(\rho_0, \rho_{\mathrm{req}}), \rho_{\max}\}. \tag{108}$$

When $\rho_{\mathrm{req}} \leq \rho_{\max}$ and the orthogonal component is active, we apply Theorem 3 with $\rho = \rho_{\mathrm{used}}$. If $\rho_{\mathrm{req}} > \rho_{\max}$, no correction within the capped orthogonal trust region can preserve the endpoint first-order descent rate while making the physical objective first-order non-increasing. This includes exact antiparallelism. The method therefore invokes the severe-conflict projection in Section E.

D.5 Positive-Scale Invariance

Proposition 6 (Positive-Scale Invariance). *Assume $g_{\mathrm{geo}} \neq 0$ and $h \neq 0$. For any $a > 0$, replacing g_{phys} by ag_{phys} leaves the orthogonal injection update unchanged, provided the same branch remains active and the radius parameters are fixed:*

$$g^\star(g_{\mathrm{geo}}, ag_{\mathrm{phys}}) = g^\star(g_{\mathrm{geo}}, g_{\mathrm{phys}}). \tag{109}$$

Proof. Positive scaling gives $h \mapsto ah$, so

$$\frac{ah}{\|ah\|} = \frac{h}{\|h\|}. \tag{110}$$

It also preserves $\cos\phi$ and $\sin\phi$, leaving ρ_{req} and ρ_{used} unchanged. Hence the correction $\rho_{\mathrm{used}}\|g_{\mathrm{geo}}\|h/\|h\|$ and the resulting update $g^\star$ are unchanged. Absolute-norm activity thresholds may change branch selection; the claim applies only when the same branch remains active. □

D.6 Why a Fixed Weighted Sum Has No Equivalent Guarantee

Consider the naive fused gradient

$$g_\lambda = g_{\mathrm{geo}} + \lambda g_{\mathrm{phys}}. \tag{111}$$

Its alignment with the endpoint gradient is

$$g_{\mathrm{geo}}^\top g_\lambda = \|g_{\mathrm{geo}}\|^2 + \lambda\|g_{\mathrm{geo}}\|\|g_{\mathrm{phys}}\|\cos\phi. \tag{112}$$

If $\cos\phi < 0$, then

$$\lambda > \frac{\|g_{\mathrm{geo}}\|}{\|g_{\mathrm{phys}}\||\cos\phi|} \tag{113}$$

makes $g_{\mathrm{geo}}^\top g_\lambda < 0$, so the corresponding raw update becomes an ascent direction for the endpoint objective. Since both the norm ratio and conflict angle vary during training, a single fixed weight does not provide the endpoint-preservation guarantee of Theorem 3.

E Severe-Conflict Projection

When the orthogonal trust region is insufficient to obtain a physics-safe update, we relax exact first-order endpoint-rate preservation and instead search for the smallest modification of the geometric gradient that is physically non-increasing.

Define the physics-safe half-space

$$\mathcal{H}_{\mathrm{phys}} = \left\{ g : g_{\mathrm{phys}}^\top g \geq 0 \right\}. \tag{114}$$

We solve

$$\min_g \frac{1}{2}\|g - g_{\mathrm{geo}}\|^2 \qquad \text{s.t.} \qquad g_{\mathrm{phys}}^\top g \geq 0. \tag{115}$$

Theorem 4 (Nearest Physics-Safe Gradient). *If $g_{\mathrm{geo}}^\top g_{\mathrm{phys}} < 0$, the unique solution of Eq. 115 is*

$$g^\star = g_{\mathrm{geo}} - \frac{g_{\mathrm{geo}}^\top g_{\mathrm{phys}}}{\|g_{\mathrm{phys}}\|^2} g_{\mathrm{phys}}. \tag{116}$$

Proof. The Lagrangian is

$$\mathcal{J}(g, \lambda) = \frac{1}{2}\|g - g_{\mathrm{geo}}\|^2 - \lambda g_{\mathrm{phys}}^\top g, \qquad \lambda \geq 0. \tag{117}$$

