ASPIRE: SADDLE-POINT DISCOVERY THROUGH SET PREDICTION AND PHYSICAL REFINEMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Predicting thermally activated diffusion and defect evolution with event-driven models requires identifying atomic rearrangement mechanisms and their activation barriers. Discovering the associated saddle points is a major computational bottleneck: multiple rearrangements may originate from one metastable state, while costly local searches can fail or repeatedly converge to the same saddle. To address this challenge, we introduce *ASPIRE* (**A**tomistic **S**addle-**P**oint **I**nference with **R**efinement for **E**vents), a framework that predicts a set of saddle candidates from a single initial atomic environment and refines them through Dimer searches on the original interatomic potential. The framework's equivariant set predictor, Ev-Quiformer, integrates (i) geometry-conditioned scalar/vector event slots for generating multiple saddle-point proposals and (ii) a decoder that maps each slot to a full atomic displacement field by combining atom, slot, and anchor-relative vectors with invariant coefficients. We also contribute two datasets: (i) BCCFE-4VACAV-4000, comprising 4,000 four-vacancy body-centered cubic iron configurations and 65,450 reference events grouped by initial state for set supervision and post-refinement evaluation; and (ii) BCCFE-1TO4VAC, comprising 5,372 configurations with one to four vacancies each. Theoretically, we establish conditions for proposal equivariance. Experimentally, ASPIRE achieves 77.20% reference-event coverage on this benchmark, compared with 75.73% for a conventional Dimer baseline, while requiring approximately half as many Dimer force evaluations. In a timing evaluation on 50 configurations, ASPIRE reduces wall time per configuration from 478.8 s to 176.3 s under the stated hardware settings. The code and dataset are available at https://anonymous.4open.science/r/ASPIRE-17B3/.

1 INTRODUCTION

Nuclear fission reactors and future fusion systems require structural materials that retain their mechanical integrity and dimensional stability under irradiation and elevated temperatures (Du et al., 2023). In iron-based alloys, irradiation-induced clusters and dislocation loops contribute to hardening (Shimodaira et al., 2018), while the growth of cavities causes swelling (Lin et al., 2021). Understanding these changes requires resolving the individual and collective atomic rearrangements that drive defect migration, recombination, and interactions. For example, two interstitial loops in irradiated iron can interact to form a new loop structure, a mechanism revealed by atomistic simulation (Xu et al., 2013). Vacancy diffusion also enables dislocation climb during creep, with migration barriers influenced by the local dislocation environment (Kabir et al., 2010).

Molecular dynamics (MD) follows atomic motion directly, but sampling rare thermally activated events requires resolving fast atomic vibrations over long waiting times. Event-based methods reach longer time scales by identifying the possible rearrangements and using their rates to select events and advance time (Voter, 2007). On-the-fly methods, including adaptive KMC (Henkelman & Jónsson, 2001), SEAKMC (Xu et al., 2012; 2015), and k-ART (Béland et al., 2011), search for these events as the simulation proceeds. In event-based simulations, saddle discovery identifies transition mechanisms and activation barriers used to estimate event rates. Saddle discovery in this setting has several challenges: (i) an initial configuration generally admits multiple possible events whose number and product configurations are unknown in advance; (ii) the recovered event catalogue affects event selection and estimated waiting times; (iii) and new configurations encountered along a tra-

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

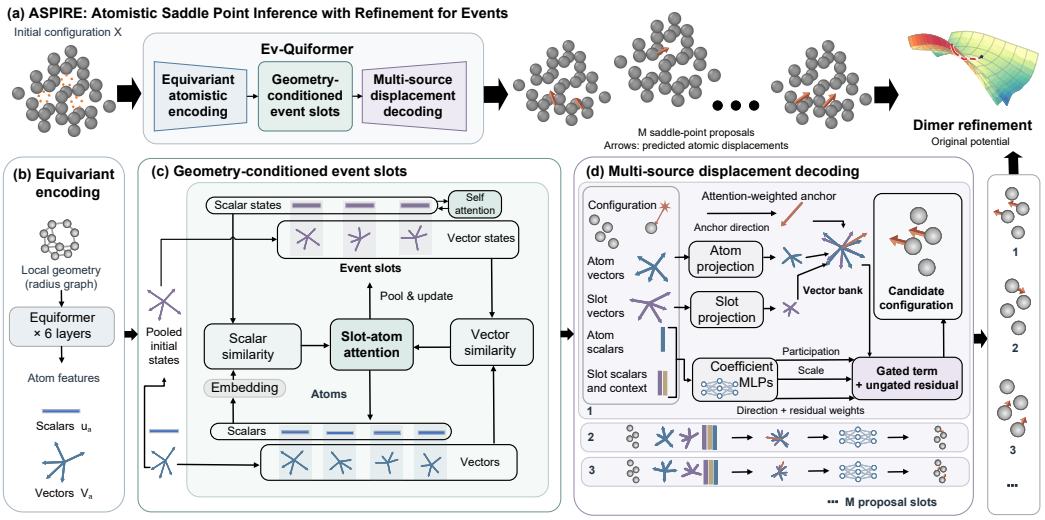


Figure 1: **ASPIRE: Atomistic Saddle-Point Inference with Refinement for Events. (a) Overall workflow.** Ev-Quiformer generates M saddle-point proposals from a single initial configuration, followed by physical refinement using Dimer searches on the original interatomic potential. Panels (b–d) detail its three neural modules. **(b) Equivariant encoding.** A six-layer Equiformer extracts invariant scalar and equivariant vector features for each atom. **(c) Geometry-conditioned event slots.** Scalar self-attention and repeated slot–atom attention update slots initialized from learned scalar seeds and pooled atom vectors. **(d) Multi-source displacement decoding.** Invariant coefficients combine atom, slot, and anchor-relative vectors into a full atomic displacement field through a gated term and an ungated residual. Adding this field to the initial configuration yields a candidate for subsequent physical refinement.

jectory may require renewed searches, making event discovery computationally costly (Hayakawa & Xu, 2021; Liang & Xu, 2023).

Existing learning-based saddle (transition-state) discovery methods differ in their required inputs and target applications. React-OT predicts transition states from reactant and product configurations (Duan et al., 2025), and OA-ReactDiff is primarily evaluated in the same double-ended setting (Duan et al., 2023). Conditioning on both endpoints does not directly address on-the-fly saddle discovery when products are unknown. ReactionAtlas instead proposes transition states from reactants and uses kinetic modeling to expand reaction networks (Gugler et al., 2026).

Our focus is efficient saddle-point prediction for multi-vacancy configurations in body-centered cubic (bcc) iron (Fe), a model system for studying defect evolution in iron-based materials. We use a fixed embedded-atom method (EAM) potential for both reference-event generation and candidate refinement. We aim to identify multiple distinct events from a single initial configuration, without supplied product states, while reducing physical search cost. A propose-refine paradigm uses learned models to generate candidates for numerical refinement, aiming to shorten the search while retaining the underlying physical model. Variants of this paradigm have been used in a range of scientific applications, including electronic-structure calculations (Schütt et al., 2019), protein structure prediction (Jumper et al., 2021), and materials discovery (Merchant et al., 2023).

Our contributions. We formulate saddle discovery in interacting defect systems as a set-learning problem, where multiple candidate rearrangements are jointly predicted from a single initial configuration without supplied product states. We introduce *ASPIRE* (Atomistic Saddle-Point Inference with Refinement for Events; Figure 1), which combines geometric prediction with physical refinement, and establish dedicated datasets and a benchmark for evaluating distinct-event recovery and physical search cost. Our contributions are threefold:

(i) Problem formulation and datasets. We make multiple rearrangements from the same initial state a joint learning target, without requiring product configurations. Our bcc Fe benchmark groups 65,450 reference events by 4,000 four-vacancy configurations, enabling set supervision and evaluation of distinct event recovery after refinement.

(ii) Framework design and theory. We propose ASPIRE, combining learned saddle proposals with physical refinement. Its neural predictor, Ev-Quiformer, integrates geometry-conditioned scalar/vector event slots and a three-source decoder to generate full atomic displacement fields, supervised by sequential matching. We establish conditions for proposal equivariance and relate geometric error to reference-event recovery under local-refinement assumptions.

(iii) Empirical evaluation. We compare ASPIRE with conventional Dimer and event-reuse baselines, alongside ablation studies. On 200 four-vacancy test configurations, ASPIRE achieves 77.20% reference-event coverage versus Dimer's 75.73%, with roughly half the recorded search-force calls under the respective search protocols.

2 Related Work

Rare-event simulation, saddle search, and reuse. Kinetic Monte Carlo (KMC) simulates transitions between metastable states by sampling events and waiting times from their rates (see Appendix A.5) (Voter, 2007). On-the-fly KMC discovers events through saddle searches during simulation, rather than requiring a predefined event catalogue. SEAKMC localizes these searches to active volumes around defects (Xu et al., 2012; 2015). The Dimer method (Henkelman & Jónsson, 1999) finds first-order saddle points without a supplied product configuration (see Appendix A.4). The kinetic database (KDB) matches local atomic geometries to stored mechanisms and supplies saddle-search initializations (Terrell et al., 2012). k-ART combines on-the-fly discovery with an event catalogue indexed by local topology for event reuse (Béland et al., 2011).

Equivariant atomistic models. Tensor Field Networks, SE(3)-Transformers, EGNN, and Equiformer encode geometric transformation rules (Thomas et al., 2018; Fuchs et al., 2020; Satorras et al., 2021; Liao & Smidt, 2022). NequIP and MACE learn interatomic energies and forces (Batzner et al., 2022; Batatia et al., 2022).

Set prediction. Set Transformer uses learnable seeds for attention-based pooling and self-attention among pooled outputs (Lee et al., 2019). DETR combines learned queries, query self-attention, cross-attention, and bipartite matching for unordered object prediction (Carion et al., 2020). Slot Attention iteratively refines slots (Locatello et al., 2020). Geometric query designs include DAB-DETR's dynamic anchor boxes (Liu et al., 2022) and Invariant Slot Attention's attention-derived slot-centric reference frames and relative-coordinate decoding (Biza et al., 2023). For one-to-many supervision, Group DETR performs matching within query groups and MS-DETR supervises multiple queries in the primary decoder (Chen et al., 2023; Zhao et al., 2024). We adapt these set-prediction designs with equivariant atomistic representations to build Ev-Quiformer for jointly predicting an unordered set of candidate saddle configurations from a single initial configuration.

3 Problem Formulation

Given an initial metastable configuration, our goal is to predict a set of saddle candidates for subsequent physical refinement, without requiring product states. Let $X \in \mathbb{R}^{N\times 3}$ be an initial configuration at a *metastable state*, a nondegenerate local minimum of the fixed interatomic potential $V : \mathbb{R}^{N\times 3} \to \mathbb{R}$, with $\nabla V(X) = 0$ and $\nabla^2 V(X) \succ 0$. Here ∇V and $\nabla^2 V$ are taken in the fixed free-coordinate space, with the surrounding environment and constraints already imposed (Appendix A.3). We consider the *open saddle configuration set prediction problem*, in which the number, identities, and outcomes of accessible events are unknown a priori.

A predictor outputs an unordered collection of displacement fields, denoted by

$$F_\theta(X) = \{\widehat{D}_m\}_{m=1}^{M}, \qquad \widehat{D}_m \in \mathbb{R}^{N\times 3}. \tag{1}$$

The corresponding proposal configurations are $\{\widehat{S}_m\}_{m=1}^{M} = \{X + \widehat{D}_m\}_{m=1}^{M}$. After refinement, the set of retained saddle configurations is defined as

$$\{S_k\}_{k=1}^{K} = \{\mathcal{R}_X(\widehat{S}_m)\}_{m=1}^{M} \setminus \{\perp\}, \qquad K \le M, \tag{2}$$

where each S_k is a configuration at a *transition state* satisfying: (i) S_k is a first-order saddle on V, that is, $\nabla V(S_k) = 0$ and $\lambda_1\big(\nabla^2 V(S_k)\big) < 0 < \lambda_2\big(\nabla^2 V(S_k)\big)$, where $\lambda_1 \le \lambda_2 \le \cdots$ are

the Hessian eigenvalues and the unit eigenvector $v_1(S_k)$ associated with λ_1 is its unstable mode; (ii) relaxing from $S_k \pm \epsilon v_1(S_k)$ reaches X and another distinct product configuration X_k^+ at a metastable local minimum. The map $\mathcal{R}_X : \mathbb{R}^{N\times 3} \to \mathbb{R}^{N\times 3} \cup \{\perp\}$ is an idealized search-and-validation map: it refines each proposed configuration on V and returns S_k only when the target saddle and connectivity conditions above hold, or $\perp$ otherwise.

We define the resulting event set as

$$\mathcal{E}_X = \{e_k\}_{k=1}^K = \left\{(X, S_k, X_k^+, \Delta E_k)\right\}_{k=1}^K, \qquad \Delta E_k = V(S_k) - V(X) > 0. \tag{3}$$

The resulting event set is determined by X and the validated saddle configurations. The objective is to recover as many distinct events as possible at low search cost, without using any product configuration X_k^+ as input. Throughout this paper, we use Dimer-based search and validation as the numerical instantiation of this framework (see Appendix A.4).

4 METHODS

Intuition Several atomic rearrangements may originate from the same metastable configuration, each associated with a saddle. We therefore predict a set of approximate saddle geometries and use them to initialize Dimer refinement on the original interatomic potential. Learning proposes alternative rearrangements, while force-based refinement seeks nearby saddle points. Proposition 1 establishes local convergence of the ideal minimum-mode iteration from a sufficiently close initialization (Luo et al., 2022).

Based on this intuition, we develop *ASPIRE* (Figure 1), integrating set prediction with physical refinement. We first describe its neural predictor, ***Ev-Quiformer***, through equivariant environment encoding (Section 4.1), geometry-conditioned event slots (Section 4.2), and three-source displacement decoding (Section 4.3). We then present the set-matching training objective (Section 4.4) and the optional confidence readout for selecting proposals for Dimer refinement (Section 4.5). The conditions and proof for proposal equivariance are given in Appendix J.2.

4.1 EQUIVARIANT ATTENTION ENVIRONMENT ENCODING

The encoder reads the initial configuration X. A directed radius graph connects atoms a, b within a maximum radius distance, with a limit to the number of neighbors per atom and edge vectors $X_a - X_b$, where $X_a \in \mathbb{R}^3$ is the a-th row of X. Six Equiformer blocks (Liao & Smidt, 2022) attend over these neighbors and return, for every atom, an invariant scalar feature $u_a \in \mathbb{R}^{128}$ and an equivariant vector feature $V_a \in \mathbb{R}^{64\times 3}$ that rotates with the input. No product configurations X_k^+ are supplied. Edge features and layer sizes are given in Appendix C.1.

4.2 GEOMETRY-CONDITIONED EVENT SLOTS

Slots. A slot m carries a scalar fingerprint initialized from a learned seed $s_m \in \mathbb{R}^{d_s}$ and an equivariant vector state $Q_m \in \mathbb{R}^{d_v\times 3}$ constructed from the encoded atomic geometry. Learned scalar seeds distinguish hypotheses, while geometry-pooled vector states retain directional information without introducing fixed learned spatial axes.

Stages. The slot module proceeds through initialization, vector–scalar updates, scalar-only updates, and final readout, using Equations (4)–(8) with the stage types summarized in Table 1. At stage t, let $\hat{s}_m$ and $\hat{u}_a$ denote the corresponding normalized slot and atomic scalar features, and let $H^{(t)}$ be the number of attention heads with head width d_h. The exact stage schedule, normalizations, initialization, dimensions, and parameter-sharing choices are given in Appendix C.2.

Slot self attention. Fingerprints optionally self-attend across slots, in place,

$$s_m^{(t)} \leftarrow (1-\eta^{(t)})s_m^{(t)} + \eta^{(t)}\ \mathrm{SA}^{(t)}\big(s_1^{(t)}, \ldots, s_M^{(t)}\big)_m, \qquad \eta^{(t)} \in \{0,1\} \tag{4}$$

Slot–atom attention. At stage t, attention head h uses scalar similarity and, when enabled, vector alignment to compute the slot–atom attention weights:

$$A_{ma}^{(t,h)} = \mathrm{softmax}_a\left[\frac{(W_q^{(t,h)}\hat{s}_m)^\top (W_k^{(t,h)}\hat{u}_a)}{\sqrt{d_h}} + \gamma^{(t,h)}\frac{\left\langle \widetilde{W}_Q^{(t,h)}Q_m^{(t)}, \widetilde{W}_V^{(t,h)}V_a\right\rangle_F}{\sqrt{3d_v}}\right]. \tag{5}$$

Each head pools a scalar and a vector context; scalar contexts are concatenated over heads,

$$\begin{gathered} z_m^{(t,h)} = \sum_a A_{ma}^{(t,h)}\, W_z^{(t,h)}\hat{u}_a, \qquad z_m^{(t)} = \left[z_m^{(t,1)};\ldots;z_m^{(t,H^{(t)})}\right] \in \mathbb{R}^{d_s}, \\ U_m^{(t,h)} = \sum_a A_{ma}^{(t,h)}\, W_V^{(t,h)}V_a \in \mathbb{R}^{d_v\times 3}. \end{gathered} \tag{6}$$

The scalar maps act on invariant features, while vector maps mix channels identically across spatial components. Their exact parameterization and sharing pattern are specified in Appendix C.2.

Update. Using the pooled scalar and vector contexts, we update the slot states according to

$$s_m^{(t+1)} = s_m^{(t)} + F^{(t)}\Big(s_m^{(t)}, z_m^{(t)}, \big(\|U_{m,v}^{(t,h)}\|_2\big)_{h,v}\Big), \qquad Q_m^{(t+1)} = Q_m^{(t)} + \sum_h W_U^{(t,h)}U_m^{(t,h)}. \tag{7}$$

$F^{(t)}$ is zero, a two-layer MLP on its three arguments, or, in the multi-head stages,

$$F^{(t)}(s,z,\cdot) = W_o^{(t)}z + \mathrm{FF}^{(t)}\big(\mathrm{LN}^{(t)}(s + W_o^{(t)}z)\big), \tag{8}$$

which ignores the vector norms. The final readout returns $A_{ma}^{(T)}$, $z_m^{(T)}$, $s_m^{(T)}$, and $Q_m^{(T)}$ for the decoder in Section 4.3.

Table 1: **Slot stage types.** Settings of Equations (4)–(8). "MLP" is a two-layer perceptron on $[s, z, \|U\|]$; "$W_o z$+FF" is Equation (8); "–" means the map is absent; γ of the vector–scalar stage is learned from initial value 1. Detailed stage counts and parameterization are given in Appendix C.2.

	Scalar attn.			Vector attn.	Update		
Stage type	H	η	N	γ	F	W_U	Updated
Initialization	1	0 or 1	Id	0	0	I	$s_m^{(t+1)} = s_m^{(t)}$, $Q_m^{(t+1)} = U_m^{(t)}$
Vector–scalar	1	0	Id	learned	MLP	learned	$s_m^{(t+1)}$, $Q_m^{(t+1)}$
Scalar-only	8	0	LN	0	$W_o z$+FF	–	$s_m^{(t+1)}$, $Q_m^{(t+1)} = Q_m^{(t)}$
Readout	1	0	Id	1	–	–	none; outputs $A_{ma}^{(T)}, z_m^{(T)}, s_m^{(T)}, Q_m^{(T)}$

4.3 Displacement decoding from three vector sources

Using the final readout outputs, with stage indices omitted, we form a bank of 32 atom-derived vectors, eight slot-derived vectors, and one direction from the anchor $b_m = \sum_a A_{ma}X_a$ to atom a:

$$\begin{gathered} \mathcal{B}_{ma} = \left[\, W_{\mathrm{atom}}V_a;\ W_{\mathrm{slot}}Q_m;\ \mathrm{unit}_\epsilon\big(X_a - b_m\big)\,\right] \in \mathbb{R}^{41\times 3}, \\ \mathrm{unit}_\epsilon(v) = \frac{v}{\max(\|v\|_2, \epsilon)}, \quad \epsilon = 10^{-8}, \end{gathered} \tag{9}$$

Here W_{atom} and W_{slot} are shared, learned, bias-free vector-channel maps.

Invariant conditioning. The decoder is conditioned on the invariant features

$$\begin{gathered} \xi_{ma} = \left[\, s_m,\ z_m,\ u'_a,\ \chi_{ma},\ \log\big(1 + \|X_a - b_m\|_2^2\big)\right], \\ \chi_{ma} = \left(\tfrac{1}{\sqrt{3}}\big\langle (G_Q Q_m)_g,\ (G_V V_a)_g\big\rangle\right)_{g=1}^{8}, \end{gathered} \tag{10}$$

where $u'_a = u_a^{(T)}$ is the readout's atomic embedding, and G_Q, G_V are learned vector-channel maps.

