Bidirectional Path Kernels for Local Split and Merge Detection in Policy Rollouts

Anonymous authors
Paper under double-blind review

Abstract

Offline policy rollouts can pass through the same local context and then follow several outgoing routes, or arrive there along several incoming routes, and neither endpoint dispersion nor a forward-only conditional law identifies this directed, path-valued object. We introduce the bidirectional local path-transition kernel (BLPTK): forward and reverse conditional laws of relative path segments, localized around a state–time context. At a declared path scale, minimum route mass and conductance tolerance, a spectral routine proposes partitions and an explicit mass-and-conductance test validates them, so every accepted partition certifies an empirical lower bound $\widehat{K}^{+}$ or $\widehat{K}^{-}$ on outgoing or incoming path multiplicity — failure to accept one has no converse interpretation. No global route labels or annotated event times enter the detector. We want these junctions located because they are where reinforcement fine-tuning quietly deletes a route, and a local partition of them is something a training objective can be constrained against. On controlled directed graphs, path features classify all 361/361 probed local graph types, versus 255/361 for endpoint features on identical anchors. Across the round-two valency sweep, orthogonally embedded routes are counted exactly at full valency (1.000 from valency 2 to 8 and at 12, 0.994 at 10), with a false-positive rate of 0.091 traced to a grader–estimator threshold mismatch; crowded embeddings fail in three separable ways, including miscounted anchors of intermediate truth, a term exactness and recall are blind to. On D3IL-Avoid diffusion-policy rollouts the forward detector recovers the branchable contexts against a withheld label it never sees (AUROC 0.874 over 300 anchors, base rate 0.387); every trajectory there branches at the same moment, so a clock is optimal by construction and we report the measurement it cannot explain — within a fixed time slice, AUROC 0.879 over 220 anchors. On a simulated 7-DoF arm whose obstacle, and therefore whose junction time, moves between episodes, a clock is no longer optimal ($\max_t p(t) = 0.410$) and the detector leads the per-slice oracle clock by +0.260 agreement over 204 anchors. Freezing those outgoing regions as a deliberately naive branch-mass constraint during reinforcement fine-tuning keeps both demonstrated routes in all five seeds and passes a predeclared gate in four, where unconstrained fine-tuning deletes one in all five and, at matched dose, fixed-time regions and fork-window KL pass in zero — a probe of whether the regions are useful, not a proposed algorithm. Time reversal matches the forward count on all 204 matched balls, validating the reverse code path but giving no independent evidence for natural merges. Throughout, "certificate" means empirical partition feasibility at a declared resolution — not a confidence bound, and not recovery of a population route graph.

1 Introduction

A rollout data set contains more structure than a cloud of endpoints. Near a behavioural junction, locally similar policy contexts may generate paths that leave along distinct routes; at a meeting point, distinct histories may enter a common continuation. The object of

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.

interest is therefore both path-valued and directed: outgoing multiplicity is not incoming multiplicity. A large scalar spread can signal variability, but it cannot say whether that variability is incoming or outgoing, how many route blocks it supports, or whether those blocks satisfy a declared mass and separation criterion. The existing offline handles are no better. A global terminal label imposes one identity on an entire trajectory, even after two routes merge; annotated event times assume the junctions are already known; endpoint features erase a split that re-merges within the observation window. We instead ask a local question: conditioned on the context at an anchor, how many well-separated, non-negligible path populations can be exhibited forward and backward?

Our answer is the bidirectional local path-transition kernel (BLPTK), a pair of conditional laws on forward and backward path segments. We turn each empirical law into a weighted path graph at a declared affinity scale; spectral embeddings propose candidate partitions, and a separate test accepts a candidate only when every block has sufficient empirical mass and sufficiently low volume-normalized conductance. The largest accepted count is a certificate in a deliberately one-sided sense: it witnesses an admissible empirical partition, so it lower-bounds the unrestricted empirical maximum, while the spectral search can miss another admissible partition and a non-fire proves nothing.

These junctions are worth locating because they are where behavioural diversity is lost. Nothing in a reinforcement objective protects a route: fine-tuning may collapse a multi-route policy onto whichever route its reward marginally prefers, and in our arm task it does so in every seed. A detector that certifies where routes separate, and which rollouts belong to which, gives such a procedure somewhere to stand — the certified regions can be frozen and used to hold route mass without dictating behaviour inside a route. That is the use we designed it for, and why its output is a partition rather than a score; Section 7.1 tests it.

The contributions are a formulation, an estimator and an evaluation. We formulate local policy behaviour as a pair of forward and reverse path kernels that separate splits, merges, crossings and corridors without persistent route identities, and show why endpoint projection is non-identifying and why the reverse law necessarily depends on data occupancy. We define resolution-dependent path multiplicity and give an estimator whose accepted counts are certified empirical lower bounds, checking the estimand itself rather than treating a small eigenvalue as a count. We evaluate direction, count, locality, resolution, finite-sample behaviour, held-out survival and a real forward split; in randomized split–remerge diamonds the path count is 2 while the endpoint count is 1 in all 12/12 serialized trials. Finally, on a simulated 7-DoF arm whose junction time varies per episode — the first environment here where a clock is not optimal by construction — the detector leads the per-slice oracle clock by $+0.260$ agreement, and freezing its certified outgoing regions as a branch-mass constraint keeps both demonstrated routes in all five seeds of reinforcement fine-tuning where unconstrained fine-tuning deletes one in all five; matched-dose fixed-time partitions and fork-window KL each fail the predeclared gate in all five.

The scope is intentionally narrower than route-graph reconstruction: counts are local and resolution-dependent and need not converge to an intrinsic population graph, and a time-reversed real trajectory set exercises the reverse implementation but is exactly feature-symmetric to the original forward run. We therefore claim real-data evidence only for forward split localization.

2 Problem setting: local directed path multiplicity

Let an offline data set contain independent rollouts $\mathcal{D} = \{C_0^i, \dots, C_{T_i}^i\}_{i=1}^N$, where the predictive context C_t includes the physical state and may include time, policy memory, observation history or a sampled action chunk when those are required for Markovianity. Our implementation forms time-indexed windows; "without annotated event times" does not mean "without a clock." For anchor (t, c) and horizon h, feature maps Ψ^+ and Ψ^- encode the forward segment $C_{t:t+h}$ and backward segment $C_{t-h:t}$.

The target is local and directed. At a declared resolution, write $K^+(c)$ for the number of outgoing path components and $K^-(c)$ for incoming ones. Their ordered pair (K^-, K^+)

names the local graph type: $(1,1)$ a corridor, $(1,\geq 2)$ a split, $(\geq 2,1)$ a merge, and $(\geq 2,\geq 2)$ a crossing or local reorganization. These are local multiplicities, not labels that must persist along a whole rollout.

Why endpoints cannot identify the target. Evaluating a path at its endpoint cannot distinguish a law concentrated on one centre path from one split equally between two disjoint paths with the same start and end, so no endpoint statistic identifies every split that remerges before the horizon; the randomized diamond experiment in Section 5 instantiates the obstruction. The incoming law obeys a Bayes reversal that depends on the rollout occupancy as well as the dynamics, so one row of a forward kernel does not identify incoming multiplicity either (Appendix C.14).

Evaluation and the conditioning trap. Synthetic truth uses single-linkage components of the noiseless path segments in the same declared feature metric, then removes components below the same mass gate; the estimator never sees those segments. Our headline metric is agreement $\mathbb{P}(\widehat{K} = K_{\text{true}})$ over all scored anchors. A conditional quantity such as $\mathbb{P}(\widehat{K} = n \mid K_{\text{true}} = n)$ is reported only with the number of anchors satisfying its condition; when that set is empty, the quantity is undefined rather than successful. For Avoid, truth is an engineered terminal-side label used only for scoring, not by the detector.

3 Bidirectional local path-transition kernels

The population object. For a closed-loop policy law, define the two Markov kernels

$$\mathbb{K}^{+}_{t,h}(c, A) = \mathbb{P}\big(\Psi^{+}(C_{t:t+h}) \in A \mid C_t = c\big)\,, \tag{1}$$

$$\mathbb{K}^{-}_{t,h}(c, A) = \mathbb{P}\big(\Psi^{-}(C_{t-h:t}) \in A \mid C_t = c\big)\,. \tag{2}$$

The reverse object is a conditional history law, not inverse robot dynamics, and both are occupancy-weighted laws of the rollout population. Endpoint conditioning is recovered by choosing $\Psi = e_h$; the converse is impossible by the construction in Section 2.

Local empirical path laws. Choose a context kernel κ_ϵ and normalized weights

$$w_i(c) = \frac{\kappa_\epsilon(c, C^i_t)}{\sum_j \kappa_\epsilon(c, C^j_t)}, \qquad \widehat{\mathbb{K}}^{\pm}_{t,h,\epsilon}(c, \cdot) = \sum_i w_i(c)\, \delta_{Y^{\pm}_i}, \tag{3}$$

where $Y^{+}_i = \Psi^{+}(C^i_{t:t+h})$ and $Y^{-}_i = \Psi^{-}(C^i_{t-h:t})$: a Nadaraya–Watson estimator of the complete conditional measure [7, 9]. Feature construction, subsampling and the refusal rule are in Appendix B.

A resolution-dependent estimand. On path-feature space, the compactly supported Wendland affinity is $\eta_\ell(y,z) = \varphi(d(y,z)/\ell)$. For a path law ν, define

$$Q_\ell(A, B) = \int_A \int_B \eta_\ell(y,z)\, \nu(dy)\nu(dz), \qquad d_{\nu,\ell}(y) = \int \eta_\ell(y,z)\nu(dz), \tag{4}$$

$$\mathrm{vol}_{\nu,\ell}(A) = \int_A d_{\nu,\ell}(y)\nu(dy), \qquad \psi_\ell(A) = \frac{Q_\ell(A, A^c)}{\mathrm{vol}_{\nu,\ell}(A)}. \tag{5}$$

ψ_ℓ is the one-sided escape probability the acceptance test applies to every block; the balanced Cheeger conductance ϕ_ℓ, which divides instead by $\min\{\mathrm{vol}_{\nu,\ell}(A), \mathrm{vol}_{\nu,\ell}(A^c)\}$, is what Section 4 is stated for and what our bipartition sweep minimizes. Pointwise $\psi_\ell \leq \phi_\ell$, and the two all-block tests coincide at every K (Appendix C.15), so the choice is one of exposition, not of strictness. For minimum route mass α and conductance tolerance δ_{cut}, the component number is

$$\mathfrak{K}_{\ell,\alpha,\delta_{\text{cut}}}(\nu) = \max\left\{ K : \begin{array}{l} \mathcal{Y} = A_1 \sqcup \cdots \sqcup A_K, \quad \nu(A_k) \geq \alpha, \\ \psi_\ell(A_k) \leq \delta_{\text{cut}} \quad \text{for every block} \end{array} \right\}, \tag{6}$$

with the ordinary one-component answer when no nontrivial partition is admissible. The declared triple $(\ell, \alpha, \delta_{\text{cut}})$ defines the question; there is no scale-free route count.

