FUNCTIONAL TANGENT GEOMETRY SELECTS ADAPTER PLACEMENT IN EQUIVARIANT INTERATOMIC POTENTIALS

Anonymous authors
Paper under double-blind review

ABSTRACT

The best adapter is determined more by placement than parameter count because different locations expose different energy–force correction spaces. We formalize these spaces with functional tangent geometry, using a Jacobian to map bounded adapter coefficients to reachable output corrections. A held-out transfer score then tests whether coefficients fitted on one configuration split represent the same correction on another. The score is invariant to adapter reparameterization when the coefficient metric transforms with it. Controlled experiments recover all four generating placements, and local molecular experiments show that tangent rankings predict constrained nonlinear adaptation. Across six rMD17 molecules, native-coordinate geometry selects the readout adapter. It remains best on untouched 1,000-frame holdouts, reducing normalized residual by 20.71% relative to the first equivariant product block. A local Ti–Al task instead selects that product block, whereas column normalization reverses the molecular ranking. Functional tangent geometry chooses adapter locations from the corrections they can express, avoiding a nonlinear sweep over placements.

1 INTRODUCTION

An adapter is specified by its trainable dimension and insertion point. The first limits how many coefficients may move; the second determines what those coefficients can do to the function. In a structured model, a scalar interaction, an equivariant product, and a nonlinear readout expose different families of output corrections even when their trainable dimensions match. Placement is therefore a functional hypothesis. A useful selection rule should test that hypothesis from the task correction before adapting every candidate.

This distinction is especially sharp for interatomic potentials. Their energy is invariant and their forces are its coordinate derivatives, so the force change induced by an adapter is a mixed coordinate–parameter derivative. Parameter-space criteria—gradient norms, Fisher scores, sparsity masks, and low-rank gradient projections—measure sensitivity or optimization geometry (Sung et al., 2021; He et al., 2022; Zhao et al., 2024), while equally sized parameter steps can induce different conservative energy–force corrections. A placement rule should compare those corrections on configurations not used to fit the rule.

We cast fixed-basis placement as held-out functional reachability. For each candidate, the pretrained energy–force Jacobian maps a coefficient ball to a reachable set of corrections. We fit the required task correction on one configuration split and measure the transferred residual on another. The resulting incompatibility score is small when the same coefficient vector both represents the correction and transfers across configurations. This held-out construction separates functional coverage from interpolation of a finite fitting set.

The reachable set has two ingredients: a tangent image and a metric that prices directions inside it. We show that the score is invariant under an invertible change of adapter coordinates when the coefficient metric transforms with those coordinates. Rescaling Jacobian columns while retaining a Euclidean coefficient ball changes the functional ellipsoid being compared. Coordinate dependence is therefore part of the model-selection question and belongs alongside the candidate modules, output norm, and trust region.

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.

This view separates three empirical claims. *Reachability* asks whether a candidate space contains the required correction. *Rank transfer* asks whether the tangent ordering predicts bounded nonlinear adaptation on unseen configurations. *Local fidelity* asks whether the realized update remains close enough to the initial tangent for that ordering to explain the endpoint. Controlled teachers test reachability directly; natural-label experiments test rank transfer and fidelity separately.

The experiments follow this logic from controlled mechanisms to natural tasks. Synthetic teachers recover all four generating placements, and strict-local molecular studies align tangent rank, nonlinear adaptation, and untouched-test performance. Across six rMD17 molecules, the native-coordinate geometry reveals a shared readout preference that persists through adaptation and 1,000-frame tests. Changing the coefficient metric reverses that preference, while a local Ti–Al study favors an earlier product block. Together, these contrasts expose task and metric dependence directly.

This work contributes a held-out energy–force tangent criterion, a reparameterization principle for its coefficient geometry, and experiments that connect controlled reachability with cross-system transfer. The resulting criterion replaces "which layer is important?" with a sharper question: under a chosen functional metric and trust radius, which fixed adapter space can transfer the correction that the task requires?

2 Related Work

Equivariant parameter-efficient adaptation. Recent work has established that parameter-efficient adaptation of geometric models must respect their representation structure. ELoRA decomposes weight updates by equivariant paths and preserves $\mathrm{SO}(3)$ equivariance (Wang et al., 2025). Magnitude-modulated equivariant adapters instead modulate feature magnitudes by tensor order and multiplicity, further restricting how pretrained equivariant features can change (Jin et al., 2026). Sparsity-promoting fine-tuning selects a small subset of parameters in $E(3)$-equivariant materials foundation models (Batatia et al., 2025) and relates the selected pattern to orbital structure (Cho et al., 2026). Equitrain evaluates LoRA-based adaptation for phonon and thermal properties across a broad materials collection (Grandel et al., 2026), while GeoAda develops equivariant adapters for controlled geometric diffusion models (Zhao et al., 2025). Broader MACE fine-tuning comparisons now include LoRA on both scalar and equivariant linear layers across chemically diverse benchmarks (Tompa et al., 2026). These methods primarily design or constrain an adaptation parameterization. Our question is complementary: given several symmetry-preserving *fixed-basis* update spaces, can their ability to realize a task-specific energy–force correction predict which space will adapt best? We compare placements in function space rather than proposing another equivariant adapter architecture.

Placement, sparsity, and low-dimensional update spaces. General parameter-efficient fine-tuning research allocates rank or trainable parameters across layers using learned importance, gradient, or Fisher scores (Zhang et al., 2023; Sung et al., 2021; He et al., 2022). Recent placement rules use differentiable search, feature alignment, representation similarity, or projected adapter-gradient energy (Nowak et al., 2024; Hayou et al., 2025; Ogawa et al., 2026; Zhang et al., 2026). Intrinsic dimension studies show that successful fine-tuning can occur in small parameter subspaces (Aghajanyan et al., 2021), and GaLore projects gradients onto low-rank matrix subspaces during optimization (Zhao et al., 2024). These approaches summarize parameters, gradients, or internal representations. For interatomic potentials, the same parameter displacement can have strongly configuration-dependent effects on energy and forces, and force changes involve mixed coordinate–parameter derivatives. Our criterion instead fits the required correction inside each joint energy–force Jacobian image and tests whether the same coefficients transfer to held-out configurations. Linearized networks have likewise been used for rapid adaptation (Maddox et al., 2021); here the linearization includes mixed coordinate–parameter derivatives and selects among symmetry-structured update spaces.

LoRA initialization and nonlinear dynamics. LoRA represents a matrix update by two trainable factors (Hu et al., 2022). Its initialization and singular-space alignment affect optimization even in matrix-factorization models (Xu et al., 2025). The zero-initialized factor convention creates a sharper issue for an initial tangent analysis: only one factor is functionally active to first order, whereas nonlinear training can activate and rotate both. Our unrestricted-LoRA experiment reveals an ordering reversal that co-occurs with substantial factor motion and subspace rotation. The matched-local experiments instead hold the active factor, coordinates, and trust region fixed. Thus our selector

concerns fixed-basis local adaptation; predicting a freely rotating rank-constrained factorization requires a path-aware reachability criterion.

Functional roles of equivariant pathways. MACE and NequIP organize features into symmetry-constrained scalar and higher-order pathways (Batatia et al., 2022; Batzner et al., 2022; Geiger & Smidt, 2022). Recent controlled modifications of scalar channels show that scalar-pathway fidelity can materially affect energy, force, and stress accuracy (Bi et al., 2026). This reinforces the broader point that equivariant pathways are not interchangeable. We do not assume that a particular tensor order or architectural location is universally preferable. Instead, held-out functional tangent incompatibility ranks candidate pathways under specified output and adapter-coordinate geometries. This perspective motivates our separation of local tangent fidelity from empirical rank transfer.

3 Adapter Placement as Functional Geometry

3.1 Coupled energy–force tangent maps

Let $x = (Z, R, C)$ denote an atomic configuration with species Z, Cartesian coordinates R, cell C, and atom count N_x. For task m and split $\sigma \in \{\mathrm{fit}, \mathrm{val}, \mathrm{test}\}$, let $\mathcal{D}_m^\sigma$ be the corresponding configurations. A pretrained interatomic potential with parameters θ_0 predicts an invariant energy $E_{\theta_0}(x)$ and conservative forces $F_{\theta_0}(x) = -\nabla_R E_{\theta_0}(x)$. Let $e_\theta(x)$ denote the energy observable used in the comparison: $e_\theta(x) = E_\theta(x)/N_x$ in the natural-label studies and $e_\theta(x) = E_\theta(x)$ in the controlled reachability interventions. For any such set $\mathcal{D}$, write

$$y_\theta(\mathcal{D}) = \mathrm{vec}_{x\in\mathcal{D}}\left(e_\theta(x), F_\theta(x)\right), \qquad \Delta_{\mathcal{D}}^\star = y^\star(\mathcal{D}) - y_{\theta_0}(\mathcal{D}), \tag{1}$$

where $\Delta_{\mathcal{D}}^\star$ is the task correction. Prediction target $o \in \{\mathrm{EF}, E, F\}$ selects the joint, energy-only, or force-only rows. We write $\|v\|_{\sigma,o} = \|(W_{\sigma,o})^{1/2} v\|_2$ for the corresponding output norm, with uniform weights within each output channel. For the joint target, natural-label experiments use aggregate channel weights $w_E : w_F = 30 : 1$, whereas the controlled interventions assign equal aggregate weight to energy and force. All targets reuse the same model derivatives. In natural-label experiments, an energy-per-atom offset is estimated on the fit split and then held fixed on validation and test configurations.