Stationarity gives

$$g = g_{\mathrm{geo}} + \lambda g_{\mathrm{phys}}. \tag{118}$$

Since $g_{\mathrm{geo}}^\top g_{\mathrm{phys}} < 0$, the unconstrained point is infeasible and the optimum lies on the boundary $g_{\mathrm{phys}}^\top g = 0$. Hence

$$\lambda = -\frac{g_{\mathrm{phys}}^\top g_{\mathrm{geo}}}{\|g_{\mathrm{phys}}\|^2}, \tag{119}$$

which yields Eq. 116. □

The fallback satisfies

$$g_{\mathrm{phys}}^{\top} g^{\star} = 0, \tag{120}$$

and

$$g_{\mathrm{geo}}^{\top} g^{\star} = \|g_{\mathrm{geo}}\|^2 \sin^2 \phi = \|g^{\star}\|^2 \geq 0. \tag{121}$$

Thus the endpoint objective remains first-order non-increasing, but its descent rate is reduced by the explicit factor $\sin^2 \phi$.

In the strictly antiparallel case

$$g_{\mathrm{phys}} = -a g_{\mathrm{geo}}, \qquad a > 0, \tag{122}$$

we have $\sin \phi = 0$ and

$$g^{\star} = 0. \tag{123}$$

No direction gives strict first-order descent for both objectives in this limiting case. Directions orthogonal to both gradients have zero first-order change in both objectives; the nearest feasible projection of g_{geo} is nevertheless the zero vector.

F Degenerate Gradient Cases

The stated fusion rule treats degenerate cases asymmetrically because the geometric endpoint objective remains the primary task.

If

$$\|g_{\mathrm{geo}}\| \leq \varepsilon, \tag{124}$$

we return a zero update rather than performing physics-only refinement.

If

$$\|g_{\mathrm{phys}}\| \leq \varepsilon, \tag{125}$$

we use $g^{\star} = g_{\mathrm{geo}}$.

If the novel component is negligible,

$$\|h\| \leq \varepsilon_{\mathrm{novel}}, \tag{126}$$

and the two gradients are non-conflicting, we again use $g^{\star} = g_{\mathrm{geo}}$, since the physical signal provides no additional direction. If instead the gradients are antiparallel or severely conflicting, we use the half-space projection of Sec. E.

G From First-Order Guarantees to Finite-Step Descent

The guarantees in the main paper concern raw-gradient first-order behavior. They can be strengthened locally under explicit smoothness and fixed-branch assumptions.

Assume that the active quotient branches remain unchanged in a neighborhood of the current parameter value and that $\nabla \mathcal{L}_{\mathrm{geo}}$ is L_{geo}-Lipschitz there. The descent lemma gives

$$\mathcal{L}_{\mathrm{geo}}(\theta - \eta g^{\star}) \leq \mathcal{L}_{\mathrm{geo}}(\theta) - \eta g_{\mathrm{geo}}^{\top} g^{\star} + \frac{L_{\mathrm{geo}} \eta^2}{2} \|g^{\star}\|^2. \tag{127}$$

In the orthogonal branch,

$$g_{\mathrm{geo}}^{\top} g^{\star} = \|g_{\mathrm{geo}}\|^2, \qquad \|g^{\star}\|^2 = (1 + \rho^2) \|g_{\mathrm{geo}}\|^2, \tag{128}$$

where ρ is the actual radius used by the update. Therefore

$$0 < \eta < \frac{2}{L_{\mathrm{geo}}(1 + \rho^2)} \tag{129}$$

is sufficient for strict decrease of the geometric objective under the ideal raw-gradient step.

The detached-force surrogate should not itself be treated as a globally smooth scalar potential across parameter updates, because the detached force, final cell, and discrete neighbor graph are refreshed

at each training step. Instead, fix the current parameter value θ_0 and freeze the final cells and neighbor-graph branches used at θ_0. For each structure define

$$R_{K,b}^{(\theta_0)}(\vartheta) := q_{K,b}(\vartheta)\, \bar{A}_{K,b}^{(\theta_0)}, \tag{130}$$

and the corresponding local fixed-cell energy

$$\mathcal{E}_{\mathrm{loc}}^{(\theta_0)}(\vartheta) = \frac{1}{B}\sum_{b=1}^{B} \frac{1}{N_b} E_\psi\Big(R_{K,b}^{(\theta_0)}(\vartheta), \bar{A}_{K,b}^{(\theta_0)}, Z_b\Big). \tag{131}$$

