WHEN REACTION GEOMETRY IS PERIODIC: UNDERSTANDING AND LEARNING TRANSITION STATES

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning transition-state (TS) geometries directly from reaction endpoints offers a route to accelerating reaction-pathway discovery, yet periodic systems pose distinct geometric challenges. Equivalent lattice representations alter endpoint displacements, local interactions can cross unit-cell boundaries, and reaction motion can be highly localized within extended structures. We characterize these effects across three periodic reaction benchmarks and introduce **SurfReact-TS**, a curated benchmark for constrained periodic surface reactions. Guided by this analysis, we develop **PeriTS**, a geometry-aware flow-matching framework that incorporates PBC-consistent endpoint conditioning, periodic interaction modeling, coordinate constraints, and reaction-focused learning. Across the three benchmarks, PeriTS or its matched One-shot counterpart achieves the lowest mean RMSD among the evaluated methods. Controlled studies further show that the value of iterative refinement depends strongly on the reaction regime, while increasing candidate breadth provides only modest gains in the settings studied. Selected-case DFT calculations provide complementary energy and force diagnostics. Overall, our results establish periodic TS prediction as a geometry-dependent learning problem in which representation, model design, and inference strategy should be adapted to the underlying reaction system.

1 INTRODUCTION

Chemical reactions transform reactants into products through changes in atomic arrangement. As these atoms rearrange, the system must pass through an energetic bottleneck between the initial and final states. The atomic configuration at this bottleneck is the *transition state* (TS). Its energy determines the activation barrier of the reaction and therefore plays a central role in reaction kinetics (Eyring, 1935). Accurately locating the TS is consequently important for understanding reaction mechanisms, estimating reaction rates, and exploring catalytic reaction pathways. Finding it, however, is substantially more difficult than optimizing a stable structure. Reactants and products are relatively stable configurations, whereas the TS is unstable and lies between them on the underlying potential-energy surface. Conventional saddle-point and reaction-path searches therefore require repeated energy and force evaluations and can become computationally expensive, particularly for large or complex systems (Henkelman & Jónsson, 1999; Henkelman et al., 2000).

Recent learning-based approaches offer a complementary strategy. Rather than constructing a TS guess entirely through iterative physical search, a model can take the reactant and product structures as endpoints and directly predict a candidate TS geometry for subsequent refinement and validation. Equivariant diffusion, optimal transport, and flow-based methods have substantially advanced this endpoint-conditioned prediction paradigm (Duan et al., 2023; 2025; Galustian et al., 2025; Luo et al., 2026). Much of this progress, however, has been developed and evaluated on finite molecular reactions and molecular benchmarks such as Transition1x (Schreiner et al., 2022).

Moving from molecules to extended periodic systems involves more than an increase in system size. The geometry of the prediction problem itself also changes. Atomic positions are equivalent under lattice translations, physically local interactions can cross cell boundaries, and the endpoint rearrangement associated with a reaction can be concentrated in a small subset of atoms. These effects influence how endpoint motion is represented, how local interactions are constructed, and which parts of a structure carry most of the observable rearrangement. Related difficulties in endpoint

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.

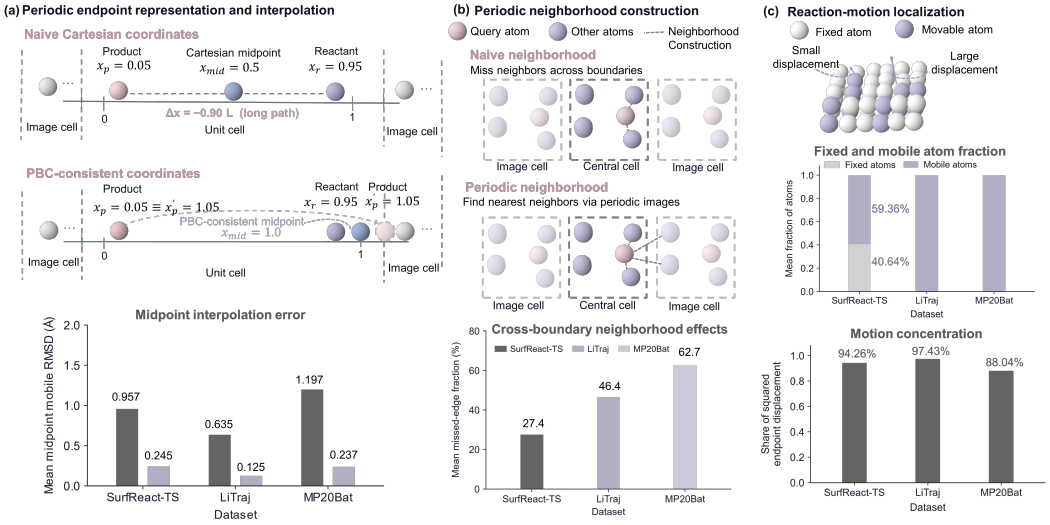

Figure 1: **What changes when reaction geometry is periodic? (a)** Periodic endpoint equivalence changes interpolation geometry, and a PBC-consistent midpoint provides a substantially better geometric prior than naive Cartesian interpolation. **(b)** Physically local neighbors can lie across unit-cell boundaries, causing non-periodic neighborhoods to miss a substantial fraction of local interactions. **(c)** Reaction-induced motion is highly localized, with the top 10% most-moving atoms accounting for most of the squared endpoint displacement. These observations motivate PBC-consistent endpoint conditioning, periodic interaction modeling, and reaction-focused learning in PeriTS.

alignment and interpolation also arise in conventional reaction-path searches for periodic surface systems (Jung et al., 2026). Fig. 1 previews these three aspects across the periodic reaction benchmarks considered in this work. Their prevalence differs substantially across systems, as quantified in Sec. 3.3.

Concurrent work on MaterialsSaddles and SaddleFlow explores endpoint-conditioned, flow-based transition-state prediction for periodic materials (Baghishov et al., 2026). Our focus is complementary: we take an analysis-driven perspective on how periodic geometry changes the learning problem, which corresponding inductive biases are empirically useful, and how additional inference computation should be allocated across periodic reaction systems.

This perspective motivates a central question: *which aspects of periodic geometry should be explicitly accounted for when learning transition-state structures?* PBC-consistent endpoint representation, periodic neighborhood construction, and reaction-focused learning are physically motivated, but physical plausibility alone does not establish their predictive value. We therefore characterize the relevant data properties and test the corresponding modeling choices separately.