Propose, then validate. For local weights w_i and affinities η_{ij}, the implementation forms

$$d_i = \sum_j w_j \eta_{ij}, \qquad S_{ij} = \frac{\sqrt{w_i w_j}\eta_{ij}}{\sqrt{d_i d_j}}, \qquad L = I - S. \tag{7}$$

Exactly disconnected positive-mass atoms are handled combinatorially. Otherwise, for each candidate K from large to small, the leading spectral subspace is row-normalized in its random-walk representation and partitioned with the basis-invariant cluster_qr procedure [3]; small disconnected atoms may be deflated during proposal and are reattached before validation. Every proposed label vector is then checked on the full weighted graph:

$$\widehat{\nu}(A_k) = \sum_{i \in A_k} w_i \geq \alpha, \qquad \widehat{\psi}(A_k) = \frac{\sum_{i \in A_k, j \notin A_k} w_i w_j \eta_{ij}}{\sum_{i \in A_k} w_i d_i} \leq \delta_{\text{cut}}. \tag{8}$$

The reported $\widehat{K}$ is the largest accepted candidate, or the ordinary count if none is accepted; a smooth score records the worst conductance and smallest-mass margins, but all count results use the hard test. Across affinity scales the implementation can require persistence; across horizons it uses existence, since routes can split and rejoin inside only some windows. Held-out validation freezes a training partition, assigns new paths by affinity mass and recomputes Eq. (8), recording zero-affinity paths as uncovered rather than silently assigning them. That checks whether an accepted partition survives new paths; it does not turn failure to find a partition into evidence of absence.

4 What is certified—and what is not

The method combines established conditional kernels, kernel regression and normalized spectral clustering [8, 2]; its guarantee comes from explicit validation, not from interpreting an eigengap as a route count. Exact-component nullity and the mass cap $\mathfrak{K} \leq \lfloor 1/\alpha \rfloor$ are stated and proved in Appendix A.4. Away from exact disconnection, small eigenvalues and low conductance are related only by Cheeger inequalities, which bound the balanced constant $\Phi_\ell = \inf_A \phi_\ell(A)$; the multiway bound controls $\max_k \psi_\ell(A_k)$ and loses factors polynomial in K [6]. Since the two all-block tests coincide (Section 3), nothing turns on which we validate against — and in either form an eigenvalue count is a proposal, not the estimand in Eq. (6).

Proposition 1 (Empirical certificate). Fix the weighted empirical path law the algorithm actually validated — the localized sample after any Monte Carlo resampling of the ball, not the underlying population law — and let $\widehat{\mathfrak{K}}$ be the maximum in Eq. (6) over all its partitions. If the algorithm returns $\widehat{K} = K$, then $\widehat{K} \leq \widehat{\mathfrak{K}}$.

This proposition is asymmetric: the spectral proposal family is a strict subset of all partitions, so $\widehat{K} = 1$ does not imply $\widehat{\mathfrak{K}} = 1$, and the finite-sample certificate does not by itself lower-bound a population route count. Weak consistency of the local conditional laws holds under standard kernel-regression conditions (Appendix C.6), and it is consistency of the conditional measure only; spectral convergence additionally requires vanishing maximum weight, growing effective local sample size, nondegenerate degrees, operator convergence and an isolated proposal eigenspace — and even those do not establish convergence of the accepted partition or count to a population route graph. We therefore offer it as background for the local-law estimator, not as partial theory of count recovery. Finally, time reversal is exact at the feature level: a backward window on a reversed rollout is a common column permutation of the corresponding forward feature, and the RMS-normalized Euclidean distance is invariant to it. That proves the symmetry used as a reverse-code-path test in Section 6; it is not a second statistical experiment.

5 Controlled identification experiments

The controlled suites separate structural questions that raw split accuracy conflates. We score every usable anchor and compute truth from noiseless path segments at the same declared feature resolution; unless stated, no statistic is conditioned on the detector firing.

Directed graph types and endpoint non-identification. The headline local-topology experiment scores a probe — a known event location where the reported (K^+, K^-) is the median over 14 anchors, 5,052 anchors behind 361 probes with 2 refusals. Because the locations are given rather than found, it establishes that the directed pair distinguishes the four local types, not that it finds them. The endpoint comparison is matched — both feature maps get the same deterministic anchor draw, asserted per probe (540/540) — and the path pair separates splits, merges, crossings and corridors across 60 geometry seeds where endpoints confuse merges with crossings and corridors with everything. In a split–remerge family that randomizes both event locations, the path estimator returns $K = 2$ and the endpoint estimator $K = 1$ in all 12 serialized variants (Appendix C.13, with an earlier unmatched run that overstated the endpoint gap by 84 probes).

State locality rather than a time window. The staggered family in Fig. 7 makes different visible lane groups encounter their own junctions at different phases. Across 10 seeds, time alone explains mean fraction 0.230 of the branchability entropy against 0.533 for state given time, a mean ratio of 2.318. Detection survives this construction: over all 2,380 anchors, agreement is 0.942 at three-way forward junctions (813 anchors at full valency), and 0.924 at eight-way junctions (442).

The baseline that matters here is a clock. Because branchability in a shared-phase geometry is a function of the time index, the sharpest competitor to a state-based detector is not a clustering method but the time index itself, in its strongest form: a lookup table on t, whose hypothesis class contains every time-only rule. A particular fit need not attain the class optimum, so we report four of increasing strength: opposite-parity anchors (0.720), the slice minus the scored anchor (0.752), the other seeds of the same cell family (0.782), and the scored anchors themselves (0.796), the last an upper bound on any function of t alone. The detector agrees with the noiseless truth on 0.930 of anchors, ahead of all four — including the oracle — by +0.134, and leads in all six cells of 2 valencies $\times$ 3 seeds. The comparison is not otherwise matched: the detector is evaluated on the same local paths it is fitted from, while the clock is supervised from a handful of labels per slice. What a time lookup cannot do is separate $K = 1$ from $K = 8$ within a slice, which is where the margin comes from: the eight-way staggered cells are mixed-valency, and pooled over their 714 anchors every one of the 84 three-way and all 141 eight-way junctions is counted exactly, with all 55 errors false positives on single-route anchors (Appendix C.16).

Resolvability and robustness, in brief. Two further suites are in the appendix because neither carries a localization claim. In the shared-phase chokepoint family, where branchability is essentially a function of the time index, the estimator returns exactly the full valency 1.000 of the time from valency 2 to 8 (840 pooled anchors) and at valency 12 (801), and 0.994 at valency 10; the residual is a false-positive rate of 0.091 whose 14 errors all occupy a single time slice, and regrading the same noiseless segments at the estimator's own declared scale removes every one (238/238 against 224/238) — a grader-versus-estimator threshold mismatch, not a detector failure (Appendix C.18). Under stress, null padding costs false-positive rate (0.109 to 0.216) while nuisance-active coordinates defeat the localizer outright, and rescoring frozen partitions on an independent path sample gives survival 0.844 to 1.000 (Appendix C.23).

5.1 Same-feature baselines, and what they cost this method

The necessary comparison holds the localized ball, the path features, the truth and the all-anchor metric fixed and changes only the method: a bare eigencount, a Gaussian mixture, agglomerative clustering, DBSCAN and a scalar path-dispersion statistic, each configured on a held-out tuning seed, over 14 cells and three evaluation seeds. The result does not favour us (Table 5). BLPTK is never strictly the most accurate: it ties for best in five cells and is beaten in the other nine, while dispersion is best or tied-best in six and agglomerative clustering in seven. More damaging for Section 3, it matches a bare eigencount to within 0.005 in nine of the fourteen cells, so propose-then-validate cannot be defended on accuracy. Two cells are degenerate — in both collinear 8-way cells the noiseless truth finds no branchable anchors, so the constant predictor scores 1.000 and every method can only lose; BLPTK

loses hardest there, firing where the declared scale cannot separate routes. Excluding them it beats the constant predictor in all twelve remaining cells, and in the staggered geometry — the family that passes the time-versus-state diagnostic — it is tied for best in five of six. What remains specific to BLPTK is the pair of directed counts, the explicit resolution declaration, and the feasibility certificate — outputs the competing methods do not produce and this metric does not score. Whether those are worth the machinery is a question this table leaves open, and we would rather leave it open than answer it with a number that cannot.

6 Forward split localization on D3IL-Avoid

We finally apply the detector to offline rollouts from a diffusion policy on D3IL-Avoid [5], giving it local state–time contexts and relative planar path segments but neither the engineered terminal-side label used for evaluation nor annotated split times. Scored against that withheld label, it finds the branchable regions: at horizon 8, forward K^+ reaches AUROC 0.874, TPR 0.810 at FPR 0.087 over all 300 anchors against a base rate of 0.387; at horizons 4 and 6 the AUROCs are 0.839 over 380 anchors and 0.815 over 340. That is what the real data was brought in to settle: whether these regions can be identified at all in real rollouts, without supervision. Anchors reuse trajectories, so an anchor-level interval is anti-conservative and we report the trajectory bootstrap instead: 0.865, 95% interval $[0.806, 0.915]$ (Table 7).

What the task cannot settle, and we can say so exactly. It cannot settle whether that skill is state information or a proxy for the clock. Let $p(t)$ be the fraction of trajectories at a junction at time t. A gate that is a function of t alone fires identically on every trajectory, so the best precision it can reach is $\max_t p(t)$ — and on Avoid every rollout starts from the same state, giving $\max_t p(t) = 1.000$: at the fork there is a moment when the whole population is branching. A clock is therefore optimal here by construction, and no state-aware detector could beat it. Measured, one does: a one-parameter predictor $-|t-t^\star|$, fitted on half the anchors and scored on the other half, reaches 0.879, 0.912 and 0.896 at the three horizons, above the detector at every one. That is a property of the task, not a deficiency of the estimator — the marginal number says the regions are found, not that they are found by geometry.

The confound-free measurement. Within a single time slice the clock is constant, so any separation there is state information no time predictor can supply. Only slices containing both classes admit an AUROC: at horizon 8, 11 of the 19 scored slices are mixed and carry 220 anchors, making this a conditional measurement on a label-selected subset, not an estimate over the rollout distribution. Restricted this way, forward K^+ attains a sample-weighted AUROC of 0.879 over those 220 anchors, and 0.795 and 0.814 at horizons 4 and 6. The two numbers answer different questions: the marginal one says the branchable regions are recovered, this conditional one says the recovery is not a clock in disguise — on the mixed slices of real rollouts the detector separates branchable from non-branchable contexts using local path geometry alone, at a fixed time. Being conditional on a label-selected subset, it does not establish junction localization across the real rollout distribution. Reversing each trajectory and running the K^- code path reproduces the forward count on all 204/204 matched balls, which validates that path but is the exact feature symmetry of Section 4 rather than independent merge evidence (Appendix C.5).

7 An environment where a clock is not optimal

Avoid's junction is time-locked ($\max_t p(t) = 1.000$), which is why Section 6 can only report a conditional measurement. This section supplies a task whose junction timing varies across rollouts, measures the detector against the strongest time-only rule, and then puts its output to work as a constraint.