Assume additionally that the potential is exactly invariant to global translation, so that the final clause of Theorem 2 applies. Then, at the current parameter point,

$$\nabla_\vartheta \mathcal{E}_{\mathrm{loc}}^{(\theta_0)}(\vartheta)\Big|_{\vartheta=\theta_0} = g_{\mathrm{phys}}. \tag{132}$$

If this local energy gradient is L_E-Lipschitz in a neighborhood of θ_0 and $g_{\mathrm{phys}}^\top g^\star > 0$, then

$$\mathcal{E}_{\mathrm{loc}}^{(\theta_0)}(\theta_0 - \eta g^\star) \le \mathcal{E}_{\mathrm{loc}}^{(\theta_0)}(\theta_0) - \eta g_{\mathrm{phys}}^\top g^\star + \frac{L_E \eta^2}{2}\|g^\star\|^2, \tag{133}$$

and therefore

$$0 < \eta < \frac{2 g_{\mathrm{phys}}^\top g^\star}{L_E \|g^\star\|^2} \tag{134}$$

is sufficient for strict descent of this local fixed-cell energy. If $g_{\mathrm{phys}}^\top g^\star = 0$, the first-order analysis gives only a zero directional derivative; without additional curvature assumptions it does not certify finite-step non-increase or strict descent.

H Differentiability and Numerical Boundaries

Periodic branch changes. The quotient Log is piecewise differentiable because periodic-image and anchor selections are discrete. All local derivative statements assume that these branches remain fixed.

Neighbor-graph topology. The MACE neighbor graph is constructed from detached positions and held fixed during a force query. Within a fixed graph branch, the potential is differentiable with respect to live positions. Crossing a cutoff or neighbor-reconstruction boundary changes the discrete graph, so the overall map is again piecewise differentiable rather than globally C^1.

Matrix exponential. For multi-step differentiable unrolling, the cell metric is evaluated as

$$G = \exp\left(\frac{S + S^\top}{2}\right), \tag{135}$$

using a matrix-exponential implementation with a stable Fréchet derivative. Since the mathematical state satisfies $S \in \mathrm{Sym}(3)$, this symmetrization is the identity in exact arithmetic. It avoids differentiating through unstable eigenvectors near repeated eigenvalues while preserving the intended state $G = \exp S$.

Zero-distance norms. Self-edges may produce exactly zero squared distances. To prevent an indeterminate $0 \times \infty$ backward path through the square-root implementation of a Euclidean norm, the backward rule returns the zero vector at zero distance. This is a valid subgradient of the Euclidean norm $\|x\|_2$ at $x = 0$ and leaves the forward value unchanged.

I Optimizer Guarantee Boundary

The exact identities in Sec. 4.3 apply to the raw fused gradient $g^\star$.

If global norm clipping simply rescales the gradient by $\kappa \in (0, 1]$,

$$g_{\mathrm{clip}} = \kappa g^\star, \tag{136}$$

then

$$g_{\mathrm{geo}}^{\top} g_{\mathrm{clip}} = \kappa g_{\mathrm{geo}}^{\top} g^{\star}, \qquad g_{\mathrm{phys}}^{\top} g_{\mathrm{clip}} = \kappa g_{\mathrm{phys}}^{\top} g^{\star}. \tag{137}$$

Thus clipping preserves the signs of both first-order margins, although the exact endpoint descent rate is scaled by κ.

AdamW, however, applies momentum, coordinate-wise preconditioning, and decoupled weight decay. Its actual parameter displacement is generally not collinear with $-g^{\star}$. Accordingly, all exact orthogonality and descent-rate statements in this work are *raw-gradient first-order guarantees*, not absolute guarantees on the final AdamW parameter displacement.