Coefficient networks and displacement. Four separately parameterized MLPs, shared across atoms and slots, give

$$\begin{aligned} p_{ma} &= \sigma\big(g_{\mathrm{act}}(\xi_{ma})\big), \quad \kappa_{ma} = \mathrm{softplus}\big(g_{\mathrm{mag}}(\xi_{ma})\big), \\ q_{ma} &= f_{\mathrm{coef}}(\xi_{ma}), \qquad q^{\mathrm{res}}_{ma} = f_{\mathrm{res}}(\xi_{ma}), \end{aligned} \tag{11}$$

where σ is the sigmoid and $q_{ma}, q^{\mathrm{res}}_{ma} \in \mathbb{R}^{41}$. The resulting displacement is

$$\widehat{D}_{m,a} = \underbrace{p_{ma}}_{\text{participation}} \underbrace{\kappa_{ma}}_{\text{scale}} \underbrace{\mathrm{unit}_{\epsilon}\big(q_{ma}^{\top}\mathcal{B}_{ma}\big)}_{\text{direction}} + \gamma_{\mathrm{res}}\,(q^{\mathrm{res}}_{ma})^{\top}\mathcal{B}_{ma}, \tag{12}$$

This gives atom a's displacement for proposal m, with learned global scale γ_{res}. The gate permits localized motion without hard sparsity, while the ungated residual accommodates surrounding relaxations. For a vacancy event, this allows the proposal to describe both the atoms moving toward vacant sites and the smaller displacements of their neighbors, without selecting the moving atoms in advance. The periodic coordinate construction is described in Appendix A.2; network widths and initialization are given in Appendix C.3.

4.4 Set matching and training objective

Let $\mathscr{E}_x = \{1, \ldots, K\}$ index the reference set of X, with saddle displacement fields $D_e = S_e - X$. Slots have no prescribed targets. Using the scalar SmoothL1 penalty ρ_β, define the proposal–target cost over all N atoms by

$$C_{me} = \ell_\beta(\widehat{D}_m - D_e), \qquad \ell_\beta(E) = \frac{1}{N}\sum_{a,d} \rho_\beta(E_{ad}). \tag{13}$$

Rectangular Hungarian matching gives the minimum-cost one-to-one first-round assignment Π_1, assigning each target to a distinct slot when $M \geq K$. We remove the matched slots and repeat against the full catalogue for at most $R_{\mathrm{match}} = 4$ rounds in total, stopping when no slots remain. Each slot is used at most once, whereas a target may supervise multiple slots across rounds. For $R_{\mathrm{match}} = 1$, the auxiliary term is omitted. The auxiliary term is also defined to be zero whenever no pair is matched after the first round, including when the first round exhausts all slots.

Let Π_+ denote all pairs matched after the first round. The objective is

$$\mathcal{L}_x = \lambda_D \left[\frac{1}{|\Pi_1|} \sum_{(m,e)\in\Pi_1} C_{me} + \frac{\lambda_+}{|\Pi_+|} \sum_{(m,e)\in\Pi_+} C_{me} \right], \qquad \lambda_D = 10, \quad \lambda_+ = 0.5. \tag{14}$$

4.5 Confidence readout for proposal selection

We hypothesize that features learned through displacement matching also encode catalogue agreement. We freeze the predictor and train a lightweight MLP *judge* on invariant latent and geometric features (Appendix E). It outputs $\widehat{p}_m \in (0, 1)$ without force evaluations; only proposals with $\widehat{p}_m \geq \tau$ undergo Dimer refinement.

For proposal $\widehat{D}_m$ and catalogue field D_e, define the moving-union root-mean-square (RMS) error by

$$\begin{aligned} U_{me} &= \{a : \max(\|\widehat{D}_{m,a}\|_2, \|D_{e,a}\|_2) \geq 0.3\,\text{Å}\}, \\ \mathrm{RMS}_U(\widehat{D}_m, D_e) &= \left(\frac{1}{|U_{me}|} \sum_{a\in U_{me}} \|\widehat{D}_{m,a} - D_{e,a}\|_2^2 \right)^{1/2}. \end{aligned} \tag{15}$$

Empty unions have infinite error. For the nearest catalogue field D_{e^*} under this RMS, set $y_m = 1$ iff the RMS is below $0.35\,\text{Å}$ and the maximum atomwise error over U_{me^*} is below $0.5\,\text{Å}$. Judge selection provides a tunable trade-off between reference-event recovery and physical search cost. The judge is trained on the training set only.

5 Experiments and Results

5.1 Datasets

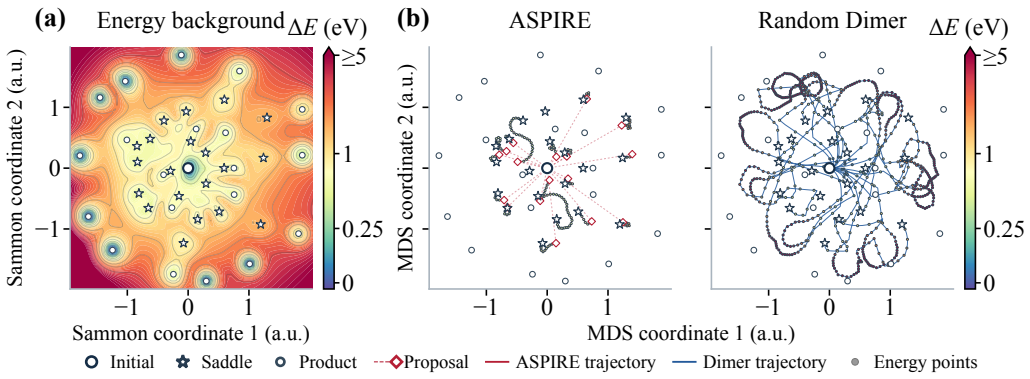


Figure 2: **Energy background, saddles, and trajectories.** **(a)** Sammon-projected view of the energy function and initial, saddle, and product configurations. **(b)** MDS-coordinate views of ASPIRE and random-start Dimer searches on one configuration. ASPIRE proposal jumps effectively predict approximate saddles, followed by light Dimer refinement, whereas random Dimer goes long paths; Each trajectory energy point denotes every five translation steps, and color denotes the energy value above initial at the point. (Appendix I). a.u.: arbitrary units.

We construct both datasets by removing one to four lattice sites from a 128,000-atom bcc Fe block and relaxing under the M07 EAM potential (Appendix A.1). Each example pairs a relaxed active volume X of radius 11 Å with a reference event set $\mathcal{E}_X^{\mathrm{ref}}$ collected through 80 random-start Dimer searches and SEAKMC validation, with barriers $\Delta E_k \in [0.1, 2.0]$ eV. These reference sets are finite and not exhaustive, but will be updated with more Dimer searches and made more complete. BCCFE-4VACAV-4000 contains 4,000 symmetry-distinct four-vacancy configurations and 65,450 events ($2 \le K \le 43$), with a 3,700/100/200 split. It is used for all evaluations in this paper. BCCFE-1TO4VAC contains 5,372 configurations with one to four vacancies and 83,372 events. It is built for future KMC trajectory simulation and analysis. Construction details are in Appendix B.

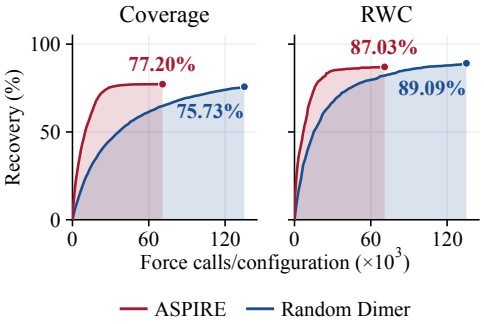


Figure 3: **Reference-event recovery versus search cost.** Coverage (left) and RWC (right) versus the mean recorded Dimer search force calls per configuration, evaluated on 200 test configurations. Curves vary the judge threshold for ASPIRE (up to 96 proposals) or the number of random-start Dimer searches (up to 80).

5.2 Evaluation protocol

Let $\mathcal{X}$ be the fixed evaluation set and $\mathcal{E}_X^{\mathrm{rec}} \subseteq \mathcal{E}_X^{\mathrm{ref}}$ the recovered references at X. Harmonic reference rates are $k_e = \nu e^{-\Delta E_e/(k_B T)} \in (0, \infty)$ with common prefactor ν. We consider events k and k' from the same X match if and only if $\max_a \|D_{k,a} - D_{k',a}\|_2 \le 0.05$ Å and $|\Delta E_k - \Delta E_{k'}| \le 0.02$ eV, where $D_k = S_k - X$ is the saddle displacement, ΔE_k the barrier, and a indexes atoms. Figure 2(a) is a visualization of the initial, saddle, and product configurations on the projected energy surface. We use coverage $\mathrm{Cov}_X = \frac{|\mathcal{E}_X^{\mathrm{rec}}|}{|\mathcal{E}_X^{\mathrm{ref}}|}$, rate-weighted coverage $\mathrm{RWC}_X = \frac{\sum_{e \in \mathcal{E}_X^{\mathrm{rec}}} k_e}{\sum_{e \in \mathcal{E}_X^{\mathrm{ref}}} k_e}$, and averaged rate-weighted coverage $\overline{\mathrm{RWC}} = \frac{1}{|\mathcal{X}|} \sum_{X \in \mathcal{X}} \mathrm{RWC}_X$ for evaluation. For a fixed reference catalogue, $\mathrm{RWC}_X \to \mathrm{Cov}_X$ as $T \to \infty$; for $\mathrm{RWC}_X > 0$, restricting to the recovered set gives $1/\mathrm{RWC}_X$ times the full-reference mean waiting time, and $\overline{\mathrm{RWC}}$ is the expected fraction recovered by the evaluated method when one event is sampled from the reference by rate at each X. These are kinetic-focused evaluation proxies, not full KMC trajectory kinetics (Appendix K). All reported RWC values use $T = 800$ K.

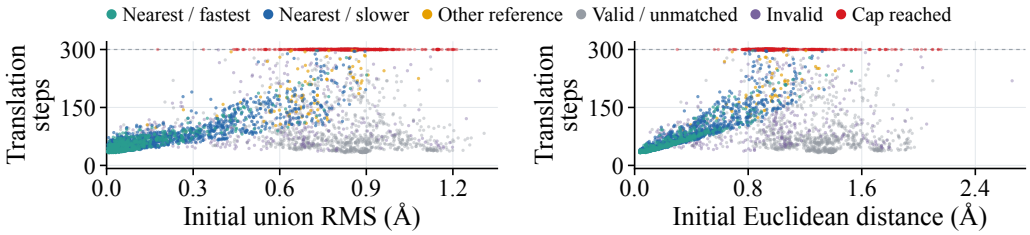


Figure 4: **Proposal geometry and refinement outcomes.** Each point represents a proposal and its subsequent Dimer search. Initial error is measured by moving-union RMS or Euclidean distance to the nearest reference saddle, and refinement effort by the number of translation steps. Colors distinguish nearest-reference recovery (fastest or slower), recovery of another reference, valid unmatched outputs, invalid outputs, and searches reaching the cap. The dashed line marks the 300-step cap.

Throughout the evaluation in this paper, the reference set $\mathcal{E}_X^{\mathrm{ref}}$ stays fixed: each output is assigned at most one matching reference, and repeated hits to the same reference are counted once (Appendix L). Throughout the experiments, each X is a SEAKMC active volume of radius 11 Å (Xu et al., 2012), treated as a global state for evaluation. Random Dimer searches are capped at 450 translation steps, whereas ASPIRE refinement is capped at 300 (Appendix F). This is because random dimer takes more translations to converge, and the main purpose of ASPIRE is to reduce geometric error and translation steps to achieve computational efficiency, as shown in Figure 2(b).

5.3 Reference-event recovery and force cost

Figure 3 compares reference-event recovery against physical search cost. ASPIRE achieves higher reference-event coverage and recovers a comparable fraction of the reference rate mass at substantially lower search cost compared to 80 random dimer searches. 80 per active volume is chosen as it is a reasonable parameter to get a relatively comprehensive set in SEAKMC deployment.

The judge controls this recovery–cost tradeoff by selecting proposals before physical refinement. The ASPIRE curves rise rapidly and then flatten, indicating the possibility of using the judge to further reduce force cost while keeping much of the event set and reference rate mass recovery. Together, these results support learned saddle proposals and selective refinement as a means of reducing physical search work while retaining high reference-event recovery.

Figure 4 shows that Geometric error is strongly associated with refinement behavior: proposals with small moving-union RMS typically converge rapidly to their nearest ground-truth saddle, whereas larger errors lead to more variable outcomes. This suggests that geometric error can serve as an effective offline proxy for model quality, consistent with the strong performance of our union-RMS-based judge.

Table 2: **Ablation of Ev-Quiformer.** Judge selection is disabled for all variants. O2M denotes the number of sequential matching rounds; Uniform denotes all-atom SmoothL1 supervision. The active-aware objective is specified in Appendix D.2. $\bar{F}_{\mathrm{search}}$ is the mean recorded Dimer search force-call count per configuration (Appendix L.5). Evaluated on 50 test configurations selected with topology-layered sampling covering all five vacancy-connectivity classes (Appendix L.1).

	Architecture			Supervision			Results		
Variant	Slots	Anchor	Self-attn.	O2M	Loss	β (Å)	Cov. (%) ↑	RWC (%) ↑	$\bar{F}_{\mathrm{search}}$ ↓
Full Ev-Quiformer	96	✓	✓	4	Uniform	0.03	**76.13**	87.34	75,768
Cross-attention only	96	✓	×	4	Uniform	0.03	75.60	84.29	71,033
Single matching round	96	✓	✓	1	Uniform	0.03	74.09	87.28	87,688
Active-aware loss	96	✓	✓	4	Active-aware	0.03	23.29	27.18	138,302
Fewer slots	32	✓	✓	4	Uniform	0.03	62.10	78.78	**26,461**
Coarser SmoothL1	96	✓	✓	4	Uniform	0.10	74.79	**87.75**	82,460
No spatial anchor	96	×	✓	4	Uniform	0.03	64.74	75.03	87,453

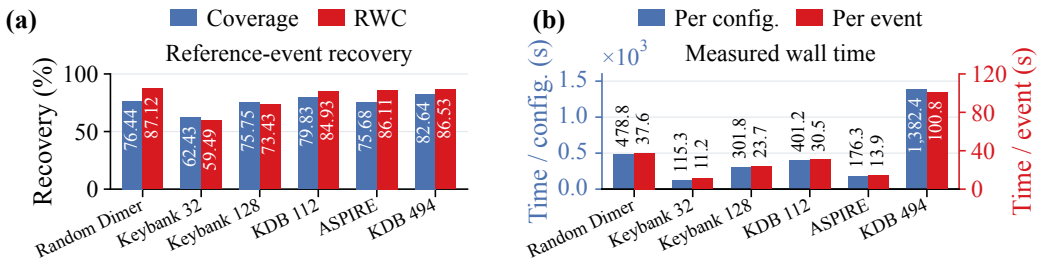


Figure 5: **Recovery and wall time compared with event reuse.** **(a)** Reference-event coverage and rate-weighted coverage. After Random Dimer, methods are ordered by increasing RWC. **(b)** Measured wall time per configuration and per recovered reference event. Random Dimer and the reuse baselines use 16 CPUs; ASPIRE uses 16 CPUs and one A100 GPU. ASPIRE uses judge-based proposal selection with threshold $\tau = 10^{-5}$, refining only proposals with $\widehat{p}_m \geq \tau$. Evaluated on 50 test configurations with topology-layered sampling

Ablation study. Table 2 examines how proposal design affects this recovery–cost tradeoff. Judge selection is disabled for all variants, so the comparison does not depend on a confidence threshold.

The full Ev-Quiformer achieves the highest coverage, 76.13%, with 87.34% RWC and 75,768 mean search force calls per configuration. Among those with similar coverage and RWC above 87%, full Ev-Quiformer reduces the most search force calls. Together, for the objective of higher coverage and RWC first, and reduced cost second, these results favor the default combination of a larger proposal set, slot interaction, repeated matching, and uniform all-atom supervision. Among the tested variants with similar RWC, the full model provides the highest coverage at the lowest force cost.

5.4 Wall time and comparison with event reuse

Force-call counts alone do not capture the lookup cost of event reuse. Cost-accounting conventions are discussed in Appendix L. We therefore compare measured wall time with *adaptations* of KDB and k-ART-style keybank baselines in Figure 5. We adapt KDB's local geometric matching and implement KeyBank using WL hashing, exact graph isomorphism, and geometric alignment. KDB transfers aligned saddle coordinates, whereas KeyBank transfers rotated displacement fields; both initialize Dimer refinement. KDB-494 denotes 494 event templates (Appendix G); KeyBank-32 and KeyBank-128 use 32 and 128 training configurations (Appendix H), respectively. Querying the test configurations was already time-consuming at these sizes. We therefore report the measured recovery–wall-time tradeoff at these operating points.

As shown in Figure 5, baselines with comparable RWC are substantially slower than ASPIRE, while faster alternatives achieve much lower RWC under the stated library and hardware settings. These methods do not contradict each other, so they can be deployed together for complementary benefits.

6 Conclusion and Limitation

We introduced ASPIRE, a propose-refine framework for saddle discovery from a single initial atomic configuration. Its Ev-Quiformer predictor uses geometry-conditioned event slots to generate a set of complete displacement fields for refinement on the original potential. On the four-vacancy bcc Fe benchmark, ASPIRE achieves higher reference-event coverage than the reported 80-start Random Dimer endpoint, at approximately half the recorded search-force cost under the respective protocols. The wall-time comparison further shows lower measured runtime than event-reuse baselines with comparable RWC under the stated hardware. These results support learning approximate saddle geometry as a means of accelerating physical event discovery.

Limitation. Our reference sets are finite and non-exhaustive, and the evaluation is restricted to four-vacancy bcc Fe under a fixed potential. Reference-relative coverage and RWC did not establish the accuracy of a full KMC trajectory. More complete reference catalogues and trajectory-level evaluation are needed to assess the resulting long-time kinetics.

AI USE STATEMENT

Generative AI tools assisted with manuscript organization and language editing, literature and citation checks, the review and revision of mathematical arguments and proofs, and feedback on the methodology and the presentation of results. The authors take responsibility for the final manuscript.

ETHICS STATEMENT

This study concerns computational modeling of vacancy events in bcc iron using atomistic simulation data. It does not involve human participants or personal data.

REPRODUCIBILITY STATEMENT

The code and dataset are available at https://anonymous.4open.science/r/ASPIRE-17B3/. Dataset construction, architecture and training, and evaluation protocols are described in Appendices B, C–D, and L, respectively. Appendices J and K provide proofs of proposal equivariance and the single-step interpretation of RWC.

REFERENCES

Mark Abraham, Andrey Alekseenko, Brian Andrews, Paul Bauer, Cathrine Bergh, Hugh Bird, Eliane Briand, Ania Brown, Yiqi Chen, Mahesh Doijade, Giacomo Fiorin, Stefan Fleischmann, Sergey Gorelov, Gilles Gouaillardet, Alan Gray, Farzaneh Jalalypour, Petter Johansson, Carsten Kutzner, Gregory Lazarski, Justin Lemkul, Magnus Lundborg, Julio Maia, Pascal T. Merz, Vedran Miletić, Dmitry Morozov, Lukas Müllender, Szilárd Páll, Andrea Pasquadibisceglie, Michele Pellegrino, Nicola Piasentin, Daniele Rapetti, Muhammad Umair Sadiq, Hubert Santuz, Michael Shirts, Tatiana Shugaeva, Alexey Shvetsov, Bálint Soproni, Philip Turner, Alessandra Villa, Yang Zhang, Berk Hess, and Erik Lindahl. Gromacs 2026.3 manual, June 2026. URL `https://doi.org/10.5281/zenodo.20845233`.

Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gábor Csányi. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. *Advances in neural information processing systems*, 35:11423–11436, 2022.

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.

Laurent Karim Béland, Peter Brommer, Fedwa El-Mellouhi, Jean-François Joly, and Normand Mousseau. Kinetic activation-relaxation technique. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 84(4):046704, 2011.

Laurent Karim Béland, Yuri N Osetsky, Roger E Stoller, and Haixuan Xu. Kinetic activation–relaxation technique and self-evolving atomistic kinetic monte carlo: Comparison of on-the-fly kinetic monte carlo algorithms. *Computational Materials Science*, 100:124–134, 2015.

Ondrej Biza, Sjoerd Van Steenkiste, Mehdi SM Sajjadi, Gamaleldin F Elsayed, Aravindh Mahendran, and Thomas Kipf. Invariant slot attention: Object discovery with slot-centric reference frames. *arXiv preprint arXiv:2302.04973*, 2023.

Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pp. 213–229. Springer, 2020.

Qiang Chen, Xiaokang Chen, Jian Wang, Shan Zhang, Kun Yao, Haocheng Feng, Junyu Han, Errui Ding, Gang Zeng, and Jingdong Wang. Group detr: Fast detr training with group-wise one-to-many assignment. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6610–6619. IEEE, 2023.

Samuel T Chill, Matthew Welborn, Rye Terrell, Liang Zhang, Jean-Claude Berthet, Andreas Pedersen, Hannes Jonsson, and Graeme Henkelman. Eon: Software for long time simulations of atomic scale systems. *Modelling and Simulation in Materials Science and Engineering*, 22(5):055002, 2014.

Murray S Daw and Michael I Baskes. Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals. *Physical review B*, 29(12):6443, 1984.

Jinlong Du, Suihe Jiang, Peipei Cao, Chuan Xu, Yuan Wu, Huaqiang Chen, Engang Fu, and Zhaoping Lu. Superior radiation tolerance via reversible disordering–ordering transition of coherent superlattices. *Nature Materials*, 22(4):442–449, 2023.

Chenru Duan, Yuanqi Du, Haojun Jia, and Heather J Kulik. Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model. *Nature computational science*, 3(12):1045–1055, 2023.

Chenru Duan, Guan-Horng Liu, Yuanqi Du, Tianrong Chen, Qiyuan Zhao, Haojun Jia, Carla P Gomes, Evangelos A Theodorou, and Heather J Kulik. Optimal transport for generating transition states in chemical reactions. *Nature Machine Intelligence*, 7(4):615–626, 2025.

Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. Se (3)-transformers: 3d roto-translation equivariant attention networks. *Advances in neural information processing systems*, 33:1970–1981, 2020.