The environment and its screens. A 7-DoF KUKA LWR reaches past an obstacle whose position along the path varies per episode. The demonstrations are solved as a two-stage trajectory nonlinear program [15], so both routes are kinematically realisable rather than drawn in task space and tracked. The environment is MDP-native: the policy observes $[q, \dot{q}, p_{\text{ee}} - p_{\text{obs}}]$ with no clock, and the detector localises on $[p - p_{\text{obs}}, \dot{p}]$. Every gate below was declared before the environment was tuned to meet it: the clock ceiling is $\max_t p(t) = 0.410$ against a gate of 0.50, the fork time has mean 20.0 and standard deviation 4.7 within a 40-step episode, and state explains 0.814 of the branchability entropy against 0.263 for time alone. Appendix C.4 has the remaining screens and the 2.6× perturbation amplification that makes a joint-space feature map worth comparing at all.

Detection, against an oracle clock. We score all 204 usable anchors at horizon $h = 20$; below $h = 16$ the one-nearest-neighbour route purity of the forward segments falls under its gate, so shorter horizons are not usable here and we do not report them as detector failures. At the declared scale $\ell = 1.64$ the detector agrees with the noiseless truth on 0.971 of anchors, against 0.672 for the trivial always-one predictor and 0.711 for the oracle clock — the per-timestep majority truth, an upper bound on every rule that is a function of t alone — a margin of +0.260 over that bound. This is the first environment here where a clock is not optimal by construction, and the reason the environment exists. As in Section 5 the comparison is not otherwise matched: the detector is evaluated on the same local paths it is fitted from, while the clock is handed the per-slice majority of the truth.

Against the same-feature baselines, each at its own oracle configuration. We ran the five methods of Table 5 on the arm's own anchors and features, each at its best configuration on the evaluation data — an upper bound no deployment reaches — against our single pre-declared $\ell = 1.64$. The bare eigencount, which matched us on the toys, does not here: 0.788 against 0.972, and unstable across seeds (0.824 down to 0.750) where we are not. A Gaussian mixture reaches 0.650, below the trivial always-one predictor's 0.670. But agglomerative single linkage at $1.0u$ and DBSCAN at $\varepsilon = 1.0u$ reach 0.959 and 0.961 — within 0.013 and 0.011 of us, and within 0.016 even tuned. We declared before running that a method-specific claim needed 0.05 over every baseline's oracle, and that gate fails. Those two are the exact-disconnection rule without the spectral proposal or the conductance test: at this resolution the routes are nearly disconnected and a linkage threshold suffices (Appendix C.20).

7.1 Freezing certified regions as a diversity constraint under fine-tuning

We freeze the detector's output — the anchors, their memberships and their base branch masses — and use it to constrain reinforcement fine-tuning of a diffusion policy behaviour-cloned on the same demonstrations, under DPPO [16] with the policy-gradient ratio formed per denoising transition.

The task must be able to punish hedging. A point obstacle that appears only in the observation makes route choice free, so we place a thin wall in the plane separating the two routes, by a rule declared before any fine-tuning run: it clears the unchanged demonstrations by 10.2 mm and blocks the straight trunk on 100% of synthetic paths. The reward is the tube term minus a collision indicator, a clearance term and the excess jerk at replan boundaries; seven calibration checks ran first, and it does not prefer a route on the demonstrations ($z = 0.97$). Appendix C.7 has the rule, the rates, the checks and the reward.

Unconstrained fine-tuning deletes a route, in every seed. Trained to a declared asymptote (tail slope below 2×10^{-4} per iteration), plain DPPO ends with every one of 400 evaluation rollouts on a single route in all five seeds: on-route M_{eff} is 1.000, collisions are 0%, the tube error 0.0031 to 0.0037 is below the demonstrations' own scatter, and the reward rises from about −3.0 to between −0.53 and −0.66. Four seeds keep route 0 and one route 1. Two controls say what this collapse is and is not. Permuting the advantages across episodes — identical update count, magnitudes, clipping and schedule, no information about behaviour — never commits but destroys the policy, leaving 0 to 17.5% of rollouts on any route and

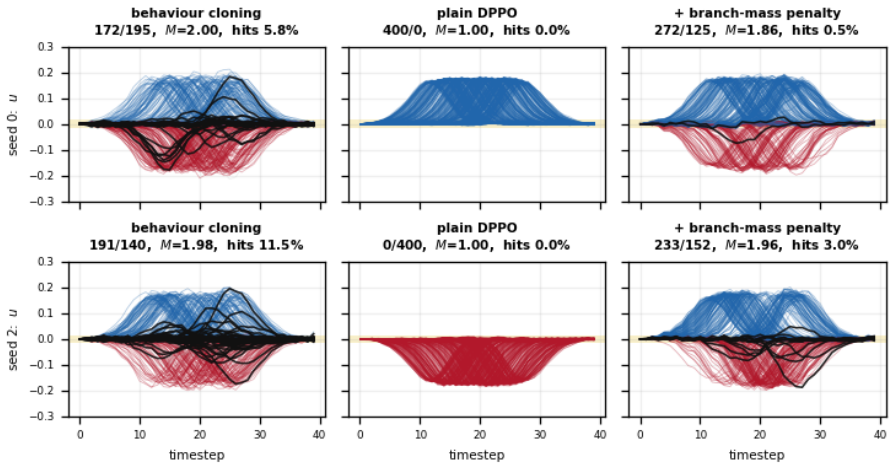

Figure 1: A frozen certified partition keeps the branch that fine-tuning deletes. Two of the five seeds; the branch coordinate u separates the two ways round the obstacle and the band is the wall. Behaviour cloning uses both routes and collides with the wall; plain DPPO ends entirely on one route with no collisions; the branch-mass penalty keeps both routes with collisions still near zero. In seed 2 (lower row) the deleted route returns as the majority. Counts are on-route rollouts out of 400.

7 to 40% colliding, so commitment is not update noise. And changing only the fine-tuning sampling stream, from the same cloned policy, flips the surviving route in one seed, so the side is broken by the stream rather than fixed by the initial policy, though nine of the ten runs chose the same route, consistent with the small route preference the cloned policy's own rollouts already carry (Appendix C.7).

The constraint, and what it buys. Over the frozen anchors the penalty is a branch-mass KL between each anchor's frozen base masses and the current policy's. A rollout the detector never saw gets a soft block membership by lifting the frozen labels through the same affinity, and the gradient is a score-function surrogate with detached clipped importance ratios — nothing differentiates through the detector or the simulator (Appendix C.4). Only the mass on each certified block is constrained, never the law within a block, so refinement inside a route is permitted while deleting one is not, and because memberships and base masses are frozen the policy cannot reduce the penalty by moving the detector. We set λ once at iteration 0 so the penalty gradient norm is a stated fraction — the dose — of the policy-gradient norm; unscaled it was 88× the objective's. At the primary dose of 1.0, both routes survive in all five seeds, with M_{eff} between 1.84 and 1.99 and the minority route holding 30 to 44% of rollouts, against 1.000 and 0% for plain fine-tuning on the same seeds. Criteria declared before the run — $M_{\text{eff}} \geq 1.5$, on-route ≥ 0.85, collisions $\leq 2\%$, an asymptoted and improved reward and a tube error below the cloned policy's — pass in four seeds; the fifth fails on collisions alone, at 3.0%. The policies keep 56 to 63% of plain fine-tuning's reward gain and tighten the tube from the cloned policy's 0.0162–0.0178 to 0.0088–0.0096, so fine-tuning still happens inside each route. Across doses, a third of the policy gradient's scale loses the branch (M_{eff} 1.16–1.80, 1/5) and three times it loses the task (0.06–0.42 of the gain kept, diverging in two seeds, 1/5); the primary dose is the knee (Table 4).

Matched controls make the certified regions load-bearing. At the same declared dose and gate, a clock-where control — fixed-time global balls with the same spectral proposal and validation — passes 0/5 seeds, as does analytic KL-to-BC restricted to the fork-window decisions, against 4/5 for BLPTK. The KL result is the cleanest separation: all five policies asymptote, improve, stay on-route, meet the collision bar and tighten the tube, yet their on-route M_{eff} is only 1.03–1.48 and the rendered minority bundles hold just 2–52 of 400 rollouts, so proximity to the cloned policy does not account for branch preservation. A privileged control using hard true-route labels and each rollout's causal pre-fork action history passes

2/5; that does not rank learned regions above truth, because its global route mass uses a high-variance self-normalised score-function estimator (Appendix C.7).

This is a proof of concept on one simulated task, five seeds and one geometry: a frozen certified partition is usable as a diversity constraint, preventing a mode deletion that unconstrained fine-tuning produces in every seed, at a stated cost in reward and a dose near the policy gradient's own scale. It is not evidence that certified partitions generally solve route deletion. The matched controls rule out two specific alternatives, not the established diversity-preserving methods — trajectory-level entropy or mutual-information objectives, mixture policies, behaviour-regularised RL — against which we do not compare; the clock control's accepted-partition coverage differs from the state-localised penalty; and the reward, neutral on the demonstrations, favours one route by about 4% under the cloned policy's own rollouts, so part of the collapse is reward-optimal (Appendix C.7).

8 Related work

Pairing a forward with a reverse conditional law has precedent we position against rather than claim: computational mechanics pairs predictive with retrodictive causal states [10], predictive state representations define state by multi-step future predictions [11], and transition-path theory is built on forward and backward committors with occupancy-weighted reactive trajectories [12]. Spectral clustering of trajectory representations with automatic cluster-number selection [13] and velocity-aware spectral clustering for unsupervised identification of dynamical systems [14] both predate this work, transition-manifold methods embed a state by its future transition density [1], and finite-time transfer operators extract slowly mixing sets from transport [4]. Our components are likewise standard: kernel regression for the local conditional measure [7, 9], the normalized path graph and its exact-component nullity [8], multiway Cheeger inequalities [6], and cluster_qr for a basis-invariant proposal from a degenerate subspace [3].

A separate line creates branches rather than detecting them — multiplicity as an environment primitive [17], online branch-budget allocation at critical decision states [18, 19], multimodal generative trajectory policies [20], and latent skill-step branching [21]. We take the rollouts as given and ask where, and in which direction, they already divide (Appendix D.1).

Against that background our claim is narrow and compositional: local conditional laws on path segments rather than states or endpoints, an explicit forward/reverse multiplicity read as a directed local graph type, and a count accepted only after an explicit mass-and-conductance feasibility check. We claim no new transition operator, no coherent-set decomposition and no new statistical theorem: the certificate is the elementary observation that a validated feasible partition lower-bounds a maximum defined by the same constraints. Mixture models, hierarchical clustering, density clustering, eigencounts and dispersion statistics are all plausible competitors once the same local path features exist, and so — wherever junctions are time-locked — is the time index itself (Section 5.1).