Let $\mathcal{S}_m$ be the candidate placements for task m. For placement $s \in \mathcal{S}_m$, slice $k \in \{1, \dots, K\}$ fixes a d-dimensional basis and a parameterization $\theta_{s,k}(a)$, $a \in \mathbb{R}^d$, with $\theta_{s,k}(0) = \theta_0$. It is *fixed-basis* when these directions do not change during selection or adaptation. The split-specific functional tangent operator is

$$J_{m,s,k}^\sigma = \left.\frac{\partial y_{\theta_{s,k}(a)}(\mathcal{D}_m^\sigma)}{\partial a}\right|_{a=0}. \tag{2}$$

Because $F = -\nabla_R E$, every force row has the form $\partial_a F = -\nabla_R(\partial_a E)$ and therefore contains a mixed coordinate–parameter derivative. Thus two placements with identical coordinate dimension can span very different conservative force corrections. We derive the score below for one task, slice, and target, suppressing m, k, o until the aggregation step; in particular, write $J_s^\sigma = J_{m,s,k}^\sigma$, $\mathcal{S} = \mathcal{S}_m$, and $\Delta_\sigma^\star = \Delta_{\mathcal{D}_m^\sigma}^\star$. Let $G_s \succ 0$ define the coefficient norm $\|a\|_{G_s}^2 = a^\top G_s a$. A placement induces the coupled fit-to-validation reachable set

$$\mathcal{R}_s(\gamma; G_s) = \left\{ \left(J_s^{\mathrm{fit}} a, J_s^{\mathrm{val}} a\right) : \|a\|_{G_s} \le \gamma \right\}. \tag{3}$$

This set, rather than the parameter count alone, is the functional object being compared across placements.

3.2 Held-out functional reachability

We divide labeled configurations into tangent-fit and tangent-validation sets. For each placement, the fit coordinates are

$$\widehat{a}_s = \operatorname*{arg\,min}_{\|a\|_{G_s}\le\gamma} \left\| J_s^{\mathrm{fit}} a - \Delta_{\mathrm{fit}}^\star \right\|_{\mathrm{fit},o}^2 + \lambda \|a\|_{G_s}^2, \tag{4}$$

where γ fixes the coefficient-space trust radius and $\lambda > 0$ makes the fit unique. The placement score is the transferred residual

$$\tau_s \equiv \tau_s^{\mathrm{val}} = \frac{\left\| J_s^{\mathrm{val}} \widehat{a}_s - \Delta_{\mathrm{val}}^{\star} \right\|_{\mathrm{val},o}}{\left\| \Delta_{\mathrm{val}}^{\star} \right\|_{\mathrm{val},o}}. \tag{5}$$

Geometrically, τ_s measures how well a point selected through the first component of $\mathcal{R}_s$ transfers through its second component. Zero denotes exact transfer and one is the residual of making no correction. Unlike a gradient norm or a fit-only residual, the score tests both functional coverage and cross-configuration transfer. The actual selector averages paired slices before choosing a placement, as defined in Equation (8). All evaluated held-out correction norms in Equation (5) are nonzero.

Proposition 1 (Gauge-consistent reparameterization)**.** *Let T_s be invertible and reparameterize $a = T_s b$. If $J_s'^{\sigma} = J_s^{\sigma} T_s$ and $G_s' = T_s^{\top} G_s T_s$, then $\mathcal{R}_s(\gamma; G_s)$, the fitted functional correction, and τ_s are unchanged. If the Jacobian is transformed but the coefficient metric is held fixed, the reachable set and ranking may change.*

The proof is a change of variables and appears in Section A. The proposition distinguishes a harmless coordinate relabeling from a change in which functional directions are cheap.

The native comparisons set $G_s = I$ and use $d = 32$ slices orthonormal in each Euclidean adapter space, equalizing optimized dimension but not native-space coverage. The score still depends on the output metric $W_{\sigma,o}$, the chosen slice, and the coefficient metric. To isolate the last dependence, the column-normalized counterfactual uses only the tangent-fit split to define

$$D_{msko,jj} = \left[\max\left\{ \left\| \left(J_{m,s,k}^{\mathrm{fit}} \right)_{:,j} \right\|_{\mathrm{fit},o}^2, \varepsilon_{\mathrm{mach}} \right\} \right]^{1/2}, \qquad \widetilde{J}_{msko}^{\sigma} = J_{m,s,k}^{\sigma} D_{msko}^{-1}. \tag{6}$$

where the floor guards zero or numerically vanishing columns. Replacing J by $\widetilde{J}$ in Equations (4) and (5) while retaining $G_s = I$, γ, λ, the output metric, and the validation split preserves the tangent image but changes the ellipsoid induced by the coefficient norm. It therefore alters functional cost geometry without using validation outcomes.

3.3 Separating nonlinear ranking from local explanation

For a matched nonlinear test, projected Adam optimizes the normalized fit residual over the same coordinates and ball while keeping the basis fixed. The nonlinear loss has no ridge term: λ belongs only to the tangent solve in Equation (4). Among the evaluated optimizer iterates, let a_s^{nl} be the one with lowest fit loss, and define $\Delta_s^{\sigma}(a) = y_{\theta_{s,k}(a)}(\mathcal{D}_m^{\sigma}) - y_{\theta_0}(\mathcal{D}_m^{\sigma})$. We measure both the nonlinear task residual and departure from the initial tangent,

$$\tau_{s,\mathrm{nl}}^{\mathrm{val}} = \frac{\| \Delta_s^{\mathrm{val}}(a_s^{\mathrm{nl}}) - \Delta_{\mathrm{val}}^{\star} \|_{\mathrm{val},o}}{\| \Delta_{\mathrm{val}}^{\star} \|_{\mathrm{val},o}}, \qquad \epsilon_s^{\mathrm{val}} = \frac{\| \Delta_s^{\mathrm{val}}(a_s^{\mathrm{nl}}) - J_s^{\mathrm{val}} a_s^{\mathrm{nl}} \|_{\mathrm{val},o}}{\| J_s^{\mathrm{val}} a_s^{\mathrm{nl}} \|_{\mathrm{val},o}}. \tag{7}$$

These quantities answer different questions. Agreement between the rankings of τ_s and $\tau_{s,\mathrm{nl}}^{\mathrm{val}}$ supports transfer of the ordering; a small $\epsilon_s^{\mathrm{val}}$ shows that the realized update remains in the local tangent regime. We call a molecule–target decision locally faithful when $\max_{s,k} \epsilon_{msko}^{\mathrm{val}} \leq 0.15$. We report ranking and local fidelity separately; all evaluated tangent corrections in the denominator of $\epsilon_s^{\mathrm{val}}$ are nonzero. Taylor's theorem and the optimizer and locality details are given in Section A.

Ordinary two-factor LoRA falls outside this fixed-basis construction when both factors train. At the standard zero-B initialization, only the fixed A_0 factor contributes to first order; subsequent motion in A can rotate the reachable subspace. We therefore study unrestricted LoRA as a boundary case, not as a matched test of Equation (5).

3.4 Experimental design

The molecular experiments use MACE-OFF23 medium. Controlled corrections and the initial natural-label studies compare four symmetry-preserving placements: three scalar layers and the first equivariant-product map. The six-molecule cohort uses six placements by adding the second product map and nonlinear readout. It contains benzene, malonaldehyde, naphthalene, salicylic acid, toluene,

and uracil, each with an 80/40 fit–validation split, three paired coordinate slices, and an untouched 1,000-frame test.

For stage $u \in \{(\tan, \mathrm{val}), (\mathrm{nl}, \mathrm{val}), (\mathrm{nl}, \mathrm{test})\}$, we aggregate the slice-level incompatibilities into a placement score and make one selection per molecule and prediction target:

$$\overline{\tau}^{u}_{mso} = \frac{1}{K}\sum_{k=1}^{K} \tau^{u}_{msko}, \qquad \widehat{s}_{mo} = \arg\min_{s \in \mathcal{S}_m} \overline{\tau}^{\tan,\mathrm{val}}_{mso}, \quad K = 3. \tag{8}$$

Test regret is $\overline{\tau}^{\mathrm{nl,test}}_{m\widehat{s}_{mo}o} - \min_s \overline{\tau}^{\mathrm{nl,test}}_{mso}$. Writing s_{p0} for product 0, the positive test gain is $\Delta\tau^{\mathrm{nl,test}}_{mo} = \overline{\tau}^{\mathrm{nl,test}}_{ms_{\mathrm{p0}}o} - \overline{\tau}^{\mathrm{nl,test}}_{m\widehat{s}_{mo}o}$, and its relative reduction divides this gain by $\overline{\tau}^{\mathrm{nl,test}}_{ms_{\mathrm{p0}}o}$. The remainder retains its run-level index $\epsilon^{\mathrm{val}}_{msko}$, with $\bar{\epsilon}^{\mathrm{val}}_{mso}$ as a descriptive slice average. Local fidelity is determined by the maximum over placements and slices; a slice mean above 0.15 necessarily places the decision outside this regime.

Each molecule contributes three correlated prediction targets evaluated over six placements and three paired slices. Molecules, rather than target-level evaluations, are the units of generalization, so pooled 18-target rank correlations are descriptive.

Additional experiments probe the interpretation of this score. A post-hoc candidate-set expansion applies the six placements to three 3BPA temperatures and the three earlier molecular tasks. A separately generated, non-nested $d = 8$ sensitivity tests whether the native winner recurs in smaller random slices; it is not an embedded truncation of the $d = 32$ bases and has no matched nonlinear adaptations. A MACE-MP-0/Ti–Al study provides a cross-domain contrast: its original four-candidate study at $\gamma = 2.5$ is prospective, whereas the $\gamma = 0.5$ and six-candidate extensions are post-hoc. We also compare fit-only residuals, gradient norms, Gauss–Newton traces, column-normalized scores, fixed-placement rules, and an exhaustive oracle. Architectures, optimizer settings, data roles, bootstrap construction, and all per-placement results appear in Section B.

4 Results

4.1 Controlled interventions recover the generating placement

A useful placement score should first recover a known source of functional change. We insert small nonlinear teachers at each of four MACE placements and ask every candidate student space to reproduce the induced correction on held-out configurations. Across five paired replicates, same-placement incompatibility ranges from 0.000 to 0.051, whereas selected off-diagonal pairs give 0.889–1.023. For the main confirmatory pair, the adverse-minus-matched gap is positive in every replicate (mean 0.934, range $[0.873, 0.985]$). Adding the teacher space to the student restores reachability in every replicate, identifying candidate geometry as the source of the separation.

Bounded nonlinear adaptation preserves this separation (Figure 1b). The matched placement reduces validation incompatibility from the tangent value 0.0510 to 0.0005; the adverse space remains at 0.9904; and the union reaches 0.0615. Both adverse-minus-matched and adverse-minus-union gaps remain positive in all five replicates. Thus the held-out tangent geometry predicts the nonlinear outcome when basis and trust region are matched.

Finite-difference, adjoint, symmetry, and Jacobian-drift checks validate the tangent construction (maximum relative discrepancy 1.15×10^{-5}; Table 4).