J Implementation Diagnostics

The theoretical identities above provide direct numerical checks of the implementation. For nonzero g_{geo} and g_{phys}, we define the normalized physical margin of a candidate gradient g as

$$m_{\mathrm{phys}}(g) := \frac{g_{\mathrm{phys}}^{\top} g}{\|g_{\mathrm{phys}}\| \, \|g_{\mathrm{geo}}\|}.$$

The normalization uses the same gradient norms for both the geometry-only and fused updates. The additional normalized physical margin introduced by the correction is

$$\Delta m_{\mathrm{phys}} := m_{\mathrm{phys}}(g^{\star}) - m_{\mathrm{phys}}(g_{\mathrm{geo}}) = \frac{g_{\mathrm{phys}}^{\top}(g^{\star} - g_{\mathrm{geo}})}{\|g_{\mathrm{phys}}\| \, \|g_{\mathrm{geo}}\|}.$$

In the orthogonal trust-region branch, Eq. 103 gives $\Delta m_{\mathrm{phys}} = \rho_{\mathrm{used}} \sin\phi$. The following identities can therefore be checked numerically:

$$\frac{g_{\mathrm{geo}}^{\top} g^{\star}}{\|g_{\mathrm{geo}}\|^2} \approx 1, \tag{138}$$

$$\frac{\|g^{\star}\|}{\|g_{\mathrm{geo}}\|} \approx \sqrt{1 + \rho_{\mathrm{used}}^2}, \tag{139}$$

$$\frac{\|h\|}{\|g_{\mathrm{phys}}\|} \approx \sin\phi, \tag{140}$$

and

$$\Delta m_{\mathrm{phys}} \approx \rho_{\mathrm{used}} \sin\phi. \tag{141}$$

For the severe-conflict branch, the corresponding checks are

$$g_{\mathrm{phys}}^{\top} g^{\star} \approx 0, \tag{142}$$

and

$$g_{\mathrm{geo}}^{\top} g^{\star} \approx \|g^{\star}\|^2. \tag{143}$$

Finally, $K = 1$ must reproduce the direct endpoint tensor exactly, whereas $K > 1$ may produce a nonzero direct-to-unrolled shift whenever the learned endpoint field is state-dependent. These identities define implementation checks, not additional optimization objectives.

K Data and Experimental Setup

Datasets. We evaluate coordinate recovery for finite local patches on $SrTiO_3$ and periodic crystal relaxation on X–Mn–O. The $SrTiO_3$ dataset comprises atomic observations collected in an industrial setting, with the original ground-truth annotations used as target structures. Atomic identities and correspondences remain fixed within each nonperiodic patch. We partition the data by spatial region and introduce buffer zones around the partition boundaries.

Physical supervision and training. A frozen MACE model provides force feedback on predicted structures. For $SrTiO_3$, the force-query environment is constructed from a lattice fitted to the input structure. Surrounding atoms remain fixed, and only forces on central atoms contribute to supervision. For periodic X–Mn–O structures, the predicted cell is held fixed during each force query, and no stress supervision is used. The returned forces are detached and used to construct physical gradients through the active paths of the unrolled predictor computation graph.

For $SrTiO_3$, the endpoint model is pretrained with geometric supervision for 100 epochs and then fine-tuned for 10 epochs. Fine-tuning uses AdamW with a learning rate of 10^{-5}, a batch size of 4, and random seed 42. In the physical supervision branch, one patch from the current batch is selected for physical feedback every 20 training batches. For X–Mn–O, training uses a learning rate of 10^{-5}, a batch size of 32, and random seed 42 for 300 epochs. Physical and audit queries are performed at every training step. The remaining optimizer and dataset-specific settings are specified in the corresponding run configurations.

Fine-tuning across backbones. For each external backbone, all variants start from the same backbone-specific pretrained checkpoint. We reinitialize the optimizer and use paired data ordering and sampling seeds across variants. Each variant receives exactly 330 parameter updates. The two variants with physical supervision query MACE every 20 updates, resulting in 16 queries per variant. The baseline makes no MACE queries.

Both physics-supervised variants use the same forward-sampling procedure and terminal force-query protocol. Local guidance propagates feedback only through the final prediction, with the preceding sampling state detached. Differentiable unrolling additionally propagates feedback through earlier sampling steps. All external-backbone variants are evaluated using the checkpoint obtained after the final fine-tuning update.

Inference. Our primary endpoint model requires a single network forward pass, whereas the adapted external backbones use 20 Euler steps. No model queries MACE during inference. For our backbone, the cross-backbone comparison reuses configurations A, B, and E from Table 3 and retains their single-forward-pass evaluation protocol. These results are reused from the main ablation rather than obtained from an additional external-backbone fine-tuning run. Auxiliary multi-step diagnostic rollouts are reported separately.