9 Limitations and conclusion

BLPTK is a local, declared-resolution detector whose strongest statement is constructive: an accepted partition witnesses a feasible empirical split or merge count, while a failure to fire shows neither that the empirical maximum is ordinary nor that no population junction exists. Our synthetic grader is itself a resolution choice, not the estimator's, and accounts for the entire residual error in the orthogonal arm (Table 2); a new domain would need the label-free protocol of Appendix C.17, which we state but have not validated. The real experiment's junction is nearly time-locked, so only its within-slice measurement is attributable to geometry, and rollout-varying timing is demonstrated only in simulation. The same-feature baselines are unfavourable (Appendix C.20). Within those limits, bidirectional conditional path laws express the missing directed object, and propose–then–validate turns local route multiplicity into an auditable lower bound.

Reproducibility statement

(Recommended by the venue; it does not count toward the page limit.)

Every number in this paper carries a source comment in the LaTeX naming the result file and the key path it was read from, and no reported number is computed inside a figure script. The detector and its gates are reproduced with nested_toy/blptk_avoid_confound.py for the real-data measurement; the arm fine-tuning with nested_toy/blptk_arm_regsweep.py and the predeclared gates in nested_toy/blptk_arm_commit_curve.py and nested_toy/blptk_arm_reg_gate.py. Gates, success criteria and protocol changes were recorded before the runs that used them.

Two limits on exact reproduction should be stated rather than discovered. Reinforcement fine-tuning is bitwise deterministic given the seed, the code and the frozen partition it starts from, because the training entry point seeds the global generator; the freezing step does not, so re-running it draws a different behaviour-cloned sample and lands on a different anchor set — regenerating one seed's frozen detector gives 110 to 132 anchors where the committed file has 100. The frozen partitions every reported run consumed are therefore kept as artifacts rather than regenerated, and results reproduce from them. Separately, the real-data driver was revised after avoid_detect.json was written and no longer regenerates its schema. Neither limit affects a detector measurement: the detection experiments draw no policy rollouts at all (Appendix D.2).

AI use statement

(This section is required by the venue and does not count toward the page limit.)

Generative AI tools were used substantially in this work, under author direction and with author responsibility for every claim.

Writing. A large language model produced the first draft of this manuscript from a prepared evidence directory and a written outline. Every numerical claim in that draft carries a source comment in the LaTeX naming the result file and key it came from, and the authors re-derived those values from the raw files independently; the drafting model also identified one error in our own prior summary (a trial count recorded as 40 that the stored results show to be 12).

Implementation and analysis. A coding assistant was used to write and modify the estimator, the experiment drivers and the analysis scripts, and to run them. All reported numbers come from serialized result files rather than from model output, and the analysis scripts that produce them are released with the paper.

Review. An independent language model, given repository access and no prior context, was used adversarially to attack the draft. It located errors that we then corrected, including a false mathematical claim about the relationship between two conductance normalizations and several overstatements of the valency results. Those corrections are reflected in the present text.

Generative AI tools were not used to generate research ideas, to fabricate or impute any experimental result, or to produce any figure from anything other than serialized data. We have reviewed all AI-assisted work and take responsibility for the final content of this paper, including its text, claims and artifacts.

References

[1] A. Bittracher, P. Koltai, S. Klus, R. Banisch, M. Dellnitz, and C. Schütte. Transition manifolds of complex metastable systems: theory and data-driven computation of effective dynamics. Journal of Nonlinear Science, 2018.

[2] R. R. Coifman and S. Lafon. Diffusion maps. Applied and Computational Harmonic Analysis, 2006.

[3] A. Damle, V. Minden, and L. Ying. Simple, direct and efficient multi-way spectral clustering. Information and Inference, 2019.

[4] G. Froyland. An analytic framework for identifying finite-time coherent sets in time-dependent dynamical systems. Physica D, 2013.

[5] X. Jia, D. Blessing, X. Jiang, M. Reuss, A. Donat, R. Lioutikov, and G. Neumann. Towards diverse behaviors: a benchmark for imitation learning with human demonstrations. International Conference on Learning Representations, 2024.

[6] J. R. Lee, S. Oveis Gharan, and L. Trevisan. Multi-way spectral partitioning and higher-order Cheeger inequalities. ACM Symposium on Theory of Computing, 2012.

[7] E. A. Nadaraya. On estimating regression. Theory of Probability and its Applications, 1964.

[8] U. von Luxburg. A tutorial on spectral clustering. Statistics and Computing, 2007.

[9] G. S. Watson. Smooth regression analysis. Sankhyā, Series A, 1964.

[10] C. J. Ellison, J. R. Mahoney, and J. P. Crutchfield. Prediction, retrodiction, and the amount of information stored in the present. Journal of Statistical Physics, 136(6):1005–1034, 2009.

[11] M. L. Littman, R. S. Sutton, and S. Singh. Predictive representations of state. In Advances in Neural Information Processing Systems, 2001.

[12] W. E and E. Vanden-Eijnden. Towards a theory of transition paths. Journal of Statistical Physics, 123(3):503–523, 2006.

[13] F. Porikli. Learning object trajectory patterns by spectral clustering. In IEEE International Conference on Multimedia and Expo, 2004.

[14] B. Fichera and A. Billard. Linearization and identification of multiple-attractor dynamical systems through Laplacian eigenmaps. Journal of Machine Learning Research, 23(294):1–35, 2022.

[15] C. E. Mower, J. Moura, N. Zamani Behabadi, S. Vijayakumar, T. Vercauteren, and C. Bergeles. OpTaS: An optimization-based task specification library for trajectory optimization and model predictive control. In IEEE International Conference on Robotics and Automation (ICRA), 2023.

[16] A. Z. Ren, J. Lidard, L. L. Ankile, A. Simeonov, P. Agrawal, A. Majumdar, B. Burchfiel, H. Dai, and M. Simchowitz. Diffusion policy policy optimization. arXiv:2409.00588, 2024.

[17] Y. Du and W. Chen. Branching reinforcement learning. ICML, 2022.

[18] S. Xing, S. Wang, C. Yang, X. Dai, and X. Ren. Lookahead tree-based rollouts for enhanced trajectory-level exploration in reinforcement learning with verifiable rewards. ICLR, 2026. arXiv:2510.24302.

[19] Spark: strategic policy-aware exploration via dynamic branching for long-horizon agentic learning. ACL, 2026. arXiv:2601.20209.

[20] X. Feng, L. Tang, C. Wang, and H. Chen. Offline reinforcement learning with generative trajectory policies. arXiv:2510.11499, 2025.

[21] D. Kim, M. Yoo, and H. Woo. Offline policy learning via skill-step abstraction for long-horizon goal-conditioned tasks. IJCAI, 2024.

A Formal details

A.1 Endpoint non-identification

Let $\Gamma_{t,h}$ be a path-segment space and $e_h(\gamma) = \gamma(t+h)$. Choose a centre path γ_0 and two separated paths γ_+, γ_- with common start x and endpoint y. The path laws $\mathbb{Q}_1 = \delta_{\gamma_0}$ and $\mathbb{Q}_2 = (\delta_{\gamma_+} + \delta_{\gamma_-})/2$ are distinct and have different local path multiplicity, yet $(e_h)_\# \mathbb{Q}_1 = (e_h)_\# \mathbb{Q}_2 = \delta_y$. Therefore no functional of the endpoint push-forward can identify every split that re-merges before the horizon.

A.2 Reverse path law from disintegration

Let $P^+_{t,t+h}$ be the forward context transition kernel and ρ_t the rollout occupancy. The joint transport measure is

$$J_{t,h}(dc, dc') = \rho_t(dc) P^+_{t,t+h}(c, dc'). \tag{9}$$

Disintegrating with respect to the second coordinate gives $J_{t,h}(dc, dc') = \rho_{t+h}(dc') P^-_{t,h}(c', dc)$, yielding Eq. (16) where densities exist. The backward path kernel is the push-forward of this conditional history law through Ψ^-. It is data-occupancy dependent even when the forward dynamics are fixed.

A.3 Derivation of weak local-law consistency

For a bounded continuous f, write $m_f(c') = \mathbb{E}[f(Y^\pm) \mid C_t = c']$. The numerator and denominator of

$$\frac{\sum_i \kappa_{\epsilon_N}(c, C^i_t) f(Y^\pm_i)}{\sum_i \kappa_{\epsilon_N}(c, C^i_t)} \tag{10}$$

are empirical averages whose expectations are kernel convolutions of $m_f(c')\rho_t(c')$ and $\rho_t(c')$. After the change of variables $u = (c' - c)/\epsilon_N$, continuity gives their normalized limits. The effective local sample condition makes both averages concentrate, and Slutsky's theorem gives the conditional expectation at c. A countable convergence-determining class plus tightness yields weak convergence in probability. Dependent windows require a mixing condition or rollout-level splitting.

A.4 Proofs of the certificate propositions

The statements are in Section 4.

Exact components. For each block, $f_k(y) = \mathbf{1}_{A_k}(y)\sqrt{d_{\nu,\ell}(y)}$ is an eigenfunction of the normalized affinity operator with eigenvalue one. The functions have disjoint support. Irreducibility makes that eigenvalue simple within each block, leaving exactly their span as the Laplacian nullspace.

Empirical certificate. The returned labels partition the empirical sample and every block has passed precisely the mass and conductance inequalities defining $\widehat{\mathfrak{K}}$. They witness feasibility at K. The unrestricted maximum is therefore at least K.

Mass cap. The blocks are disjoint, each has mass at least α, and their masses sum to one.

B Implementation details

The code uses relative path features: the anchor state is subtracted before flattening the segment. Euclidean distance is divided by the square root of feature length, making a declared distance comparable across horizons. Optional arc-length resampling instead walks outward from the anchor in either direction and samples at equal distance increments; a trajectory shorter than the declared arc is clamped at its last observed point. Feature

normalization, an arc horizon when used, and a scale unit are declared on a training split, never fitted per query.

The implementation localizes on the full population before subsampling. If the neighborhood is too large, it draws with replacement according to the kernel weights and aggregates multiplicities; if a fixed-radius ball has insufficient support, it refuses the anchor. An adaptive k-nearest-neighbor mode instead declares the context resolution by k.

The experiments call the dense certified-count path in nested_toy/blptk.py. Its steps are:

1. Form relative forward or backward path features, using only declared feature coordinates, and RMS-normalize pairwise distance by flattened feature length.
2. Declare one path-distance scale unit on a training subset. Multiply it by the predeclared ℓ at every anchor in that slice.
3. Localize on the complete context population with a Wendland radius or adaptive nearest-neighbor neighborhood. If subsampling is necessary, sample with replacement and aggregate multiplicity weights.
4. Build the compact-support path affinity and weighted normalized Laplacian in Eq. (7).
5. Enumerate feasible coarsenings when the graph has only a small number of exact atoms. Otherwise deflate sub-mass atoms for the proposal, extract each candidate spectral subspace with cluster_qr, reattach deflated mass, and validate on the full graph.
6. Return the largest candidate for which every block passes Eq. (8). Flag multiple feasible coarsenings as partition ambiguity even when the count agrees.

The default certified path does not accept or reject a partition from an eigenvalue threshold. Eigenvectors generate candidates; empirical mass and volume-normalized conductance decide.

Resolution declaration used in the round-two sweeps. The serialized driver uses horizon 20, context radius 0.12, relative path scale 2.5, conductance tolerance 0.005, angular perturbation 0.45, and context band 0.20 for the headline valency rows. The mass gate is adjusted with target valency by the experiment driver; this avoids capping a requested count before data are observed.