4.2 Strict-local tangent rankings predict natural-label adaptation

We next ask when the ranking remains faithful to its local tangent explanation. Under the original four-placement candidate set, prospective paracetamol and the label-blind azobenzene and ethanol replications achieve placement-rank correlation $\rho_{\mathrm{place}} = 1.0$, top-1 agreement, and maximum slice remainder at most 0.0342. The selected equivariant-product placement remains optimal on three untouched 1,000-frame tests (top-1 3/3, mean $\rho_{\mathrm{place}} = 1.000$, mean regret 0.0000) and reduces force MAE by 3.31%, 2.87%, and 2.06%, respectively. These studies occupy the local regime: tangent ranking, nonlinear ranking, and first-order fidelity agree.

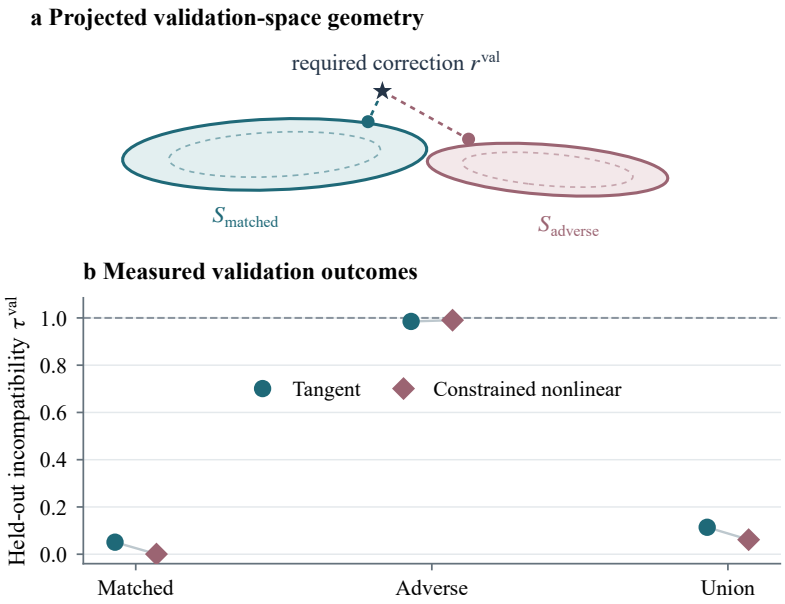


Figure 1: **Functional reachability predicts controlled adaptation. a**, Schematic projected validation images (not data). The star is the required correction, dashed segments indicate residual distances, and inner contours indicate smaller coefficient radii. **b**, Held-out tangent and bounded nonlinear incompatibility for the matched, adverse, and union controls. Points are means over five paired replicates; the dashed line marks no correction.

At larger radii, rank agreement can persist after the tangent approximation has degraded. The three-temperature 3BPA study at $\gamma = 5$ has $\rho_{\text{place}} = 1.0$ but a maximum remainder of 0.2006; reducing the radius to $\gamma = 2.5$ lowers this value to 0.1062. The four-placement Ti–Al study likewise preserves its ordering at $\gamma = 2.5$ while reaching a remainder of 0.3867; at $\gamma = 0.5$, the maximum slice remainder falls to 0.0458. Ranking can therefore remain useful beyond the radius at which it offers an accurate first-order explanation.

4.3 A NATIVE-METRIC READOUT POLICY TRANSFERS BEYOND LINEARIZATION

We next evaluate all six placements on a cohort of six rMD17 molecules. Under the native Euclidean adapter-coordinate metric, the readout block has the lowest $\overline{\tau}^{\tan,\text{val}}$ for all six molecules and, descriptively, all 18 molecule–target selections. Relative to the first equivariant-product placement (product 0), its tangent incompatibility is lower by 0.233 on average, with target-level gains from 0.167 to 0.317.

The selected placement remains best at both nonlinear outcome stages. After constrained adaptation is averaged over the paired slices, the readout has zero regret and improves over product 0 by 0.201 on average. Untouched 1,000-frame tests retain all 18 choices, with mean incompatibility gain 0.200 and mean relative residual reduction 20.71% (11.60%–29.57%). The reduction is 21.72% for joint energy–force prediction, 16.56% for energy, and 23.84% for forces. Table 1 and Figure 2 show the per-molecule results and outcome stages.

These normalized gains correspond to physical-error reductions. For joint adaptation, the readout lowers mean test energy MAE from 5.93 to 4.87 meV/atom and force MAE from 321.22 to 232.95 meV Å^{-1}; product 0 reaches 6.17 meV/atom and 311.39 meV Å^{-1} under the same training and evaluation procedure (Table 13).

The ordering remains predictive even though local linearization has degraded. All 108 placement–target means reach the $\gamma = 2.5$ trust boundary; 23/108 slice-averaged remainders exceed 0.15,

including 17/18 selected means, and the maximum is 1.284. The six-molecule result therefore shows that endpoint ranking can transfer outside the defined local regime, where the tangent no longer accurately reconstructs the endpoint.

Table 1: **A native-metric tangent ranking survives constrained adaptation and untouched tests.** Values average three paired slices over three related targets per molecule. Improvements are relative to fixed product 0; at $\gamma = 2.5$, 17/18 selected adaptations are nonlocal.

A. Six-molecule rMD17 cohort						
Molecule	$\Delta\tau^{\mathrm{tan,val}}$	$\Delta\tau^{\mathrm{nl,val}}$	$\Delta\tau^{\mathrm{nl,test}}$	Test reduction	NL top-1	Test top-1
Benzene	0.205	0.193	0.188	19.55%	3/3	3/3
Malonaldehyde	0.245	0.222	0.199	20.69%	3/3	3/3
Naphthalene	0.253	0.227	0.240	24.80%	3/3	3/3
Salicylic acid	0.280	0.237	0.241	24.78%	3/3	3/3
Toluene	0.204	0.160	0.157	16.23%	3/3	3/3
Uracil	0.211	0.166	0.176	18.20%	3/3	3/3
Mean / total	**0.233**	**0.201**	**0.200**	**20.71%**	**18/18**	**18/18**
B. Effects by prediction target						
Target	$\Delta\tau^{\mathrm{tan,val}}$	$\Delta\tau^{\mathrm{nl,val}}$	$\Delta\tau^{\mathrm{nl,test}}$	Test reduction	NL top-1	Test top-1
Joint	0.219	0.213	0.211	21.72%	6/6	6/6
Energy	0.249	0.155	0.159	16.56%	6/6	6/6
Force	0.230	0.234	0.231	23.84%	6/6	6/6
C. Cross-system contrast under the same six placements						
Task family	Candidate set		Tangent winner	NL winner	Targets	Top-1
Six rMD17 groups	same 6 placements		Readout	Readout	18 related	18/18
Ti–Al, post-hoc $\gamma = 0.5$	same 6 architectural roles		Product 0	Product 0	1	1/1

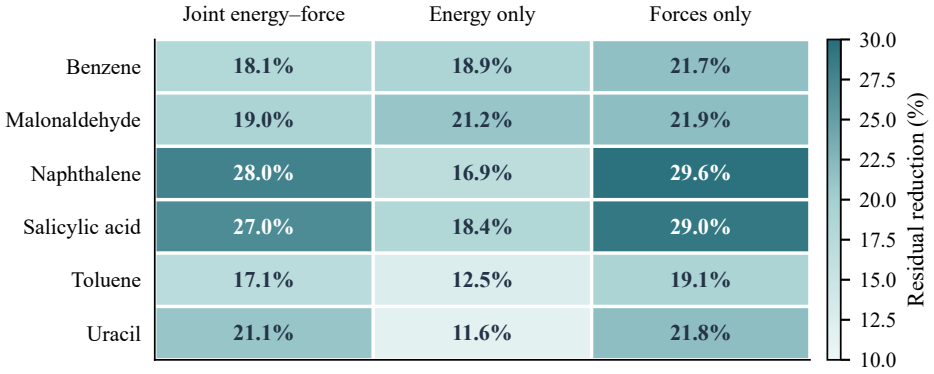


Figure 2: **Native-metric readout selection across molecules.** Cell values are the untouched-test relative reduction in normalized residual from the selected readout adapter versus fixed product 0. Each cell averages three paired slices; the three target views within a molecule share configurations.

4.4 What the cohort teaches about placement

The molecular cohort reveals a shared readout preference. The unanimous choice reveals a stable cohort-level architectural preference, and an always-readout rule reproduces every molecular selection. A six-placement expansion on paracetamol, azobenzene, ethanol, and the three 3BPA temperatures reinforces this pattern: the tangent and constrained rankings have $\rho_{\mathrm{place}} = 1.0$ and select the readout on all six tasks. The mean nonlinear gain over product 0 is 0.165, but none of these

tasks is local under the full candidate family (maximum remainders 0.871–0.992). For this cohort, selection discovers a shared policy that can subsequently be reused across molecules.

The coefficient metric changes the reachable set. Separately generated, non-nested $d = 8$ random slices reproduce every $d = 32$ native-coordinate winner. Only $d = 32$ has nonlinear outcomes. In contrast, column normalization changes every winner, selecting the early scalar placement in 18/18 cases, consistent with Proposition 1 when the coefficient metric is not co-transformed. Fit-only, gradient-norm, and Gauss–Newton criteria all recover the readout in 18/18. This cohort therefore establishes metric sensitivity but does not separate held-out transfer from these simpler selectors. Equal 32-coordinate slices also cover 0.625%–50% of the native spaces, so unequal fractional coverage may contribute to the preference. Component costs and hardware-separated timings are reported in Table 15.

A local Ti–Al task favors an earlier product block. After the prospective four-placement experiment at $\gamma = 2.5$, we extended the Ti–Al study post hoc at $\gamma = 0.5$ to the same six architectural placements. Its tangent analysis covers three prediction targets, while constrained nonlinear adaptation covers the joint target only. Product 0 is the joint tangent and nonlinear winner, respectively 0.9375 and 0.9430, compared with 0.9725 and 0.9754 for product 1 and 0.9726 and 0.9759 for the readout. The maximum across the six placement-mean remainders is 0.0307 (below 0.031). This post-hoc reversal provides a cross-system contrast to the molecular readout preference and shows that the task can change the best functional geometry.