L Evaluation Metrics and Matching Criteria

Aggregation and units. We compute each metric per structure and then average equally over the evaluation set. Coordinate, bond, and lattice errors are reported in Å, volume errors in $Å^3$, and relative volume errors and match rates in percent. Lower errors and higher match rates are better. Coordinate and bond errors use the prescribed atom correspondences.

Coordinate and bond errors. For a structure with N atoms, coordinate MAE is

$$\text{Coord.} = \frac{1}{3N} \sum_{i=1}^{N} \|\delta r_i\|_1. \tag{144}$$

For finite $SrTiO_3$ patches, $\delta r_i = \hat{r}_i - r_i$ without translation or rotation alignment. For periodic X–Mn–O (Ong et al., 2013), δr_i is the residual after selecting the closest corresponding reference image in a target-cell periodic-image search; no rigid alignment or atom reassignment is applied. Bond MAE is the mean absolute difference between predicted and reference distances over a fixed reference edge set $\mathcal{E}$:

$$\text{Bond} = \frac{1}{|\mathcal{E}|} \sum_{e \in \mathcal{E}} |\hat{\ell}_e - \ell_e|. \tag{145}$$

For $SrTiO_3$, this set contains reference Ti–O and Sr–O pairs within 2.5 and 3.2 Å, respectively, without periodic images. For X–Mn–O, we use the fixed reference periodic neighbor graph with a 6 Å cutoff. Predicted atoms are first assigned corresponding periodic images using the reference metric; edge lengths then use the fixed reference edge indices and image shifts with each structure's own cell. The same reference edges are used for every method.

Lattice and volume errors. Let $\hat{A}$ and A denote predicted and reference row-basis cell matrices. Raw and aligned lattice MAEs are

$$\begin{aligned} \text{Lat.}-\text{R} &= \frac{1}{9}\sum_{a,b}|\hat{A}_{ab}-A_{ab}|, \\ \text{Lat.}-\text{A} &= \frac{1}{9}\sum_{a,b}|(\hat{A}Q^{\star}-A)_{ab}|, \qquad Q^{\star}=\arg\min_{Q\in\mathrm{SO}(3)}\|\hat{A}Q-A\|_F. \end{aligned} \tag{146}$$

The alignment permits a proper rotation, without basis permutation or lattice reduction; Lat.-A is not a lattice-angle error. With $\hat{V} = |\det \hat{A}|$ and $V = |\det A|$, we report $\text{Vol.} = |\hat{V} - V|$ and $\text{Rel.V} = 100|\hat{V} - V|/V$. For finite $SrTiO_3$ patches, cells are inferred separately from predicted and reference coordinates by least-squares affine fits to the same input-derived lattice prior, including a translation term. These are primitive-cell diagnostics, not predicted periodic cells or crop volumes. Periodic X–Mn–O metrics use the supplied predicted and reference cells directly.

Matching criteria. For M finite patches, a patch matches only if every corresponding atom is within 0.1 Å of its reference coordinate:

$$\text{Match}_{\text{patch}} = \frac{100}{M}\sum_{m=1}^{M}\mathbf{1}\left[\max_i \|\hat{r}_{m,i}-r_{m,i}\|_2 \le 0.1\,\text{Å}\right]. \tag{147}$$

This criterion uses the full valid patch, without alignment, symmetry matching, or periodic-supercell reconstruction. For periodic X–Mn–O, Match is the percentage of structures accepted by `pymatgen.StructureMatcher`, with `ltol=0.2`, `stol=0.3`, `angle_tol=5`, and `primitive_cell=True` Degenerate or orientation-inconsistent predicted cells are counted as unmatched. Because matching permits volume rescaling, volume accuracy is evaluated separately. The finite-patch and periodic match rates have different definitions and are not compared across tasks.

M Visual Comparisons of Structural Recovery

Figure 4 presents three selected examples from each dataset. For $SrTiO_3$, the predicted atomic arrangements closely match the reference structures, correcting the positional deviations visible in the inputs. For X–Mn–O, all three selected examples show lower coordinate errors than their inputs, although residual differences in atomic positions and cell geometry remain. These examples illustrate successful aspects of structural recovery and complement the aggregate results reported in the main text. The examples were selected from low-coordinate-error $SrTiO_3$ cases and X–Mn–O cases showing improvement over the input; they do not constitute a random sample of either dataset.