C Additional quantitative results

C.1 Round-two valency sweep, decomposed

Agreement over all anchors is the headline metric, but on its own it is saturated in the orthogonal arm and conflates two different behaviours in the crowded arms. We therefore give it beside the size of the full-valency conditioning set, exactness on that set, and the false-positive rate. Exactness is undefined, not perfect, where the conditioning set is empty.

C.2 Localizing the constant false positives

The orthogonal arm returns the identical integer 224/238 in every cell and every seed. We re-ran single cells keeping each anchor's label to find out where the 14 errors sit. They are one whole time slice, at every valency and both re-run seeds; the corridor slices and the post-junction slices are exact.

C.3 Held-out validation table

C.4 The arm environment and the branch-mass constraint, in full

Screens, with every gate declared before the environment was tuned to meet it:

Table 1: Forward round-two valency sweep at context band 0.20, pooled over 10 geometry seeds. "Exact" is $\mathbb{P}(\widehat{K} = n \mid K_{\text{true}} = n)$ and is shown only where its conditioning set is nonempty. The orthogonal arm is exact at every valency; what degrades in the crowded arms is the conditioning set and the false-positive rate, not the count on the routes that are resolved.

embedding	valency	agreement	scored	full-valency	exact	FPR
orthogonal	2	0.941	2,380	840	1.000	0.091
orthogonal	4	0.941	2,380	840	1.000	0.091
orthogonal	8	0.941	2,380	838	1.000	0.091
orthogonal	10	0.939	2,380	827	0.994	0.091
orthogonal	12	0.940	2,380	801	1.000	0.092
half-orthogonal	2	0.895	2,297	669	1.000	0.148
half-orthogonal	4	0.883	2,364	678	1.000	0.163
half-orthogonal	8	0.697	2,351	138	1.000	0.174
half-orthogonal	10	0.646	2,344	0	—	0.213
half-orthogonal	12	0.647	2,334	0	—	0.212
collinear	2	0.941	2,378	840	1.000	0.091
collinear	4	0.881	2,354	696	1.000	0.169
collinear	8	0.629	2,137	0	—	0.371
collinear	10	0.618	1,941	0	—	0.382
collinear	12	0.613	1,927	0	—	0.387

Table 2: Per-slice errors for the orthogonal cell at valency 8, seed 0. Truth is the noiseless single-linkage count at the absolute threshold SEP = 0.35; the estimator uses $\ell \cdot u$, the declared scale from a data quantile. Only slice $t = 60$ disagrees, and it disagrees in the direction the mismatch predicts. Regrading the same segments at the estimator's scale removes every error in all eight re-run cells (238/238 against 224/238), which is what identifies the residual as a threshold mismatch rather than a detector failure; it is not an independent accuracy measurement. Reproduce with nested_toy/blptk_fp_slice.py.

time slices	anchors each	truth	estimate
8, 12, 16	14	8	8
20–56	14	1	1
60	14	**1**	**8**
64, 68, 72	14	8	8

Table 3: Partition survival is conditioned on training acceptance; agreement is over all anchors. Each cell is run over five geometry seeds and pooled, with ± the standard error over seeds — the seed redraws the routes, so it is independent in a way anchors within one geometry are not. Held-out scoring lowers agreement in the two clean orthogonal cells and raises it in the two crowded collinear ones: it changes the decision rather than uniformly improving it.

cell	anchors	accepted	survival	agreement in	agreement out
orthogonal 3	1,190	490	1.000 ± 0.000	0.941	0.941
orthogonal 4	1,190	490	1.000 ± 0.000	0.940	0.940
orthogonal 6	1,190	490	0.988 ± 0.002	0.935	0.930
orthogonal 8	1,190	490	0.927 ± 0.012	0.915	0.893
collinear 4	1,187	490	0.937 ± 0.017	0.875	0.894
collinear 6	1,142	412	0.844 ± 0.038	0.638	0.696

- clock ceiling $\max_t p(t) = 0.410$ against a gate of 0.50. The fork time has mean 20.0 and standard deviation 4.7 within a 40-step episode, spanning steps 11 to 28, so branchability is not a function of the time index;
- state explains 0.814 of the branchability entropy, against 0.263 for time alone;
- the routes leave at 89.9° and re-merge to a worst terminal error of 0.014;

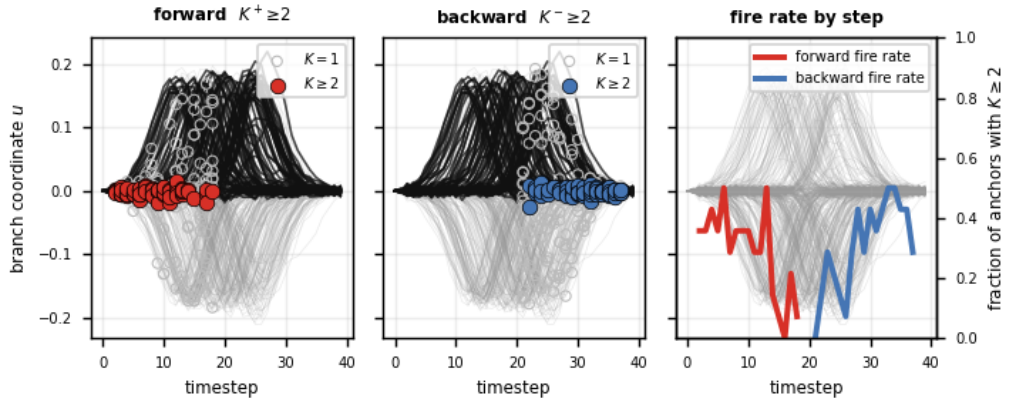

Figure 2: Both directions on arm rollouts, with no route labels and no fork times. Forward $K^+ \geq 2$ fires while a fork is still open, backward $K^- \geq 2$ where the routes have re-joined, and the two fire rates separate in time (right). The obstacle position varies per episode, so these events are not at a fixed step; the detector localises on the end effector relative to the obstacle. Declared $\ell = 1.64$, $h = 20$, $\alpha = 0.25$ on 400 rollouts.

- 0/300 IK failures and manipulability within $[0.065, 0.272]$, so neither route passes near a singularity.

The trajectory program also matters for the ablation below: it amplifies a 1 cm task-space perturbation by 2.6× in joint space, against 115× for a greedy per-timestep solve. Without that, a joint-space feature map would be comparing noise.

The prediction registered before this run was that a task-space path feature map is flat under a posture nuisance that a joint-space map cannot quotient. Injecting null-space posture noise of standard deviation σ_{p}, with each map read at its own declared scale, task space holds 0.971 across the first three levels and 0.966 at the fourth, while joint space falls to 0.559 — below the trivial always-one predictor. Reading both at one shared ℓ instead showed joint space failing at every nuisance level, which was the scale, not the feature map:

σ_{p}	0.00	0.03	0.15	0.30
task space ($\ell = 1.64$)	0.971	0.971	0.971	0.966
joint space ($\ell = 1.18$)	0.966	0.951	0.657	0.559
trivial always-one	0.672	0.672	0.681	0.676

Let $\mathcal{A}$ be the frozen anchors, ω_a the localisation weight of anchor a, s_a^+ its forward score, $r_{a,i,k}$ the frozen membership of rollout i in block k of a's certified partition, and $\bar{\pi}_{0,a,k}$ its base mass. With $w_{a,i}$ the localisation weights of the current rollouts and $\rho_{a,i}$ a clipped importance ratio over a's own decision window,

$$R_{\mathrm{graph}}(\theta) = \sum_{a \in \mathcal{A}} \omega_a \, s_a^+ \sum_k \bar{\pi}_{0,a,k} \log \frac{\bar{\pi}_{0,a,k}}{\bar{\pi}_{\theta,a,k}}, \qquad \bar{\pi}_{\theta,a,k} = \frac{\sum_i w_{a,i} \, \rho_{a,i} \, r_{a,i,k}}{\sum_i w_{a,i} \, \rho_{a,i}}, \tag{11}$$

and the fine-tuning objective is the policy-gradient surrogate minus $\lambda R_{\mathrm{graph}}$.

What $r_{a,i,k}$ is for a rollout the detector has never seen. The rollouts i are drawn from the current policy, so they carry no label from the frozen partition and need an out-of-sample rule. Membership is the Nadaraya–Watson lift of the frozen cluster indicators through the same Wendland affinity η_ℓ that built the graph, at the frozen absolute scale:

$$r_{a,k}(y) = \frac{\sum_{j \,:\, z_{a,j} = k} w_{a,j}^{\mathrm{ref}} \, \eta_\ell\big(y, \, Y_{a,j}^{\mathrm{ref}}\big)}{\sum_j w_{a,j}^{\mathrm{ref}} \, \eta_\ell\big(y, \, Y_{a,j}^{\mathrm{ref}}\big)}, \tag{12}$$

so memberships are soft, lie on the simplex, and are consistent with the components they came from — unlike a nearest-centroid rule, which would misassign non-convex blocks. A

query path with no affinity to any reference path falls back to the uniform $1/K$, which is the conservative choice: an uncovered rollout contributes equally to every block and so cannot be used to satisfy the constraint preferentially. Frozen for the whole run are the reference features Y_a^{ref}, their weights, the labels z_a, K, ℓ, the kernel, the anchor set with its context scale, and the base masses π_0; nothing in Eq. (12) depends on θ.

How the gradient is taken. It is not taken through Eq. (12). There is no pathwise differentiation through the simulator, the detector or the membership: $\bar{\pi}_{\theta,a,k}$ depends on θ only through which paths are sampled, so the update is a score-function surrogate on the decision log-probabilities within the anchor's own window,

$$\widehat{\nabla}_\theta R_{\mathrm{graph}} = -\sum_a s_a^+ \, o_a \sum_i \tilde{w}_{a,i} \left[\sum_k (r_{a,i,k} - \bar{\pi}_{\theta,a,k}) \, \frac{\bar{\pi}_{0,a,k}}{\bar{\pi}_{\theta,a,k}} \right] \nabla_\theta \log \pi_\theta(\mathrm{window}_a), \tag{13}$$

where $(r - \bar{\pi})$ is the self-normalisation correction, $\tilde{w}_{a,i} \propto w_{a,i}\rho_{a,i}$ are the normalised weights, and o_a is the fraction of current rollouts with nonzero localisation weight at a. The importance ratio $\rho_{a,i}$ is detached and clipped to $[0, 10]$ over that window, correcting the reuse of one PPO batch across epochs exactly as the main objective is corrected. The bar in $\bar{\pi}$ denotes the stabilised mass $(\pi + \varepsilon/K)/(1 + \varepsilon)$ with $\varepsilon = 0.05$, applied to both arguments so the KL stays finite when a block empties. Anchors are subsampled (24 per iteration) and an anchor with $o_a < 0.05$ is skipped, the policy having left that region entirely.