Stefan Gugler, Max Eissler, Khaled Kahouli, and Klaus-Robert Müller. Reactionatlas: Ab origine exploration of chemical reaction networks with machine learning. *arXiv preprint arXiv:2606.30778*, 2026.

Sho Hayakawa and Haixuan Xu. Saddle point sampling using scaled normal coordinates. *Computational Materials Science*, 200:110785, 2021. ISSN 0927-0256. doi: https://doi.org/10.1016/j.commatsci.2021.110785. URL https://www.sciencedirect.com/science/article/pii/S0927025621005127.

Graeme Henkelman and Hannes Jónsson. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *The Journal of chemical physics*, 111(15): 7010–7022, 1999.

Graeme Henkelman and Hannes Jónsson. Long time scale kinetic monte carlo simulations without lattice approximation and predefined event table. *The Journal of Chemical Physics*, 115(21): 9657–9666, 2001.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *nature*, 596(7873):583–589, 2021.

Mukul Kabir, Timothy T Lau, David Rodney, Sidney Yip, and Krystyn J Van Vliet. Predicting dislocation climb and creep from explicit atomistic details. *Physical review letters*, 105(9):095501, 2010.

Johannes Kästner and Paul Sherwood. Superlinearly converging dimer method for transition state search. *The Journal of chemical physics*, 128(1), 2008.

Juho Lee, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. Set transformer: A framework for attention-based permutation-invariant neural networks. In *International conference on machine learning*, pp. 3744–3753. PMLR, 2019.

Tao Liang and Haixuan Xu. Saddle point search with dynamic active volume. *Computational Materials Science*, 228:112354, 2023. ISSN 0927-0256. doi: https://doi.org/10.1016/j.commatsci.2023.112354. URL https://www.sciencedirect.com/science/article/pii/S0927025623003488.

Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. *arXiv preprint arXiv:2206.11990*, 2022.

Yan-Ru Lin, Arunodaya Bhattacharya, Da Chen, Ji-Jung Kai, Jean Henry, and Steven J Zinkle. Temperature-dependent cavity swelling in dual-ion irradiated fe and fe-cr ferritic alloys. *Acta Materialia*, 207:116660, 2021.

Shilong Liu, Feng Li, Hao Zhang, Xiao Yang, Xianbiao Qi, Hang Su, Jun Zhu, and Lei Zhang. Dab-detr: Dynamic anchor boxes are better queries for detr. *arXiv preprint arXiv:2201.12329*, 2022.

Francesco Locatello, Dirk Weissenborn, Thomas Unterthiner, Aravindh Mahendran, Georg Heigold, Jakob Uszkoreit, Alexey Dosovitskiy, and Thomas Kipf. Object-centric learning with slot attention. *Advances in neural information processing systems*, 33:11525–11538, 2020.

Yue Luo, Xiangcheng Zheng, Xiangle Cheng, and Lei Zhang. Convergence analysis of discrete high-index saddle dynamics. *SIAM Journal on Numerical Analysis*, 60(5):2731–2750, 2022.

Lorenzo Malerba, Mihai-Cosmin Marinica, N Anento, C Björkas, H Nguyen, C Domain, F Djurabekova, Pär Olsson, K Nordlund, A Serra, et al. Comparison of empirical interatomic potentials for iron applied to radiation damage studies. *Journal of Nuclear Materials*, 406(1): 19–38, 2010.

M-C Marinica, Fran Willaime, and J-P Crocombette. Irradiation-induced formation of nanocrystallites with c 15 laves phase structure in bcc iron. *Physical review letters*, 108(2):025501, 2012.

Michael J Mehl, David Hicks, Cormac Toher, Ohad Levy, Robert M Hanson, Gus Hart, and Stefano Curtarolo. The aflow library of crystallographic prototypes: part 1. *Computational Materials Science*, 136:S1–S828, 2017.

Amil Merchant, Simon Batzner, Samuel S Schoenholz, Muratahan Aykol, Gowoon Cheon, and Ekin Dogus Cubuk. Scaling deep learning for materials discovery. *Nature*, 624(7990):80–85, 2023.

Vıctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *International conference on machine learning*, pp. 9323–9332. PMLR, 2021.

Kristof T Schütt, Michael Gastegger, Alexandre Tkatchenko, K-R Müller, and Reinhard J Maurer. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nature communications*, 10(1):5024, 2019.

Masaki Shimodaira, Takeshi Toyama, Kenta Yoshida, Koji Inoue, Naoki Ebisawa, Keiko Tomura, Toshimasa Yoshiie, Milan J Konstantinović, Robert Gérard, and Yasuyoshi Nagai. Contribution of irradiation-induced defects to hardening of a low-copper reactor pressure vessel steel. *Acta Materialia*, 155:402–409, 2018.

Rye Terrell, Matthew Welborn, Samuel T Chill, and Graeme Henkelman. Database of atomistic reaction mechanisms with application to kinetic monte carlo. *The Journal of Chemical Physics*, 137(1), 2012.

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018.

George H Vineyard. Frequency factors and isotope effects in solid state rate processes. *Journal of Physics and Chemistry of Solids*, 3(1-2):121–127, 1957.

Arthur F Voter. Introduction to the kinetic monte carlo method. In *Radiation effects in solids*, pp. 1–23. Springer, 2007.

Haixuan Xu, Yuri N Osetsky, and Roger E Stoller. Self-evolving atomistic kinetic monte carlo: fundamentals and applications. *Journal of Physics: Condensed Matter*, 24(37):375402, 2012.

Haixuan Xu, Roger E. Stoller, Yury N. Osetsky, and Dmitry Terentyev. Solving the puzzle of <100> interstitial loop formation in bcc iron. *Phys. Rev. Lett.*, 110:265503, Jun 2013. doi: 10.1103/PhysRevLett.110.265503. URL https://link.aps.org/doi/10.1103/PhysRevLett.110.265503.

Haixuan Xu, Roger E. Stoller, Laurent K. Béland, and Yuri N. Osetsky. Self-evolving atomistic kinetic monte carlo simulations of defects in materials. *Computational Materials Science*, 100:135–143, 2015. ISSN 0927-0256. doi: https://doi.org/10.1016/j.commatsci.2014.12.026. URL https://www.sciencedirect.com/science/article/pii/S0927025614008738. Special Issue on Advanced Simulation Methods.

Chuyang Zhao, Yifan Sun, Wenhao Wang, Qiang Chen, Errui Ding, Yi Yang, and Jingdong Wang. Ms-detr: Efficient detr training with mixed supervision. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17027–17036, 2024. doi: 10.1109/CVPR52733.2024.01611.

SUPPLEMENTARY MATERIAL

Appendices A–F describe the physical background, dataset construction, neural prediction and screening, and numerical refinement. Appendices G–I describe the event-reuse baselines and trajectory visualization. Appendices J and K provide the theoretical analyses, and Appendix L gives the evaluation protocols and cost accounting. The complete appendix contents follow on two pages.

APPENDIX CONTENTS

A Physical Background and Coordinate Conventions **16**
- A.1 bcc Fe, vacancies, and the fixed interatomic potential . . . 16
- A.2 Periodic images and a local Cartesian representation . . . 16
- A.3 Active volumes, surrounding atoms, and free coordinates . . . 18
- A.4 Dimer saddle searches and learned initialization . . . 18
- A.5 Transition rates and kinetic Monte Carlo . . . 19
- A.6 Interface to the main-text learning problem . . . 20

B Dataset Construction and bcc-Symmetry Deduplication **21**
- B.1 Physical construction and recorded examples . . . 21
- B.2 Four-vacancy benchmark construction . . . 21
- B.3 One-to-four-vacancy collection . . . 22
- B.4 Reference-event records and corpus distinction . . . 23
- B.5 The 48 cubic operations used for configuration deduplication . . . 23
- B.6 Equivalence of vacancy patterns and an explicit canonical key . . . 24
- B.7 Configuration deduplication, event matching, and data splitting . . . 25

C Architecture and Training Details **25**
- C.1 Encoder . . . 25
- C.2 Slot refinement and cross-attention . . . 26
- C.3 Decoder details . . . 27

D Displacement Supervision and Active-Aware Matching **28**
- D.1 Displacement fields and the uniform objective . . . 28
- D.2 The active-aware ablation . . . 28
- D.3 Sequential assignment and loss aggregation . . . 29
- D.4 Predictor optimization and checkpoint scope . . . 29

E Initial-State-Only Neural RMS Judge **29**
- E.1 Frozen predictor and inference inputs . . . 29
- E.2 Atom features, pooling, and network widths . . . 30
- E.3 The ten geometric descriptors, in implementation order . . . 30
- E.4 Offline labels and judge training . . . 31
- E.5 Deployment and proposal identifiers . . . 32

F Native Dimer Refinement and Validation **32**
- F.1 Enabled path and coordinate convention . . . 32
- F.2 Random axes, proposal loading, and orientation reset . . . 32
- F.3 Curvature estimation, rotation, and force projection . . . 33
- F.4 Two-stage translation and fine-step controls . . . 34
- F.5 Search loop, stopping, and cap handling . . . 34
- F.6 Native two-sided relaxation and connectivity . . . 35
- F.7 Returned saddle geometry and the numerical map . . . 35
- F.8 Search-force accounting and run configuration . . . 35

SUPPLEMENTARY MATERIAL

APPENDIX CONTENTS (CONTINUED)

G KDB Event-Reuse Adaptation **36**
G.1 Stored mechanisms and the scope of the baseline . . . 36
G.2 Mobile atoms and the retained local template . . . 36
G.3 Graph matching and geometric screening . . . 36
G.4 Aligned state transfer and displacement extraction . . . 37
G.5 Global candidate deduplication and Dimer submission . . . 37

H Rooted-Topology KeyBank Adaptation **38**
H.1 Rooted graphs and directed library entries . . . 38
H.2 Exact graph mappings and necessary distance checks . . . 38
H.3 Orthogonal fit and direct displacement transfer . . . 38
H.4 Mapping limit, sorting, and candidate deduplication . . . 39

I Shared Multiscale Embedding of Search Trajectories **39**
I.1 Illustrative sample and one shared coordinate system . . . 39
I.2 Configuration features, landmarks, and radial preprocessing . . . 39
I.3 Six classes of distance constraints . . . 40
I.4 Block normalization and weighted stress . . . 41
I.5 Initialization, optimization, and reproducibility . . . 41
I.6 Common display alignment, markers, and energy colors . . . 42

J Theoretical Analysis **42**
J.1 Local convergence motivation . . . 42
J.2 Proof of Geometric Consistency . . . 42

K Single-Step Reference Interpretation of RWC **43**
K.1 Reference rates and restricted one-step models . . . 43
K.2 Average recovery probability and the high-temperature limit . . . 44
K.3 Metric interpretation and conventions . . . 44

L Physical Evaluation, Event Matching, and Cost Accounting **44**
L.1 Fixed 50-configuration evaluation set and topology coverage . . . 44
L.2 Numerical acceptance and reference-event matching . . . 45
L.3 Distinct thresholds serve distinct operations . . . 46
L.4 Reference metrics and experiment-specific settings . . . 46
L.5 Recorded search-force count . . . 47
L.6 Wall time, retrieval size, and experiment records . . . 47

A Physical Background and Coordinate Conventions

This section develops the physical and representational background for the learning problem. We use $X \in \mathbb{R}^{N\times 3}$ for the initial configuration, V for the fixed interatomic potential, S_e for a reference saddle, $D_e = S_e - X$ for its displacement field, and X_e^+ for the connected product minimum. Parent-crystal, periodic-cell, and search-coordinate notation specifies how these quantities are constructed.

A.1 bcc Fe, vacancies, and the fixed interatomic potential

Parent lattice. In conventional cubic axes, an ideal body-centered-cubic (bcc) lattice with lattice parameter a_0 can be written as

$$\Lambda_{\mathrm{bcc}} = a_0\mathbb{Z}^3 \ \cup \ a_0\big(\mathbb{Z}^3 + (\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2})\big)\,. \tag{16}$$

There are two lattice sites per conventional cubic cell. The elemental bcc structure has space group $Im\bar{3}m$ (No. 229) (Mehl et al., 2017). An equivalent integer-coordinate description is

$$\Lambda_{\mathrm{bcc}} = \frac{a_0}{2}\left\{z \in \mathbb{Z}^3 : z_1 \equiv z_2 \equiv z_3 \pmod 2\right\}. \tag{17}$$

Thus the integer coordinates are either all even or all odd, enabling exact comparison of parent vacancy-site patterns. Relaxed atomic positions and saddle candidates are represented by continuous Cartesian coordinates.

Vacancies and competing rearrangements. A vacancy is an unoccupied site of the parent lattice. Equation (16) gives eight nearest-neighbor vectors from a lattice site,

$$\frac{a_0}{2}(\pm 1,\pm 1,\pm 1), \qquad d_{\mathrm{nn}} = \frac{\sqrt{3}}{2}a_0. \tag{18}$$

These directions describe nearest-neighbor atom–vacancy exchange. Multiple vacancies alter local coordination and induce relaxation of the surrounding Fe atoms, producing rearrangements with different saddle geometries and barriers. SEAKMC searches directly in atomic coordinate space to discover the corresponding mechanisms (Xu et al., 2012).

The prediction target is a complete $N \times 3$ displacement field, which describes both motion toward vacant sites and surrounding atomic relaxation. This representation accommodates varying numbers of participating atoms and multiple accessible rearrangements for the same vacancy configuration.

M07 and the EAM energy. The main-text experiments use the M07 embedded-atom-method (EAM) potential for Fe. The NIST repository identifies the standard M07 file with the Fe potential developed by Marinica in 2007 using a point-defect-oriented fitting database, and associates it with the studies of Malerba et al. (2010) and Marinica et al. (2012) . For a single-element EAM model, the energy of a physical configuration Y has the form

$$V_{\mathrm{full}}(Y) = \sum_i F_{\mathrm{emb}}\left(\sum_{j\neq i} f_{\mathrm{dens}}(r_{ij})\right) + \frac{1}{2}\sum_i\sum_{j\neq i}\phi_{\mathrm{pair}}(r_{ij}), \tag{19}$$

where r_{ij} is evaluated with the physical boundary convention, F_{emb} is the embedding function, f_{dens} contributes to the local density, and ϕ_{pair} is the pair interaction (Daw & Baskes, 1984). Here Y distinguishes the physical system from the local array X supplied to the predictor.

The neural model proposes candidate geometries, while the fixed Fe potential supplies energies and forces for initial relaxation, Dimer refinement, and barrier evaluation. This division connects learned proposals to a consistent physical energy surface throughout the benchmark.

A.2 Periodic images and a local Cartesian representation

Periodic images are translated copies. A periodic simulation cell represents a crystal together with translated copies of that cell (Abraham et al., 2026). Let H_{sys} be the nonsingular matrix whose rows are the physical cell vectors, and use row vectors for positions. Then r and $r + nH_{\mathrm{sys}}$ represent the same indexed atom for an allowed integer lattice shift n. In a partially periodic system, components of n along nonperiodic axes are zero.

For a displacement row vector v, the minimum-image operation is

$$\mathrm{MIC}_{H_{\mathrm{sys}}}(v) = v - n^* H_{\mathrm{sys}}, \qquad n^* \in \operatorname*{arg\,min}_{n \in \mathbb{Z}^3_{\mathrm{per}}} \|v - nH_{\mathrm{sys}}\|_2. \tag{20}$$

It chooses a shortest representative among the allowed translations. For an orthogonal, fully periodic cell, this becomes

$$\mathrm{MIC}_{H_{\mathrm{sys}}}(v) = \big(vH_{\mathrm{sys}}^{-1} - \mathrm{round}(vH_{\mathrm{sys}}^{-1})\big)H_{\mathrm{sys}}, \tag{21}$$

apart from image ties. The row-vector convention gives the inverse H_{sys}^{-1}. For a general skew cell, Equation (20) defines the shortest-image search. At exact ties, preprocessing uses a fixed image-selection convention compatible with the stated symmetry conditions.

From stored positions to X. A local representation places atoms near a defect in a common image even when their stored coordinates straddle the periodic box. Let c be a local reference point, also written as a row vector, and let $i(a)$ identify local atom a in the physical configuration Y. A suitable local representation is

$$X_a = \mathrm{MIC}_{H_{\mathrm{sys}}}\big(Y_{i(a)} - c\big). \tag{22}$$

Alternatively, an already unwrapped local configuration can be supplied directly. In either case, X is the resulting Cartesian array with fixed atom order; cell information and c specify its construction.

Pairwise Cartesian differences agree with minimum-image displacements under the following sufficient geometric condition. Let

$$\ell_{\min} = \min_{n \in \mathbb{Z}^3_{\mathrm{per}} \setminus \{0\}} \|nH_{\mathrm{sys}}\|_2,$$

and suppose all represented positions satisfy $\|X_a\|_2 \le r_{\mathrm{AV}}$. If

$$4r_{\mathrm{AV}} < \ell_{\min}, \tag{23}$$

then all pair differences in the local array are their unique shortest periodic representatives. Indeed, $\|X_a - X_b\|_2 \le 2r_{\mathrm{AV}}$, whereas any nonzero lattice translation l gives $\|X_a - X_b - l\|_2 \ge \ell_{\min} - 2r_{\mathrm{AV}} > 2r_{\mathrm{AV}}$. This criterion makes the relation between the local Cartesian chart and minimum-image geometry explicit in terms of the actual cell dimensions.

Operations used by the main-text model. After the local representation is fixed, the encoder uses $X_a - X_b$ in its 6 Å radius graph. The decoder uses

$$b_m = \sum_a A_{ma} X_a, \qquad X_a - b_m, \qquad \sum_a A_{ma} = 1. \tag{24}$$

The anchor is the attention-weighted mean in the supplied local Cartesian representation. Preparing a consistent local image before this averaging preserves the intended neighborhood geometry; for example, raw positions $0.49L$ and $-0.49L$ straddle a periodic boundary and have nearest-image separation $0.02L$.

Displacement targets and atom correspondence. The initial and saddle arrays must refer to the same atoms in compatible representations before forming

$$D_e = S_e - X, \qquad \widehat{S}_m = X + \widehat{D}_m. \tag{25}$$

Image tracking along a search gives a consistent physical displacement. For a localized event, compatible shortest displacements can be selected relative to the initial atoms; trajectories winding around the cell use path unwrapping.

With consistent atom correspondence and image representation, assignment and regression use the full Cartesian residual $\widehat{D}_m - D_e$ directly. The same atom ordering is retained in X, S_e, and their displacement field.

Three different meanings of "image." Periodic images are translated representatives of the same atom. The reflections and inversion used in cubic-symmetry deduplication are spatial transformations of an entire vacancy pattern. The two images of a Dimer are nearby trial configurations in configuration space. Each operation therefore has a defined role in data preparation or numerical search.

A.3 Active volumes, surrounding atoms, and free coordinates

SEAKMC localizes event searches to active volumes surrounding defects, reducing the coordinates involved in a search while retaining an atomistic description of the rearrangement (Xu et al., 2012; 2015; Liang & Xu, 2023). In this paper, X represents an active volume of radius $r_{\mathrm{AV}} = 11\,\text{Å}$. Its N local atoms are selected from the 128,000-atom parent block before vacancy removal. The active-volume radius controls the search region, the 6 Å graph cutoff controls neural neighborhoods, and the potential's interaction cutoff controls physical interactions.

The active-volume model incorporates its prescribed surrounding environment. With fixed surroundings, a local trial array Z can be embedded into the parent configuration by a map ι_X. The local potential is

$$V(Z) = V_{\mathrm{full}}(\iota_X Z) - C_X, \tag{26}$$

where C_X is any configuration-independent energy offset. Interactions across the active-volume boundary are included according to the physical model. The offset cancels from $V(S_e) - V(X)$. This physical locality model specifies the forces and barriers used for refinement (Béland et al., 2015).

When some represented atoms are fixed during refinement, let $\mathcal{F}_X$ index the free atoms. An admissible search displacement then has $D_a = 0$ for $a \notin \mathcal{F}_X$, and the search acts on vectorized free coordinates. Any additional linear constraints are imposed before taking the gradient and Hessian of the restricted potential. This is the coordinate interpretation used by the local-convergence proposition.

The free-coordinate representation specifies the nondegeneracy conditions. A translation-invariant whole-system potential has uniform translation zero modes; fixing the environment or removing these modes supplies the space in which a positive-definite minimum Hessian or nondegenerate index-one saddle is defined. Motion of the free atoms relative to a fixed shell remains part of the local physical problem.

A.4 Dimer saddle searches and learned initialization

What Dimer searches for. The Dimer method seeks an index-one saddle using first derivatives of the potential, without requiring a product minimum as input (Henkelman & Jónsson, 1999). Work in the same vectorized free coordinates as above, and denote a trial point by y and its force by $f(y) = -\nabla V(y)$. Dimer combines ascent along an unstable direction with relaxation in transverse directions to locate an index-one saddle.

Two configurations estimate directional curvature. A symmetric description of a Dimer consists of a center y, a unit orientation n, and a half-separation $h_{\mathrm{dim}} > 0$:

$$y^+ = y + h_{\mathrm{dim}} n, \qquad y^- = y - h_{\mathrm{dim}} n, \qquad \|n\|_2 = 1. \tag{27}$$

The orientation assigns a displacement direction to every free coordinate. A finite difference of the forces estimates a Hessian–vector product,

$$B_{h_{\mathrm{dim}}}(y, n) = -\frac{f(y^+) - f(y^-)}{2h_{\mathrm{dim}}} \longrightarrow \nabla^2 V(y) n \quad \text{as } h_{\mathrm{dim}} \to 0, \tag{28}$$

with estimated curvature $\kappa_{h_{\mathrm{dim}}} = n^\top B_{h_{\mathrm{dim}}}(y, n)$. This explains how curvature information can be obtained without assembling a full Hessian. The symmetric formula describes the method's principle; implementations can use force reuse or one-sided estimates to reduce the number of force evaluations (Kästner & Sherwood, 2008).