In this work, we introduce **SurfReact-TS**, a curated benchmark of periodic surface reactions, and study it alongside two public periodic TS datasets. Together, these systems allow us to examine how periodic geometric effects vary across different reaction environments.

Motivated by this analysis, we introduce **PeriTS**, a geometry-aware flow-matching model for periodic transition-state prediction. PeriTS combines PBC-consistent endpoint conditioning, periodic interaction modeling, known coordinate constraints, and reaction-focused learning. Controlled comparisons evaluate which of these design choices provide reliable gains and which remain system dependent. We also construct a separately trained one-shot predictor with the same representation, enabling a controlled comparison between one-shot and iterative prediction.

We further study two complementary inference strategies: *candidate breadth* and *refinement depth*. Candidate breadth produces multiple alternative TS estimates, whereas refinement depth updates a candidate through successive flow evaluations. We compare their prediction quality under explicitly specified inference budgets and distinguish candidate-set coverage from predictions that can be selected or aggregated without the reference TS. Neural function evaluations (NFE) count network

calls within a stated architecture and solver; equal NFE alone does not guarantee equal wall-clock cost across different predictors.

Our contributions are threefold. **1)** We formulate and systematically study periodic TS prediction across distinct reaction regimes, and introduce **SurfReact-TS**, a new benchmark for constrained periodic surface reactions. **2)** We introduce **PeriTS**, a geometry-aware flow model tailored to periodic TS prediction through PBC-consistent endpoint conditioning, periodic interaction modeling, coordinate constraints, and reaction-focused learning. Together with its matched one-shot counterpart, our approach achieves the lowest mean RMSD across all three benchmarks among the evaluated methods. **3)** We identify when periodic modeling and iterative refinement provide additional gains, and contrast refinement depth with candidate breadth.

2 Related Work

2.1 Learning-based Transition-State Prediction

Learning-based TS prediction has progressed from direct geometric regression to generative modeling. Early approaches predict a single TS structure directly from reaction information, including distance-based models and TSNet (Pattanaik et al., 2020; Jackson et al., 2021). More recent methods formulate TS prediction generatively: OA-ReactDiff uses equivariant diffusion (Duan et al., 2023), React-OT adopts an optimal-transport formulation (Duan et al., 2025), and flow-based approaches such as GoFlow and TS-DFM perform iterative generation through learned vector fields (Galustian et al., 2025; Luo et al., 2026). These methods have been developed predominantly for finite molecular reactions. In our experiments, React-OT, and TS-DFM serve as representative direct, transport-based, and flow-based baselines, respectively.

2.2 Transition States in Periodic Systems

Periodic TS calculations have traditionally relied on iterative reaction-path or saddle-point searches. Recent methods such as CatTSunami accelerate these searches with pretrained machine-learning potentials, while NEBscape automates surface-reaction workflows including endpoint alignment, interpolation, and NEB optimization (Wander et al., 2025; Jung et al., 2026). These approaches accelerate or automate physical TS search rather than directly predicting a TS geometry from reaction endpoints. Direct learned TS generation for periodic materials remains comparatively less explored. Concurrent work on MaterialsSaddles and SaddleFlow independently investigates flow-based saddle prediction from periodic reactant–product pairs (SciLM, 2026). Our work focuses instead on how periodic geometry changes the learning problem, which corresponding modeling choices are empirically useful across systems, and how these choices can be incorporated into **PeriTS**; we additionally introduce **SurfReact-TS** as a benchmark for periodic surface TS prediction.

3 Transition-State Prediction in Periodic Systems

In this section, we first formalize the periodic TS-prediction task, then introduce the three benchmarks and characterize the geometric properties highlighted in Fig. 1.

3.1 Task Definition

We consider endpoint-conditioned TS-geometry prediction in a fixed periodic cell. For each reaction, the reactant and product contain the same N atom-mapped atoms, with Cartesian coordinates $R, P \in \mathbb{R}^{N\times 3}$ and atomic species Z. The periodic cell is described by a lattice matrix $H \in \mathbb{R}^{3\times 3}$, whose rows are the three lattice vectors. An integer vector $\mathbf{n} \in \mathbb{Z}^3$ indexes a periodic image, with Cartesian translation $\mathbf{n}H$. The indicator $\mathbf{p} \in \{0,1\}^3$ specifies which lattice directions are periodic. Some systems additionally provide coordinate constraints, represented by a mobility mask $M \in \{0,1\}^{N\times 3}$. If no constraints are supplied, we set $M = \mathbf{1}$. The mask specifies which coordinate components are allowed to move and does not indicate which atoms are chemically active.

The prediction task is

$$(R, P, Z, H, \mathbf{p}, M) \longmapsto \widehat{X}_{\mathrm{TS}} \in \mathbb{R}^{N\times 3}. \tag{1}$$

Here, $\widehat{X}_{\mathrm{TS}}$ denotes the predicted TS geometry.

3.2 Periodic Reaction Benchmarks

We study three periodic reaction benchmarks spanning distinct reaction environments. LiTraj (Dembitskiy et al., 2025) provides the `nebDFT2k` benchmark for lithium-ion migration, while the MP20Bat subset of MaterialsSaddles (Baghishov et al., 2026) contains bulk-material transition events. We preserve the official partitions of both public benchmarks.

We additionally introduce **SurfReact-TS**, a curated benchmark of periodic Pt-surface reactions involving C/H adsorbates. Unlike the two public benchmarks, SurfReact-TS contains constrained slab structures in which localized adsorbate rearrangements occur within a larger surface environment. Dataset preprocessing and partition details are provided in Appendix A.

3.3 Periodic Geometry Diagnostics

Fig. 1 highlights three geometric properties that distinguish the periodic benchmarks considered in this work. We report reaction-averaged statistics over the training and validation partitions.

Periodic endpoint representation. For corresponding atoms in the reactant R and product P, we compute the minimum-image displacement $d = \mathrm{MIC}_{H,\mathbf{p}}(P - R)$, where H is the lattice matrix, $\mathbf{p}$ specifies the periodic directions, and $\mathrm{MIC}_{H,\mathbf{p}}$ selects the shortest displacement among periodically equivalent images. We then apply the supplied mobility mask M through the projection Π_M and construct the PBC-consistent midpoint

$$X_{\mathrm{mid}} = R + \tfrac{1}{2}\Pi_M(d).$$

Here, Π_M removes displacement components that are not allowed to move.