Only the mass on each certified block is constrained, never the law within a block, so refinement inside a route is permitted while deleting a route is not. Because the memberships and base masses are frozen and the gradient does not flow through them, the policy cannot reduce the penalty by moving the detector — only by putting branch mass back. We set λ once, at iteration 0, so that the penalty gradient norm is a stated fraction — the dose — of the policy-gradient norm; at $\lambda = 1$ unscaled the penalty gradient was 88× the objective's.

C.5 Reverse code-path validation on Avoid

For implementation validation, we reverse each trajectory and run the K^- code path. At horizon 6 the reversed score has AUROC 0.837, TPR 0.700, and FPR 0.038 over 360 anchors. More importantly, on the same anchor and local ball, the reversed K^- count equals the original K^+ count on all 204/204 matched cases. This equality is the exact feature symmetry proved in Section 4. The reversed run contains no new trajectories and therefore no independent statistical evidence for natural merges. The real-data claim of this paper is forward split localization at a fixed time, on one task whose junction is time-locked.

C.6 Weak consistency statement

Proposition 2 (Weak consistency of the local laws). Assume independent rollout windows; a context kernel κ that is nonnegative, bounded, integrable and of nonzero integral, applied as $\kappa_\epsilon(c, \cdot) = \kappa((c - \cdot)/\epsilon)$; an interior context point with positive continuous density; continuity of $c' \mapsto \mathbb{E}[f(Y^\pm) \mid C_t = c']$ for every bounded continuous f; uniform tightness nearby; $\epsilon_N \downarrow 0$; and $N\epsilon_N^{d_c} \to \infty$. Then

$$\int f \, d\widehat{\mathbb{K}}^\pm_{t,h,\epsilon_N}(c) \xrightarrow{p} \int f \, d\mathbb{K}^\pm_{t,h}(c). \tag{14}$$

The proof is the standard ratio-of-kernel-averages argument. This is weak consistency of the conditional measure only. Spectral convergence additionally requires vanishing maximum weight, growing effective local sample size, nondegenerate degrees, operator convergence, and an isolated proposal eigenspace. Even those conditions do not establish convergence of the accepted partition or count to a population route graph, especially where path support is a union of intersecting strata.

C.7 The wall rule and reward calibration

Wall capture rates, the KL match, and the privileged control. The selected wall catches 59% of late departures and 47% of mid-passage switches in the synthetic scan used to choose

Table 4: Dose response, five seeds per row, everything else identical. The dose is the penalty gradient norm as a fraction of the policy-gradient norm at iteration 0. Ranges are over seeds. Every column is a property of the regularised run at that dose: "on-route rate" is the fraction of the 400 evaluation rollouts lying on either demonstrated route, "on-route M_{eff}" the effective number of routes among those, "minority share" the fraction of on-route rollouts on the less-used route, and "gain kept" the fraction of paired plain fine-tuning's reward gain retained — so the plain row is 1.00 by definition, marked (ref). "passing" counts seeds meeting all five criteria declared before the runs. A third of the policy gradient's scale loses the branch; three times it loses the task, diverging in two seeds.

dose	on-route rate	on-route M_{eff}	minority share	collisions	gain kept	passing
0.0 (plain)	1.000	1.000	0.000	0%	1.00 (ref)	— collapse
0.3	0.993–1.000	1.16–1.80	0.03–0.27	0–0.5%	0.65–0.80	1/5
1.0	0.963–0.993	1.84–1.99	0.30–0.44	0.5–3.0%	0.56–0.63	4/5
3.0	0.895–0.968	1.85–1.99	0.30–0.45	1.5–5.8%	0.06–0.42	1/5

it. Because analytic KL-to-BC has zero gradient at the cloned policy, its predeclared dose match is taken on a discarded clone after one PPO step; the trained policy still starts at BC. The privileged control uses hard true-route labels and each rollout's causal pre-fork action history; besides the 2/5 seeds that pass, two more retain both routes but miss the collision or convergence criteria. Unlike the local soft memberships, its hard global route mass is estimated by a self-normalised score function, so it is privileged in labels but is not an empirical upper bound on optimisation reliability. Across the ten plain and stream-varied runs, nine chose the same route, consistent with the preference the cloned policy's own rollouts already carry ($+0.11$, $z = 2.03$, because cloning tracks that route slightly more precisely). The null control is also a partial one: it degrades the policy rather than isolating reward maximisation from optimisation pressure, which would need a trust-region-anchored variant applied to both arms.

The task must be able to punish hedging. A point obstacle that appears only in the observation makes route choice free: a policy can drift toward one route and still switch. We therefore place a thin wall in the plane that separates the two routes, chosen by a rule declared before any fine-tuning run: among candidate walls, take the one whose first-percentile clearance on the unchanged demonstrations exceeds 5 mm and which maximises the smaller of the late-departure and mid-passage-switch capture rates. The selected wall clears the demonstrations by 10.2 mm, blocks the straight trunk on 100% of synthetic paths, and catches 59% of late departures and 47% of mid-passage switches. Because the demonstrations are unchanged, the behaviour-cloned policies, the frozen detectors and every detector number above carry over untouched.

The reward is the pointwise tube term minus a collision indicator, a clearance term inside a 5 mm margin, and the excess jerk at replan boundaries. Seven calibration checks were run before any fine-tuning: the demonstrations lose 0.024 to the constraint terms, a colliding rollout loses more than one standard deviation of the tube reward, and the reward does not prefer a route on the demonstrations ($z = 0.97$).

C.8 The directed local-topology experiment, in full

C.9 Dose response, in full

C.10 Resolution: the declared window and the decomposition

BLPTK answers a declared question. For route features Ψ, the affinity scale must be large enough to connect within-route variation and small enough not to connect distinct routes:

$$\text{within-route spread under } \Psi \;<\; \ell \;<\; \text{smallest between-route separation under } \Psi. \tag{15}$$

The embedding, path horizon, context neighborhood, mass gate, and effective sample size all affect whether this window is populated.

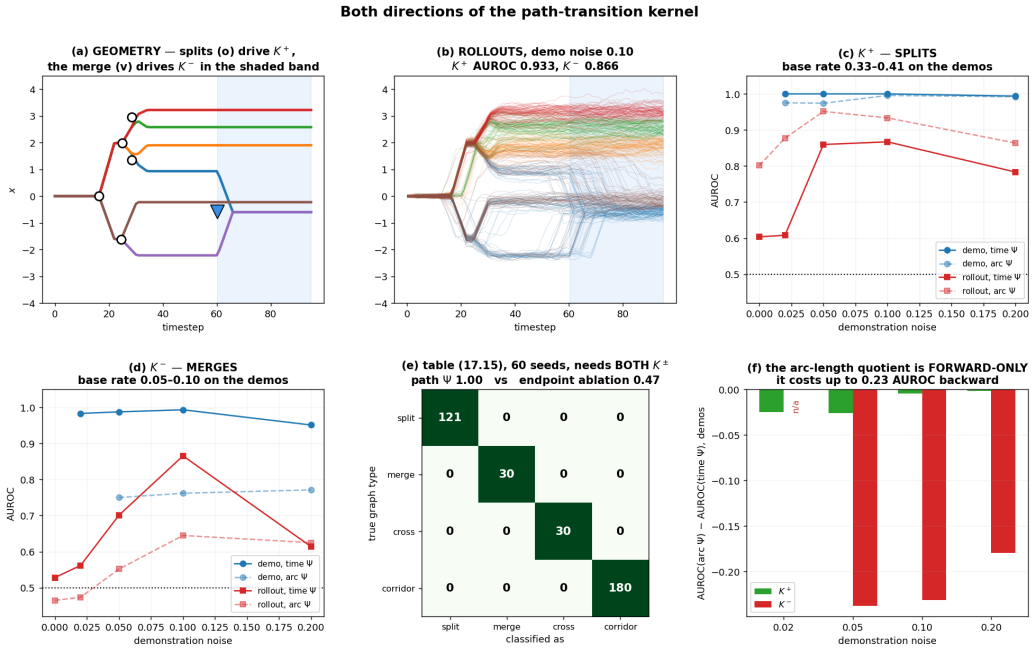

Figure 3: Directed local route multiplicity. The forward path law responds to splits and the reverse path law to merges. On the directed-graph scan, the path-valued pair classifies 361/361 probes: 121 splits, 30 merges, 30 crossings, and 180 corridors. A probe is placed at a known event location and reports the median (K^+, K^-) over 14 anchors (5,052 anchors in total, of which 2 were refused), so this experiment measures classification given the location, not localization. Replacing each path by its endpoint, on the same anchors and the same local balls, yields 255/361 correct — but only 15/30 merges and 92/180 corridors. The endpoint result is a matched feature-map ablation — both maps score identical anchors, asserted on all 540 probes — and not a same-feature clustering baseline.

The agreement number is saturated; the decomposition is the result. All-anchor agreement alone is misleading here, and its own constancy is what exposes that. Through valency 8 every orthogonal cell returns the identical integer 224/238 in all 10 geometry seeds, with true-positive rate 1.000 and false-positive rate $14/154 = 0.091$; at valencies 10 and 12 the constancy breaks, with 5 and 2 seeds respectively at 223/238. An error rate that barely moves with the geometry is not sampling noise, so we decompose it. On the anchors whose noiseless truth is the full valency, the estimator returns exactly that valency 1.000 of the time from valency 2 to 8 (840 pooled anchors, 838 at valency 8) and at valency 12 (801 anchors); at valency 10 it is $822/827 = 0.994$. The residual is therefore false positives except at valency 10, where five full-valency anchors are also miscounted, and Table 2 localizes the false positives: in the re-run cells all 14 occupy a single time slice. That slice is one grid step before the noiseless truth re-declares multiplicity, and the cause is a resolution mismatch between grader and estimator rather than a detector failure. The truth label thresholds path separation at an absolute SEP = 0.35, whereas the estimator is given $\ell \cdot u$, the declared scale times the unit returned by a data quantile. We test that explanation rather than assert it, by regrading the same noiseless segments at the estimator's own scale: every error disappears, in all eight re-run cells, at valencies 2, 4, 8 and 12 and both seeds, giving 238/238 against 224/238. The disagreement is therefore entirely accounted for by the threshold difference. This is not an independent accuracy measurement — regrading at the estimator's own scale cannot fail in the estimator's disfavour — so we report both numbers. What it does show is that grading a declared-resolution estimator at an undeclared resolution manufactures a fixed 5.9% error, which is the inconsistency this paper argues against.

The embedding sweep, the two-sided scale test and the finite-sample edge are in Appendix C.18, with Fig. 4.

C.11 Two further propositions

Proposition 3 (Exact components). Suppose $\mathcal{Y} = A_1 \sqcup \cdots \sqcup A_K$ up to ν-null sets, each block has positive ν-measure, the degree $d_{\nu,\ell}$ is positive ν-almost everywhere, cross-block affinity vanishes, and the affinity operator restricted to each block is irreducible in the sense that no block splits further into two sets of positive measure with zero affinity across them. Then the normalized population Laplacian has nullity exactly K.

Proposition 4 (Mass cap). $\mathfrak{K}_{\ell,\alpha,\delta_{\mathrm{cut}}}(\nu) \le \lfloor 1/\alpha \rfloor$.