Unions and rotating factors require richer objects. A separate rank-4 four-placement union uses 10,016 parameters (0.70% of full fine-tuning) and obtains validation force MAE $8.21/13.49/25.39$ versus $7.23/12.24/23.16$ meV Å^{-1} for full tuning at 300/600/1,200 K. This shows that composing update spaces can close much of the performance gap, but does not identify the composition from single-placement scores. For ordinary zero-B rank-4 LoRA with both factors free, initial fixed-A incompatibility is instead anticorrelated with endpoint regret ($\rho_{\mathrm{LoRA}} = -0.6268$). The reversal co-occurs with initially silent factor growth and subspace rotation, motivating a path-aware functional object for freely rotating LoRA (Equation (13), Table 18, and Figure 6).

5 Discussion

Adapter placement chooses a family of function changes accessible within a bounded coefficient norm. This explains why parameter count alone is a weak guide: equal-dimensional updates at different modules induce different conservative energy–force corrections. In controlled experiments, these functional sets recover the generating placement. In strict-local molecular tasks, they also predict which bounded adaptation transfers to unseen configurations.

The coefficient norm and radius are part of the placement hypothesis. By Proposition 1, transforming the coefficient metric with the coordinates preserves the score. Column normalization keeps a Euclidean ball and therefore changes the reachable ellipsoid, reversing every molecular winner. The native metric is meaningful when it matches the deployed parameterization and trust radius. Other scientific settings may instead call for regularized Fisher or Gauss–Newton metrics, or for output-space costs.

The molecular results separate a local explanation from a transferable policy. The readout choice survives constrained adaptation and untouched tests even after most endpoints leave the accurately linearized regime. Because the same choice wins for every molecule, the cohort supports a shared readout policy; instance-wise selection adds no benefit within this cohort. Ti–Al provides the contrasting case: inside a local radius, the first product block wins. A broader collection of heterogeneous tasks is needed to test when tangent geometry predicts this switch before adaptation.

The output norm is equally scientific. Energy and force arise from one potential but differ in units, noise, and downstream value. Agreement across joint, energy-only, and force-only objectives shows that the cohort preference survives these three choices. Applications with competing objectives could instead compare a family of metrics or retain a Pareto set of placements.

Fixed bases define the second limit. When both LoRA factors train, an initially silent factor becomes active and the reachable subspace can rotate. The initial tangent then describes the beginning of a

path, not the complete update. Predicting such adaptation requires transported or integrated Jacobian images. Likewise, a union of placements is a compositional space that must be evaluated as a new functional object.

The same comparison can be formulated for differentiable multi-output models with fixed candidate bases and a held-out correction. Interatomic potentials make the geometry especially visible because energy and force changes must remain conservative. Across other structured models, the central question remains the same: which affordable update space can express a task correction and carry it to unseen inputs?

The evidence supports a separation between screening and confirmation. Held-out tangents compare candidate update spaces before adaptation and can eliminate a placement whose reachable corrections are poorly aligned with the task. They do not replace the nonlinear test: the molecular experiments confirm transfer after constrained training, whereas Ti–Al shows that the preferred placement can change with the task. A useful selection procedure should therefore state the coefficient metric and radius, rank placements on held-out functional changes, and then verify the chosen update on untouched configurations. This sequence preserves the geometric interpretation without treating a local score as a guarantee about an unrestricted optimization path.

6 Conclusion

An adapter location defines a geometry of reachable function changes. Held-out functional tangents compare those geometries directly in the model outputs, turning placement into a testable scientific hypothesis. Controlled and strict-local experiments connect the score to nonlinear adaptation, while the molecular cohort reveals a transferable readout policy and the Ti–Al contrast shows that the winner can change with the task. Adapter placement should therefore be chosen through the correction a module can express, the metric that prices that correction, and the way the update space evolves during training.

The experiments also identify the evidence needed for such a choice. A training split determines whether a candidate tangent can fit the desired correction, a held-out split tests whether the same coefficients transfer, and bounded nonlinear adaptation tests whether the local ordering survives actual training. Controlled interventions establish the geometric mechanism; the molecular cohort and Ti–Al task then show when the resulting preference is shared and when it is task-specific. Taken together, these stages turn adapter placement from a convention into an empirical question about which functional changes remain stable across configurations.

The resulting comparison is deliberately made in output space. It can therefore place equally sized adapters on the same scientific scale even when their parameters live in different representations. When the ranking changes under a new coefficient metric, that change identifies a different family of admissible corrections. Stating the metric, radius, and held-out configurations is thus part of the placement claim itself. This perspective also separates adapter placement from adapter compression. Rank and parameter count describe the size of an update, but not the direction of the energy–force changes it can produce. Equal-sized adapters can therefore have different held-out reachability, while a larger space may devote capacity to corrections irrelevant to the task. Functional comparison makes the order of decisions explicit: first specify the physical output and coefficient metric, then identify a location whose correction transfers, and only then choose the rank needed within that location. This explicit specification also makes placement results comparable across tasks: a change in the selected module can then be traced to the desired functional correction rather than to an unreported change of coordinates. It makes placement reproducible as a comparison of physical corrections rather than parameter layouts.

AI use statement

OpenAI Codex was used for coding assistance and language editing. The authors reviewed all AI-assisted content and take full responsibility for the manuscript, implementation, and reported results.

Reproducibility statement

The Supplement specifies the tangent construction, data partitions, statistical analysis, and nonlinear evaluation. The anonymous release contains the method implementation, evaluation settings, aggregate tables, and figure data. Licensed datasets and pretrained model weights are referenced to their official sources rather than redistributed.

References

Armen Aghajanyan, Sonal Gupta, and Luke Zettlemoyer. Intrinsic dimensionality explains the effectiveness of language model fine-tuning. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 7319–7328. Association for Computational Linguistics, 2021. doi: 10.18653/v1/2021.acl-long.568. URL https://aclanthology.org/2021.acl-long.568/.

Ilyes Batatia, Dávid Péter Kovács, Gregor N. C. Simm, Christoph Ortner, and Gábor Csányi. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. In *Advances in Neural Information Processing Systems*, volume 35, 2022. doi: 10.52202/068431-0830. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/4a36c3c51af11ed9f34615b81edb5bbc-Abstract-Conference.html.

Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin Marin Elena, Dávid Péter Kovács, Janosh Riebesell, Xavier R. Advincula, Mark Asta, Matthew Avaylon, William J. Baldwin, et al. A foundation model for atomistic materials chemistry. *The Journal of Chemical Physics*, 163(18):184110, 2025. doi: 10.1063/5.0297006. URL https://doi.org/10.1063/5.0297006.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. $E(3)$-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13 (1):2453, 2022. doi: 10.1038/s41467-022-29939-5. URL https://www.nature.com/articles/s41467-022-29939-5.

Jia Bi, Alin Marin Elena, and Samuel Pinilla. Scalar-pathway fidelity improves physical accuracy in short-range equivariant interatomic potentials. *arXiv preprint arXiv:2606.15892*, 2026. doi: 10.48550/arXiv.2606.15892. URL https://arxiv.org/abs/2606.15892.

Youngwoo Cho, Seunghoon Yi, Wooil Yang, Sungmo Kang, Young-Woo Son, Jaegul Choo, Joonseok Lee, Soo Kyung Kim, and Hongkee Yoon. Robust and interpretable adaptation of equivariant materials foundation models via sparsity-promoting fine-tuning. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/60dc26558762425a465cb0409fc3dc52-Abstract-Conference.html.

Mario Geiger and Tess E. Smidt. e3nn: Euclidean neural networks. *arXiv preprint arXiv:2207.09453*, 2022. doi: 10.48550/arXiv.2207.09453. URL https://arxiv.org/abs/2207.09453.

Jonas Grandel, Philipp Benner, and Janine George. Parameter-efficient fine-tuning of machine-learning interatomic potentials for phonon and thermal properties. *arXiv preprint arXiv:2604.01017*, 2026. doi: 10.48550/arXiv.2604.01017. URL https://arxiv.org/abs/2604.01017.

Soufiane Hayou, Nikhil Ghosh, and Bin Yu. PLoP: Precise LoRA placement for efficient finetuning of large models. *arXiv preprint arXiv:2506.20629*, 2025. doi: 10.48550/arXiv.2506.20629. URL https://arxiv.org/abs/2506.20629.

Shwai He, Liang Ding, Daize Dong, Jeremy Zhang, and Dacheng Tao. SparseAdapter: An easy approach for improving the parameter-efficiency of adapters. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pp. 2184–2190. Association for Computational Linguistics, 2022. doi: 10.18653/v1/2022.findings-emnlp.160. URL https://aclanthology.org/2022.findings-emnlp.160/.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

Dian Jin, Yancheng Yuan, and Xiaoming Tao. Magnitude-modulated equivariant adapter for parameter-efficient fine-tuning of equivariant graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 498–506, 2026. doi: 10.1609/aaai.v40i1.37013. URL https://ojs.aaai.org/index.php/AAAI/article/view/37013.

Wesley Maddox, Shuai Tang, Pablo Moreno, Andrew Gordon Wilson, and Andreas Damianou. Fast adaptation with linearized neural networks. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 2737–2745. PMLR, 2021. URL https://proceedings.mlr.press/v130/maddox21a.html.

Aleksandra I. Nowak, Otniel-Bogdan Mercea, Anurag Arnab, Jonas Pfeiffer, Yann Dauphin, and Utku Evci. Towards optimal adapter placement for efficient transfer learning. *arXiv preprint arXiv:2410.15858*, 2024. doi: 10.48550/arXiv.2410.15858. URL https://arxiv.org/abs/2410.15858.

Keith Ando Ogawa, Bruno Lopes Yamamoto, Lucas Lauton de Alcantara, Lucas Pellicer, Rosimeire Pereira Costa, Edson Bollis, Anna Helena Reali Costa, and Artur Jordao. Layer-wise LoRA fine-tuning: A similarity metric approach. *arXiv preprint arXiv:2602.05988*, 2026. doi: 10.48550/arXiv.2602.05988. URL https://arxiv.org/abs/2602.05988.

Yi-Lin Sung, Varun Nair, and Colin A. Raffel. Training neural networks with fixed sparse masks. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/cb2653f548f8709598e8b5156738cc51-Abstract.html.