For example, endpoint fractional coordinates 0.95 and 0.05 appear to differ by -0.90 in the stored cell, whereas the equivalent product image at 1.05 gives the shorter periodic displacement $+0.10$. Compared with the Cartesian midpoint, the PBC-consistent midpoint reduces mean movable-atom RMSD from $0.957/0.635/1.197$ Å to $0.245/0.125/0.237$ Å on SurfReact-TS, LiTraj, and MP20Bat, respectively (Fig. 1a). This shows that periodic endpoint representation provides a substantially better geometric prior.

Periodic neighborhoods. Atoms on opposite sides of a unit cell can be close physical neighbors through periodic images, even when their stored coordinates are far apart. Using a 5 Å cutoff, we find that a graph constructed without PBC misses on average 27.4%, 46.4%, and 62.7% of the local neighbor connections identified by the corresponding periodic graph in SurfReact-TS, LiTraj, and MP20Bat, respectively (Fig. 1b). This motivates explicit periodic interaction modeling.

Reaction-motion localization. Although periodic structures may contain many atoms, only a small subset often undergoes substantial endpoint rearrangement. The top 10% most-moving atoms account for 94.3%, 97.4%, and 88.0% of the total squared endpoint displacement in SurfReact-TS, LiTraj, and MP20Bat, respectively (Fig. 1c). Together with the endpoint and neighborhood diagnostics above, this motivates the PBC-consistent endpoint conditioning, periodic interaction modeling, and reaction-focused learning used in PeriTS.

4 Geometry-Aware Transition-State Prediction

The diagnostics in Sec. 3.3 motivate PBC-consistent endpoint conditioning, periodic interaction modeling, and reaction-focused learning. We incorporate these considerations into a common representation and an endpoint-conditioned scalar–vector backbone. On this foundation, we study a one-shot residual predictor and the iterative flow predictor PeriTS. Both start from the same periodic midpoint X_{mid} and retain endpoint conditioning throughout prediction. They share the architectural design, but have independently trained parameters and different training objectives. Fig. 2 summarizes their inference procedures.

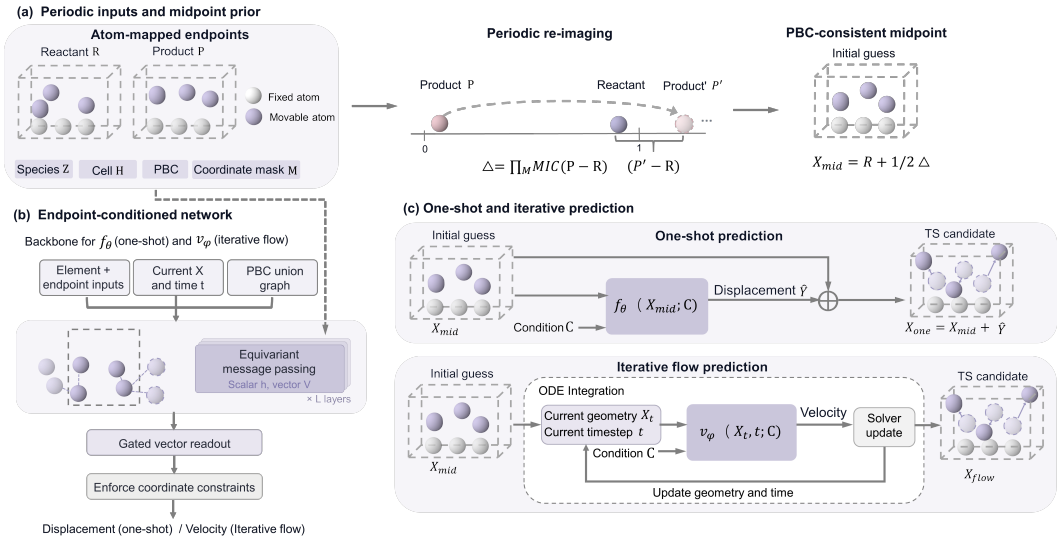

Figure 2: **Geometry-aware periodic transition-state prediction.** **(a)** Atom-mapped periodic endpoints are re-imaged using the minimum-image convention and projected onto the allowed coordinate directions to construct the PBC-consistent midpoint X_{mid}. **(b)** Endpoint features, current geometry, time conditioning, and a periodic union graph are processed by an equivariant scalar–vector backbone, with the output projected to satisfy coordinate constraints. **(c)** The one-shot model predicts a single residual from X_{mid}, whereas PeriTS iteratively refines the geometry by integrating an endpoint-conditioned velocity field.

4.1 Periodic Conditioning

As defined in Sec. 3.1, each reaction is represented by $\mathcal{C} = (R, P, Z, H, \mathbf{p}, M)$. As illustrated in Fig. 2(a), we first re-image the product relative to the reactant using the minimum-image convention,

$$P' = R + \mathrm{MIC}_{H,\mathbf{p}}(P - R), \qquad \Delta = \Pi_M(P' - R), \qquad X_{\mathrm{mid}} = R + \tfrac{1}{2}\Delta. \tag{2}$$

Here, P' is a periodically equivalent representation of the product, chosen so that its displacement from R is minimum-image consistent. The projection Π_M then removes displacement components forbidden by the mobility mask M. Its implementation is given in Appendix C.1. The resulting X_{mid} initializes both the one-shot and flow prediction routes.

To capture interactions that may be relevant at different stages of the reaction, we construct periodic neighbor graphs for R, P', and X_{mid}, and take their union. Two atoms are connected if their minimum-image distance is within the cutoff r_{c} in any of the three structures. This retains pairs that are close before or after the reaction, including those associated with bond breaking or formation even when they are not close at the midpoint. The resulting edge set is kept fixed during prediction, while its distances and directions are updated from the current geometry as prediction proceeds.

4.2 Endpoint-Conditioned Equivariant Backbone

Fig. 2(b) shows the backbone used to process the periodic reaction representation. We use a PaiNN-style (Schütt et al., 2021) equivariant message-passing network with L layers. In each layer, atom i aggregates information from its neighbors in the periodic union graph and updates a scalar state $h_i^\ell \in \mathbb{R}^d$ and a vector state $V_i^\ell \in \mathbb{R}^{d\times 3}$. The scalar state contains rotation-invariant information such as atomic species, time, and endpoint-derived descriptors, whereas the vector state contains d three-dimensional channels carrying directional geometric information from the current structure and reaction endpoints.