Rotation and translation are different updates. At fixed y, the orientation is adjusted toward a low-curvature mode. In the small-separation limit, minimizing $n^\top \nabla^2 V(y) n$ over the unit sphere gives a tangential gradient proportional to

$$(I - nn^\top)\nabla^2 V(y) n, \tag{29}$$

which can be estimated using Equation (28). A translation update then moves the center. In a negative-curvature region, the minimum-mode-following effective force is

$$f_{\mathrm{eff}}(y, n) = f(y) - 2n\big(n^\top f(y)\big) = -(I - 2nn^\top)\nabla V(y). \tag{30}$$

The force component parallel to n is reversed while perpendicular components are retained. If n is the unstable mode near an index-one saddle, this reverses the downhill motion along that mode and leaves relaxation in stable directions. Rotation and translation are repeated, with numerical step control and convergence tests. Additional controls handle positive-curvature escape from the initial basin (Chill et al., 2014).

Random and learned initialization. The conventional path initializes the search direction randomly, using the axis sampler in Appendix F. ASPIRE supplies $\widehat{S}_m = X + \widehat{D}_m$ as the proposed center configuration. In free coordinates, this gives

$$y_0 = \mathrm{vec}\Big(\widehat{S}_m|_{\mathcal{F}_X}\Big). \tag{31}$$

The Dimer also requires an orientation. For displacement-derived initialization, its general form is

$$n_0 = \frac{\mathsf{T}_X(\widehat{D}_m)}{\|\mathsf{T}_X(\widehat{D}_m)\|_2}, \qquad \|\mathsf{T}_X(\widehat{D}_m)\|_2 > 0, \tag{32}$$

where T_X collects the orientation preprocessing and restriction to search coordinates. In the supplied Cartesian path, the bridge removes the mean displacement for the initial axis, retains the full displacement for the center preload, and can reset the axis during loading (Appendix F). The center preload specifies a candidate geometry, while orientation updates estimate the local unstable direction.

Proposition 1 analyzes the exact minimum-mode iteration with its specified step size. The resulting local convergence region provides a mathematical rationale for geometric initialization. Appendix F gives the finite-difference orientation updates and search controls used in numerical refinement.

Termination, saddle validation, and connectivity. The target in Section 3 combines stationarity, the index-one condition, and downhill connectivity from $S_e \pm \epsilon v_1(S_e)$ to X and a distinct minimum X_e^+. Together these conditions define the transition tuple and barrier

$$e = (X, S_e, X_e^+, \Delta E_e), \qquad \Delta E_e = V(S_e) - V(X) > 0. \tag{33}$$

The main-text idealized map $\mathcal{R}_X$ encodes these target conditions. The native Dimer/SEAKMC implementation uses force/step stopping, two-sided relaxation, final flags, and cap checks as its operational acceptance protocol (Appendix F).

Search work in this benchmark. The main-text evaluation caps ASPIRE refinement at 300 translation steps and random-start Dimer at 450. The predictor has $M = 96$ slots; with judge selection, fewer than M proposals may be submitted. The random baseline uses up to 80 searches per configuration. The slot count specifies the proposal pool, the translation cap controls each search, and force-call counts measure its computational work, including orientation updates.

The Dimer search-force ledger counts calls routed through its force routine for recorded searches, including failed and duplicate searches. It excludes model inference and force evaluations attributed to separate setup and downhill relaxation. Wall time is reported separately. Dimer separation, rotation limits, force tolerances, and step-size controls are run-specific settings described in the numerical refinement protocol.

A.5 Transition rates and kinetic Monte Carlo

From a saddle to an event rate. Transition-state theory associates a rate with escape through a transition state. Under its harmonic approximation, an event rate has Arrhenius form

$$k_e = \nu_e \exp\left(-\frac{\Delta E_e}{k_{\mathrm{B}}T}\right). \tag{34}$$

The prefactor ν_e accounts for the local vibrational contribution. For d unconstrained vibrational degrees of freedom after zero modes have been removed, the classical harmonic expression is

$$\nu_e = \frac{1}{2\pi}\frac{\prod_{i=1}^{d}\omega_i(X)}{\prod_{i=2}^{d}\omega_i(S_e)}, \tag{35}$$

where ω_i are angular frequencies from the mass-weighted Hessian and the unstable saddle mode is omitted from the denominator (Vineyard, 1957). The benchmark uses the common-prefactor convention $k_e = \nu \exp[-\Delta E_e/(k_{\mathrm{B}}T)]$ for reference weights. The common factor cancels from RWC and relative event probabilities; absolute waiting times retain its scale.

State selection and the clock. The jump-process description assumes that relaxation within a metastable basin is fast relative to escape, so that event rates can be treated as constant while the system remains in that state. KMC advances the system through discrete transitions using their event rates (Voter, 2007). For a nonempty event set $\mathcal{E}_X$ with total rate

$$\Lambda_X = \sum_{e \in \mathcal{E}_X} k_e, \tag{36}$$

a residence-time step draws independent $u_1, u_2 \sim \mathrm{Uniform}(0,1)$ and uses

$$\begin{aligned} e^* = \min\left\{ j : \sum_{e=1}^{j} k_e \geq u_1 \Lambda_X \right\}, \qquad X \leftarrow X^+_{e^*}, \\ \Delta t = -\frac{\log u_2}{\Lambda_X}. \end{aligned} \tag{37}$$

Event indices can be ordered arbitrarily for the cumulative sum. The event probability is k_e/Λ_X, and the physical waiting time is exponential with mean $1/\Lambda_X$. Event selection and clock advancement use independent random draws (Voter, 2007).

On-the-fly discovery and SEAKMC. When the accessible products are unknown, an on-the-fly method first constructs an event catalogue at the current state. SEAKMC combines defect identification, local active-volume searches, transition processing, and KMC event selection (Xu et al., 2012; 2015). After a selected transition, the affected local environments and event catalogues are updated. Repeated event discovery therefore motivates efficient proposal generation across the configurations encountered during simulation.

ASPIRE targets the proposal-and-refinement stage of this process. Its experiments evaluate distinct reference-event recovery and search efficiency from fixed initial configurations. These measurements assess the saddle-discovery component that supplies mechanisms for downstream event-based modeling.

Why distinct events and rate mass both matter. Each distinct physical exit channel contributes to the total escape rate. Repeated searches reaching one event count as one recovered identity, while symmetry-related exits with equal rates can contribute separately. Combining channels into a common catalogue entry therefore retains their summed rates or multiplicities (Appendix B.7).

For a fixed reference set and its recovered subset,

$$\Lambda_X^{\mathrm{ref}} = \sum_{e \in \mathcal{E}_X^{\mathrm{ref}}} k_e, \qquad \Lambda_X^{\mathrm{rec}} = \sum_{e \in \mathcal{E}_X^{\mathrm{rec}}} k_e, \qquad \mathrm{RWC}_X = \frac{\Lambda_X^{\mathrm{rec}}}{\Lambda_X^{\mathrm{ref}}}. \tag{38}$$

Keeping the original reference rates, a one-step model restricted to the recovered subset has mean waiting time larger by $1/\mathrm{RWC}_X$ when $\mathrm{RWC}_X > 0$. Also, a rate-weighted draw from the reference set belongs to the recovered subset with probability RWC_X. These are the single-state interpretations used in the main text. An empty recovered set has no finite escape time under that restricted model.

The benchmark measures distinct reference-event recovery and search work across held-out states. Coverage summarizes recovered event identities, while RWC quantifies their reference escape-rate mass. These metrics provide complementary measures of the quality and efficiency of saddle discovery.

A.6 Interface to the main-text learning problem

The physical environment, fixed potential, and consistent Cartesian representation define the input X and targets $D_e = S_e - X$. The predictor maps X to $\{\widehat{D}_m\}_{m=1}^{M}$ without product endpoints. The

Table 3: **Dataset records.** Each state is one retained local AV. The two collections use their respective AV-selection protocols and have complementary roles in evaluation and future trajectory studies.

Dataset	AV states	Events	Vacancies	Use
BCCFE-4VACAV-4000	4,000	65,450	4	Reported evaluation
BCCFE-1TO4VAC	5,372	83,372	1–4	Future trajectory studies

Table 4: **Reference-catalogue generation settings.** Dataset collection uses the following protocol; evaluation settings are specified in Appendix F.

Setting	Value
Point-defect-reduction cutoff, `DCut4PDR`	4.0 Å
`DActive`, `DBuffer`, `DFixed`	11, 1, 11 Å
Dimer starts, `NSearch`	80 per searched AV
Buffer search, `SearchBuffer`	False
Maximum translation iterations, `NMax4Trans`	900
Trial / maximum step size	0.015/0.05 Å
Dimer separation	0.005 Å
Local relaxation	Enabled
Barrier window	$[0.1, 2.0]$ eV
Connectivity checking	Enabled
Prefactor calculation, `CalPrefactor`	False

encoder's $X_a - X_b$, decoder's $X_a - b_m$, full SmoothL1 residual, and reference matcher use this shared representation, connecting the physical task to geometric learning and refinement.

B Dataset Construction and bcc-Symmetry Deduplication

B.1 Physical construction and recorded examples

Both collections start from a periodic $40\times40\times40$ conventional-cell bcc Fe block containing 128,000 atoms before vacancy removal. One to four ideal lattice sites are removed, and the parent is relaxed through SEAKMC under the M07 EAM potential. The lattice parameter is approximately $a_0 = 2.85528$ Å and the cubic box length is approximately 114.2112 Å; the stored cell specifies the exact geometry. Forces and energies are evaluated with LAMMPS using `eam/fs` and `M07_eam.fs`. Positions and energies use Å and eV, respectively; the physical background is given in Appendix A.1.

The learning input is the relaxed local active-volume (AV) array X of radius 11 Å. Each example pairs X with a variable-size, unordered reference event set collected through 80 random-start Dimer searches and SEAKMC validation, retaining barriers in $[0.1, 2.0]$ eV. Table 3 gives the dataset sizes and their roles in this paper. All reported evaluations use BCCFE-4VACAV-4000; BCCFE-1TO4VAC is constructed for future KMC trajectory simulation and analysis.

Reference-collection settings. The common generation settings are summarized in Table 4. The active, buffer, and fixed settings describe the physical AV construction; they are distinct from the encoder's graph cutoff. With `SearchBuffer=False`, the search coordinates comprise the active region.

Reference collection uses a 900-translation limit. The main-text evaluation uses caps of 300 for ASPIRE refinement and 450 for conventional Dimer (Appendix F); search records belong to their respective collection or evaluation stage.

B.2 Four-vacancy benchmark construction

Vacancy topology enumeration and parent sampling. For BCCFE-4VACAV-4000, vacancies occupy ideal bcc sites represented as $(a_0/2)(i, j, k)$, with the three integers either all even or all odd. Candidate four-site patterns have maximum pairwise vacancy distance strictly below 11 Å.

Table 5: **Four-vacancy parent sampling.** Counts are for the recorded PDR fragmentation signatures used in dataset construction.

PDR fragmentation signature	Selected parents
4	35
$3+1$	583
$2+2$	246
$2+1+1$	1,254
$1+1+1+1$	1,882
Total	4,000

Identifying patterns under periodic lattice translations and all 48 operations of the cubic point group O_h gives 21,297 symmetry-distinct candidate topologies. The symmetry operations and a mathematical canonical key are described below.

Parent sampling selects 4,000 configurations using seed `20260818`. All candidates in the point-defect-reduction (PDR) fragmentation classes 4, $3+1$, and $2+2$ are retained; the remaining classes are sampled across 24 diameter bins. The selected class counts are given in Table 5. The seed and stratification specify parent selection for the 4,000-configuration corpus.

PDR denotes point-defect reduction in the defect-detection procedure. The table records the parent-sampling categories; selection of the fixed 50-state evaluation subset is described separately in Appendix L.1.

One-AV Wigner–Seitz construction. For each relaxed parent, Wigner–Seitz (WS) detection against the full perfect reference first determines the possible PDR groups. A target group is selected that reproduces itself when the other vacancy markers are hidden. The physical parent retains all four vacancies. The case-specific WS reference is formed by removing the non-target vacancy sites from the perfect reference, so only the target-group vacancies appear as WS differences. This yields exactly one selected AV per parent, with all four physical vacancies inside its active sphere. The WS matching radius is $1.03785\,\text{Å}$, obtained from `DCut4Def` $\times$ `cutdefectmax` $= 0.3 \times 3.4595\,\text{Å}$, and `NPredef=0`.

Selecting one AV preserves its reference labels during collection. The reduced WS reference defines defect detection while retaining the physical parent and potential-energy surface. The resulting local-environment corpus contains 65,450 connected events, with 2–43 per AV and a mean of 16.36.

B.3 One-to-four-vacancy collection

Candidate coverage. The initial candidate bank contains 8,192 ideal vacancy arrangements: 1, 95, 2,048, and 6,048 with one, two, three, and four vacancies, respectively. Candidates are filtered under translations and all 48 O_h operations. Coverage includes compact clusters, fragmented motifs, spatially separated arrangements, and directly constructed vacancy-hop endpoints. Hop-related descendants retain their source family assignment. This designed candidate bank provides broad structural coverage for the subsequent AV construction and release filtering.

Native CN construction and independent WS audit. Production AVs use native coordination-number (CN) defect detection, `NPredef=1`, native PDR, and the shared $11/1/11$ Å settings. WS is run only on a deep copy of the relaxed full parent against the full perfect reference to audit vacancy content. An eligible native AV must contain all intended parent vacancies in its active region, have the expected vacancy count, and have no WS interstitial. Eligibility is evaluated at the native center against the full perfect reference; parents without an eligible native AV are recorded as ineligible.

The primary AV maximizes the minimum vacancy-to-active-boundary margin, with randomized ties within $0.25\,\text{Å}$ bands. Approximately one ninth of topology families may contribute a second distinct native AV, with at most two searched AVs per parent. This policy builds a structurally diverse collection with explicit parent and family provenance.

Table 6: **One-to-four-vacancy release.** Counts summarize the final AV collection after construction and filtering.

Vacancies	Released AVs	Events
1	1	12
2	38	492
3	1,121	16,073
4	4,212	66,795
Total	5,372	83,372

Full-geometry deduplication and event collection. Before Dimer search, complete active, buffer, and fixed coordinates are compared under translations and all 48 cubic operations using species-preserving bijective matching at 0.001 Å tolerance. A shared registry prevents concurrent searches of congruent geometries. Subset selection preserves the native coordinates, atom regions, AV indices, and center provenance. Native per-AV search and postprocessing use 80 starts, and labels are archived before cross-AV deletion. The collection wrapper stops before KMC event selection and propagation.

The released collection contains 5,372 deduplicated AVs from 4,812 parent IDs and 83,372 events, distributed as in Table 6. The number of AV examples is therefore different from the number of distinct parent IDs.

B.4 Reference-event records and corpus distinction

Each example groups the retained reference events by their shared initial AV, providing direct supervision for the one-to-many prediction task. Core labels include initial, saddle, and final positions; barriers; energy biases; connectivity; and event/search identifiers. Vacancy topology, AV selection, and parent/family metadata record provenance. Compatible atom ordering and coordinate representations are retained when forming $D_e = S_e - X$, as described in Appendix A.2. Vacancy patterns record construction provenance. Reference rates use the common-prefactor convention, consistent with `CalPrefactor=False` during collection.

The full BccFe-1to4Vac representation retains complete AV geometry and region masks. Its compact representation stores active-atom arrays and keeps the static buffer/fixed coordinates once in `context_pos0`; original ordering and metadata are retained under `source__*`. The provided `compact_io.load_full_av` reader reconstructs the full representation, preserving the input X and displacement fields used by the learning objective.

The 80-start collection protocol and barrier window define the observed reference catalogue for each state. Deduplication retains distinct events, and evaluation reports recovery relative to these catalogues.

The corpora provide complementary local-environment collections. The four-vacancy benchmark has one WS-selected AV per parent; the one-to-four-vacancy collection uses native-CN AVs with an independent WS audit and can retain a second AV. Each corpus is deduplicated using its construction protocol.

B.5 The 48 cubic operations used for configuration deduplication

Configuration deduplication uses the 48 operations of the full cubic point group $O_h = m\bar{3}m$, comprising 24 proper rotations and 24 improper operations. In conventional cubic coordinates, it has the explicit matrix representation

$$\mathcal{G}_{\mathrm{bcc}} = \{G = \mathrm{diag}(\varsigma_1, \varsigma_2, \varsigma_3) C_\pi : \varsigma_i \in \{-1, +1\},\ \pi \in \mathfrak{S}_3\}\,, \qquad |\mathcal{G}_{\mathrm{bcc}}| = 2^3\, 3! = 48, \tag{39}$$

where C_π is a 3×3 coordinate-permutation matrix. This is different from the $N \times N$ atom-permutation matrix P used in the main-text equivariance statement.

The six coordinate orders are

$$\begin{array}{l} (x, y, z),\ (x, z, y),\ (y, x, z), \\ (y, z, x),\ (z, x, y),\ (z, y, x). \end{array} \tag{40}$$

For each order, all eight independent sign choices are included. Every matrix in Equation (39) is orthogonal. Its determinant is $(\varsigma_1\varsigma_2\varsigma_3)\,\mathrm{sgn}(\pi)$, so 24 matrices have determinant $+1$ and 24 have determinant -1. The latter include reflections and inversion combined with rotations; for example, $\mathrm{diag}(-1,1,1)$ reflects one coordinate and $-I$ is inversion.

These matrices preserve the bcc lattice: a coordinate permutation and sign changes preserve the all-even or all-odd condition in Equation (17). The point group is finite, while lattice translations belong to the space-group description and are handled separately. A symmetric vacancy pattern can have fewer than 48 distinct transformed representatives because multiple group operations may yield the same unordered set.

B.6 Equivalence of vacancy patterns and an explicit canonical key

Let n_v be the number of initial vacancies and let $\mathcal{V} = \{v_1, \ldots, v_{n_\mathrm{v}}\} \subset \Lambda_\mathrm{bcc}$ be their unordered parent-lattice positions, represented as row vectors. The vacancy-pattern reduction uses one common cubic operation and a lattice translation. In a consistent local representation, the equivalence criterion is

$$\mathcal{V}' \sim \mathcal{V} \quad \Longleftrightarrow \quad \exists G \in \mathcal{G}_\mathrm{bcc},\ t \in \Lambda_\mathrm{bcc} : \mathcal{V}' = \{vG^\top + t : v \in \mathcal{V}\}. \tag{41}$$

A common transform acts on the whole pattern and preserves its geometry. Translation removal defines the comparison key; the stored AV retains its native coordinate representation.

For completeness, one exact canonicalization of Equation (41) is as follows. Write $z_i = 2v_i/a_0 \in \mathbb{Z}^3$. For each of the 48 matrices and each choice of vacancy j as a translation anchor, form the unordered relative set

$$\mathcal{Z}_{G,j} = \{(z_i - z_j)G^\top : i = 1, \ldots, n_\mathrm{v}\}. \tag{42}$$

Sort its rows lexicographically, concatenate the sorted triples, and take the lexicographically smallest resulting integer tuple:

$$\mathrm{key}_\mathrm{bcc}(\mathcal{V}) = \min_{\substack{G \in \mathcal{G}_\mathrm{bcc} \\ 1 \le j \le n_\mathrm{v}}} \mathrm{lex}(\mathrm{sort}_\mathrm{row}(\mathcal{Z}_{G,j}))\,. \tag{43}$$

Sorting handles the absence of vacancy labels; enumeration of all anchors removes dependence on which vacancy was chosen as the origin. For four vacancies, this construction inspects at most $48 \times 4 = 192$ candidate relative sets, accounting for all four choices of translation anchor.

To see why the key works, a lattice translation cancels in every $z_i - z_j$, a vacancy reordering only permutes candidate anchors and sorted rows, and applying a cubic operation permutes the 48 group elements by composition. Hence equivalent patterns have equal keys. Conversely, equality of two such sorted relative sets supplies a common cubic transformation and translation between their parent vacancy positions. This construction therefore realizes the stated equivalence without a floating-point tolerance on ideal lattice coordinates.

Equation (43) gives an explicit canonicalization of the stated equivalence relation. A fixed-center implementation can enumerate 48 centered transforms using its chosen origin convention. The canonical construction provides a mathematical reference for these implementations.

Local representation and relaxed coordinates. Before canonicalization, the parent vacancy sites are expressed in the consistent local Cartesian representation described in Appendix A.2.

The integer key acts on parent vacancy sites. The one-to-four-vacancy collection additionally compares relaxed full-AV geometries at 0.001 Å tolerance (Appendix B.3). Together, canonical vacancy patterns and full-geometry checks distinguish configuration identity at their respective stages.

Transforming a complete event record. When changing the representative of a configuration, the same spatial operation and atom correspondence must be applied to its entire record. For atom permutation P, cubic operation G, and row translation t,

$$\begin{aligned} X' &= PXG^\top + \mathbf{1}t, \qquad & S'_e &= PS_eG^\top + \mathbf{1}t, \\ (X^+_e)' &= PX^+_eG^\top + \mathbf{1}t, \qquad & D'_e &= PD_eG^\top. \end{aligned} \tag{44}$$

All atom-indexed masks and any physical boundary data must transform consistently. Because the EAM expression depends on interatomic distances, these transformations preserve its energies

and barriers when the complete physical environment is transformed consistently, as follows from Equation (19). The neural transformation law is established separately in Appendix J.2.