C.12 The eigencount at its own oracle configuration

Against a bare eigencount at its own oracle configuration. Table 5 showed a bare eigencount matching this method on the toys. Here it does not. Giving the eigencount its best (ℓ, δ) cell on the same anchors and the same local balls, it reaches 0.824 in task space against the detector's declared 0.971 (the detector's own oracle scale would give 0.975), a gap of +0.152; in joint space the gap is +0.275, and at $h = 16$ it is +0.163. We report the oracle configuration rather than a fixed threshold because a tight threshold makes the eigencount refuse to split, which is exactly the failure the detector would then be credited for avoiding.

These are single-seed figures, measured on evaluation seed 0 before the full baseline suite of Appendix C.20 existed, and the two reconcile. Over that suite's three evaluation seeds the same comparison gives 0.788 against 0.972, a wider gap of +0.185, because the eigencount is unstable across seeds (0.824, 0.789, 0.750) where the declared configuration is not (0.971, 0.975, 0.971). The three-seed figure is the one Section 7 quotes; the single-seed one is kept here because the joint-space and $h = 16$ cells were only ever run at one seed.

C.13 The directed local-topology experiment, in detail

Directed graph types and endpoint non-identification. Figure 3 is the headline local-topology experiment. Its unit is a probe: a known event location at which the reported (K^+, K^-) is the median over 14 anchors, 5,052 anchors behind 361 probes with only 2 refusals. Medians over 14 anchors are more robust than the single-anchor scores used everywhere else in this paper, and the probe locations are given rather than found, so this experiment establishes that the directed pair distinguishes the four local types — not that it finds them. The placement is moreover type-dependent: splits are probed slightly before the event, merges after it, and crossings 72% through the convergence transient. The endpoint comparison is matched: both feature maps are handed the same deterministic anchor draw, so they score identical anchors on identical local balls, which we assert per probe rather than assume (540/540). This matters: an earlier unmatched run, in which the two passes consumed each other's random draws, scored endpoints at 171/361 and so overstated the gap by 84 probes. The full path pair separates splits, merges, crossings, and corridors across 60 geometry seeds, whereas endpoints confuse merges with crossings and ordinary corridors with all non-corridor types. A separate split–remerge family randomizes the two event locations. In all 12 serialized variants, the path estimator returns $K = 2$, the endpoint estimator returns $K = 1$, and every trial passes this predicted contrast.

The noise ladder and the fork-free null are in Appendix C.16.

C.14 Endpoint non-identification and the reverse law

Why endpoints cannot identify the target. Let e_h evaluate a path at its endpoint. A law concentrated on one centre path and a law split equally between two disjoint paths can have the same start and endpoint; both push forward through e_h to the same endpoint law. Hence no endpoint statistic can distinguish them in general. The randomized diamond experiment in Section 5 instantiates this obstruction. Even a complete forward endpoint kernel asks only which futures leave c. The incoming law obeys the Bayes reversal

$$p^-_{t,h}(c', c) = \frac{\rho_t(c) p^+_{t,h}(c, c')}{\rho_{t+h}(c')}, \tag{16}$$

when densities exist. It depends on the occupancy ρ as well as the dynamics, so one row of a forward kernel does not identify incoming multiplicity.

C.15 The two block tests coincide

Both are volume-normalized, and they are not the same quantity. The balanced form ϕ_ℓ is the Cheeger conductance, and it is the one the inequalities of Section 4 are stated for; our bipartition sweep minimizes it. The one-sided form ψ_ℓ is the escape probability of a single step of the induced reversible walk out of A, it is the quantity appearing in higher-order multiway bounds, and it is what the acceptance test applies to every block. Pointwise $\psi_\ell \le \phi_\ell$, with equality on any block carrying at most half the total volume. The two all-block tests are nonetheless equivalent at every K:

Proposition 5 (The two block tests coincide). For any partition $\{A_k\}_{k=1}^K$ of $\mathcal{Y}$, $\max_k \phi_\ell(A_k) = \max_k \psi_\ell(A_k)$.

Proof. At most one block, say A_b, has $\mathrm{vol}_{\nu,\ell}(A_b) > \frac{1}{2}\mathrm{vol}_{\nu,\ell}(\mathcal{Y})$; for every other block the two agree. It therefore suffices that $\phi_\ell(A_b) \le \max_{k\neq b} \psi_\ell(A_k)$. Since $A_b^c = \bigsqcup_{k\neq b} A_k$ and $Q_\ell(A_b, A_b^c) = \sum_{k\neq b} Q_\ell(A_b, A_k) \le \sum_{k\neq b} Q_\ell(A_k, A_k^c)$, the mediant inequality gives

$$\phi_\ell(A_b) = \frac{Q_\ell(A_b, A_b^c)}{\mathrm{vol}_{\nu,\ell}(A_b^c)} \le \frac{\sum_{k\neq b} Q_\ell(A_k, A_k^c)}{\sum_{k\neq b} \mathrm{vol}_{\nu,\ell}(A_k)} \le \max_{k\neq b} \psi_\ell(A_k). \qquad \square$$

So the choice between them is one of exposition, not of strictness: we write the estimand with ψ_ℓ because that is the multiway quantity and what the code computes, and keep ϕ_ℓ only because the Cheeger inequalities of Section 4 are stated for it.

C.16 Noise ladder and fork-free null

The noise ladder tests the two directions and a fork-free null. At demonstration noise 0.05, forward time-parametrized paths have AUROC 1.000, TPR 1.000, and FPR 0.000 over 267 anchors; the reverse run has AUROC 0.988, TPR 1.000, and FPR 0.021 over 295 anchors. The matched fork-free null fires at rate 0.000 in both directions at that rung.

Counts, not just detections. The eight-way staggered cells are mixed-valency: the truth takes the values 1, 3, and 8 as an anchor sees a group fork or the full junction. Pooled over their 714 anchors, every one of the 84 three-way junctions and all 141 eight-way junctions is counted exactly, no anchor is detected as branchable but assigned the wrong count, and all 55 errors are false positives on single-route anchors. A time lookup is not representationally incapable of emitting 3 and 8 — in these cells the two occur at different slices, so a table could in principle separate them. What it cannot do is separate $K = 1$ from $K = 8$ within a slice, which is where the detector's margin comes from. Reproduce with nested_toy/blptk_clock_baseline.py.

C.17 Choosing the resolution without labels, and when to abstain

Every experiment in this paper declares $(h, \epsilon, \ell, \alpha, \delta_{\mathrm{cut}})$ before scoring, and the synthetic truth is used only to evaluate. That is not a deployment procedure, because a new domain has no noiseless route truth and no branchability label. The protocol we would follow there uses no labels at any step.

1. Fix the physical quantities first. The horizon h is the duration over which a route distinction is meant to be visible, and the context scale ϵ is the neighbourhood within which contexts are meant to be exchangeable. Both are properties of the question, chosen before any path graph is built, and both are reported.
2. Set the affinity unit from the data, not the median. Take ℓ in units of a low quantile ($q = 0.1$) of pairwise path distances in the local ball. At a fork the median lands in the empty gap between routes and silently selects a scale that cannot separate them.

3. Scan a declared range rather than a point. Report $\widehat{K}$ across a range of ℓ fixed in advance, and accept a count only where it is stable across adjacent scales as well as passing the mass and conductance tests at that scale. A count that exists at exactly one scale is not reported as a detection.
4. Set α from the smallest route worth asserting, noting that it caps resolvable valency at $\lfloor 1/\alpha \rfloor$ (Appendix A.4); set δ_{cut} from the separation the application treats as a real gap. Both are declarations about what counts as a route, not tuning knobs.
5. Abstain rather than report when no stable accepted interval exists, when the local ball is refused for insufficient occupancy, or when out-of-sample coverage is poor — that is, when a held-out path sample lands outside the affinity support of the fitted partition. The uniform fallback of Eq. (12) makes poor coverage visible rather than silently assigning mass.
6. Report the resolution with the count. A count without its declared triple is not interpretable, and no number in this paper should be read as scale-free.

We have not validated this protocol on a domain where truth is unavailable, which is the honest statement: it is the procedure the paper's own declarations instantiate, not a rule shown to transfer. Figure 4 shows how sharply the answer moves with affinity scale and sample size, which is the reason the declaration, the stability requirement and the abstention rule are all necessary rather than decorative.

C.18 Resolution sweeps, in full

What these sweeps measure, and what they do not. They run on the shared-phase chokepoint family, not the staggered one of Section 5. There branchability is essentially a property of the time index — the noiseless truth takes a single value within every scored time slice at valencies 2, 4 and 8, and at valency 12 only 3 of 17 slices are mixed — so a clock would separate branchable from non-branchable anchors and these sweeps are not evidence that the detector localizes junctions. What they measure is the count returned at a junction and the width of the resolvable window, questions a clock cannot answer at all; localization rests on the staggered family of Section 5, on the within-slice real-data measurement of Section 6, and on Section 7. The declared question is two-sided — the affinity scale must be large enough to connect within-route variation and small enough not to connect distinct routes (Eq. (15)) — and the embedding, horizon, neighborhood, mass gate and sample size all decide whether that window is populated.

Embedding, not a universal valency ceiling. At context band 0.20 across 10 seeds, mean all-anchor orthogonal-route agreement is 0.941 across 2,380 total anchor evaluations at valency 2 and 0.940 across 2,380 at valency 12. Over the same endpoints of the sweep, mean collinear agreement falls from 0.941 across 2,378 total anchor evaluations to 0.613 across 1,927; the mean for the half-orthogonal embedding moves from 0.895 across 2,297 total anchor evaluations to 0.647 across 2,334. This is consistent with Eq. (15): adjacent collinear routes crowd while orthogonal separation stays fixed. It does not establish a general embedding law.

Where the crowded embeddings actually fail. Full-valency exactness stays at 1.000 in the crowded embeddings too, so quoting it alone would be the conditioning trap of Section 2. Three things degrade instead, and all three must be reported. First the conditioning set: in the half-orthogonal embedding the number of full-valency anchors falls from 669 at valency 2 to 0 at valencies 10 and 12, and the collinear set is already empty at valency 6. Second the false-positive rate, which rises from 0.148 to 0.212 in the half-orthogonal embedding and from 0.091 to 0.387 in the collinear one. Third — and invisible to all of the above — miscounts on branchable anchors whose truth is neither 1 nor the full valency. Full-valency exactness cannot see them by construction. At half-orthogonal valency 8 there are 396 such miscounts beside 316 false positives, and at valency 10 all 420 branchable anchors are miscounted. So crowding does not merely stop the declared resolution from separating routes; where it partially separates them the count is wrong, and a report of exactness-plus-FPR alone would still conceal that. Table 1 gives the decomposition including this term.