For a directed edge $j \to i$, the vector message is

$$m_{j\to i}^{V,\ell} = a_{ij}^\ell \odot W^\ell V_j^\ell + \sum_{Q\in\{X,R,P',X_{\mathrm{mid}}\}} b_{ij}^{\ell,Q} \otimes \widehat{r}_{ij}^{\,Q},$$

where $\widehat{r}_{ij}^{Q}$ is the minimum-image unit direction from atom i to atom j in geometry Q, and a_{ij}^{ℓ} and $b_{ij}^{\ell,Q}$ are learned scalar coefficients. The first term propagates directional information already carried by the neighboring atom, while the second introduces geometric directions from the current structure, reactant, product, and midpoint. Because these directions transform as vectors while their learned weights are scalars, the message update preserves rotational equivariance.

After L message-passing layers, the vector channels are combined into a three-dimensional output for each atom and projected with Π_M to satisfy the supplied coordinate constraints. As shown in Fig. 2(c), the same backbone design is used to predict a residual displacement in the one-shot route and a velocity in PeriTS. Fixed atoms remain in the graph and provide environmental context, while their constrained output components are removed.

Additional feature and architecture details are provided in Appendix C.

4.3 One-Shot and Iterative Prediction

Fig. 2(c) illustrates the two prediction routes built on the same PBC-consistent initialization X_{mid} and backbone design. In the one-shot route, the network predicts a single residual displacement $\widehat{y}$ from the midpoint,

$$\widehat{y} = f_\theta(X_{\mathrm{mid}}; \mathcal{C}), \qquad X_{\mathrm{one}} = X_{\mathrm{mid}} + \widehat{y}. \tag{3}$$

The predicted displacement is projected onto the allowed coordinate directions, yielding one TS candidate from a single backbone evaluation.

PeriTS instead refines the geometry iteratively through an endpoint-conditioned velocity field. Starting from $X_0 = X_{\mathrm{mid}}$, the geometry evolves according to

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = v_\phi(X_t, t; \mathcal{C}), \qquad X_0 = X_{\mathrm{mid}}, \qquad X_{\mathrm{flow}} = X_1. \tag{4}$$

At each integration step, the network takes the current geometry X_t, the integration variable t, and the reaction condition $\mathcal{C}$, and predicts a velocity used to update the structure. Here, $t \in [0, 1]$ is the integration variable of the learned flow. Coordinate constraints are enforced throughout the integration. For inference, we integrate Eq. 4 using the explicit midpoint method, a second-order Runge–Kutta scheme. Details are described in Appendix C.

The two routes share the same representation and backbone design but are trained independently with different objectives. Given a reference TS geometry $X^\star$, we define its residual from the periodic midpoint as

$$y^\star = \Pi_M[\mathrm{MIC}_{H,\mathbf{p}}(X^\star - X_{\mathrm{mid}})]\,.$$

The one-shot model is trained to predict $y^\star$ directly. For PeriTS, we additionally train on intermediate states y_t drawn from ideal, perturbed, or rollout-generated trajectories. At time t, the corresponding corrective target is

$$u_t = \frac{y^\star - y_t}{\max(1 - t, \tau_{\min})},$$

so that the velocity field learns to correct the current state toward the reference TS rather than only follow an ideal interpolation. Losses are normalized within each reaction and emphasize movable, reaction-relevant atoms, following the localized-motion observation in Sec. 3.3. More training details are provided in Appendix C.

5 Experiments

The geometric analysis in Fig. 1 suggests that periodicity affects endpoint representation, local interactions, and where reaction motion is concentrated. We now ask whether explicitly modeling these effects improves TS prediction, and whether iterative refinement provides benefits beyond a strong one-shot correction. We further complement the geometric evaluation with DFT-based energy diagnostics on randomly sampled SurfReact-TS test cases.

Table 1: Periodic all-atom test RMSD in Å (lower is better). Mean and median are computed over reaction-level RMSDs. Bold indicates the lowest value in each column.

	SurfReact-TS		LiTraj		MP20Bat	
Method	Mean	Median	Mean	Median	Mean	Median
PBC midpoint	0.2112	0.1659	0.1312	0.1026	0.2358	0.1871
React-OT	0.1744	0.1145	0.1220	0.0953	0.2066	0.1273
TS-DFM	0.1439	0.0797	0.1301	0.1037	0.3416	0.2836
One-shot	0.1174	0.0845	**0.1213**	**0.0897**	0.1327	0.0816
PeriTS	**0.1022**	**0.0724**	0.1236	0.0946	**0.1283**	**0.0786**

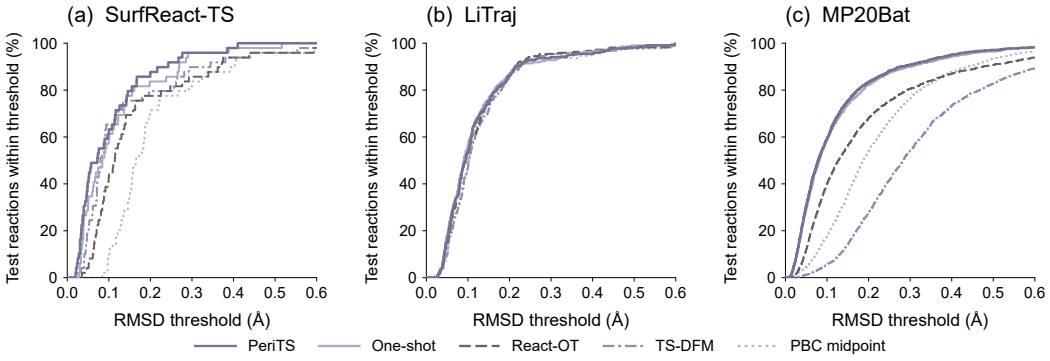

Figure 3: **Test-set distributions of periodic all-atom RMSD.** Curves show the fraction of reactions whose error is no greater than the horizontal-axis threshold, using the primary configurations in Tab. 1. Higher curves indicate better geometric agreement at a given threshold. Errors above the displayed interval remain in the denominator.

5.1 Experimental Setup

Datasets and evaluation. We use the partitions introduced in Sec. 3.2 and train a separate model for each benchmark. Our primary metric is periodic all-atom RMSD, computed using minimum-image displacements between predicted and reference TS coordinates. We compute RMSD for each reaction and report the mean and median across reactions. For SurfReact-TS, fixed atoms are included in the primary metric. In the periodic ablation study, Tab. 2 additionally reports RMSD over movable atoms only.