Tamás Lajos Tompa, Eszter Varga-Umbrich, Ilyes Batatia, Alin M. Elena, Noam Bernstein, and Gábor Csányi. Fine-tuning MLIP foundation models: Strategies for accuracy and transferability. *arXiv preprint arXiv:2606.12704*, 2026. doi: 10.48550/arXiv.2606.12704. URL https://arxiv.org/abs/2606.12704.

Chen Wang, Siyu Hu, Guangming Tan, and Weile Jia. ELoRA: Low-rank adaptation for equivariant GNNs. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 63113–63135. PMLR, 2025. URL https://proceedings.mlr.press/v267/wang25al.html.

Ziqing Xu, Hancheng Min, Lachlan Ewen MacDonald, Jinqi Luo, Salma Tarmoun, Enrique Mallada, and Rene Vidal. Understanding the learning dynamics of LoRA: A gradient flow perspective on low-rank adaptation in matrix factorization. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 4636–4644. PMLR, 2025. URL https://proceedings.mlr.press/v258/xu25h.html.

Qingru Zhang, Minshuo Chen, Alexander Bukharin, Nikos Karampatziakis, Pengcheng He, Yu Cheng, Weizhu Chen, and Tuo Zhao. AdaLoRA: Adaptive budget allocation for parameter-efficient fine-tuning. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=lq62uWRJjiY.

Suoxin Zhang, Run He, Di Fang, Xiang Tan, Kaixuan Chen, and Huiping Zhuang. Rethinking adapter placement: A dominant adaptation module perspective. *arXiv preprint arXiv:2605.06183*, 2026. doi: 10.48550/arXiv.2605.06183. URL https://arxiv.org/abs/2605.06183.

Jiawei Zhao, Zhenyu Zhang, Beidi Chen, Zhangyang Wang, Anima Anandkumar, and Yuandong Tian. GaLore: Memory-efficient LLM training by gradient low-rank projection. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 61121–61143. PMLR, 2024. URL https://proceedings.mlr.press/v235/zhao24s.html.

Wanjia Zhao, Jiaqi Han, Siyi Gu, Mingjian Jiang, James Y. Zou, and Stefano Ermon. GeoAda: Efficiently finetune geometric diffusion models with equivariant adapters. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-3172. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/88237ac4e9941b1be5c6d3c1ad408184-Abstract-Conference.html.

A Functional Reachability and Locality

The reachability calculation compares a required held-out correction with the functional images exposed by candidate placements. A single coefficient vector is fitted to the required correction and transferred unchanged to held-out configurations. Validation incompatibility tests whether the candidate space covers that correction, while the nonlinear remainder tests whether the local tangent still describes the adapted model. The derivation below makes these conditions precise; the later protocol details specify adapter spaces, data partitions, optimization, and statistical estimands.

Using the energy observable e_θ defined in the main text, the split-indexed joint output is

$$y_\theta(\mathcal{D}) = \mathrm{vec}_{x\in\mathcal{D}}\left(e_\theta(x), F_\theta(x)\right). \tag{9}$$

Suppressing task, slice, and target indices as in the main derivation, define the realized correction by $\Delta_s^\sigma(a) = y_{\theta_{s,k}(a)}(\mathcal{D}_m^\sigma) - y_{\theta_0}(\mathcal{D}_m^\sigma)$. Projected Adam generates feasible iterates $\{a_s^{(t)}\}_{t=0}^{T-1}$ and uses

$$\ell_s^{\mathrm{fit}}(a) = \frac{\left\|\Delta_s^{\mathrm{fit}}(a) - \Delta_{\mathrm{fit}}^\star\right\|_{\mathrm{fit},o}^2}{\left\|\Delta_{\mathrm{fit}}^\star\right\|_{\mathrm{fit},o}^2}, \qquad t_s^\star \in \arg\min_{0\le t<T} \ell_s^{\mathrm{fit}}(a_s^{(t)}), \quad a_s^{\mathrm{nl}} = a_s^{(t_s^\star)}. \tag{10}$$

There is no ridge term in ℓ_s^{fit}; λ is used only for the tangent fit. The nonlinear held-out incompatibility is then

$$\tau_{s,\mathrm{nl}}^{\mathrm{val}} = \frac{\left\|\Delta_s^{\mathrm{val}}(a_s^{\mathrm{nl}}) - \Delta_{\mathrm{val}}^\star\right\|_{\mathrm{val},o}}{\left\|\Delta_{\mathrm{val}}^\star\right\|_{\mathrm{val},o}}. \tag{11}$$

We quantify departure from the initial tangent by

$$\epsilon_s^{\mathrm{val}} = \frac{\left\|\Delta_s^{\mathrm{val}}(a_s^{\mathrm{nl}}) - J_s^{\mathrm{val}} a_s^{\mathrm{nl}}\right\|_{\mathrm{val},o}}{\left\|J_s^{\mathrm{val}} a_s^{\mathrm{nl}}\right\|_{\mathrm{val},o}}. \tag{12}$$

If the Jacobian is L-Lipschitz inside the coordinate ball, Taylor's theorem bounds the unnormalized remainder by $L\|a\|_2^2/2$. This motivates the locality criterion but does not determine a universal cutoff. We define the local regime for a molecule–target decision by $\max_{s,k} \epsilon_{msko}^{\mathrm{val}} \le 0.15$; all evaluated denominators are nonzero. Some tables also report the descriptive slice mean $\bar{\epsilon}_{mso}^{\mathrm{val}}$. A mean above 0.15 lies outside the local regime, whereas a mean below 0.15 does not determine the maximum. Reporting rank and remainder separately distinguishes predictive ordering from local first-order fidelity.

Proof of Proposition 1. Set $a = T_s b$. Because T_s is invertible, this is a bijection between coefficient vectors. Moreover,

$$\|b\|_{G_s'}^2 = b^\top T_s^\top G_s T_s b = \|a\|_{G_s}^2, \qquad J_s'^\sigma b = J_s^\sigma T_s b = J_s^\sigma a.$$

The trust region, ridge penalty, fit residual, and validation correction are therefore identical after the change of variables. Since $\lambda > 0$ and $G_s \succ 0$, the fit is unique, so the fitted functional correction and τ_s coincide. If J_s^σ is transformed while G_s is not, the coefficient balls correspond only when T_s is an isometry of G_s; otherwise their images under J_s^σ are different reachable ellipsoids. □

The initial tangent of ordinary two-factor LoRA has a different scope. For a matrix M, rank-q LoRA writes $M' = M + (\alpha/q)AB$, with $A \in \mathbb{R}^{m\times q}$ and $B \in \mathbb{R}^{q\times n}$. The factor rank $q = 4$ is distinct from the fixed-basis slice dimension $d = 32$. At the usual initialization $(A_0, B_0 = 0)$,

$$\mathrm{d}M'|_{(A_0,0)} = \frac{\alpha}{q} A_0 \,\mathrm{d}B. \tag{13}$$

Changes in A are silent to first order, so the measured image contains only fixed-A_0 changes in B. Training both factors instead yields the finite update $(\alpha/q)(A_0\Delta B + \Delta A\Delta B)$ and can rotate it away from that image. The differential exactly characterizes the first-order admissible directions; finite fixed-basis updates additionally require the empirical locality check. The unrestricted experiment records how activation of A co-occurs with changes in the optimization path and final functional space.

B Experimental Details

B.1 Adapter spaces and data partitions

The candidate set is defined by architectural role before any downstream outcome is evaluated. In matched-local experiments, every candidate is restricted to 32 orthonormal directions in the initially active LoRA factor; the same directions define the tangent fit and the nonlinear update. In the unrestricted benchmark, both rank-4 factors train. Table 2 reports the MACE-OFF23 dimensions and parameter counts. The initial studies use the first four rows. The six-molecule cohort uses all six architectural roles; MACE-MP-0/Ti–Al uses four roles in its prospective $\gamma = 2.5$ experiment and all six only in a later post-hoc extension, with dimensions inherited from the corresponding pretrained model.

Table 2: **Adapter spaces.** "Initial coordinates" is the size of the first-order-active B factor before drawing the 32-dimensional matched-local slice. "LoRA parameters" counts both trainable factors in the unrestricted MACE-OFF23 rank-4 benchmark.

Placement	Architectural role	Initial coordinates	LoRA parameters
Early scalar	First interaction, scalar-pathway input	4×64	288
Middle scalar	First interaction, scalar-pathway output	4×512	2,304
Late scalar	Second interaction, scalar-pathway output	$4\times1{,}280$	5,376
Equivariant product 0	First product, equivariant linear map	1,024	2,048
Equivariant product 1	Second product, equivariant linear map	512	1,024
Nonlinear readout	Final nonlinear readout map	64	576
Four-placement union	All four predeclared roles	–	10,016

All wrappers act only on representation-compatible scalar or equivariant-linear paths. Zero initialization therefore preserves the base model, and every finite adapter perturbation respects the transformation rule of the wrapped module. The same definitions are used across temperatures, molecules, and Ti–Al.

Only 300 K is used to choose slice dimension, optimizer, trust radius, or learning rate. For paracetamol, all analysis choices precede the first model output, and the 1,000-frame test remains untouched until after placement selection. Its 12 adapted coordinate vectors are reused without retraining, with the energy offset estimated solely from the original 80 fitting configurations. Azobenzene and ethanol use the same construction; each contributes 12 internal fits and 12 test evaluations. The six-molecule cohort extends this design to joint, energy-only, and force-only prediction targets and to six placements. In total, 108 Jacobian runs yield 324 target-specific tangent scores, followed by 324 constrained fits and 324 untouched-test evaluations. These are computation counts, not independent evidence points. The Ti–Al split is stratified over low, medium, and high distortions. Its original four-placement analysis at $\gamma = 2.5$ is prospective; $\gamma = 0.5$ and the later six-placement extension are post-hoc sensitivities. The 3BPA result at $\gamma = 2.5$ is likewise a single smaller-radius sensitivity to the original $\gamma = 5$ analysis.

B.2 Optimization and statistical estimands

The controlled reachability matrix uses eight orthonormal directions per placement, a teacher-parameter norm of 0.02, ridge 10^{-10}, and five paired replicates. Each replicate uses three configurations to construct an excluded teacher direction, four to fit, and two to validate. Its bounded nonlinear follow-up uses the inherited middle-teacher/late-student pair, Adam for 240 steps at learning rate

Table 3: **Data partitions and evaluation stages.** Fit/validation counts in the third column are those used for the tangent and matched-local analyses; ordinary-LoRA performance controls use the train/validation counts in the second column.