B.7 Configuration deduplication, event matching, and data splitting

Configuration deduplication. The 48-operation rule groups cubic-symmetry-equivalent initial vacancy environments into configuration classes. Each retained state then has its own reference set of saddle mechanisms and connected products.

Event identity within a fixed state. The main-text reference matcher compares events from the same represented X using their full atomwise displacements and barriers:

$$\max_a \|D_{e,a} - D_{e',a}\|_2 \le 0.05\,\text{Å}, \qquad |\Delta E_e - \Delta E_{e'}| \le 0.02\,\text{eV}. \tag{45}$$

This is the event-matching criterion in Section 5, with no additional minimization over cubic operations. It is distinct from both initial-configuration deduplication and the moving-union RMS used for judge labels. During recovery evaluation the reference catalogue stays fixed, an output receives at most one matching reference identity, and repeated hits to that identity count once. Fixed reference identities define the recovery denominator, providing a consistent basis for comparing outputs across methods.

Retaining spatial exit channels preserves their contribution to escape rates. For an isolated vacancy, nearest-neighbor exchanges in different directions can be related by cubic symmetry and have equal barriers. They move the vacancy to different positions and contribute separately to the total escape rate. Event-level evaluation retains these channels and their multiplicities.

Grouping for the four-vacancy train/monitor/test split. All targets sharing an initial state belong to the same split. The reported BCCFE-4VACAV-4000 split contains 3,700 training, 100 monitoring, and 200 held-out states. The 200-state recovery evaluation uses that test split; the fixed 50-state ablation and wall-time subset is described in Appendix L.1. Both evaluation populations are drawn from the four-vacancy test bank.

State-level grouping keeps the reference events of each configuration together. Parent-corpus construction and assignment to training, monitoring, and test partitions are separate data-preparation stages.

For a split intended to separate symmetry-equivalence classes, all symmetry-equivalent copies must remain in one group. If $\mathcal{K}_{\text{train}}$, $\mathcal{K}_{\text{monitor}}$, and $\mathcal{K}_{\text{test}}$ are the configuration-key sets, the integrity condition is

$$\mathcal{K}_{\text{train}} \cap \mathcal{K}_{\text{monitor}} = \mathcal{K}_{\text{train}} \cap \mathcal{K}_{\text{test}} = \mathcal{K}_{\text{monitor}} \cap \mathcal{K}_{\text{test}} = \varnothing. \tag{46}$$

This criterion defines symmetry-level split integrity and can be checked directly from the released configuration-key sets.

One-to-four-vacancy release assignments. The separate BCCFE-1TO4VAC release manifest supplies 4,171/543/658 train/validation/test AV assignments. The generation record reports that no parent ID or family ID is shared across these partitions. These release assignments support future studies using the broader vacancy collection; the reported experiments use the four-vacancy split. The single one-vacancy AV supplies an elementary reference case within the release.

C Architecture and Training Details

C.1 Encoder

The encoder uses the represented configuration X defined in the main text, after the background coordinate preparation in Appendix A.2. A directed edge $b \to a$ is present for a nonzero separation within 6 Å, with at most 128 neighbors per atom. Its features are

$$\begin{gathered} r_{ba} = X_a - X_b, \qquad d_{ba} = \|r_{ba}\|_2, \qquad \widehat{r}_{ba} = r_{ba}/d_{ba}, \\ e_{ba} = \left(\phi(d_{ba}),\, \bigoplus_{\ell=0}^{2} Y^{(\ell)}(\widehat{r}_{ba}) \right), \end{gathered} \tag{47}$$

where ϕ is a 128-dimensional Gaussian radial expansion and the spherical harmonics extend through degree two. Equation (47) uses the prepared Cartesian edge vector directly for graph inclusion, distances, and angular features.

Species embeddings are augmented by an equivariant edge-degree embedding. Every atomic representation contains

$$h_a = h_a^{(0)} \oplus h_a^{(1)} \oplus h_a^{(2)}, \qquad h_a^{(\ell)} \in \mathbb{R}^{C_\ell \times (2\ell+1)}, \qquad (C_0, C_1, C_2) = (128, 64, 32). \tag{48}$$

The last axis transforms under the degree-ℓ representation of $R \in \mathrm{SO}(3)$, giving $128 + 3(64) + 5(32) = 480$ components per atom. All benchmark atoms are Fe.

Suppressing the layer index, each of the six Equiformer blocks constructs an edge representation using equivariant source and destination projections and a radially conditioned depthwise tensor product (Liao & Smidt, 2022):

$$t_{ba} = \mathrm{TP}_{\psi(\phi(d_{ba}))}\left(W_{\mathrm{src}} h_b + W_{\mathrm{dst}} h_a,\ Y(\widehat{r}_{ba})\right). \tag{49}$$

Here ψ predicts invariant tensor-product weights. For each of four heads, scalar channels provide an invariant logit, while equivariant projection, gating, and a further separable tensor product provide the value $v_{ba}^{(h)}$:

$$\begin{aligned} s_{ba}^{(h)} = w_h^\top \,\mathrm{act}(P_h^{(0)} t_{ba}), \qquad \alpha_{ba}^{(h)} = \mathrm{softmax}_{b \in \mathcal{N}(a)}\, s_{ba}^{(h)}, \\ m_a^{(h)} = \sum_{b \in \mathcal{N}(a)} \alpha_{ba}^{(h)} v_{ba}^{(h)}. \end{aligned} \tag{50}$$

The heads are concatenated and equivariantly projected. Normalization, residual connections, and an equivariant feed-forward update complete the block. Attention-weight dropout is 0.2 during training; Equation (50) gives the pre-dropout weights. The output features used in the main text are $u_a \in \mathbb{R}^{128}$ and $V_a \in \mathbb{R}^{64\times 3}$.

The equivariance result in Appendix J.2 assumes atom-reindexing-compatible graph construction and neighbor truncation, including a compatible convention when the neighbor cap splits a distance tie.

C.2 Slot refinement and cross-attention

The main model has $M = 96$, $d_s = 256$, and $d_v = 16$. We use the main-text stage index throughout: initialization is $t = 0$, the two vector–scalar stages are $t = 1, 2$, the three scalar-only stages are $t = 3, 4, 5$, and the final readout is $T = 6$. The state entering initialization is the learned scalar seed $s_m^{(0)} = s_m$ and the zero vector state $Q_m^{(0)} = 0$. Each slot acquires directional information through attention-weighted pooling of the encoded atomic vectors.

Initialization. The optional slot self-attention in Equation (4) acts in place on the scalar seeds. It is enabled for the full-model setting and disabled for the cross-attention-only ablation in Table 2. With these scalar states, the initialization attention and vector readout are

$$\begin{aligned} A_{ma}^{(0,1)} &= \mathrm{softmax}_a\left[\frac{(W_q s_m^{(0)})^\top (W_k u_a^{(0)})}{\sqrt{d_s}}\right], \\ U_m^{(0,1)} &= \sum_a A_{ma}^{(0,1)} W_V V_a, \qquad s_m^{(1)} = s_m^{(0)}, \qquad Q_m^{(1)} = U_m^{(0,1)}. \end{aligned} \tag{51}$$

Here $u_a^{(0)}$ is an MLP embedding $128 \to 256 \to 256$ of u_a. The zero initial vector state and $\gamma^{(0,1)} = 0$ reduce the main-text template to this scalar-attention bootstrap.

Two vector–scalar updates. For $t = 1, 2$, write $u_a^{(t)} = u_a^{(0)}$. The shared scalar maps are W_q, W_k, W_z, and the shared atomic vector maps are W_K, W_V. The slot-vector maps, scalar update MLPs, vector update maps, and compatibility coefficients are stage specific. Omitting only the single head index, the main-text attention becomes

$$A_{ma}^{(t)} = \mathrm{softmax}_a\left[\frac{(W_q s_m^{(t)})^\top (W_k u_a^{(t)})}{\sqrt{d_s}} + \frac{\gamma^{(t)}}{\sqrt{3 d_v}} \left\langle \widetilde{W}_Q^{(t)} Q_m^{(t)}, W_K V_a \right\rangle_F\right]. \tag{52}$$

The maps W_q, W_k, W_z, W_K, W_V are shared with initialization where applicable. Initial logits use the scalar term with zero vector-compatibility coefficient, while each update-stage $\gamma^{(t)}$ starts at one. The contexts and updates are

$$\begin{aligned} z_m^{(t)} &= \sum_a A_{ma}^{(t)} W_z u_a^{(t)}, & U_m^{(t)} &= \sum_a A_{ma}^{(t)} W_V V_a, \\ s_m^{(t+1)} &= s_m^{(t)} + F^{(t)}\Big(s_m^{(t)}, z_m^{(t)}, (\|U_{m,v}^{(t)}\|_2)_{v=1}^{d_v}\Big), & Q_m^{(t+1)} &= Q_m^{(t)} + W_U^{(t)} U_m^{(t)}. \end{aligned} \tag{53}$$

The scalar update is a two-layer MLP. The invariant vector norms communicate geometric information to scalar fingerprints; vector-channel mixing preserves orientation. The scalar query/key/value maps in these stages and all vector-channel maps are bias free.

Scalar-only updates and readout. The three scalar-only stages use an independent embedding $u'_a \in \mathbb{R}^{256}$ from another $128 \to 256 \to 256$ MLP, with eight heads, pre-normalization, and the residual feed-forward update of Equation (8). They leave $Q_m^{(t)}$ unchanged. In the attention notation of the main text, $\hat{s}_m = \mathrm{N}^{(t)}(s_m^{(t)})$ and $\hat{u}_a = \mathrm{N}^{(t)}(u_a^{(t)})$, with the appropriate separately parameterized normalizations; $\mathrm{N}^{(t)}$ is identity in the unnormalized stages and layer normalization in the scalar-only stages.

The final single-head readout at $T = 6$ has its own $W_q, W_k, W_z, \widetilde{W}_Q, \widetilde{W}_V$. It recomputes attention with $\gamma^{(T)} = 1$ and returns $A_{ma}^{(T)}, z_m^{(T)}, s_m^{(T)}, Q_m^{(T)}$ for decoding, retaining the final slot states. Stage-specific vector maps mix channels identically across the three spatial components and use zero spatial bias.

C.3 Decoder details

Omitting the final stage index as in the main text, the anchor is

$$b_m = \sum_a A_{ma} X_a, \qquad \sum_a A_{ma} = 1. \tag{54}$$

The offset $X_a - b_m$ encodes each atom's position relative to the attention-weighted anchor. Computing this mean in the prepared local Cartesian environment gives the decoder a geometry-conditioned spatial reference.

The atom-derived generators are the 32 rows of $W_{\mathrm{atom}} V_a$, and the slot-derived generators are the eight rows of $W_{\mathrm{slot}} Q_m$. Higher-order encoder channels contribute through equivariant message passing. Together with the anchor-relative direction, they form the 41-row vector bank $\mathcal{B}_{ma}$ in Equation (9), combining three sources of geometric information for displacement synthesis.

The decoder uses the invariant conditioning

$$\begin{aligned} \xi_{ma} &= \left[s_m, z_m, u'_a, \chi_{ma}, \log(1 + \|X_a - b_m\|_2^2)\right], \\ \chi_{ma} &= \left(\frac{1}{\sqrt{3}} \langle (G_Q Q_m)_g, (G_V V_a)_g \rangle\right)_{g=1}^{8}. \end{aligned} \tag{55}$$

The coordinates are numerically expressed in the fixed input length unit. With $s_m, z_m, u'_a \in \mathbb{R}^{256}$, $\xi_{ma} \in \mathbb{R}^{777}$. The maps G_Q and G_V have channel widths $16 \to 8$ and $64 \to 8$, respectively.

The four MLPs follow Equation (11). Gate and scale networks $g_{\mathrm{act}}, g_{\mathrm{mag}}$ have widths $777 \to 256 \to 1$; coefficient networks $f_{\mathrm{coef}}, f_{\mathrm{res}}$ have widths $777 \to 256 \to 256 \to 41$. Hidden activations are SiLU, and the outputs are p_{ma}, κ_{ma}, q_{ma}, and q_{ma}^{res}.

The output layers of $g_{\mathrm{act}}, g_{\mathrm{mag}}, f_{\mathrm{res}}$ start with zero weights and biases $-2, -3.9, 0$, respectively. The learned residual scale $\gamma_{\mathrm{res}} = \sigma(\vartheta)$ starts at 0.05. Initial participation is $\sigma(-2) \approx 0.12$, the magnitude parameter is $\mathrm{softplus}(-3.9) \approx 0.02\,\text{Å}$, and the residual output is zero. The realized displacement magnitude also depends on the directional combination.

The residual lies outside the activity gate in Equation (12), allowing small-participation atoms to express surrounding relaxation. Radial normalization uses $v / \max(\|v\|_2, \epsilon)$ with $\epsilon = 10^{-8}$, giving norm $\min(1, \|v\|_2/\epsilon)$ and zero output at $v = 0$. This stabilized geometric operation preserves the rotation law used in the equivariance proof.

D DISPLACEMENT SUPERVISION AND ACTIVE-AWARE MATCHING

D.1 DISPLACEMENT FIELDS AND THE UNIFORM OBJECTIVE

For each initial configuration $X \in \mathbb{R}^{N\times 3}$, the predictor produces M displacement fields $\widehat{D}_m$, supervised by K reference fields $D_e = S_e - X$. All fields share atom correspondence and Cartesian units of Å. Training uses the full residual $\widehat{D}_m - D_e$ directly; inference takes the initial configuration as input.

The scalar SmoothL1 penalty and its all-atom reduction are

$$\rho_\beta(t) = \begin{cases} t^2/(2\beta), & |t| < \beta, \\ |t| - \beta/2, & |t| \geq \beta, \end{cases} \qquad \ell_\beta(E) = \frac{1}{N}\sum_{a=1}^{N}\sum_{d=1}^{3}\rho_\beta(E_{ad}). \tag{56}$$

The objective sums the three coordinate penalties and averages over all N atoms, including those with zero reference displacement. The main Uniform model uses $C_{me} = \ell_\beta(\widehat{D}_m - D_e)$ for both assignment and matched regression, with $\beta = 0.03$ Å.

D.2 THE ACTIVE-AWARE ABLATION

The active-aware ablation changes both the proposal–reference assignment cost and matched displacement loss. For each reference event, define

$$\begin{gathered} g_{e,a} = \|D_{e,a}\|_2, \qquad a_{ma} = \|\widehat{D}_{m,a}\|_2, \qquad a_e^* = \arg\max_a g_{e,a}, \\ \mathcal{A}_e = \begin{cases} \{a : g_{e,a} \geq \delta_{\mathrm{act}}\}, & \text{if this set is nonempty,} \\ \{a_e^*\}, & \text{otherwise,} \end{cases} \qquad \delta_{\mathrm{act}} = 0.05\ \text{Å}. \end{gathered} \tag{57}$$

The active set is defined from each event's reference displacement. The fallback ensures a nonempty set, with tied maxima resolved by the first maximum in stored atom order.

Let the coordinatewise penalty at one atom be

$$E_{me,a} = \sum_{d=1}^{3}\rho_\beta(\widehat{D}_{m,ad} - D_{e,ad}).$$

The five enabled components are

$$\begin{aligned} \ell_{\mathrm{all},me} &= \frac{1}{N}\sum_a E_{me,a}, & \ell_{A,me} &= \frac{1}{|\mathcal{A}_e|}\sum_{a\in\mathcal{A}_e} E_{me,a}, \\ \ell_{\max,me} &= E_{me,a_e^*}, & \ell_{\mathrm{mag},me} &= \frac{1}{|\mathcal{A}_e|}\sum_{a\in\mathcal{A}_e}\rho_\beta(a_{ma} - g_{e,a}), \\ \ell_{\cos,me} &= \frac{1}{|\mathcal{A}_e|}\sum_{a\in\mathcal{A}_e}\omega_{m,a}\big[1 - \cos_{\epsilon_{\cos}}(\widehat{D}_{m,a}, D_{e,a})\big]. \end{aligned} \tag{58}$$

Here the cosine and its detached amplitude gate are

$$\begin{gathered} \cos_{\epsilon_{\cos}}(u,v) = \frac{\langle u, v\rangle}{\max(\|u\|_2, \epsilon_{\cos})\max(\|v\|_2, \epsilon_{\cos})}, \qquad \epsilon_{\cos} = 10^{-4}, \\ \omega_{m,a} = \mathrm{stopgrad}\left[\mathrm{clip}\left(\frac{a_{ma}}{\max(\delta_{\mathrm{act}}, 10^{-6})}, 0, 1\right)\right]. \end{gathered} \tag{59}$$

The stabilizers use the supplied length convention. The detached gate reduces directional supervision near a zero prediction. Each active mean is normalized by its event's active-atom count, and the maximum-atom term evaluates the atom with the largest reference displacement.

The enabled pair objective is

$$C_{me}^{\mathrm{active}} = \ell_{\mathrm{active},me} = 0.25\ell_{\mathrm{all},me} + \ell_{A,me} + \ell_{\max,me} + 0.5\ell_{\mathrm{mag},me} + 0.1\ell_{\cos,me}. \tag{60}$$

The all-atom term remains present at reduced weight. The active variant retains $\beta = 0.03$ Å and the same sequential matching and configuration-level aggregation as Uniform. The 0.05 Å training activity cutoff is distinct from the judge's 0.3 Å moving-atom cutoff and its score threshold τ.

D.3 Sequential assignment and loss aggregation

Here $K = |\mathcal{E}_X^{\mathrm{ref}}|$ counts reference targets, following the training notation of Section 4.4. Let $\mathcal{O}_r$ denote the slots available at round r, starting with $\mathcal{O}_1 = \{1, \dots, M\}$. For a nonempty reference set, define the rectangular one-to-one assignment family

$$\mathfrak{M}(\mathcal{O}_r, [K]) = \{\Pi \subseteq \mathcal{O}_r \times [K] : |\Pi| = \min(|\mathcal{O}_r|, K), \quad \deg_\Pi(m) \le 1, \quad \deg_\Pi(e) \le 1\}. \tag{61}$$

At each round, Hungarian matching solves

$$\Pi_r \in \operatorname*{arg\,min}_{\Pi \in \mathfrak{M}(\mathcal{O}_r, [K])} \sum_{(m,e)\in\Pi} C_{me}, \qquad \mathcal{O}_{r+1} = \mathcal{O}_r \setminus \{m : (m,e) \in \Pi_r \text{ for some } e\}. \tag{62}$$

Use $C_{me} = \ell_{\mathrm{all},me}$ for Uniform and $C_{me} = C_{me}^{\mathrm{active}}$ for the active-aware ablation. The detached cost matrix determines the discrete assignment; regression then propagates gradients through the predicted fields of the selected pairs.

The procedure performs at most $R_{\mathrm{match}} = 4$ rounds, stopping when no slots remain. The full reference set is available again at every round. A slot is used at most once, whereas each target can receive one slot per round. With finite pair costs and $K > 0$, the total number of matched pairs is

$$\sum_r |\Pi_r| = \min(M, R_{\mathrm{match}} K). \tag{63}$$

The number of supervised slots adapts to the reference-set size: for example, $K = 2$ gives eight matched slots when $M = 96$. When $M \ge K$, the first round covers every target, and subsequent rounds assign the remaining slots against the full reference set. Each round solves its own residual assignment problem.

Primary, duplicate, and batch normalization. Write $\Pi_+ = \bigcup_{r\ge 2} \Pi_r$. Using the chosen pair objective C_{me}, define

$$L_1 = \frac{1}{|\Pi_1|} \sum_{(m,e)\in\Pi_1} C_{me}, \qquad L_+ = \begin{cases} |\Pi_+|^{-1} \displaystyle\sum_{(m,e)\in\Pi_+} C_{me}, & |\Pi_+| > 0, \\ 0, & |\Pi_+| = 0. \end{cases} \tag{64}$$

The configuration loss uses the main-text notation $\mathcal{L}_x$ and aggregation

$$\mathcal{L}_x = \lambda_D(L_1 + \lambda_+ L_+), \qquad \lambda_D = 10, \quad \lambda_+ = 0.5. \tag{65}$$

The primary and duplicate terms are averaged separately before applying the 0.5 duplicate weight. All later-round pairs contribute to the common duplicate mean. With one matching round or an empty duplicate set, $L_+ = 0$; an empty reference set gives zero loss. Batch reduction weights each configuration equally.

Repeated supervision gives a reference event multiple assigned proposals across rounds. Assigned pairs provide displacement gradients through the shared predictor, while distinct event recovery is measured after Dimer refinement and reference matching.

D.4 Predictor optimization and checkpoint scope

The predictor training recipe uses AdamW with learning rate 8×10^{-5}, weight decay 10^{-5}, batch size two, gradient clipping at five, and a 120-epoch cosine schedule. The data split comprises 3,700 training, 100 monitoring, and 200 held-out configurations. Monitoring-set first-round displacement loss selects the predictor checkpoint. Downstream recovery and search-force metrics evaluate its physical-search performance. Appendix E specifies the frozen Full-model checkpoint and monitoring procedure for the judge.

E Initial-State-Only Neural RMS Judge

E.1 Frozen predictor and inference inputs

The neural judge scores proposals using only the initial configuration, its frozen encoder features, and the complete predicted displacement pool. For the Full-model pairing described here, the predictor checkpoint is epoch 103, loaded with strict parameter matching and kept in evaluation mode.