Compared methods. We compare the PBC midpoint with periodic adaptations of React-OT (Duan et al., 2025) and TS-DFM (Luo et al., 2026), together with our One-shot predictor and PeriTS. PeriTS uses nearest-image neighborhoods on SurfReact-TS and LiTraj and the multi-image configuration on MP20Bat described in Appendix E.3. Comparison with SaddleFlow and additional periodic adaptations are provided in Appendix D.

5.2 Transition-State Geometry Prediction

Tab. 1 summarizes the primary test results. PeriTS achieves the lowest mean and median RMSD on SurfReact-TS and MP20Bat, while One-shot performs best on LiTraj. The gain from iterative refinement is substantial for SurfReact-TS, more modest for MP20Bat, and absent on LiTraj, indicating that its benefit depends on the reaction regime.

Fig. 3 shows the corresponding error distributions. The curves further reveal that no method dominates at every geometric tolerance. For example, TS-DFM has more very-low-error predictions on SurfReact-TS, whereas PeriTS provides better overall error statistics.

Table 2: Periodic ablations on SurfReact-TS. RMSD in Å, lower is better.

	All atoms		Movable atoms	
Configuration	Mean	Median	Mean	Median
PeriTS	**0.1022**	**0.0724**	**0.1321**	**0.0928**
Cartesian midpoint	0.1277	0.0974	0.1652	0.1256
Non-periodic graph	0.1087	0.0818	0.1407	0.1047

Table 3: Mean periodic all-atom test RMSD (Å) versus inference depth. $*$ denotes the primary PeriTS setting.

Method	NFE	SurfReact-TS	LiTraj	MP20Bat
One-shot	1	0.1174	0.1213	0.1327
PeriTS	2	0.1037	0.1246	0.1387
	4	0.1032	0.1250	0.1298
	6^*	0.1022	0.1236	0.1283
	8	0.1032	0.1250	0.1283
	12	0.1023	0.1252	0.1275
	24	0.1017	0.1248	0.1276

5.3 Which Periodic Modeling Choices Matter?

We isolate two components of PeriTS on SurfReact-TS while keeping the remaining settings fixed. The *Cartesian-midpoint* variant replaces the PBC-consistent initialization with the midpoint of the stored endpoint coordinates. The *non-periodic-graph* variant retains the PBC-consistent initialization but removes periodic image connections from message passing. These interventions separately test periodic endpoint initialization and periodic neighborhood construction.

Both interventions degrade prediction, with a substantially larger effect from replacing the PBC-consistent midpoint. The same trend holds when evaluation is restricted to movable atoms, supporting both periodic initialization and periodic message passing in SurfReact-TS. Additional ablation details are provided in Appendix E.2.

5.4 Prediction Quality versus Inference Budget

Tab. 3 examines how PeriTS changes as the number of neural function evaluations (NFE) increases. The effect differs substantially across datasets. On SurfReact-TS, only small improvements are obtained beyond a few refinement steps, whereas MP20Bat benefits more clearly from deeper iterative refinement. On LiTraj, increasing NFE does not consistently reduce the error, and the One-shot predictor remains more accurate than the evaluated PeriTS settings. These results show that the value of iterative refinement depends on the reaction system.

Candidate breadth versus refinement depth. We further evaluate a six-candidate extension of the One-shot predictor on SurfReact-TS and MP20Bat. Candidate breadth provides only modest improvements over the corresponding One-shot anchors, whereas iterative refinement yields substantially larger gains in both settings. This suggests that, when refinement is beneficial, repeatedly updating a candidate can be more effective than generating additional alternatives. Full candidate-generation and selection details are provided in Appendix E.4.

5.5 Single-Point DFT Assessment

To complement the geometric evaluation, we perform single-point DFT calculations on 13 randomly sampled SurfReact-TS test reactions. All predicted geometries are evaluated at fixed atomic positions without structural relaxation. We measure the absolute energy deviation from the reference TS,

$$\varepsilon_E = \left| E_{\mathrm{DFT}}(\widehat{X}_{\mathrm{TS}}) - E_{\mathrm{DFT}}(X^\star) \right|,$$

Table 4: Candidate-breadth study. Mean periodic all-atom test RMSD (Å, lower is better).

Prediction	SurfReact-TS	MP20Bat
One-shot anchor	0.1174	0.1327
Six-candidate fusion	0.1163	0.1322
PeriTS	0.1022	0.1283

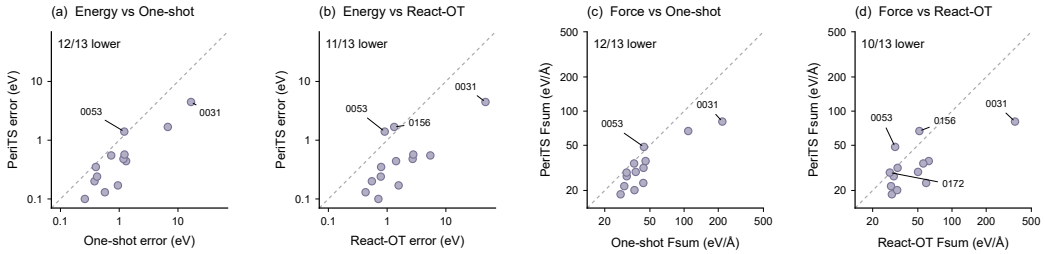

Figure 4: **Single-point DFT assessment on 13 randomly sampled SurfReact-TS test reactions. (a,b)** Absolute DFT energy errors for PeriTS versus One-shot and React-OT. **(c,d)** Corresponding aggregate force measures F_{sum}. Each point represents the same reaction under two methods, with points below the diagonal favoring PeriTS.

together with the aggregate force measure F_{sum}.

Fig. 4 compares PeriTS with One-shot and React-OT reaction by reaction. PeriTS yields lower energy error than One-shot in 12 of 13 cases and React-OT in 11 of 13 cases. For F_{sum}, PeriTS is lower in 12 of 13 and 10 of 13 cases, respectively. These paired results provide complementary evidence that improved geometric predictions are often accompanied by improved DFT energy and force agreement. They do not by themselves establish first-order saddle character or endpoint connectivity. Calculation details and the definition of F_{sum} are provided in Appendix E.5.