System/task	Train/ validation	Tangent fit/ validation	Evaluation stage
3BPA, 300 K	$320/80$	$60/40$ from a disjoint 100	Development and original $\gamma = 5$ analysis
3BPA, 600 K	$250/130$	$80/40$ from a disjoint 120	Temperature transfer
3BPA, 1,200 K	$250/130$	$80/40$ from a disjoint 120	Temperature transfer
Paracetamol	–	$80/40$ from a preassigned 120	Prospective molecular validation
Paracetamol external	–	untouched 1,000 frames	Prospective test; no retraining
Azobenzene	–	$80/40$ from a preassigned 120	Label-blind prospective replication
Azobenzene external	–	untouched 1,000 frames	Prospective test; no retraining
Ethanol	–	$80/40$ from a preassigned 120	Label-blind prospective replication
Ethanol external	–	untouched 1,000 frames	Prospective test; no retraining
Six-molecule cohort	–	$80/40$ per molecule	Three-target placement selection
Six-molecule tests	–	untouched 1,000 frames each	External confirmation; no retraining
Ti–Al/GPAW, $\gamma = 2.5$	–	$84/36$ from a preassigned 120	Prospective; original four placements
Ti–Al/GPAW, $\gamma = 0.5$	–	same $84/36$	Post-hoc; original four placements
Ti–Al/GPAW six-role tangent	–	same $84/36$	Post-hoc; three prediction targets
Ti–Al/GPAW six-role nonlinear	–	same $84/36$	Post-hoc; joint target only

10^{-4}, gradient clipping at 100, and trust radius $\gamma = 0.04$ per placement. These interventions use raw total energy and equal aggregate energy–force channel weights. Nonlinear optimization starts from the tangent solution and retains the evaluated iterate with minimum normalized fit loss. Matched, adverse, and union controls use identical data windows within each replicate.

The real-label matched-local experiments use 32 directions, ridge 10^{-12}, three paired random-slice identities, and zero-start Adam for 80 steps at learning rate 10^{-2} with unit gradient clipping and projection after every step; the evaluated iterate with minimum normalized fit loss is retained. The nonlinear loss has no ridge term, while λ regularizes only the tangent solve. Tangent-warm starts are secondary checks. The original trust radii are $\gamma = 5.0$ for 3BPA and $\gamma = 2.5$ for paracetamol and Ti–Al. The locality criterion is $\max_{s,k} \epsilon^{\mathrm{val}}_{msko} \leq 0.15$. The main six-molecule study uses $\gamma = 2.5$ and applies this identical optimization to all six placements and all three prediction targets.

The rank estimand is Spearman correlation across condition means: 12 temperature-by-placement means for 3BPA and four placement means for each initial transfer system. In the six-molecule analysis, the molecule is the unit of generalization and joint, energy-only, and force-only are related prediction targets. Descriptive 95% intervals use 10,000 paired bootstrap draws over the three slice identities, applying the same resampled identities to every placement. They quantify sensitivity to these three slices, not population-level uncertainty.

Ordinary-LoRA controls use AdamW, a 30:1 energy-per-atom/force objective, at most 200 epochs, and patience 20. Fixed LoRA uses learning rate 10^{-3}; full-update methods search $\{10^{-5}, 3 \times 10^{-5}, 10^{-4}\}$ on 300 K, select 10^{-4}, and transfer it unchanged. All 48 predefined runs are retained. The 288-parameter random-layer control matches the selected early-scalar layer, not the 10,016-parameter union.

C Extended Results and Mechanistic Analysis

C.1 Numerical and controlled checks

Table 4: **Numerical verification of the joint tangent operator.** Finite-difference entries use the best step in the sweep; the symmetry entry is the maximum over translation, rotation, permutation, and net-force checks.

Check	Value
Equivalence at zero initialization	2.840×10^{-19}
Energy JVP finite-difference relative error	1.149×10^{-5}
Force JVP finite-difference relative error	3.751×10^{-8}
JVP/VJP adjoint relative error	7.346×10^{-17}
Maximum measured symmetry discrepancy	6.401×10^{-15}
Maximum validation Jacobian drift	1.245×10^{-7}

Table 5: **Controlled bounded-nonlinear experiment.** Values are mean validation incompatibilities over five paired replicates; both adverse nonlinear gaps are positive in every replicate.

Condition	Tangent τ^{val}	Nonlinear τ^{val}
Matched placement	0.050954	0.000503
Adverse student	0.985147	0.990381
Union control	0.114029	0.061531

Across the full teacher–student matrix, mean same-placement incompatibility is 0.000–0.051, versus 0.889–1.023 for the selected off-diagonal pairs. For the prespecified middle-teacher/late-student pair, the adverse-minus-matched gap is positive in all five replicates (mean 0.9342, range $[0.8726, 0.9852]$); maximum adverse validation-Jacobian drift is 1.245×10^{-7}.

C.2 Cross-system locality and matched-local outcomes

Table 6: **Cross-system rank and locality results.** Intervals are descriptive 95% paired-slice bootstrap intervals. Rank or top-1 agreement does not imply locality. External tests reuse preselected coordinates and therefore bear on rank transfer rather than local fidelity. Post-hoc sensitivities are shown separately from the corresponding larger-trust-radius experiments.

Experiment	Analysis	γ	ρ [95% interval]	Locality statistic / outcome
Unrestricted 3BPA LoRA	Path rotation	–	-0.6268 $[-0.6690, -0.5423]$	Evolving basis
3BPA original	Original analysis	5.0	1.000 $[0.979, 1.000]$	$0.2006 > 0.15$
3BPA sensitivity	Post-hoc trust sensitivity	2.5	1.000 $[0.979, 1.000]$	$0.1062 < 0.15$
Paracetamol validation	Prospective	2.5	1.000 $[1.000, 1.000]$	$0.0342 < 0.15$
Paracetamol external	Prospective test	2.5	1.000 $[1.000, 1.000]$	top-1 match
Azobenzene/ethanol validation	Label-blind replication	2.5	1.000 $[0.976, 1.000]$	$0.0332 < 0.15$
Azobenzene/ethanol external	Untouched tests	2.5	1.000	$2/2$ top-1
Six-molecule cohort	Empirical rank/test	2.5	–	max condition mean 1.284; $23/108$ means > 0.15
Earlier-task six-placement expansion	Post-hoc	2.5	1.000	$0/6$ tasks all-local
Six-molecule $d = 8$	Dimension sensitivity	2.5	–	$18/18$ matches $d = 32$
Ti–Al original	Prospective analysis	2.5	1.000 $[0.800, 1.000]$	$0.3867 > 0.15$
Ti–Al sensitivity	Post-hoc trust sensitivity	0.5	1.000 $[0.800, 1.000]$	$0.0458 < 0.15$

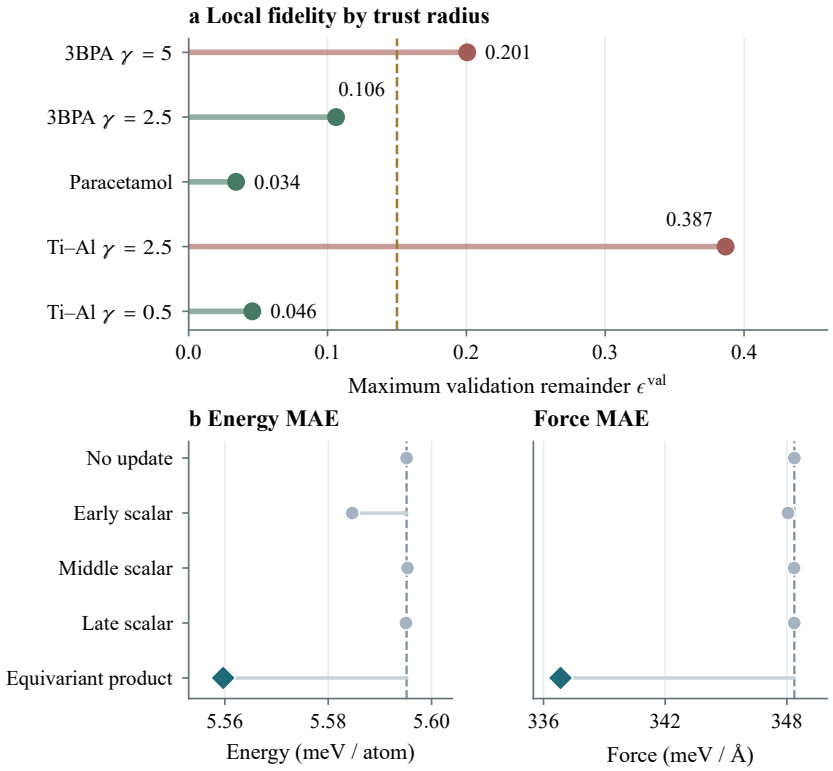


Figure 3: **Matched-local fidelity and transfer. a**, Maximum validation linearization remainder for five task–radius conditions; the dashed line is the 0.15 locality limit. **b**, Untouched paracetamol energy and force MAE after reusing the preselected adapted coordinates. Each horizontal segment compares a placement with the unadapted baseline (dashed line); the diamond identifies the best equivariant-product placement. Values are three-slice means.

Table 7: **All 3BPA $\gamma = 2.5$ sensitivity conditions.** Each row is the mean over three paired random slices. This is the single post-hoc smaller-radius sensitivity; the prespecified $\gamma = 5$ experiment is reported separately in Table 6.

Task	Placement	Tangent τ^{val}	Nonlinear τ^{val}	$\epsilon^{\mathrm{val}}_{\mathrm{max}}$
300 K	Early scalar	0.996713	0.997333	0.000488
	Middle scalar	0.999884	0.999906	0.000030
	Late scalar	0.999944	0.999958	0.000014
	Equivariant product	0.932660	0.939398	0.089650
600 K	Early scalar	0.996993	0.997555	0.000417
	Middle scalar	0.999895	0.999916	0.000045
	Late scalar	0.999950	0.999960	0.000014
	Equivariant product	0.923750	0.930363	0.106156
1,200 K	Early scalar	0.997309	0.997765	0.000425
	Middle scalar	0.999903	0.999921	0.000116
	Late scalar	0.999953	0.999962	0.000013
	Equivariant product	0.906844	0.916662	0.036785

Table 8: **Prospective paracetamol and original four-candidate Ti–Al conditions.** Tangent and nonlinear columns are three-slice means; $\epsilon_{\max}^{\mathrm{val}}$ is the largest slice remainder. The $\gamma = 2.5$ analysis is prospective and the $\gamma = 0.5$ sensitivity is post-hoc. The later six-placement Ti–Al extension is reported separately in Table 9.