Its scalar features $u_a \in \mathbb{R}^{128}$, vector channels $V_a \in \mathbb{R}^{64\times 3}$, and proposals $\widehat{D}_m$ are detached from judge optimization. This frozen-feature readout scores the proposal pool without force evaluations. Reference displacements supply offline training and monitoring labels; inference uses the initial state and its predicted proposals.

E.2 Atom features, pooling, and network widths

Let $a_{ma} = \|\widehat{D}_{m,a}\|_2$. To avoid confusing the judge feature with the decoder coefficients q_{ma}, denote it by ζ_{ma}:

$$\zeta_{ma} = \Big[u_a,\ \big(\langle \widehat{D}_{m,a}, V_{a,v}\rangle\big)_{v=1}^{64},\ a_{ma},\ \log(1+a_{ma})\Big] \in \mathbb{R}^{194}. \tag{66}$$

Numerical amplitudes in logarithms are expressed in the fixed Å convention. A shared atom network produces $h^J_{ma} = \phi_J(\zeta_{ma}) \in \mathbb{R}^{128}$, where

$$\phi_J : 194 \xrightarrow{\text{Linear}} 192 \xrightarrow{\text{SiLU, Dropout(0.2)}} 192 \xrightarrow{\text{Linear}} 128 \xrightarrow{\text{SiLU}} 128.$$

A learned affine map $w_J^\top h + b_J$ gives the attention score. The complete readout is

$$\begin{aligned} \alpha^J_{ma} &= \mathrm{softmax}_a\big[w_J^\top h^J_{ma} + b_J + 3\log(1+a_{ma})\big], \\ v^J_m &= \left[\sum_a \alpha^J_{ma} h^J_{ma},\ \max_a h^J_{ma},\ g_m\right] \in \mathbb{R}^{266}, \\ t^J_m &= \psi_J(v^J_m), \qquad \widehat{p}_m = \sigma(t^J_m), \end{aligned} \tag{67}$$

where the maximum is channelwise and $g_m \in \mathbb{R}^{10}$ is defined below. The output network is

$$\psi_J : 266 \xrightarrow{\text{Linear}} 128 \xrightarrow{\text{SiLU, Dropout(0.2)}} 128 \xrightarrow{\text{Linear}} 1.$$

Both dropout layers are disabled for scoring. The head's sigmoid output provides the score used for threshold selection.

E.3 The ten geometric descriptors, in implementation order

All descriptors are computed *before* filtering, from the full proposal pool of the configuration. Define

$$a^{\max}_m = \max_a a_{ma}, \qquad j^*_m = \arg\max_a a_{ma}, \qquad \mathcal{U}_m = \{a : a_{ma} \ge 0.3\,\text{Å}\}.$$

The descriptor vector is

$$g_m = (f_{\mathrm{mmax}}, f_{\mathrm{band}}, f_{\mathrm{njump}}, f_{\mathrm{nhalo}}, f_{\mathrm{nclust}}, f_{\mathrm{second}}, f_{\mathrm{sum}}, f_{\mathrm{red}}, f_{\mathrm{rank}}, f_{\mathrm{coher}})_m.$$

The following definitions specify the raw features supplied to the head; no feature standardization is inserted.

Amplitude and atom counts (features 1–4 and 7).

$$\begin{aligned} f_{\mathrm{mmax},m} &= a^{\max}_m, \\ f_{\mathrm{band},m} &= \max\{0,\ 0.8\,\text{Å} - a^{\max}_m,\ a^{\max}_m - 1.6\,\text{Å}\}, \\ f_{\mathrm{njump},m} &= \big|\{a : a_{ma} \ge 0.5\,\text{Å}\}\big|, \\ f_{\mathrm{nhalo},m} &= \big|\{a : 0.05\,\text{Å} \le a_{ma} < 0.5\,\text{Å}\}\big|, \\ f_{\mathrm{sum},m} &= \sum_a a_{ma}. \end{aligned} \tag{68}$$

The count features retain absolute atom counts, and $f_{\mathrm{sum},m}$ sums their displacement amplitudes.

Moving-atom clusters (features 5–6). Construct an undirected graph on $\mathcal{U}_m$, connecting two moving atoms when their *initial* positions are separated by at most 3 Å. Let $\mathcal{C}_{m,1}, \dots, \mathcal{C}_{m,c_m}$ be its connected components and let $P_{m,j} = \max_{a\in\mathcal{C}_{m,j}} a_{ma}$ be their peak amplitudes. Then $f_{\mathrm{nclust},m} = c_m$ and $f_{\mathrm{second},m}$ is the second-largest component peak. An empty moving set gives $(0,0)$, and fewer than two components gives $f_{\mathrm{second},m} = 0$. The epoch-103 judge computes distances as $\|X_a - X_b\|_2$ in the prepared nonperiodic local-coordinate representation of the active volume.

Pool-relative redundancy and rank (features 8–9). Let $\mathcal{H}_m = \{n \neq m : a_n^{\max} > a_m^{\max}\}$. The redundancy distance is

$$f_{\mathrm{red},m} = \begin{cases} \min\limits_{n \in \mathcal{H}_m} \left(\dfrac{1}{N} \sum\limits_a \|\widehat{D}_{m,a} - \widehat{D}_{n,a}\|_2^2 \right)^{1/2}, & \mathcal{H}_m \neq \varnothing, \\ 5\,\text{Å}, & \mathcal{H}_m = \varnothing. \end{cases} \tag{69}$$

This feature measures all-atom Cartesian RMS separation within the proposal pool. The strict $>$ test selects higher-amplitude proposals. Rank is the zero-based position in the descending sort of $a_m^{\max}$, ranging from 0 to $M-1$, with ties following the array-sort ordering.

Directional coherence (feature 10). When $|\mathcal{U}_m| > 1$ and $a_m^{\max} > 0$,

$$f_{\mathrm{coher},m} = \frac{1}{|\mathcal{U}_m|} \sum_{a \in \mathcal{U}_m} \left\langle \frac{\widehat{D}_{m,a}}{\max(a_{ma}, 10^{-8})}, \frac{\widehat{D}_{m,j_m^*}}{a_m^{\max}} \right\rangle; \tag{70}$$

otherwise it equals one. The mean includes the maximum-displacement atom itself. j_m^* is the first maximum in the stored atom order.

Geometric invariance and tie conventions. Norms, dot products, local distances, and atom pooling preserve rotations under the predictor's transformation laws. Exact atom-permutation invariance also requires compatible handling of tied maximum amplitudes in coherence; proposal-rank ties follow pool order. The redundancy and rank features provide context from the complete proposal pool, so scores reflect both individual geometry and the alternatives proposed for the same state.

E.4 Offline labels and judge training

For each proposal–reference pair, use the main-text moving union

$$U_{me} = \{a : \max(\|\widehat{D}_{m,a}\|_2, \|D_{e,a}\|_2) \geq 0.3\,\text{Å}\}.$$

For nonempty unions, compute

$$r_{me}^U = \left(\frac{1}{|U_{me}|} \sum_{a \in U_{me}} \|\widehat{D}_{m,a} - D_{e,a}\|_2^2 \right)^{1/2}, \qquad d_{me}^U = \max_{a \in U_{me}} \|\widehat{D}_{m,a} - D_{e,a}\|_2. \tag{71}$$

Empty unions are ineligible, equivalently having infinite RMS. First select the reference e_m^* of smallest r_{me}^U, keeping the first reference in catalogue order at a tie. Then set

$$y_m = \mathbf{1}\{r_{me_m^*}^U < 0.35\,\text{Å} \text{ and } d_{me_m^*}^U < 0.5\,\text{Å}\}. \tag{72}$$

If no eligible reference exists, $y_m = 0$. Both tests are strict and apply to the RMS-nearest reference selected above. These geometric-accuracy labels train proposal screening; the final event evaluation uses the all-atom displacement and barrier criteria.

Head training and checkpoint selection. Gradient updates use training configurations only, as stated in Section 4.5. The stage-1 head-training recipe uses the first 200 configurations of `train_strict` for those updates and the first 50 monitoring configurations for head selection; no held-out test configurations enter either role. One optimizer step processes all slots of one configuration, and the binary cross-entropy is averaged over those slots.

Let π_+ be the positive-label fraction obtained by concatenating all training labels. The positive-class weight is

$$w_+ = \frac{1 - \pi_+}{\max(\pi_+, 0.001)}, \qquad 0 < \pi_+ < 1. \tag{73}$$

For one configuration, the logit loss is

$$\mathcal{L}_J = -\frac{1}{M} \sum_m \left[w_+ y_m \log \sigma(t_m^J) + (1 - y_m) \log(1 - \sigma(t_m^J)) \right].$$

The implementation uses binary cross-entropy with logits for numerical stability. The head uses AdamW with learning rate 8×10^{-4} and weight decay 10^{-4}. Torch and Python random seeds are reset to zero, and training configurations are shuffled each epoch. Training lasts at most 40 epochs. After each epoch, monitoring ROC AUC is computed from the concatenated proposal scores and labels. Strictly improved AUC saves the head; six consecutive non-improving epochs stop training. The best saved head is restored with strict parameter matching and placed in evaluation mode.

E.5 Deployment and proposal identifiers

Compute the ten descriptors on the complete proposal pool, evaluate the head, and refine exactly the indices

$$I_X(\tau) = \{m : \widehat{p}_m \geq \tau\}. \tag{74}$$

Selection retains the original slot/search identifiers and uses descriptors computed on the full pool. The default score threshold $\tau = 10^{-5}$ is used in the 50-configuration wall-time comparison (Figure 5). The 200-configuration recovery curves vary this threshold, while the ablations in Table 2 disable selection. The sigmoid readout is used as a geometric proposal-selection score, with physical recovery assessed after refinement.

F Native Dimer Refinement and Validation

F.1 Enabled path and coordinate convention

The physical search uses the native SEAKMC Dimer implementation on the fixed interatomic potential. The geometric background of Dimer is given in Appendix A.4; here we specify the supplied numerical path. Searches use Cartesian active-atom coordinates with the native fixed/buffer surroundings. The options `SNC`, `ApplyMass`, `SearchBuffer`, `LocalRefine`, and `EarlyStop` are disabled, while `LocalRelax` is enabled. The main Dimer search refines every submitted proposal; `LocalRefine=False` specifies that its additional local-refinement branch is disabled.

The paper writes atomic fields as $N \times 3$ arrays. The native search stores their transpose as $3 \times N$ arrays. Below, y and n denote the vectorized active-coordinate center and orientation, respectively, and $f(y) = -\nabla V(y)$ denotes the physical force in that same coordinate space. The potential includes fixed surroundings, while n spans the active coordinates. Native magnitude operations use $\mathrm{mag}_{\mathrm{native}}$, corresponding to the `mat_mag` helper and its stopping convention.

F.2 Random axes, proposal loading, and orientation reset

For a random-start search, the native initial-axis sampler draws an array with independent entries from $[-0.5, 0.5)$. The wrapper sets

$$\mathrm{seed}(c, i) = (2026091400 + 10000c + i) \bmod 2^{32}, \tag{75}$$

before the native draw, where c is the configuration ordinal and i is the search identifier. Angle checking against previous axes follows the native settings, completing the initial-axis sampling protocol.

For a learned or transferred field $D \in \mathbb{R}^{N\times 3}$, the bridge transposes the full field for preload and forms a separate translation-removed direction. In paper notation,

$$\bar{D} = \frac{1}{N}\sum_a D_a, \qquad \widetilde{n}(D) = \frac{\mathrm{vec}(D - \mathbf{1}\bar{D})}{\|D - \mathbf{1}\bar{D}\|_F}. \tag{76}$$

The helper accepts finite fields whose translation-removed norm exceeds 10^{-12}. Mean removal defines the axis; the complete field D preloads the candidate center $X + D$.

Native loading pads or restricts the field to the configured active coordinates and applies any type/fixed-coordinate scaling. Write its actual center field as D^{exec}, and let n_{pre} be the native axis immediately before loading. The bridge supplies $\widetilde{n}(D)$; any intervening native axis handling remains governed by the job settings. Let S denote the native type-scaling operator on the axis, or the

identity when no such scaling is enabled. The loading/reset rule is

$$n_0 = \begin{cases} \mathrm{unit}\big(\mathsf{S}\big[r_{\mathrm{reset}} n_{\mathrm{pre}} + \mathrm{vec}(D^{\mathrm{exec}})\big]\big), & \texttt{ResetVN04Preload=True}, \\ \mathrm{unit}(\mathsf{S} n_{\mathrm{pre}}), & \text{otherwise.} \end{cases} \tag{77}$$

The configured coefficient is $r_{\mathrm{reset}} = 0.2$ in the implementation's numerical Å convention. In the unscaled case with no intervening axis change, this reduces to normalizing $0.2\widetilde{n}(D) + \mathrm{vec}(D)$. This initializes the orientation from the proposal. Subsequent centering, rescaling, and angle controls follow the resolved job settings.

On an admissible, already represented displacement field D, the complete initialization in the conditional analysis is written

$$\mathcal{I}_X(D) = \big(X + D,\ n_0(D)\big). \tag{78}$$

Here $n_0(D)$ includes the configured orientation processing. The admissible domain $\mathcal{D}_X$ enforces coordinate constraints and a finite, nonzero processed orientation. The recovery analysis applies to the executed field and this configured initialization through its loading condition and local-success assumption.

F.3 Curvature estimation, rotation, and force projection

The native implementation reuses the center force to estimate endpoint curvature. At center y, let $f_0 = f(y)$ and $f_1 = f(y + h_{\mathrm{dim}} n)$. With `LocalRefine=False`, it computes

$$f_2 = 2f_0 - f_1, \qquad c = \frac{\langle f_2 - f_1, n\rangle}{2h_{\mathrm{dim}}}, \qquad f_{\mathrm{rot}} = \frac{(f_1 - f_2) - n\langle f_1 - f_2, n\rangle}{2h_{\mathrm{dim}}}. \tag{79}$$

The reused-force estimate f_2 supplies the opposite-endpoint approximation. The curvature c and tangential force drive the native orientation update, with separation $h_{\mathrm{dim}} = 0.005$ Å.

Native rotation and force projection. The rotation update maintains a conjugate direction, with a restart based on the current and previous rotation forces. Its coefficient has the form

$$\gamma(f, f_{\mathrm{old}}) = \begin{cases} \dfrac{\langle f, f - f_{\mathrm{old}}\rangle}{\langle f_{\mathrm{old}}, f_{\mathrm{old}}\rangle}, & |\langle f, f_{\mathrm{old}}\rangle| \le \frac{1}{2}\langle f_{\mathrm{old}}, f_{\mathrm{old}}\rangle, \\ & \langle f_{\mathrm{old}}, f_{\mathrm{old}}\rangle \neq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{80}$$

For rotation, the accumulated direction is normalized to a search vector u. A forward angular sample updates the orientation and search vector by

$$(n, u) \longmapsto \big(n\cos\theta + u\sin\theta,\ u\cos\theta - n\sin\theta\big), \qquad \theta = \pi/4.$$

Writing the directional rotation-force samples as r_1, r_2, the return stage applies the incremental angle

$$\theta_{\mathrm{inc}} = \begin{cases} -\frac{1}{2}\arctan(r_1/r_2), & r_2 \neq 0, \\ -\frac{1}{2}\pi, & r_2 = 0 \end{cases} + \tfrac{1}{2}\pi\,\mathbf{1}\{r_2 > 0\} - \tfrac{1}{4}\pi. \tag{81}$$

Native normalization and enabled axis controls follow this update. Rotation-force tests and the rotation counter use `FMin4Rot`, `FThres4Rot`, and `NMax4Rot`.

For translation, write $f_{\parallel} = n\langle f_0, n\rangle$. The effective force is curvature dependent:

$$f_{\mathrm{eff}} = \begin{cases} f_0 - 2f_{\parallel}, & c < 0, \\ -f_{\parallel}, & c \ge 0 \text{ and the horizon mode is off}, \\ f_0 - f_{\parallel}, & c \ge 0 \text{ and the horizon mode is on.} \end{cases} \tag{82}$$

The horizon switch follows the native curvature, energy, and initial-well logic, selecting the appropriate positive-curvature branch.

Table 7: **Native numerical controls and defaults.** Defaults identify the implementation's parameter families. Each experiment uses its resolved settings and the reported translation cap.

Quantity / native key	Value	Scope
`DimerSep`	0.005 Å	supplied setting/default
Ordinary trial / maximal step	0.015/0.05 Å	Random defaults
Fine trial / maximal step	0.003/0.01 Å	preload defaults
`FConv`	10^{-6}	native force-magnitude threshold
`FMin4Rot` / `FThres4Rot`	0.001/0.1	native rotation tests
`NMax4Rot`	3	native rotation counter setting
`Tol4Connect`	0.1 Å	maximum coordinate component
`NMax4Trans`	300	ML and event reuse
`NMax4Trans`	450	Random, up to 80 starts

F.4 Two-stage translation and fine-step controls

The translation direction is also updated by a restarted conjugate-force rule using Equation (80) and is normalized to u. Let t_f and t_m be the current trial and maximal translation settings. The forward stage moves the current center by $t_f u$. With effective forces before and after that trial, define

$$a_1 = \langle f_{\mathrm{before}}, u\rangle, \qquad a_2 = \langle f_{\mathrm{trial}}, u\rangle, \qquad c_t = \frac{a_1 - a_2}{t_f}.$$

The second stage adds a further displacement $t_c u$. Its native rule is

$$t_* = \frac{(a_1 + a_2)/2}{c_t + 10^{-12}} \quad \text{when } c_t \geq 0,$$
$$t_c = \begin{cases} t_m, & c_t < 0, \\ t_m - t_f, & c_t \geq 0,\ |t_*| > t_m,\ t_* \geq 0, \\ t_f - t_m, & c_t \geq 0,\ |t_*| > t_m,\ t_* < 0, \\ t_* - t_f/2, & c_t \geq 0,\ |t_*| \leq t_m. \end{cases} \tag{83}$$

The combined center displacement is $(t_f + t_c)u$, with t_c determined by the trial forces and the configured maximal-step parameter in this two-stage update.

A preloaded search uses the native `FineStep` settings. The supplied ordinary trial/maximal defaults are $0.015/0.05$ Å, while the fine-step defaults are $0.003/0.01$ Å. When the decay style is not fixed, the source updates them in the form

$$t_f(j) = t_f^{(0)} d(j) + t_{\min}, \qquad t_m(j) = t_m^{(0)} d(j) + t_{\min} r_{\min},$$

with the corresponding ordinary or fine-step parameter family. The decay factor is either $d(j) = d_0^{\lfloor j/J \rfloor}$ or $d_0^{j/J}$ under the configured style. The resolved job settings select the ordinary or fine-step family and its decay parameters for each search.

F.5 Search loop, stopping, and cap handling

The loop calls `dimer_force` before the first preload to establish the initial energy. At the first iteration, an available preload is then loaded, the runner is prepared again, and another `dimer_force` call evaluates the preloaded configuration. Subsequent iterations alternate the native rotation and translation states. Rotation completion advances the native `NTSITR` counter, which supplies the recorded translation count.

In the enabled branch, the force-convergence flag `FO_CON` uses

$$\mathrm{mag}_{\mathrm{native}}(\mathtt{FC}) < \mathtt{FConv},$$

after the appropriate projection/update, and `ST_CON` requires `NTSITR > IgnoreTransSteps`. The search converges when both flags hold. The enabled branch uses these force and step flags for convergence; energy convergence is recorded separately. Native relaxation and connectivity checks complete the acceptance protocol.

The cap check is applied when NTSITR >= NMax4Trans; it sets the maximum-translation flag and terminates the search. Cap-reaching outputs are invalidated in native finalization. The downstream scorer also requires a translation count strictly below the configured cap. Optional native displacement, barrier, and angle checks remain governed by the job configuration.

F.6 Native two-sided relaxation and connectivity

With LocalRelax enabled and the additional refinement/early-stop branches disabled, the two relaxation seeds are constructed from the last search segment. In row-coordinate notation, let Y_c denote XCR, Y_0 denote X0, and r_{rel} denote DRATIO. The seeds are

$$Z_+^{(0)} = Y_0 + r_{\mathrm{rel}}(Y_c - Y_0), \qquad Z_-^{(0)} = Y_0 - r_{\mathrm{rel}}(Y_c - Y_0). \tag{84}$$

Each seed is relaxed by the native SPSRELAX runner with the configured fixed surroundings. Relaxed active coordinates are shifted back relative to the initial configuration by the native periodic reboxing convention. Thus the final search segment provides both seeds for the native two-sided connectivity test.

The implementation first identifies the farther and nearer relaxed branches using their overall native displacement magnitude from X. For those selected branches, it then computes

$$d_{\mathrm{far}} = \max_{a,d} |(Z_{\mathrm{far}} - X)_{ad}|, \qquad d_{\mathrm{near}} = \max_{a,d} |(Z_{\mathrm{near}} - X)_{ad}|.$$

If $d_{\mathrm{far}} < \texttt{Tol4Connect}$, it marks the event invalid and disconnected. Otherwise the event is connected when $d_{\mathrm{near}} < \texttt{Tol4Connect}$; when this latter test fails, the relaxation routine can return a valid but disconnected result. Final native barrier and displacement cuts can additionally invalidate an output. Connectivity uses a maximum coordinate-component tolerance; reference matching uses the all-atom Euclidean displacement gate.

F.7 Returned saddle geometry and the numerical map

The wrapper returns the final active saddle coordinates from XCR.T, the recorded barrier BARR, native valid and connected flags, the cap flag, translation count, and search-force count. The saddle field for scoring is

$$S_{\mathrm{out}} = \texttt{saddle_active_coords}, \qquad D_{\mathrm{out}} = S_{\mathrm{out}} - X. \tag{85}$$

Scoring uses this saved saddle geometry and its associated barrier. Native XDISP records the displacement of a selected downhill branch seed during finalization.