6 Discussion and Conclusion

Periodic geometry fundamentally reshapes the transition-state prediction problem, affecting endpoint representation, local interactions, and the distribution of reaction motion. We introduce **SurfReact-TS** and develop **PeriTS**, a geometry-aware framework that incorporates PBC-consistent conditioning, periodic interactions, coordinate constraints, and reaction-focused learning.

Our results further show that the value of iterative refinement depends strongly on the reaction regime, while increasing candidate breadth provides only modest gains in the settings studied. Together with the DFT diagnostics, these findings suggest that periodic TS learning should align representation, model design, and inference strategy with the geometry of the underlying reaction system.

AI use statement

In this work, we used generative AI tools to assist with parts of coding and manuscript writing, primarily to improve research and writing efficiency. All AI-assisted code, analyses, and text were manually reviewed and verified by the authors. Scientific conclusions were based on the implemented methods and experimental results. We take full responsibility for the final content of this work.

Ethics statement

This work does not involve human subjects, personal data, or human-participant experiments. Our primary ethical considerations concern the provenance, documentation, and release of scientific datasets and code. SurfReact-TS will be released only under appropriate permissions and licensing,

with its curation and data-processing procedures documented. We take responsibility for ensuring that all released artifacts respect the terms of their original data sources.

Reproducibility statement

To support reproducibility, we provide detailed descriptions of dataset preprocessing, periodic geometry construction, model architecture, training and inference procedures, baseline adaptations, evaluation metrics, and additional experimental protocols in the Appendix. The complete source code, configuration files, data-processing scripts, and instructions for obtaining the public benchmarks will be made publicly available upon acceptance. SurfReact-TS, together with its fixed data splits and preprocessing metadata, will also be released subject to the corresponding redistribution permissions and licensing requirements.

References

Ilgar Baghishov, Sung Hoon Jung, and Graeme Henkelman. MaterialsSaddles: 34 million transition states and a flow-matching saddle-point predictor for materials. `https://huggingface.co/datasets/SciLM/MaterialsSaddles`, 2026. Submitted to the Datasets & Benchmarks Track, NeurIPS 2026.

Artem D Dembitskiy, Innokentiy S Humonen, Roman A Eremin, Dmitry A Aksyonov, Stanislav S Fedotov, and Semen A Budennyy. Benchmarking machine learning models for predicting lithium ion migration. *npj Computational Materials*, 11(1):131, 2025.

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.

Henry Eyring. The activated complex in chemical reactions. *The Journal of chemical physics*, 3(2): 107–115, 1935.

Leonard Galustian, Konstantin Mark, Johannes Karwounopoulos, Maximilian P-P Kovar, and Esther Heid. Goflow: efficient transition state geometry prediction with flow matching and e (3)-equivariant neural networks. *Digital Discovery*, 4(12):3492–3501, 2025.

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, Blas P Uberuaga, and Hannes Jónsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of chemical physics*, 113(22):9901–9904, 2000.

Riley Jackson, Wenyuan Zhang, and Jason Pearson. Tsnet: predicting transition state structures with tensor field networks and transfer learning. *Chemical Science*, 12(29):10022–10040, 2021.

Hyunwook Jung, Emanuel Colombi Manzi, Tiago J Goncalves, Vanessa J Bukas, Sandip De, Johannes T Margraf, Karsten Reuter, and Hendrik H Heenen. From global optimization to transition state search: an automated workflow for surface reaction barriers. *npj Computational Materials*, 12(1):300, 2026.

Yufei Luo, Xiang Gu, and Jian Sun. Generative flow model on distance geometry for predicting transition states of chemical reactions. *Nature Communications*, 2026.

Lagnajit Pattanaik, John B Ingraham, Colin A Grambow, and William H Green. Generating transition states of isomerization reactions with deep learning. *Physical Chemistry Chemical Physics*, 22(41):23618–23626, 2020.

Mathias Schreiner, Arghya Bhowmik, Tejs Vegge, Jonas Busk, and Ole Winther. Transition1x-a dataset for building generalizable reactive machine learning potentials. *Scientific Data*, 9(1):779, 2022.

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

SciLM. SaddleFlow: Generative ai for transition-state structures in periodic materials. GitHub repository, 2026. https://github.com/SciLM-ai/SaddleFlow.

Brook Wander, Muhammed Shuaibi, John R Kitchin, Zachary W Ulissi, and C Lawrence Zitnick. Cattsunami: Accelerating transition state energy calculations with pretrained graph neural networks. *ACS Catalysis*, 15(7):5283–5294, 2025.

A Dataset Preprocessing

A.1 SurfReact-TS

SurfReact-TS contains 330 curated periodic Pt-surface reactant–TS–product triplets involving C/H adsorbates. The benchmark is curated from 338 available triplets, with eight records removed due to endpoint/reference duplication, duplicate reactions, or flagged reference inconsistencies. We preserve the supplied atom ordering, periodic cells, coordinate constraints, and reference TS geometries.

All experiments use a fixed grouped train/validation/test split of 232/49/49 reactions, with records linked during curation kept within the same partition. The supplied reference geometries are treated as geometry labels and do not constitute dataset-wide certification of first-order saddle character or endpoint connectivity.

A.2 Public Benchmarks

LiTraj. For the `nebDFT2k` subset of LiTraj (Dembitskiy et al., 2025), we preserve the official data splits. For each trajectory, the first and last images define the reactant and product, and the highest-energy image is used as the reference TS geometry. The supplied atom ordering and periodic cell are preserved, and all atoms are treated as movable.

MP20Bat. For MP20Bat (Baghishov et al., 2026), we preserve the official splits and use the supplied reactant–saddle–product triplets directly. The supplied atom ordering and periodic cells are retained, and all atoms are treated as movable.

B Periodic Geometry Analysis Details

The statistics in Fig. 1 are computed on the training and validation partitions and averaged over reactions.

Periodic midpoint error. For corresponding endpoint atoms, the minimum-image displacement is $d = \mathrm{MIC}_{H,\mathbf{p}}(P - R)$, where the minimum-image operation selects the shortest periodically equivalent displacement under lattice H and periodic directions $\mathbf{p}$. The PBC-consistent midpoint is

$$X_{\mathrm{mid}} = R + \tfrac{1}{2}\Pi_M(d),$$

where Π_M removes displacement components forbidden by the supplied mobility mask. The Cartesian baseline replaces d with the direct displacement $P - R$. Midpoint error is the RMSD between the resulting midpoint and the reference TS over movable atoms, using minimum-image distances.