System	Setting	Placement	Tangent τ^{val}	Nonlinear τ^{val}	$\epsilon_{\max}^{\mathrm{val}}$
Paracetamol	Prospective	Early scalar	0.998894	0.999116	0.000474
		Middle scalar	0.999945	0.999955	0.000062
		Late scalar	0.999977	0.999982	0.000022
		Equivariant product	0.960270	0.967134	0.034151
Ti–Al	$\gamma = 2.5$	Early scalar	0.998551	0.998813	0.000059
		Middle scalar	0.999522	0.999647	0.000023
		Late scalar	0.999931	0.999938	0.000012
		Equivariant product	0.756188	0.765841	0.386695
Ti–Al	$\gamma = 0.5$	Early scalar	0.999679	0.999740	0.000015
		Middle scalar	0.999882	0.999911	0.000005
		Late scalar	0.999988	0.999988	0.000002
		Equivariant product	0.937502	0.943007	0.045845

Table 9: **Post-hoc six-placement Ti–Al extension at $\gamma = 0.5$.** The tangent analysis contains joint, energy, and force prediction targets, while constrained adaptation evaluates the joint target only. Values below are joint-target placement means over three slices. The last column is a placement-mean remainder, so its maximum (0.030722) is distinct from the largest slice remainder (0.045845) in the original four-candidate sensitivity in Table 8.

Placement	Tangent τ^{val}	Nonlinear τ^{val}	Mean remainder
Early scalar	0.999679	0.999740	0.000011
Middle scalar	0.999882	0.999911	0.000003
Late scalar	0.999988	0.999988	0.000001
Equivariant product 0	0.937502	0.943007	0.030722
Equivariant product 1	0.972477	0.975361	0.007158
Nonlinear readout	0.972644	0.975863	0.021354

Table 10: **Untouched 1,000-frame paracetamol test.** Values are means over three preselected adapted coordinate vectors; no external frame is used for training, model selection, or energy-offset fitting.

Placement	External τ	Energy MAE (meV/atom)	Force MAE (meV/Å)
No update	1.000000	5.59516	348.370
Early scalar	0.999129	5.58465	348.059
Middle scalar	0.999954	5.59535	348.354
Late scalar	0.999982	5.59504	348.364
Equivariant product	0.967777	5.55968	336.841

Table 11: **Original four-placement label-blind replications.** Each row averages the same three paired slice identities. Internal nonlinear fits use the 80/40 fit–validation split; test values reuse those coordinates on 1,000 untouched frames without retraining.

Molecule	Placement	$\tau^{\mathrm{tan,val}}$	$\tau^{\mathrm{nl,val}}$	$\tau^{\mathrm{nl,test}}$	Force MAE	$\epsilon_{\max}$
Azobenzene	Early scalar	0.999319	0.999461	0.999501	354.569	0.000517
	Middle scalar	0.999944	0.999956	0.999952	354.787	0.000055
	Late scalar	0.999980	0.999985	0.999984	354.798	0.000007
	Eq. product	0.955976	0.963511	0.962653	344.634	0.028538
Ethanol	Early scalar	0.998973	0.999214	0.999201	369.910	0.000269
	Middle scalar	0.999935	0.999948	0.999947	370.194	0.000069
	Late scalar	0.999977	0.999981	0.999980	370.207	0.000010
	Eq. product	0.975724	0.979292	0.979975	362.581	0.033232

Force MAE is in meV $\mathrm{Å}^{-1}$. The selected placement is determined only by the internal tangent column.

Table 12: **All 18 related six-molecule prediction targets.** Positive values are gains over the fixed first equivariant-product map. The held-out tangent score selects the nonlinear readout in every row under the native coordinate metric; that selection is also top-ranked after matched nonlinear adaptation and on the untouched 1,000-frame test. External reduction is the relative decrease in normalized residual. Targets within a molecule share configurations. At $\gamma = 2.5$, these rows describe rank transfer outside the local regime.

Molecule	Target	$\Delta\tau^{\mathrm{tan,val}}$	$\Delta\tau^{\mathrm{nl,val}}$	$\Delta\tau^{\mathrm{nl,test}}$	Test reduction
Benzene	Joint	0.167	0.174	0.175	18.13%
	Energy	0.268	0.196	0.181	18.86%
	Force	0.178	0.208	0.209	21.66%
Malonaldehyde	Joint	0.204	0.191	0.184	19.02%
	Energy	0.317	0.257	0.202	21.16%
	Force	0.214	0.218	0.211	21.90%
Naphthalene	Joint	0.283	0.279	0.271	27.96%
	Energy	0.177	0.107	0.162	16.86%
	Force	0.298	0.295	0.287	29.57%
Salicylic acid	Joint	0.279	0.260	0.262	26.96%
	Energy	0.272	0.170	0.178	18.37%
	Force	0.290	0.280	0.282	29.01%
Toluene	Joint	0.186	0.178	0.166	17.09%
	Energy	0.229	0.103	0.120	12.51%
	Force	0.197	0.200	0.185	19.09%
Uracil	Joint	0.198	0.198	0.205	21.15%
	Energy	0.234	0.095	0.112	11.60%
	Force	0.202	0.203	0.211	21.84%
Mean		**0.233**	**0.201**	**0.200**	**20.71%**

Table 13: **Physical errors on untouched tests after joint-target adaptation.** Values are means over three paired slices. The selected nonlinear readout, fixed product 0 map, and unadapted pretrained model are evaluated on the same 1,000 frames, with the energy gauge inherited from the fit split. Energy MAE is in meV/atom and force MAE in meV $\mathrm{Å}^{-1}$.

	Energy MAE			Force MAE		
Molecule	Base	Fixed product	Selected	Base	Fixed product	Selected
Benzene	4.37	4.23	**3.26**	234.16	226.44	**181.46**
Malonaldehyde	7.33	8.12	**6.52**	405.40	391.32	**315.51**
Naphthalene	5.51	5.49	**4.46**	289.48	279.95	**182.73**
Salicylic acid	6.21	6.45	**5.14**	351.36	342.77	**240.00**
Toluene	5.49	5.68	**4.23**	282.19	273.97	**219.83**
Uracil	6.68	7.05	**5.62**	364.71	353.91	**258.16**
Mean	5.93	6.17	**4.87**	321.22	311.39	**232.95**

C.3 Six-molecule outcomes and robustness checks

Figures 4 and 5 visualize the aggregate outcomes of the selected fixed-basis readout adapter. They use the same six-molecule and target-level values as Tables 12 and 13; no individual configuration or atom-wise field is inferred from aggregate errors.

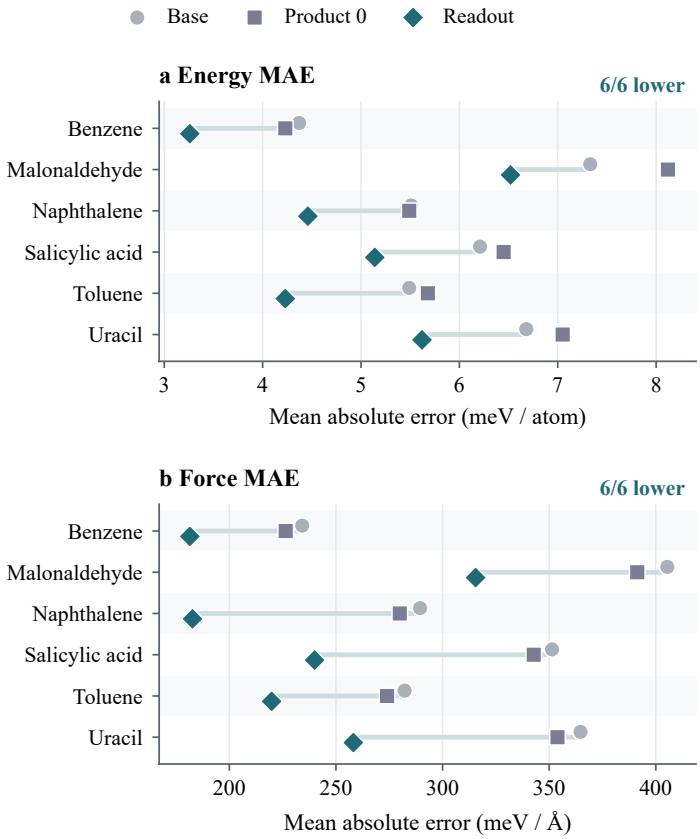


Figure 4: **Physical error profile of the selected solution. a**, Energy and **b**, force mean absolute error on 1,000 untouched frames per molecule after joint-target adaptation. Diamonds show the selected readout solution, squares the fixed first equivariant-product placement, and circles the unadapted base model. Values are means over three paired slices; lower is better. Lines connect the readout and base outcomes to make the improvement magnitude visible.

The following post-hoc checks examine candidate-set expansion, tangent dimension, and the coordinate metric without changing the reported selections.

Table 14: **Post-hoc six-placement extension on earlier tasks.** The candidate set was expanded after the original four-placement outcomes were known. All values average three slices. "Gain" is product-0 nonlinear incompatibility minus the selected readout value. None of the six tasks satisfies the all-placement 0.15 locality criterion.