Let $\mathcal{R}_X^{\mathrm{num}}$ denote this configured initialization, search, and native acceptance pipeline. It accepts a configuration $X + D$ and returns the numerical saddle output when the native flags and cap checks pass, or $\perp$ otherwise. For completed search indices I,

$$A_I = \{\mathcal{R}_X^{\mathrm{num}}(X + D_m^{\mathrm{exec}}) : m \in I,\ \mathcal{R}_X^{\mathrm{num}}(X + D_m^{\mathrm{exec}}) \neq \perp\}. \tag{86}$$

The main-text ideal map specifies the saddle and connectivity target. The numerical map implements the configured search and native acceptance checks. Its connection to reference recovery is formalized by the local-success assumption for this map and the final reference matcher.

F.8 Search-force accounting and run configuration

Each entry to Dimer.dimer_force increments the search counter once per routine invocation. The ledger covers preloading and search work, including failed, duplicate, and capped searches with completed records. LAMMPS initialization and native SPSRELAX force calls lie outside this routine.

Reproduction uses the native helpers together with each run's resolved step-size and decay settings, axis controls, fixed-coordinate masks, displacement cuts, and timer boundaries. The specifications above document the numerical path and default parameter families.

G KDB EVENT-REUSE ADAPTATION

G.1 STORED MECHANISMS AND THE SCOPE OF THE BASELINE

The KDB comparison adapts the kinetic-database approach (Terrell et al., 2012). Insertion and querying retain the upstream geometric kernel at commit `a03dd2f`, including periodic clumping, graph mapping, feasibility tests, and overlap checks.

The template library is built exclusively from `train_strict`. A stored mechanism contains its training reactant X^{lib}, saddle S^{lib}, product $X^{+,\mathrm{lib}}$, mobile-atom indices, periodic clump orders, barriers, and source/event identifiers. The in-memory adapter retains every successfully inserted template and records failed insertions separately. Candidate deduplication is performed after querying as specified below.

KDB-494 denotes 494 stored event templates. Every template is queried in both directions: one directed entry uses the training reactant as its minimum, and the other uses the training product. Both share the stored saddle and swap their minimum/product roles and clump order. Thus the 494-template library provides 988 directed query entries. Reverse barriers equal the forward barrier minus the stored product energy bias. Lookup and ranking use the target initial state and the training-library records.

G.2 MOBILE ATOMS AND THE RETAINED LOCAL TEMPLATE

Let MIC_H denote the periodic displacement convention for the stored cell H, with row-vector coordinates as in the background. For each training atom, define

$$\begin{aligned} m_a = \max\{&\| \mathrm{MIC}_H(S_a^{\mathrm{lib}} - X_a^{\mathrm{lib}})\|_2,\ \| \mathrm{MIC}_H(S_a^{\mathrm{lib}} - X_a^{+,\mathrm{lib}})\|_2, \\ &\| \mathrm{MIC}_H(X_a^{+,\mathrm{lib}} - X_a^{\mathrm{lib}})\|_2\}. \end{aligned} \tag{87}$$

The mobile set consists of atoms with $m_a > 0.7\,\text{Å}$. If it is empty, insertion retains the atom with largest reactant-to-product displacement. The local template retains the mobile atoms and their neighbors in any of the reactant, saddle, or product states. The neighbor threshold uses the upstream elemental radii:

$$r_{ab}^{\mathrm{neigh}} = (r_a^{\mathrm{elem}} + r_b^{\mathrm{elem}})(1 + n_f), \qquad n_f = 0.2. \tag{88}$$

The upstream routine applies its periodic unwrapping/clumping procedure and size checks using these elemental radii.

G.3 GRAPH MATCHING AND GEOMETRIC SCREENING

For the target initial state X, the upstream kernel builds the reactant graph and searches subgraph mappings for the directed library entries. Each surviving mapping is subjected to the upstream periodic ordered clump construction, geometric alignment, and feasibility/overlap checks.

The acceleration layer caches target distances and applies the same necessary pair-distance condition earlier during mapping. For a newly mapped atom and each already mapped atom, at least one of the upstream periodic-image distances must satisfy

$$\left|d_{\mathrm{target}}^{(j)} - d_{\mathrm{template}}\right| < 2d_c, \qquad d_c = 0.3\,\text{Å}. \tag{89}$$

The image index j ranges over the upstream `atomAtomPermDistance` candidates. Caching and JIT execution accelerate this pair-distance check while preserving its image candidates; surviving mappings pass the full `feasibility()` path.

Orthogonal alignment and the geometric gate. After correspondence and periodic clumping have been established, let A and B be the centered template-minimum and mapped target-coordinate matrices. Compute

$$A^\top B = U\Sigma W^\top, \qquad \mathsf{O} = UW^\top. \tag{90}$$

The orthogonal map O acts on row vectors from the right and permits both rotations and reflections when aligning a stored template to the query configuration.

The alignment is computed by an SVD least-squares fit and then screened by the maximum atomwise residual,

$$d_{\mathrm{fit}} = \max_a \|(A\mathsf{O} - B)_a\|_2 \le d_c. \tag{91}$$

The SVD supplies the least-squares alignment, and d_{fit} supplies the maximum-residual query score and acceptance gate.

For a nearly coplanar template, the kernel also tries the normal-flipped alignment. The coplanarity test is

$$\frac{\sigma_{\min}(A^\top A)}{n_{\mathrm{tpl}}} < d_c^2,$$

where n_{tpl} is the retained template size. The alternative is obtained by changing the sign of the final row of $W^\top$ before forming O. Alignments passing the distance gate subsequently undergo the upstream product/mobile overlap checks.

G.4 Aligned state transfer and displacement extraction

Let μ_{tpl} and μ_X be the fitted template-minimum and mapped-target centroids, and let $\iota(a)$ be the target atom mapped to template atom a. For a surviving unconstrained mapping, the aligned saddle coordinates are placed into a copy of the target initial state:

$$\widetilde{S}_{\iota(a)} = (S_a^{\mathrm{lib}} - \mu_{\mathrm{tpl}})\mathsf{O} + \mu_X, \qquad \widetilde{S}_b = X_b \quad \text{for unfilled atoms } b. \tag{92}$$

The training product is transferred in the same way. Native constraints, periodic reboxing, and overlap checks remain part of the upstream path. The wrapper extracts the saddle and product displacement fields relative to X. In the supplied cell representation, its extraction is

$$u = \widetilde{S} - X, \qquad D^{\mathrm{cand}} = u - \mathrm{rint}(uH^{-1})H, \tag{93}$$

and analogously for the product. This fractional-rounding extraction uses the supplied cell representation.

Before the final reboxing, the mapped saddle displacement decomposes as

$$\begin{aligned}\widetilde{S}_{\iota(a)} - X_{\iota(a)} = (S_a^{\mathrm{lib}} - X_a^{\mathrm{lib}})\mathsf{O} \\ + \big[(X_a^{\mathrm{lib}} - \mu_{\mathrm{tpl}})\mathsf{O} + \mu_X - X_{\iota(a)}\big].\end{aligned} \tag{94}$$

The bracket is the initial-state fitting residual. KDB coordinate transfer therefore combines the transformed stored displacement with this alignment correction. KeyBank uses the direct displacement transfer described below.

G.5 Global candidate deduplication and Dimer submission

Entry-internal deduplication is disabled in the query path. After all directed entries have been queried, candidates are sorted by increasing geometric fit score and greedily deduplicated. Let D^{cand} and $D^{+,\mathrm{cand}}$ be one candidate's saddle and product displacements. It is discarded when an already retained candidate j satisfies both

$$\max_a \|D_a^{\mathrm{cand}} - D_{j,a}^{\mathrm{cand}}\|_2 \le 0.3\,\text{Å}, \qquad \max_a \|D_a^{+,\mathrm{cand}} - D_{j,a}^{+,\mathrm{cand}}\|_2 \le 0.3\,\text{Å}. \tag{95}$$

These fields use the target atom correspondence. Score-ordered greedy deduplication compares each candidate against retained candidates, using the transferred saddle and product predictions from the training library. Source, event, and direction identifiers are preserved.

Retained saddle displacements initialize native Dimer searches with a 300-translation cap. Physical acceptance and final recovery use the common protocol, including the all-atom $0.05\,$Å displacement and $0.02\,\mathrm{eV}$ barrier gates. Library size, directed entries, returned proposals, and recovered references summarize successive pipeline stages.

H ROOTED-TOPOLOGY KEYBANK ADAPTATION

H.1 ROOTED GRAPHS AND DIRECTED LIBRARY ENTRIES

KeyBank adapts the rooted-topology catalogue idea of k-ART (Béland et al., 2011) to retrieve proposals for the common native Dimer pipeline. KeyBank-32 and KeyBank-128 use libraries built from 32 and 128 `train_strict` configurations, respectively.

For a root atom r in a configuration, retain

$$\mathcal{N}_r = \{a : \|X_a - X_r\|_2 \le 5\,\text{Å}\}, \qquad p_a = X_a - X_r \quad (a \in \mathcal{N}_r). \tag{96}$$

The local graph has an atom-species/root label $(z_a, \mathbf{1}\{a = r\})$ on each vertex and an undirected edge when

$$\|p_a - p_b\|_2 < 3.145\,\text{Å}, \qquad a \neq b. \tag{97}$$

The neighborhood radius is inclusive and the bond cutoff is strict. A four-iteration Weisfeiler–Lehman hash with node labels identifies lookup buckets; exact graph mappings and geometric checks then resolve candidates.

KeyBank operates directly on local active-volume coordinates, with the stated assumption that the represented neighborhoods remain within a consistent cell image.

Forward and reverse library entries. For every training event, construct a forward entry with initial state X^{lib} and a reverse entry with initial state $X^{+,\text{lib}}$. In either direction, let X^{dir} be the chosen initial state, $X^{+,\text{dir}}$ the opposite minimum, and S^{lib} the shared saddle. The full directed fields are

$$D^{\text{lib}} = S^{\text{lib}} - X^{\text{dir}}, \qquad D^{+,\text{lib}} = X^{+,\text{dir}} - X^{\text{dir}}.$$

Choose the training root as the atom with largest $\|D_a^{\text{lib}}\|_2$, using the stored-order first maximum at a tie. The rooted graph is constructed from the directed initial state X^{dir}. An entry is retained only when the saddle-displacement tail outside its 5 Å neighborhood satisfies

$$\max_{a \notin \mathcal{N}_r} \|D_a^{\text{lib}}\|_2 < 0.3\,\text{Å}, \tag{98}$$

with an empty tail assigned value zero. Entries with equality at the threshold are rejected. The stored record contains local initial coordinates, the colored graph, saddle and product displacements within the neighborhood, source/event identifiers, direction, and tail maximum. The entry filter applies the stated saddle-displacement tail criterion.

H.2 EXACT GRAPH MAPPINGS AND NECESSARY DISTANCE CHECKS

At query time, *every* target atom is considered as a root. Target root selection uses no reference moving atoms, saddle, or product. For entries in the corresponding WL bucket, exact colored graph isomorphism is tested using the VF2 matcher. The root label must match the root label and chemical species must match.

The geometric matcher additionally prunes a candidate correspondence (i, j) against every already mapped pair (a, b) unless

$$\big|\|p_i - p_a\|_2 - \|q_j - q_b\|_2\big| \le 0.600001\,\text{Å}. \tag{99}$$

The 0.6 Å scale is twice the 0.3 Å atomwise fit tolerance, with the implementation's small numerical allowance. This pair-distance filter precedes the final fit gate. Candidate vertices are ordered by root status, root-relative distance, and index, followed by exact colored-isomorphism and geometric tests.

H.3 ORTHOGONAL FIT AND DIRECT DISPLACEMENT TRANSFER

For an exact graph mapping, reorder the target local coordinates to the template order. Subtract each local point set's own centroid, obtaining A and B, and form

$$A^\top B = U\Sigma W^\top, \qquad \mathsf{O} = UW^\top.$$

The orthogonal fit permits both rotations and reflections. Its acceptance criterion is

$$\max_a \|(A\mathsf{O} - B)_a\|_2 \leq 0.3\,\text{Å}. \tag{100}$$

The maximum residual on the least-squares fit serves as both acceptance criterion and candidate score.

KeyBank transfers stored displacement vectors by the fitted orthogonal map. With target indices $\iota(a)$,

$$\begin{aligned} D^{\text{cand}}_{\iota(a)} &= D^{\text{lib}}_a \mathsf{O}, \quad & D^{+,\text{cand}}_{\iota(a)} &= D^{+,\text{lib}}_a \mathsf{O}, \\ D^{\text{cand}}_b &= 0, & D^{+,\text{cand}}_b &= 0 \quad \text{for all unfilled target atoms } b. \end{aligned} \tag{101}$$

These fields directly transfer the library motion to matched target atoms, with zero displacement assigned to the remaining atoms.

H.4 Mapping limit, sorting, and candidate deduplication

The query has a global mapping-enumeration protection limit of 100,000 per target configuration, accumulated across roots and library entries. The counter increments for each yielded graph mapping before the alignment gate. When it exceeds the limit, enumeration stops and `truncated=True` records that retrieval ended at the mapping limit.

Candidates are sorted lexicographically by geometric fit score, source, event, reverse-direction flag, and query root. They are then greedily deduplicated against already retained candidates using both the saddle and product conditions in Equation (95), with $0.3\,\text{Å}$ for each. The transferred library product provides the product field used in these candidate-to-candidate comparisons.

Retained fields initialize native Dimer searches capped at 300 translations. Native acceptance and common reference matching determine recovery using the all-atom $0.05\,\text{Å}$ displacement and $0.02\,\text{eV}$ barrier criteria. Retrieval uses the graph, fit, tail, and deduplication settings above.

I Shared Multiscale Embedding of Search Trajectories

Complementary views of the energy landscape and search paths. Figure 2(a) presents the Sammon-projected energy background with initial, saddle, and product configurations. Panel (b) compares ASPIRE and conventional Dimer trajectories in one shared MDS coordinate system. The multiscale construction for panel (b) is detailed below.

I.1 Illustrative sample and one shared coordinate system

Figure 2(b) embeds 33 saved successful searches: 17 ASPIRE and 16 Random trajectories, totaling 4,994 frames. This selected sample visualizes search geometry; the quantitative recovery and cost experiments use their separately specified evaluation populations.

Both methods, all retained frames, and the geometric landmarks share one embedding with fixed atom identities and ordering. For a trajectory with stored initial configuration $X^{(\ell)}$, a single translation

$$t_\ell = \frac{1}{N}\sum_a (X_a - X^{(\ell)}_a)$$

aligns that initial configuration to the shared native state X. Agreement is checked after the shift, which is then applied unchanged to every full-atom frame of that trajectory.

I.2 Configuration features, landmarks, and radial preprocessing

For an aligned configuration Z, define

$$\mathbf{f}(Z) = \frac{\text{vec}(Z - X)}{\sqrt{N}}. \tag{102}$$

Consequently,

$$\|\mathbf{f}(Z_i) - \mathbf{f}(Z_j)\|_2 = \left(\frac{1}{N} \sum_a \|Z_{i,a} - Z_{j,a}\|_2^2 \right)^{1/2}. \tag{103}$$

This metric is the Cartesian RMS over all corresponding atoms. Coordinates and energies retain their recorded atom correspondence.

The reference list contains the native initial state, followed by 41 separately supplied landmarks: another initial-state record, 20 saddles, and 20 products. In the code's zero-based reference list, `refs[2:22]` selects the 20 saddle landmarks. Exactly equal feature rows across landmarks and trajectories share one two-dimensional variable, with first-occurrence ordering preserved. Associated energy records are retained separately for coloring.

Radial compression before dimensionality reduction. Let the reference radius be the median saddle distance from the initial configuration in the full feature space,

$$r_{\mathrm{ref}} = \mathrm{median}_{e=1}^{20} \|\mathbf{f}(S_e^{\mathrm{landmark}})\|_2, \qquad r_{\mathrm{map}} = 0.4 r_{\mathrm{ref}}. \tag{104}$$

The displayed selected map first applies the radial transformation

$$\mathbf{G}(\mathbf{f}) = \begin{cases} r_{\mathrm{map}} \operatorname{asinh}\left(\dfrac{\|\mathbf{f}\|_2}{r_{\mathrm{map}}} \right) \dfrac{\mathbf{f}}{\|\mathbf{f}\|_2}, & \|\mathbf{f}\|_2 > 0, \\ 0, & \mathbf{f} = 0. \end{cases} \tag{105}$$

The common radial transformation acts in the full $3N$-dimensional feature space before the shared two-dimensional fit. Fitting targets are distances between the transformed features $\mathbf{G}_i$. The display thus provides a compressed geometric view of the trajectories, with axes expressed in arbitrary units and physical search performance reported by the recovery and force-call metrics.

I.3 Six classes of distance constraints

Let the deduplicated feature nodes be indexed by i. The edge collection is constructed once from the original features $\mathbf{f}_i$ with random seed 20260924, and is kept fixed when changing the fitted metric. It contains the following six classes.

Global pairs. Generate 60,000 pairs by sampling a trajectory uniformly and then a saved frame uniformly within that trajectory, independently for the two endpoints. Self-pairs are removed. Uniform trajectory selection balances the contribution of the 33 selected successful searches at this stage.

Frame–reference pairs. Connect every saved frame of each trajectory to all reference-list entries. Each trajectory forms a separate normalized block with initial importance $1/33$.

Adjacent-step pairs. For each trajectory, connect all consecutive saved frames. The input check requires consecutive recorded translation-step indices. Each trajectory's adjacent-step block has initial importance $1/33$.

Temporal-window pairs. For each trajectory and each lag in $\{2, 4, 8, 16, 32\}$ shorter than the trajectory, connect frames at that lag. Each available block has initial importance $1/(33 \times 5)$.

Reference–reference pairs. Include all unordered pairs of reference-list entries, removing pairs whose node identifiers are identical. The block has initial importance one.

Original-space nearest neighbors. Connect each deduplicated feature node to five nearest neighbors in the original Euclidean feature space. This couples nearby configurations across trajectories and methods, with initial block importance one. Neighbor selection uses the original Euclidean feature distances.

Each block removes self-pairs and retains its own weights. Contributions from multiple blocks together define the weighted geometric constraints.

I.4 Block normalization and weighted stress

The optimizer uses the internal numerical scale $s_{\mathrm{num}} = 0.05$. Define original and compressed target distances in those internal coordinates by

$$d_{ij}^0 = \frac{\|\mathbf{f}_i - \mathbf{f}_j\|_2}{s_{\mathrm{num}}}, \qquad d_{ij} = \frac{\|\mathbf{G}_i - \mathbf{G}_j\|_2}{s_{\mathrm{num}}}. \tag{106}$$

For a block B, the base weighting is

$$b_B(d) = \begin{cases} 1/\max(d, 2\times 10^{-4}), & B \text{ is frame–reference or nearest-neighbor,} \\ 1, & \text{otherwise.} \end{cases} \tag{107}$$

The 2×10^{-4} stabilizing floor is expressed in the optimizer's internally scaled distance units.

Let α_B be the initial block importance above, with $\alpha_B = 1$ for the global block. Original-space weights are

$$w_{ij,B}^0 = \frac{\alpha_B b_B(d_{ij}^0)}{\max\Big(\sum_{(u,v)\in B} b_B(d_{uv}^0)(d_{uv}^0)^2, 10^{-15}\Big)}. \tag{108}$$

When substituting the compressed metric, the implementation preserves $\alpha_B^{\mathrm{eff}} = \sum_{(i,j)\in B} w_{ij,B}^0 (d_{ij}^0)^2$, which equals the intended α_B when the normalization floor is inactive, and recomputes

$$w_{ij,B} = c_B \frac{\alpha_B^{\mathrm{eff}} b_B(d_{ij})}{\max\Big(\sum_{(u,v)\in B} b_B(d_{uv}) d_{uv}^2, 10^{-15}\Big)}, \qquad c_B = \begin{cases} 12, & \text{adjacent step,} \\ 6, & \text{temporal window,} \\ 1, & \text{all other blocks.} \end{cases} \tag{109}$$

These block weights balance trajectory, temporal, local, and global geometric constraints in the shared fit.

The two-dimensional variables $Y_i \in \mathbb{R}^2$ minimize the weighted metric stress

$$\mathcal{S}(Y) = \sum_B \sum_{(i,j)\in B} w_{ij,B}\big(\|Y_i - Y_j\|_2 - d_{ij}\big)^2. \tag{110}$$

The implementation accumulates the two opposite endpoint contributions of each edge using the stabilized expression

$$2w_{ij,B}\frac{\widetilde{r}_{ij} - d_{ij}}{\widetilde{r}_{ij}}(Y_i - Y_j), \qquad \widetilde{r}_{ij} = \max(\|Y_i - Y_j\|_2, 10^{-12}).$$

Numerical stress evaluation also uses $\widetilde{r}_{ij}$. Above the 10^{-12} floor, the expression is the stress gradient; the floor stabilizes updates for coincident or nearly coincident nodes. Fitting uses geometric constraints from the selected successful trajectories and landmark coordinates. Energies provide the separate color overlay described below.

I.5 Initialization, optimization, and reproducibility

The selected map starts from a saved shared raw-metric MDS embedding. That baseline fits the uncompressed distances using the same edge construction, with three stages of adjacent-step multipliers $2, 8, 24$ and temporal-window multipliers equal to half those values. Each baseline stage allows up to 900 L-BFGS-B iterations. Its saved initial coordinates come from the common Sammon map in this workflow.