Periodic neighborhood omission. At the PBC-consistent midpoint, we construct periodic and non-periodic neighbor graphs using the same 5 Å cutoff. Periodic edges are image-resolved. For each reaction, we report the fraction of periodic edges that are absent when only the stored, non-periodic coordinates are used, and then average this fraction across reactions.

Reaction-motion concentration. For each atom, we compute the squared magnitude of its constraint-projected minimum-image endpoint displacement, $q_i = \|\Delta_i\|_2^2$. After ranking atoms by q_i, we report

$$c_{10} = \frac{\sum_{i \in \text{Top } 10\%} q_i}{\sum_i q_i},$$

the fraction of total squared endpoint motion carried by the top 10% most-moving atoms. The values in Fig. 1 are averages of c_{10} across reactions.

C Implementation Details

C.1 Periodic Constraints

For a Cartesian displacement $D \in \mathbb{R}^{N\times 3}$, the mobility mask is applied in lattice coordinates as

$$\Pi_M(D) = \big[(DH^{-1}) \odot M\big]H,$$

where H is the lattice matrix, M specifies the allowed coordinate components, and $\odot$ denotes element-wise multiplication. Thus, DH^{-1} converts the displacement to lattice coordinates, the mask removes constrained components, and multiplication by H maps the result back to Cartesian coordinates. When no constraint mask is supplied, we use $M = \mathbf{1}$.

C.2 Backbone Details

The scalar–vector backbone follows the architecture described in Sec. 4.2. Scalar features combine atomic embeddings, mobility information, endpoint-derived descriptors, and time conditioning. Vector features encode the periodic endpoint displacement, the current residual from X_{mid}, and directions relative to the reaction endpoints. Pairwise distances are represented using Gaussian radial basis functions.

For the main models, we use six interaction layers with scalar/vector width 160, 24 radial basis functions, a 5 Å cutoff, neighbor cap 48, and dropout 0.05. The periodic graph retains the nearest periodic image for each selected atom pair. Fixed atoms remain in the graph and contribute to message passing, while the output projection prevents motion along constrained coordinates.

C.3 Training and Inference Settings

All main models use six scalar–vector message-passing layers with hidden width 160, 24 Gaussian radial basis functions, a 5 Å cutoff, neighbor cap 48, and dropout 0.05. Models are trained with AdamW using an initial learning rate of 10^{-4} and exponential-moving-average weights. The batch size is 8 for SurfReact-TS and LiTraj and 16 for MP20Bat.

For the SurfReact-TS flow configuration, we use $\sigma = 0.04$ Å, $p_{\text{r}} = 0.30$, $J = 2$ detached rollout steps, and $\tau_{\min} = 0.10$. The auxiliary-loss weights are $\lambda_{\text{g}} = 0.10$, $\lambda_{\text{pair}} = 0.10$, $\lambda_{\text{clash}} = 0.01$, and $\lambda_{\text{swap}} = 0.05$.

For inference, PeriTS integrates the learned velocity field with the explicit midpoint method. For a step of size h from geometry X_t,

$$k_1 = v_\phi(X_t, t; \mathcal{C}),$$

$$X_{t+\frac{h}{2}} = X_t + \frac{h}{2}k_1, \qquad k_2 = v_\phi\left(X_{t+\frac{h}{2}}, t + \frac{h}{2}; \mathcal{C}\right),$$

and

$$X_{t+h} = X_t + hk_2,$$

with coordinate constraints applied to each update. This is a second-order Runge–Kutta scheme requiring two velocity-field evaluations per integration step.

Table 5: MP20Bat graph-mode comparison at NFE $= 6$. Periodic all-atom test RMSD in Å (lower is better).

Graph mode	Mean	Median
Nearest-image	0.1327	0.0801
Multi-image	**0.1283**	**0.0786**

Table 6: MP20Bat RMSD under the SaddleFlow evaluation protocol (Å, lower is better).

Method	Mean	Median
SaddleFlow (reported)	0.2495	0.1125
PeriTS (rescored)	0.1283	0.0786

C.4 Periodic Graph Configurations

SurfReact-TS and LiTraj use nearest-image periodic neighborhoods with a neighbor cap of 48 for both One-shot and PeriTS. For MP20Bat, the primary One-shot model retains this nearest-image configuration, whereas the primary PeriTS model uses an uncapped multi-image graph.

We use the multi-image variant specifically for MP20Bat because image multiplicity is substantially more pronounced in this benchmark. At a 5 Å cutoff, retaining only one image per atom pair omits on average 0.00%, 1.75%, and 32.50% of image-resolved midpoint edges in SurfReact-TS, LiTraj, and MP20Bat, respectively. Thus, nearest-image neighborhoods already capture nearly all such connections in SurfReact-TS and LiTraj, whereas a substantial fraction is lost in MP20Bat.

The multi-image graph retains distinct lattice images of an atom pair when multiple images fall within the cutoff, with each image contributing its own distance and direction during message passing. To isolate this choice on MP20Bat, we compare nearest-image and multi-image models without a neighbor cap and with the remaining settings matched:

The observed improvement supports using the multi-image configuration for the primary MP20Bat PeriTS model.

D Baseline Comparisons

D.1 Primary Baselines

To evaluate molecular TS predictors on periodic systems, we adapt React-OT (Duan et al., 2025) and TS-DFM (Luo et al., 2026) to preserve the supplied atomic species, atom correspondence, periodic cell, periodic directions, coordinate constraints, and dataset partitions. Reference TS coordinates are used for training and evaluation only and are not provided as inference inputs.

React-OT. We retain the original React-OT model and ODE sampling procedure while adding periodic distance handling and coordinate constraints. The reported configuration uses NFE $= 10$.

TS-DFM. We retain the distance-based flow formulation of TS-DFM using periodic endpoint distances. Predicted distances are subsequently reconstructed into Cartesian coordinates under periodic boundary conditions. The reported configuration uses NFE $= 40$ for the flow component.

SaddleFlow. We additionally compare with the reported MP20Bat result of SaddleFlow (SciLM, 2026), a concurrent flow-based method developed directly for periodic materials. Because we do not independently reproduce its trained checkpoint, we use the repository-reported SaddleFlow result and rescore our saved PeriTS predictions with the official SaddleFlow evaluation protocol on the same MP20Bat test set.