Task	Tangent winner	Nonlinear winner	Gain	Max. remainder
3BPA 300 K	Readout	Readout	0.0875	0.973
3BPA 600 K	Readout	Readout	0.0925	0.992
3BPA 1,200 K	Readout	Readout	0.1145	0.940
Paracetamol	Readout	Readout	0.1959	0.926
Azobenzene	Readout	Readout	0.3668	0.871
Ethanol	Readout	Readout	0.1317	0.950
Mean	Readout	Readout	0.1648	–

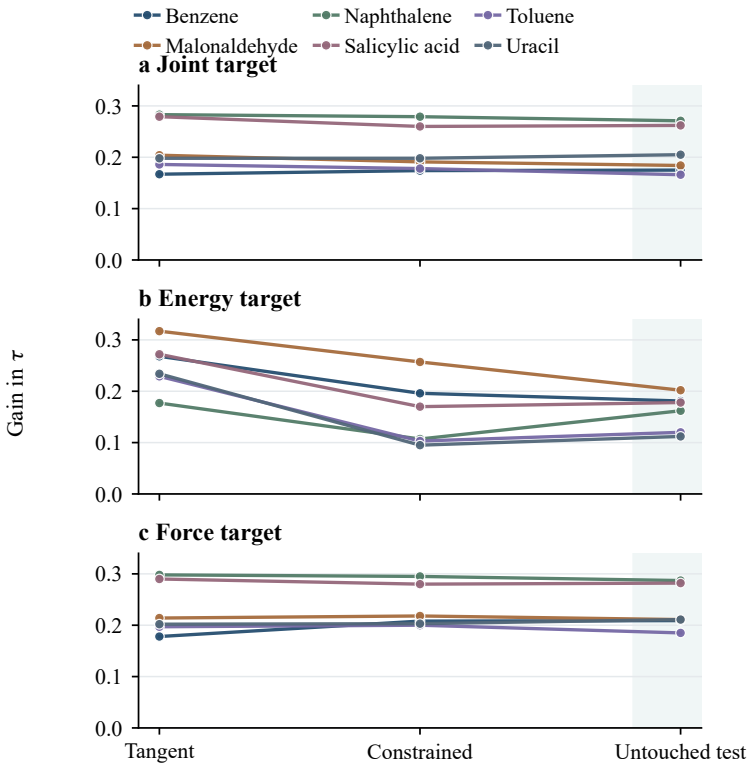


Figure 5: **How the readout solution changes across evaluation stages.** Each line is one molecule, showing the readout's incompatibility gain over fixed product 0 for the joint, energy, or force target. The first value is the held-out tangent prediction; the second is measured after constrained nonlinear adaptation; the third is measured on the untouched 1,000-frame test. Positive gains favor readout. The three target views within a molecule share configurations and are not independent replications.

Table 15: **Tangent-dimension, metric, and component-cost analysis.** The $d = 8$ analysis contains 108 Jacobian runs and no nonlinear fits. Its A40 tangent timing and the H200 $d = 32$ nonlinear timing are reported separately because the hardware differs.

Check	Value	Interpretation
$d = 8$ versus $d = 32$ native-metric winner	$18/18$	Tangent only
Column-normalized versus native winner	$0/18$	Metric dependent
$d = 8$ tangent operators per selection	48 JVPs	Six placements
Tangent wall time per placement	46.09 s	A40; $d = 8$
Tangent peak memory	22.05 GB	A40; $d = 8$
Nonlinear work per placement/target	80 steps	H200; $d = 32$
Nonlinear wall time per placement/target	34.38 s	H200; $d = 32$
Nonlinear peak memory	22.10 GB	H200; $d = 32$

We also compare the held-out selector with fit-only and raw-count rankings on the same constrained-adaptation outcomes.

Table 16: **Direct placement-selector comparisons on constrained outcomes.** Each entry gives Spearman ρ against zero-start nonlinear validation incompatibility and taskwise top-1 matches. "Fit only" ranks the tangent residual on the fitting configurations. "Raw count" ranks the negative number of first-order-active coordinates, encoding the heuristic that a larger exposed space should perform better. Comparator definitions and values are computed after the primary analysis; no model is retrained. Every candidate trains 32 coordinates, although native-space coverage differs.

Analysis	n	Held-out tangent		Fit only		Raw count
		ρ	Top-1	ρ	Top-1	ρ / Top-1
3BPA, $\gamma = 5$	12	1.000	3/3	0.993	3/3	-0.777 / 0/3
3BPA, $\gamma = 2.5$	12	1.000	3/3	1.000	3/3	-0.777 / 0/3
Paracetamol, $\gamma = 2.5$	4	1.000	1/1	1.000	1/1	-0.800 / 0/1
Ti–Al, $\gamma = 2.5$	4	1.000	1/1	1.000	1/1	-0.800 / 0/1
Ti–Al, $\gamma = 0.5$	4	1.000	1/1	1.000	1/1	-0.800 / 0/1

For the six-molecule cohort, native-metric held-out incompatibility, fit-only incompatibility, gradient norm, and Gauss–Newton trace all select the readout for all 18 related prediction targets. Column-normalized incompatibility selects early scalar in 18/18, giving 0/18 agreement with the native metric. All 108 condition means have the trust-region constraint active. Equal sliced dimension therefore does not imply invariance to the coordinate metric or equal fractional coverage.

C.4 Efficiency controls and unrestricted LoRA

Table 17: **3BPA performance comparison.** Entries are validation force MAE in meV Å^{-1} (mean over three seeds). The random-layer LoRA control is parameter-matched to the 288-parameter development-selected early-scalar layer, not to the placement union.

Task	Union LoRA	Random-layer LoRA	Selected-layer full	Full fine-tuning
300 K	8.21	18.71	19.59	7.23
600 K	13.49	25.50	26.64	12.24
1,200 K	25.39	38.07	38.70	23.16
Parameters	10,016	288	512	1,428,368

The unrestricted comparison contains all 36 task–placement–seed runs. Normalized regret uses the taskwise three-seed oracle and the median initial validation objective as its baseline. The resulting initial-tangent association is -0.6268 with descriptive interval $[-0.6690, -0.5423]$; taskwise values are -0.80, -0.80, and -0.40 at 300, 600, and 1,200 K. No condition or seed is excluded.

Table 18: **Factor motion and path geometry for unrestricted LoRA.** Growth and minimum canonical correlation are means over nine runs per scalar placement. Path correlations compare the stated post-hoc summary with the original normalized regret across the 12 condition means.

Quantity	Estimate	Population standard deviation
Early-scalar A-norm growth	474.3	46.4
Middle-scalar A-norm growth	343.0	25.6
Late-scalar A-norm growth	402.5	45.6
Early minimum canonical correlation	0.188	0.094
Middle minimum canonical correlation	0.062	0.038
Late minimum canonical correlation	0.036	0.036
Final affine tangent vs. regret	-0.1268	–
Union of path tangents vs. regret	-0.4437	–
Final realized state vs. regret	-0.0986	–

The initially silent A factor moves substantially at all three scalar placements, and its final column space has little overlap with the initial space. All three post-hoc path summaries remain anticorrelated with the endpoints. Factor growth and subspace rotation therefore co-occur with, and are consistent with, departure from the fixed initial tangent; they do not identify a unique causal mechanism.

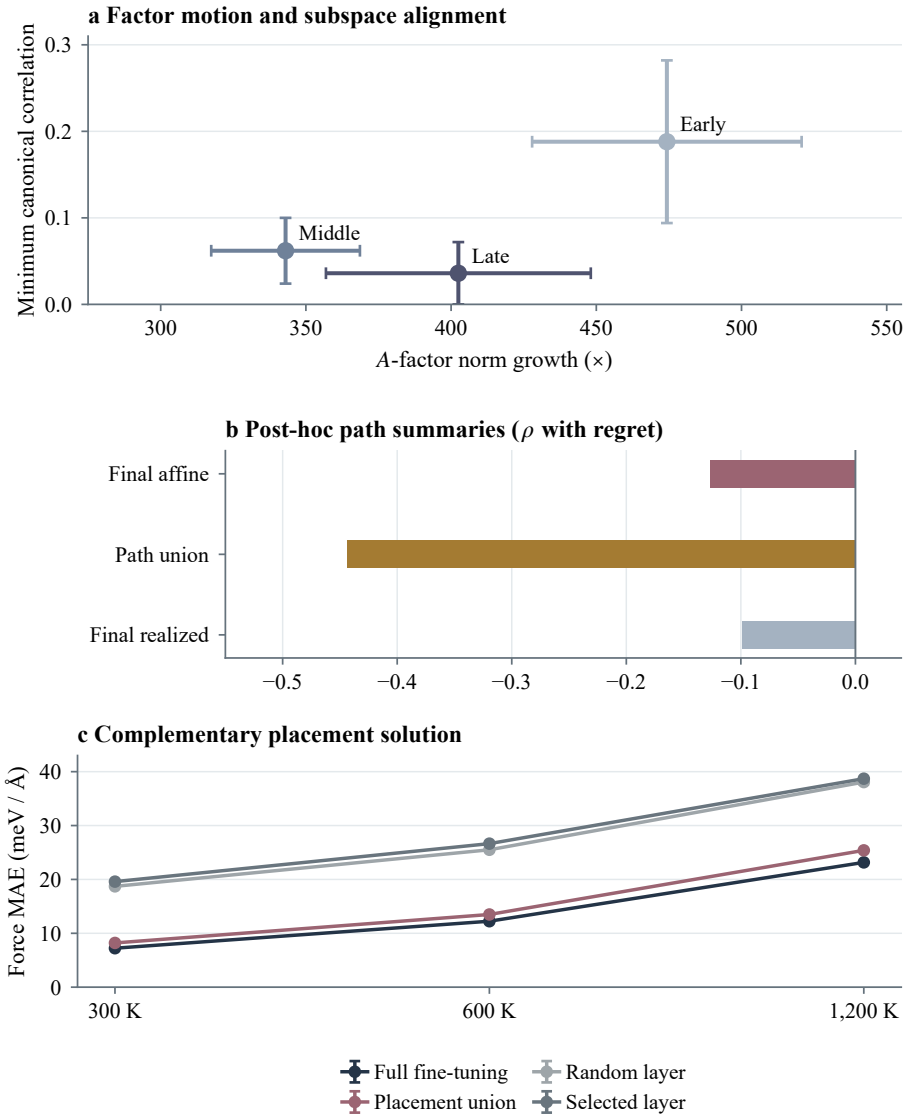


Figure 6: **Path geometry and complementary placement. a**, Unrestricted-LoRA A-factor growth versus initial/final subspace alignment; error bars are population standard deviations across nine runs per placement. **b**, Post-hoc path correlations with endpoint regret. **c**, Validation force MAE for union, full fine-tuning, and two controls; error bars show population standard deviations over three seeds.