For the displayed map, the saved two-dimensional baseline is radially compressed with r_{map} and divided by s_{num} for initialization. The compressed-space distances and weights in Equations (106)–(109) are then optimized in two consecutive L-BFGS-B stages, each allowing 800 iterations. The optimization settings are `ftol`$= 2 \times 10^{-10}$, `gtol`$= 10^{-6}$, `maxcor`$= 12$, and `maxls`$= 30$. The initial-state node is recentered to the origin after each stage.

Both selected-map stages ran to their 800-iteration limits. Reproduction uses the saved baseline and initial coordinates together with the frames, landmarks, sampled edges, and hyperparameters. Final stress and distance distortion are evaluated from the saved optimization outputs.

I.6 Common display alignment, markers, and energy colors

After optimization, a single common orthogonal Procrustes alignment of the landmark coordinates places the map in the orientation of the saved baseline. No fitted scale factor is applied. The display coordinates are

$$y_i^{\text{display}} = \frac{s_{\text{num}}}{r_{\text{ref}}}(Y_i - Y_{\text{initial}})\mathsf{O}_{2D}, \tag{111}$$

so the axes are labeled in arbitrary units (a.u.). Both methods share this transformation and scale; panel view limits can provide a closer view of the ASPIRE region alongside the wider Random trajectories.

Lines retain all saved frames. Markers highlight translation-step indices divisible by five, together with each trajectory's first and last frames.

J Theoretical Analysis

J.1 Local convergence motivation

Proposition 1 (Local convergence, Luo et al., 2022, Theorem 4.2)**.** *In fixed, vectorized free coordinates, let s be a nondegenerate first-order saddle of V. Suppose $V \in C^2$ and, throughout $B(s, R_0)$, its Hessian is L_3-Lipschitz, has exactly one negative eigenvalue, and satisfies $\mu \le |\lambda_i| \le L$ for all i, where $0 < \mu < L$. We take $L_3 > 0$. Consider the ideal minimum-mode iteration*

$$y_{j+1} = y_j - \frac{2}{L+\mu}\big(I - 2v_1(y_j)v_1(y_j)^\top\big)\nabla V(y_j), \tag{112}$$

where $v_1(y_j)$ is an exact unit eigenvector associated with the lowest Hessian eigenvalue. Writing $r_j = \|y_j - s\|_2$, if $r_0 < \min\{R_0, \rho\}$, $\rho = 2\mu/L_3$, then the iterates remain in the neighborhood and converge geometrically to s, with $r_j \le \left(\frac{L+\mu}{L+3\mu}\right)^j \frac{\rho r_0}{\rho - r_0}$.

J.2 Proof of Geometric Consistency

The positions and vector-channel rows use the main-text row-vector convention. The input transformation is

$$X' = PXR^\top, \qquad R \in \mathrm{SO}(3), \tag{113}$$

where P reindexes atoms and all atom-indexed auxiliary features consistently. We establish rotation and atom-permutation equivariance for deterministic inference on the prepared Cartesian representation.

Lemma 1 (Attention and weighted readouts)**.** *Let $A = \mathrm{softmax}_{\mathrm{row}}(L)$. If $L' = LP^\top$, scalar atomic features satisfy $Z' = PZ$, and each vector channel satisfies $V' = PVR^\top$, then*

$$A' = AP^\top, \qquad A'Z' = AZ, \qquad A'V' = AVR^\top. \tag{114}$$

The vector identity also applies to $X' = PXR^\top$.

Proof. Row-wise softmax commutes with a permutation of its entries. Thus $A' = AP^\top$. The two readout identities follow by multiplication and $P^\top P = I$. A vector-channel projection acts on channels, so it commutes with right multiplication by $R^\top$. □

Lemma 2 (Slot transformation law)**.** *Suppose encoder scalars are invariant under rotations, encoder vector channels rotate covariantly, and both are reindexed by P. The slot stages and final readout satisfy, for each fixed slot m,*

$$s'_m = s_m, \qquad z'_m = z_m, \qquad Q'_m = Q_m R^\top, \qquad A' = AP^\top. \tag{115}$$

Proof. The learned scalar seeds and their scalar self-attention are unchanged by input rotations and atom permutations. The zero initial vector state is covariant. At initialization, scalar query–key logits only reindex across atoms, so Lemma 1 gives the required vector law for $Q_m^{(1)}$.

Assume the transformation laws hold at a vector–scalar stage. The scalar part of Equation (5) is invariant, and the vector part is invariant because

$$\langle UR^\top, VR^\top \rangle_F = \langle U, V \rangle_F.$$

Consequently, the attention weights reindex with the atoms. The pooled scalar context is invariant, and the vector context rotates by $R^\top$. Vector norms are invariant, so the MLP update preserves the scalar law; bias-free channelwise vector maps preserve the vector law. This proves the induction step for both vector–scalar stages.

The subsequent scalar-only stages use invariant fingerprints and atomic scalars. Their headwise dot products, normalizations, attention-weighted sums, and feed-forward updates therefore preserve scalar invariance and leave the vector state unchanged. The final readout uses invariant scalar products and vector inner products again. Applying Lemma 1 to its contexts yields Equation (115). □

Proposition 2 (Proposal equivariance)**.** *Assume deterministic inference, rotation- and permutation-compatible graph construction and neighbor selection, an equivariant encoder, the stated slot updates, invariant scalar decoder conditioning, bias-free channelwise vector maps, and the radial normalization in Equation (9). Then every fixed slot satisfies*

$$\widehat{D}_m(PXR^\top) = P\widehat{D}_m(X)R^\top, \qquad \widehat{S}_m(PXR^\top) = P\widehat{S}_m(X)R^\top. \tag{116}$$

Proof. Corresponding Cartesian edges rotate as $r'_{ba} = r_{ba}R^\top$ after atom reindexing. Their lengths and radial features are unchanged, and their spherical harmonics transform in the appropriate rotation representations. Compatible neighbor selection preserves the same reindexed graph. The equivariant encoder therefore supplies the scalar and vector laws required by Lemma 2.

The final attention satisfies $A' = AP^\top$. Hence

$$b'_m = \sum_a A'_{ma} X'_a = b_m R^\top.$$

The offsets $X_a - b_m$ rotate by $R^\top$ under the same atom reindexing. The 32 atom-derived vectors and eight slot-derived vectors also rotate, since their learned maps only mix vector channels. The stabilized direction satisfies

$$\mathrm{unit}_\epsilon(vR^\top) = \frac{vR^\top}{\max(\|v\|_2, \epsilon)} = \mathrm{unit}_\epsilon(v)R^\top,$$

including at $v = 0$. Thus every row of $\mathcal{B}_{ma}$ is covariant.

All scalar entries of ξ_{ma} are invariant: they are invariant scalar representations, inner products, or squared norms. The shared coefficient networks therefore produce the same $p_{ma}, \kappa_{ma}, q_{ma}, q^{\mathrm{res}}_{ma}$ at corresponding atom–slot pairs. Both the gated directional term and the ungated residual in Equation (12) rotate covariantly. This proves the displacement law. Adding $X' = PXR^\top$ gives the configuration law. □

Representation and implementation conditions. The deterministic inference setting uses atom-reindexing-compatible graph truncation, disabled dropout, and a common Cartesian chart. The arithmetic anchor is evaluated in that chart as described in Appendix A.2, with common origin changes handled during input preparation. These conditions establish the neural proposal transformation law.

K Single-Step Reference Interpretation of RWC

K.1 Reference rates and restricted one-step models

Fix a state X and a finite, nonempty reference set $\mathcal{E}_X^{\mathrm{ref}}$ with positive rates k_e. For a nonempty subset $G \subseteq \mathcal{E}_X^{\mathrm{ref}}$, write

$$\Lambda_G = \sum_{e \in G} k_e, \qquad p_G(e) = \frac{k_e}{\Lambda_G} \quad (e \in G), \qquad \Delta t_G \sim \mathrm{Exp}(\Lambda_G). \tag{117}$$

These are the single-step event distribution and waiting time described in Appendix A.5 (Voter, 2007), providing a kinetic interpretation of reference recovery at a fixed state.

Proposition 3 (Single-step interpretation of RWC)**.** *Let the recovered reference subset $\mathcal{E}_X^{\mathrm{rec}}$ be nonempty and retain the same rates as the full reference set. Extend its event distribution by zero outside that subset. Then*

$$\mathrm{TV}(p_{\mathrm{ref}}, p_{\mathrm{rec}}) = 1 - \mathrm{RWC}_X, \qquad \frac{\mathbb{E}[\Delta t_{\mathrm{rec}}]}{\mathbb{E}[\Delta t_{\mathrm{ref}}]} = \frac{1}{\mathrm{RWC}_X}, \tag{118}$$

where $p_{\mathrm{ref}} = p_{\mathcal{E}_X^{\mathrm{ref}}}$ and $p_{\mathrm{rec}} = p_{\mathcal{E}_X^{\mathrm{rec}}}$.

Proof. Put $c = \Lambda_{\mathrm{rec}}/\Lambda_{\mathrm{ref}} = \mathrm{RWC}_X \in (0, 1]$. On the recovered subset, $p_{\mathrm{rec}}(e) = p_{\mathrm{ref}}(e)/c \geq p_{\mathrm{ref}}(e)$. Consequently,

$$\begin{aligned}\mathrm{TV}(p_{\mathrm{ref}}, p_{\mathrm{rec}}) &= \frac{1}{2}\left[\sum_{e \in \mathcal{E}_X^{\mathrm{rec}}} (p_{\mathrm{rec}}(e) - p_{\mathrm{ref}}(e)) + \sum_{e \notin \mathcal{E}_X^{\mathrm{rec}}} p_{\mathrm{ref}}(e)\right] \\ &= \frac{1}{2}[(1-c) + (1-c)] = 1 - c.\end{aligned}$$

The exponential mean is $1/\Lambda_G$, so the ratio of waiting-time means is $\Lambda_{\mathrm{ref}}/\Lambda_{\mathrm{rec}} = 1/c$. □

K.2 Average recovery probability and the high-temperature limit

Sample one reference event E_X at each fixed evaluation state using its normalized reference rates. Then

$$\Pr(E_X \in \mathcal{E}_X^{\mathrm{rec}}) = \mathrm{RWC}_X.$$

For the fixed evaluation bank $\mathcal{X}$, linearity of expectation gives

$$\mathbb{E}\left[\frac{1}{|\mathcal{X}|}\sum_{X \in \mathcal{X}} \mathbf{1}\{E_X \in \mathcal{E}_X^{\mathrm{rec}}\}\right] = \frac{1}{|\mathcal{X}|}\sum_{X \in \mathcal{X}} \mathrm{RWC}_X = \overline{\mathrm{RWC}}. \tag{119}$$

Linearity of expectation makes this identity valid for dependent samples as well. Evaluation configurations receive uniform weight, consistently with the main-text metric.

Under the main-text common prefactor, $k_e = \nu \exp[-\Delta E_e/(k_{\mathrm{B}}T)]$. For a fixed finite reference set and a fixed recovered subset, every exponential approaches one as $T \to \infty$. Therefore

$$\lim_{T \to \infty} \mathrm{RWC}_X = \frac{|\mathcal{E}_X^{\mathrm{rec}}|}{|\mathcal{E}_X^{\mathrm{ref}}|} = \mathrm{Cov}_X.$$

This formal limit describes how the weights of a fixed catalogue become uniform under the common-prefactor convention. With event-specific prefactors, the limiting normalized weights are proportional to those prefactors.

K.3 Metric interpretation and conventions

The proposition uses fixed reference rates. Numerical barriers enter the matching gate, while matched events retain their reference rates in the sums. For an empty recovered set, RWC is zero and the restricted exit distribution and finite mean waiting time are undefined. Aggregating event identities into product-basin labels can decrease total variation, yielding a corresponding comparison at the basin level.

RWC measures the fraction of reference escape-rate mass recovered at each state. A common scaling of all rates preserves this fraction and rescales the clock, separating relative event recovery from absolute waiting times. Averaging RWC over the fixed state bank summarizes this reference-relative recovery across configurations.

L Physical Evaluation, Event Matching, and Cost Accounting

L.1 Fixed 50-configuration evaluation set and topology coverage

The ablation and wall-time experiments use the same fixed set $\mathcal{X}_{50}$ of 50 held-out configurations. It augments an existing 30-configuration evaluation set with 20 additional cases selected to broaden

Table 8: **Vacancy-connectivity coverage of the fixed evaluation subset.** The existing 30 cases are augmented by 20 cases to obtain the 50-case set used for both ablation and wall-time comparisons. Counts report the configuration selection in each connectivity class.

Component-size signature	Existing 30	Added 20	Final 50
4	0	2	2
$3+1$	4	4	8
$2+2$	0	5	5
$2+1+1$	11	5	16
$1+1+1+1$	15	4	19
Total	30	20	50

topology coverage. The set is drawn from the held-out configuration bank; the judge uses a separately defined 50-configuration monitoring subset.

The five vacancy-connectivity classes are the component-size signatures recorded as `fragmentation_signature`: 4, $3+1$, $2+2$, $2+1+1$, and $1+1+1+1$. For example, $3+1$ denotes one connected component of three vacancies and one isolated vacancy, whereas $1+1+1+1$ denotes four separate vacancy components. These signatures summarize connectivity in the dataset's vacancy graph. Each class can contain multiple geometric arrangements and fine-grained topologies. Encoder, judge, and KeyBank atom-neighbor graphs are defined in their respective method sections.

After excluding the existing 30 configurations from the held-out bank, the augmentation included all remaining cases in classes 4 and $2+2$. It prioritized the available rare target-PDR label combinations $(1,3)$, $(0,3)$, and $(0,1,2,3)$, then filled the remaining positions deterministically by configuration ID within the relevant connectivity classes until 20 additional cases had been selected. The target-PDR combinations refer to the stored `target_pdr_group` metadata; they complement the vacancy component-size signatures in offline subset construction. At query time, the predictor, judge, and event-reuse methods use their specified initial-state inputs.

The 20-case augmentation uses topology, target-label coverage, and configuration ID as its selection criteria. Together with the existing 30 cases, it forms a fixed evaluation set covering all five vacancy-connectivity classes. Every compared method uses this same designed subset.

L.2 Numerical acceptance and reference-event matching

The evaluation state is the relaxed 11 Å SEAKMC active volume X in the offline protocol of Section 5. All methods use the same fixed local reference catalogue $\mathcal{E}_X^{\mathrm{ref}}$, so coverage measures event recovery for that active-volume configuration.

For a completed search, an output is eligible for reference matching when the native valid and connected flags are true, the maximum-translation flag is false, and the recorded translation count is strictly below its configured cap. These operational acceptance criteria follow Appendix F; the reported pipeline uses them without an additional independent Hessian-index check.

Saddle geometry, recorded barriers, and reference identities. The output saddle coordinates are the saved `saddle_active_coords`, obtained from native `XCR.T`. Set $D_{\mathrm{out}} = S_{\mathrm{out}} - X$ and use the associated recorded barrier ΔE_{out}. Native `XDISP` records the branch-dependent finalization displacement. For reference e from the same X, the eligible set is

$$\mathcal{M}_X(S_{\mathrm{out}}) = \big\{ e : \max_{1\le a\le N} \|D_{\mathrm{out},a} - D_{e,a}\|_2 \le 0.05\,\text{Å}, \; |\Delta E_{\mathrm{out}} - \Delta E_e| \le 0.02\,\mathrm{eV} \big\}. \tag{120}$$

This main-text criterion uses both non-strict inequalities and compares all corresponding atoms in the fixed Cartesian representation. Matching uses the recorded saddle displacement and barrier directly.

The matcher $J_X(S_{\mathrm{out}})$ assigns at most one reference identity in $\mathcal{M}_X(S_{\mathrm{out}})$, or $\perp$ if the set is empty. Distinct recovery is the set of assigned reference identities: repeated hits to one reference count

Table 9: **Thresholds in the implementation.** "Reference" denotes the target configuration's reference catalogue. Event-reuse methods retrieve stored mechanisms from their training libraries.

Operation	Criterion / threshold	Target-reference use
Active-aware training	Reference displacement $\geq 0.05\,\text{Å}$; largest-reference-atom fallback	Training only
Judge movers	Proposal/reference union uses $\geq 0.3\,\text{Å}$	Offline labels only
Judge positive label	Nearest-union RMS $< 0.35\,\text{Å}$ and union maximum atom error $< 0.5\,\text{Å}$	Offline labels only
Judge selection	$\widehat{p}_m \geq \tau$ after full-pool features	None
KDB / KeyBank retrieval	Maximum aligned atom error $\leq 0.3\,\text{Å}$	None
Reuse candidate deduplication	Both saddle and product maximum field differences $\leq 0.3\,\text{Å}$	None
Native connectivity	Maximum coordinate component $< 0.1\,\text{Å}$ for the return branch, with the distinct-branch test	Initial X only
Final reference recovery	All-atom maximum Euclidean error $\leq 0.05\,\text{Å}$ and barrier error $\leq 0.02\,\text{eV}$	Evaluation only

Table 10: **Settings of the main-text evaluation populations.** The 50-state ablation and wall-time studies use the same held-out subset, but different judge settings. Proposal count, number of submitted searches, and translation cap are distinct quantities.

Experiment	States	Proposal pool and selection	Translation cap
Recovery curves	200	ASPIRE: up to 96 proposals, varying judge threshold; Random: up to 80 starts	300 / 450
Ablations	50	Judge disabled for every variant; slot counts and objectives as in Table 2	300
Wall time	50	ASPIRE: $\widehat{p}_m \geq 10^{-5}$; reuse library sizes as in Section 5.4	300; Random 450

once. The reference catalogue remains fixed and accepted unmatched outputs receive no reference-coverage credit. Accepted unmatched outputs are candidates for further mechanism analysis and expansion of the reference catalogue.

L.3 Distinct thresholds serve distinct operations

Training activity, judge labels, retrieval, numerical connectivity, and reference matching are separate definitions. Their use is summarized below to make clear which operations depend on reference data. KDB mobile cutoffs and KeyBank graph/tail criteria are specified in their respective sections. Dataset deduplication applies the 48 bcc operations to initial configurations, while event recovery uses the fixed-X matching predicate above.

L.4 Reference metrics and experiment-specific settings

Coverage and RWC use the recovered-reference subset and the reference rates, exactly as in Section 5:

$$\begin{aligned} \mathrm{Cov}_X &= \frac{|\mathcal{E}_X^{\mathrm{rec}}|}{|\mathcal{E}_X^{\mathrm{ref}}|}, \qquad \mathrm{RWC}_X = \frac{\sum_{e\in\mathcal{E}_X^{\mathrm{rec}}} k_e}{\sum_{e\in\mathcal{E}_X^{\mathrm{ref}}} k_e}, \\ \overline{\mathrm{RWC}} &= \frac{1}{|\mathcal{X}|}\sum_{X\in\mathcal{X}} \mathrm{RWC}_X, \qquad k_e = \nu \exp\!\left(-\frac{\Delta E_e}{k_{\mathrm{B}}T}\right). \end{aligned} \tag{121}$$

All reported RWC values use $T = 800\,\text{K}$ and the common prefactor ν, as stated in the main text. The average weights evaluation configurations uniformly. Matched outputs retain their assigned reference rates, and the common prefactor cancels from RWC (Appendix A.5). The reference catalogue, temperature, and prefactor convention are fixed across judge-threshold and random-start comparisons.

L.5 RECORDED SEARCH-FORCE COUNT

For the fixed evaluation bank $\mathcal{X}$, let I_X index recorded completed searches and $F_{X,m}$ count entries to `Dimer.dimer_force` in search m. The mean search-force cost is

$$\bar{F}_{\text{search}} = \frac{1}{|\mathcal{X}|} \sum_{X \in \mathcal{X}} \sum_{m \in I_X} F_{X,m}. \tag{122}$$

The ledger includes recorded force calls from failed, capped, duplicate, and accepted-unmatched searches. Each entry counts one invocation of the search-force routine for the configuration.

The search-force ledger excludes LAMMPS initialization, native `SPSRELAX` force work outside the routine, neural inference, and catalogue lookup.

ASPIRE and reuse searches use a 300-translation cap; the Random baseline uses up to 80 searches with a 450-translation cap. Preloaded searches use the native fine-step parameter family. The reported comparisons therefore evaluate these specified operating protocols. Each translation may involve multiple force calls for orientation updates.

L.6 WALL TIME, RETRIEVAL SIZE, AND EXPERIMENT RECORDS

Wall time is reported separately from search-force counts. In the main-text wall-time comparison, Random and event-reuse baselines use 16 CPUs, while ASPIRE uses 16 CPUs and one A100 GPU. This comparison uses the shared 50-state subset, with ASPIRE judge threshold $\tau = 10^{-5}$ as stated in Figure 5. The timing results complement search-force counts with recorded elapsed time under these hardware settings.

KDB-494 contains 494 stored templates with 988 directed query entries; KeyBank-32/128 designate libraries from 32/128 training configurations. Retained proposals undergo native Dimer refinement. KeyBank's truncation flag identifies retrievals stopped at the mapping limit. Detailed baseline construction appears in Appendices G and H.

The experiment populations comprise the 200-configuration recovery test, the shared 50-configuration ablation and wall-time subset of Appendix L.1, the judge's 200 training/50 monitoring configurations, and the 33 selected successful MDS trajectories. Each result is interpreted with its state selection, checkpoints, thresholds, library, Dimer settings, scorer, and force/timing protocol. The MDS sample provides a complementary illustration of search geometry.