Table 7: Additional adapted baselines on SurfReact-TS. Periodic all-atom test RMSD in Å (lower is better).

Adaptation	Mean	Median
TSNet (Jackson et al., 2021)	0.2436	0.1948
AEFM + React-OT prior	0.9674	0.7278

Table 8: Movable-atom test RMSD for the SurfReact-TS periodic interventions at NFE $= 6$. Each value is a reaction-averaged RMSD in Å. This supplements the all-atom main table.

Configuration	Mean	Median
Flow reference	0.132054	0.092802
Cartesian midpoint source	0.165161	0.125560
Non-periodic message graph	0.140700	0.104731

D.2 Additional Periodic Adaptations

We report two additional SurfReact-TS adaptations that are not included in the primary comparison because their inference procedures differ from the main baselines.

The AEFM adaptation uses React-OT predictions as initialization and is therefore a hybrid periodic adaptation rather than the native inference procedure.

E Additional Evaluation Details

E.1 Geometric Evaluation Protocol

For reaction j, periodic all-atom RMSD is

$$e_j = \left[\frac{1}{N_j} \sum_{i=1}^{N_j} \left\| \mathrm{MIC}_{H_j,\mathbf{p}_j}(\widehat{x}_{ji} - x^{\star}_{ji}) \right\|_2^2 \right]^{1/2}.$$

We use the supplied atom correspondence without additional alignment or remapping, and report the mean and median of the reaction-level RMSDs. For SurfReact-TS, movable-atom RMSD uses the same definition restricted to atoms allowed to move. LiTraj and MP20Bat contain only movable atoms in our conversion, so the two metrics coincide for those datasets.

E.2 Periodic Ablation Details

The SurfReact-TS periodic ablations use the same data split, backbone, training objective, coordinate constraints, and NFE $= 6$ setting as the reference PeriTS model. The Cartesian-midpoint variant replaces only the PBC-consistent initialization with the midpoint of the stored endpoint coordinates. The non-periodic-graph variant keeps the PBC-consistent initialization but removes periodic image connections from message passing. All checkpoints are selected using validation performance.

E.3 Nearest-Image versus Multi-Image Neighborhoods

Periodic neighborhoods and multiple-image neighborhoods are distinct choices. A minimum-image graph accounts for cross-boundary neighbors but need not retain multiple lattice images of one neighbor within the cutoff. For MP20Bat, the no-cap comparison changes graph mode while retaining the remaining recorded configuration. At NFE $= 6$, mean test RMSD is 0.132688 Å for the no-cap nearest-image model and 0.128305 Å for the no-cap multi-image model. The capped primary Flow configuration has mean RMSD 0.131083 Å. The no-cap pair is the relevant comparison for image multiplicity; comparing the capped primary model directly with the no-cap multi-image model changes both graph mode and the cap. These single-run results show a modest observed improvement, not evidence that multi-image neighborhoods are universally superior.

Table 9: Six-candidate One-shot results. Periodic all-atom test RMSD in Å (lower is better).

	SurfReact-TS		MP20Bat	
Prediction	Mean	Median	Mean	Median
One-shot anchor	0.1174	0.0845	0.1327	0.0816
Hard selection	0.1161	0.0813	0.1322	0.0813
Uniform fusion	0.1164	0.0844	0.1322	0.0813
Soft fusion	0.1163	0.0841	0.1322	0.0814

E.4 Six-Candidate One-Shot Prediction

We extend a frozen One-shot predictor with five additional equivariant prediction heads while retaining the original One-shot prediction as the sixth candidate. All candidates share the same backbone representation. A learned scorer uses endpoint and candidate information to support candidate selection without access to the reference TS at inference.

We evaluate three aggregation strategies: selecting the highest-scoring candidate, uniformly averaging the candidates, and a soft-fusion rule selected on the validation set. All selection and fusion rules are fixed before test evaluation.

The six-candidate variants provide only modest improvements over the corresponding One-shot anchors. Because all candidates share a single frozen backbone evaluation, $K = 6$ does not correspond to NFE $= 6$. We therefore interpret this experiment as a comparison of candidate breadth with sequential refinement rather than as a compute-matched comparison.

E.5 DFT Calculation Details

We perform single-point DFT calculations on the 13 randomly sampled SurfReact-TS test reactions used in Fig. 4. All predicted geometries are evaluated at fixed atomic positions without structural relaxation, using the same electronic-structure settings for One-shot, PeriTS, React-OT, and the corresponding reference geometry.

For a predicted TS geometry $\widehat{X}_{\mathrm{TS}}$, we report the absolute energy deviation from the reference TS,

$$\varepsilon_E = \left| E_{\mathrm{DFT}}(\widehat{X}_{\mathrm{TS}}) - E_{\mathrm{DFT}}(X^\star) \right|,$$

and the summed atomic force magnitude,

$$F_{\mathrm{sum}} = \sum_{i=1}^{N} \|\mathbf{F}_i\|_2.$$

The force statistic is used as a comparative diagnostic on the same reaction and is not interpreted as a saddle-point convergence criterion.

All calculations use [DFT code and version] with [exchange–correlation functional], [pseudopotential], an energy cutoff of [XX] eV, and [k-point setting]. [Add smearing, spin/dipole treatment, and electronic convergence criteria if applicable.]

Table 10: Single-point DFT diagnostics for the 13 SurfReact-TS test reactions used in Fig. 4.

	Energy deviation (eV)			F_{sum} (eV/Å)		
Case	One-shot	PeriTS	React-OT	One-shot	PeriTS	React-OT
0013	1.31	0.44	1.42	43.89	31.64	33.15
0053	1.22	1.40	0.91	44.27	48.39	31.53
0062	0.38	0.20	0.55	29.71	21.83	29.02
0336	1.18	0.48	2.72	45.78	36.32	62.41
0156	6.72	1.68	1.32	108.70	66.77	51.59
0043	0.42	0.24	0.78	31.22	26.68	30.68
0069	0.73	0.55	5.44	37.48	29.14	50.17
0106	1.23	0.57	2.79	36.27	34.58	55.87
0172	0.40	0.35	0.79	31.16	28.75	28.32
0245	0.95	0.17	1.57	43.71	23.28	59.18
0299	0.26	0.10	0.71	27.58	18.50	29.52
0321	0.57	0.13	0.43	36.55	20.14	32.88
0031	16.68	4.47	47.23	217.19	80.90	359.01