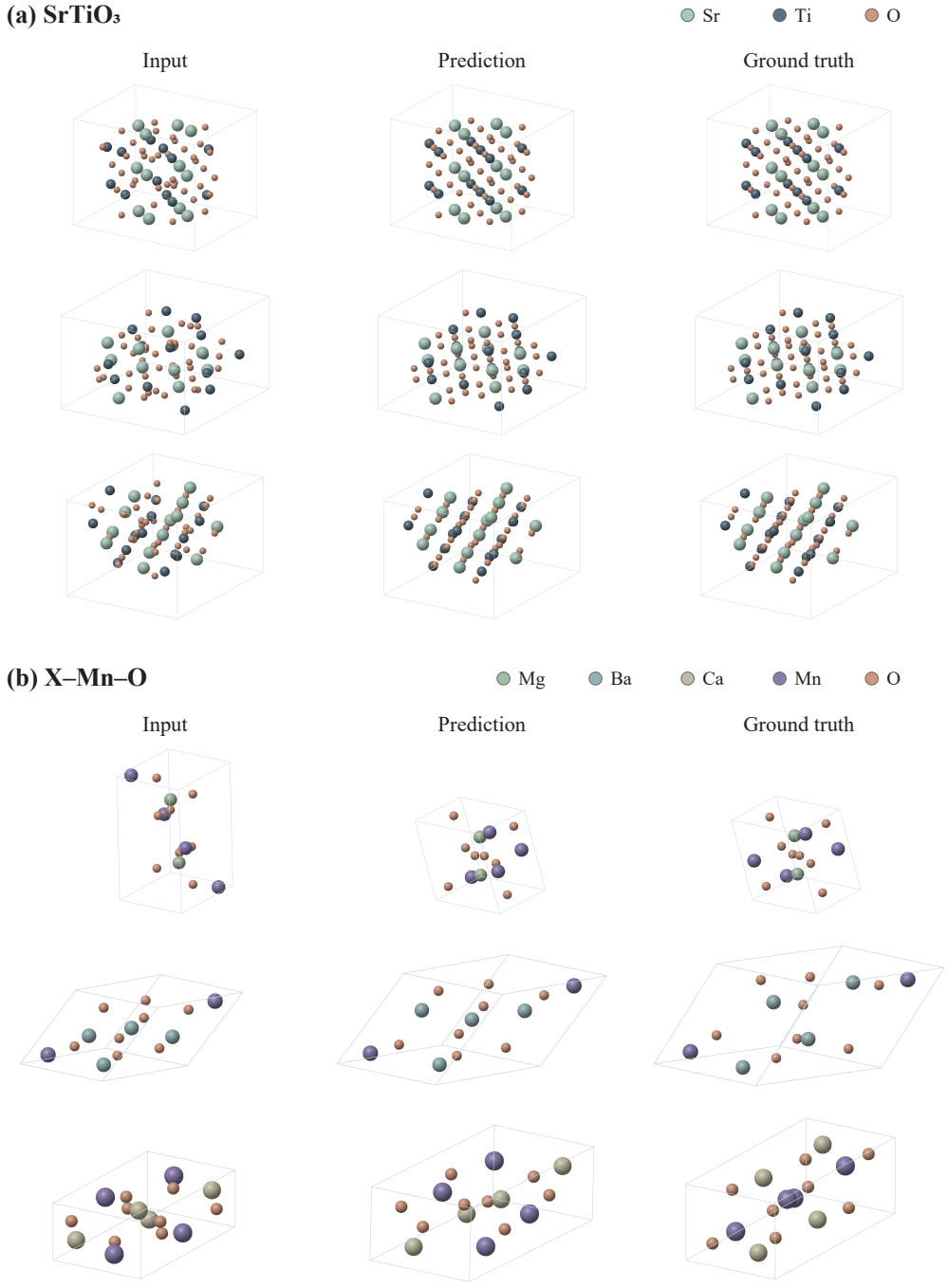


Figure 4: Visual comparisons of structural recovery on (a) $SrTiO_3$ and (b) X–Mn–O. Each dataset includes three selected examples. Columns show the input structure, the CrySPFlow prediction, and the reference structure, respectively. Atom colours identify the elements indicated in the legends.