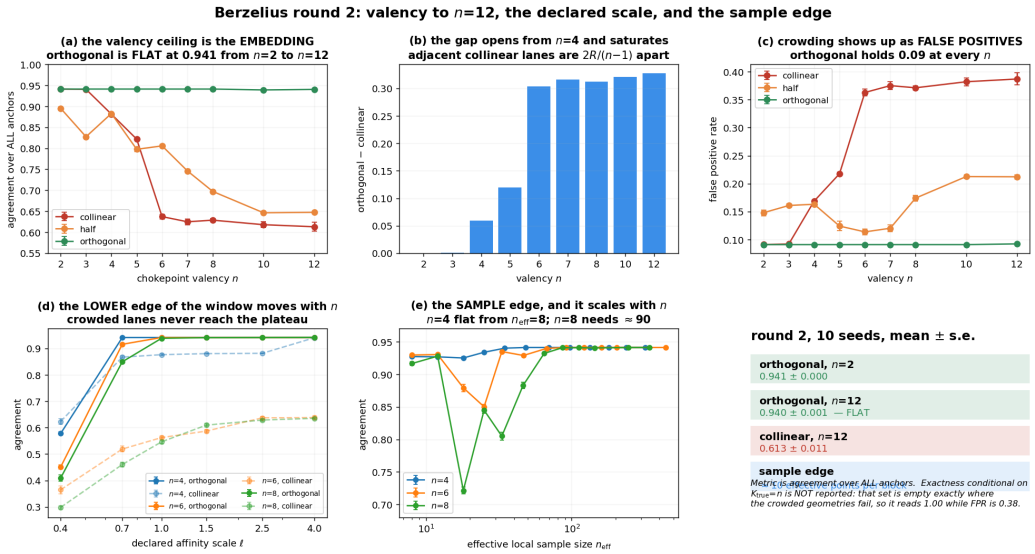

Figure 4: Valency, affinity scale, and finite-sample edge. Means and standard errors use the round-two runs and all-anchor agreement. The plot deliberately omits conditional exactness: in crowded cells the set $K_{\text{true}} = n$ can be empty.

Both sides of the declared scale matter. For orthogonal four-way paths, mean agreement rises from 0.578 at $\ell = 0.4$ to 0.941 at $\ell = 0.7$, with 2,380 anchors at each scale. For orthogonal eight-way paths the corresponding values are 0.409, 0.849, and 0.938 at $\ell = 0.4$, 0.7, and 1.0, again over 2,380 anchors per scale. The measured collinear eight-way sweep never exceeds 0.635 agreement over 2,137 anchors, and its full-valency conditioning set is empty at every recorded scale. That last fact is why conditional-on-full-valency exactness would hide rather than diagnose the failure.

The finite-sample edge. For eight-way paths at resample budget 18, all-anchor agreement is 0.722 over 2,856 anchors, while only 3 anchors have full eight-way truth. At budget 110, mean effective sample size is 86.4, agreement is 0.941 over 2,856 anchors, and the full-valency set contains 912 anchors. Four-way paths already score 0.927 at effective size 8 over 2,748 anchors, with a full-valency set of 607. These curves are merely consistent with roughly ten effective observations per block in this family; they do not define a general sample threshold.

C.19 Same-feature baselines, in full

BLPTK is never strictly the most accurate cell. It ties for the best score in five cells and is beaten in the other nine. A scalar dispersion statistic is best or tied-best in six cells and agglomerative clustering in seven. On raw all-anchor agreement over the same features, this method is not the one to use.

The certificate rarely changes the eigencount. More damaging for the construction of Section 3: BLPTK matches a bare eigencount to within 0.005 in nine of the fourteen cells, and where propose-then-validate does move the answer the gain is small (staggered 4-way forward 0.961 against 0.944, 6-way forward 0.943 against 0.926, collinear 8-way reverse 0.798 against 0.778). Validation is what makes $\widehat{K}$ a certificate rather than a guess, and that status is not something agreement can reward — but neither can we claim it buys accuracy here.

Where the comparison is degenerate, and where it is real. In both collinear 8-way cells the noiseless truth finds no branchable anchors at the declared resolution, so the constant predictor $\widehat{K} \equiv 1$ scores 1.000 and every method can only lose. BLPTK loses hardest (0.625 and 0.798): it fires most where the declared scale cannot separate routes, which is the failure already measured in Appendix C.18. Excluding those two, BLPTK beats the

Table 5: Same-feature baselines. Agreement $\mathbb{P}(\widehat{K} = K_{\text{true}})$ over all anchors, mean over three evaluation seeds, each method configured on a held-out tuning seed. "const" is the trivial predictor $\widehat{K} \equiv 1$; a cell where it scores 1.000 contains no branchable anchors at the declared resolution and can only be lost. BLPTK is never strictly best; it ties for best in five cells, all staggered.

cell	const	BLPTK	eigen	GMM	agglo	DBSCAN	disp
collinear 4 +	0.703	0.939	0.938	0.597	0.941	0.942	0.936
collinear 4 −	0.841	0.943	0.933	0.537	1.000	0.947	0.988
collinear 8 +	1.000	0.625	0.620	0.119	0.926	0.773	0.976
collinear 8 −	1.000	0.798	0.778	0.121	0.945	0.870	1.000
orthogonal 4 +	0.647	0.941	0.941	0.706	1.000	0.941	0.980
orthogonal 4 −	0.842	0.947	0.947	0.575	0.946	0.947	0.999
orthogonal 8 +	0.647	0.941	0.941	0.569	1.000	1.000	0.971
orthogonal 8 −	0.842	0.947	0.947	0.287	0.946	0.947	1.000
staggered 3 +	0.653	0.937	0.937	0.866	0.957	0.957	0.966
staggered 3 −	0.699	1.000	1.000	0.882	1.000	0.971	0.965
staggered 4 +	0.685	0.961	0.944	0.824	0.961	0.961	0.860
staggered 4 −	0.694	1.000	1.000	0.855	1.000	1.000	0.947
staggered 6 +	0.699	0.943	0.926	0.745	0.943	0.943	0.899
staggered 6 −	0.733	0.960	0.960	0.776	0.960	0.960	0.930

constant predictor in all twelve remaining cells. The one place the comparison is both non-degenerate and favourable is the staggered geometry — the family that passes the time-versus-state diagnostic — where BLPTK is tied for best in five of six cells, losing only the three-way forward cell to dispersion (0.937 against 0.966).

What this does and does not settle. It settles that local route multiplicity, on these toys, is largely recoverable by generic clustering of the same path features, and that our contribution cannot be argued from predictive accuracy. What remains specific to BLPTK is the pair of directed counts, the explicit resolution declaration, and the feasibility certificate — outputs the competing methods do not produce and which this metric does not score. Whether those are worth the machinery is a question this table leaves open, and we would rather leave it open than answer it with a number that does not say so.

C.20 Same-feature baselines on the arm, in full

The suite of Table 5 rerun on the arm, using that file's committed fairness protocol unchanged so that the selection rule predates these results: every method sees the identical localised ball, path features and declared scale unit at each anchor, distances are taken on $Y/\sqrt{p}$ so a threshold is numerically the same quantity as ℓu, and an anchor enters only where the full estimator could report. All four seeds give exactly 204 anchors with none skipped, and the trivial always-one predictor scores 0.670 at a base rate of 0.330 — the same anchor set Section 7 reports on. tuned selects each method's configuration on a held-out tuning seed and applies it unchanged; oracle selects on the evaluation seeds themselves and is not achievable in practice.

method	configurations	tuned	oracle
BLPTK, declared $\ell = 1.64$	—	—	**0.972**
BLPTK, own grid	8	0.974	0.974
DBSCAN	15	0.958	0.961
agglomerative	18	0.956	0.959
dispersion	4	0.884	0.897
bare eigencount	72	0.788	0.788
Gaussian mixture	22	0.650	0.650

Three things are worth drawing out. The declared configuration costs almost nothing against our own oracle (0.972 against 0.974), so the pre-declaration is not what loses the comparison. The winning baselines are agg_single_1.0 and db_1.0_0.08 — single linkage and density

connection at one scale unit, with the same minimum-mass filter — which is the exact-disconnection branch of Section 3 stripped of the spectral proposal and the conductance test. And this comparison is cleaner than the toy one in a way that counts against us: on the toys the truth is itself a noiseless single-linkage count at a threshold, so agglomerative's showing there is partly the grader's own algorithm applied to noisier data, whereas the arm's truth is the route-label mass count and carries no such confound.

The gate, declared before the run, required the declared configuration to beat every baseline's oracle by 0.05; it fails on both. We report that rather than adding cells or re-tuning, and what remains specific to the method on this task is the directed pair, the declared resolution, the feasibility certificate, and the frozen partition of Section 7.1 — none of which this metric scores.

C.21 Same-feature baselines at an oracle configuration

Table 5 reports each method with its configuration chosen on a held-out tuning seed. Table 6 instead gives every method the configuration that maximizes its own score on the evaluation cell, an upper bound no deployable procedure attains. The ordering does not improve for us: BLPTK remains never strictly best, and dispersion reaches 1.000 on both collinear eight-way cells — where, as noted in Section 5.1, the truth contains no branchable anchors and never firing is the correct answer.

Table 6: Oracle configurations: each method tuned on the cell it is scored on. An upper bound for every column, reported so that the tuned comparison in Table 5 cannot be read as a tuning artifact.

cell	BLPTK	eigen	GMM	agglo	DBSCAN	disp
collinear 4 +	0.939	0.938	0.597	0.942	0.942	0.980
collinear 4 −	0.943	0.933	0.537	1.000	0.947	0.988
collinear 8 +	0.625	0.620	0.119	0.926	0.773	1.000
collinear 8 −	0.798	0.778	0.130	0.945	0.870	1.000
orthogonal 4 +	0.941	0.941	0.706	1.000	0.941	1.000
orthogonal 4 −	0.947	0.947	0.575	0.947	0.947	1.000
orthogonal 8 +	0.943	0.941	0.569	1.000	1.000	1.000
orthogonal 8 −	0.947	0.947	0.287	0.947	0.947	1.000
staggered 3 +	0.947	0.937	0.866	0.957	0.957	1.000
staggered 3 −	1.000	1.000	0.882	1.000	1.000	1.000
staggered 4 +	0.961	0.957	0.824	0.961	0.961	0.903
staggered 4 −	1.000	1.000	0.855	1.000	1.000	1.000
staggered 6 +	0.943	0.926	0.745	0.943	0.943	0.922
staggered 6 −	0.960	0.960	0.776	0.960	0.960	1.000

C.22 Avoid horizon sweep

Table 7: All-anchor metrics on the forward Avoid split, ground truth mode_sep. The time-reversed run is a unit test of the K^- code path, not independent evidence: reversal is an exact feature symmetry, and the reversed count matches the forward one on 204/204 matched balls. Uncertainty at $h = 8$ from the trajectory bootstrap is $[0.806, 0.915]$; resampling anchors instead gives $[0.833, 0.915]$, which is anti-conservative because anchors reuse trajectories.

run	scored anchors	AUROC	TPR	FPR
forward, $h = 4$	380	0.839	0.729	0.069
forward, $h = 6$	340	0.815	0.660	0.049
forward, $h = 8$	300	0.874	0.810	0.087

C.23 High-dimensional failure modes and held-out validation, in full

Null padding, nuisance-active coordinates, and fibers. High ambient dimension is not one phenomenon, and the three cases fail in different ways. Genuinely constant padding leaves

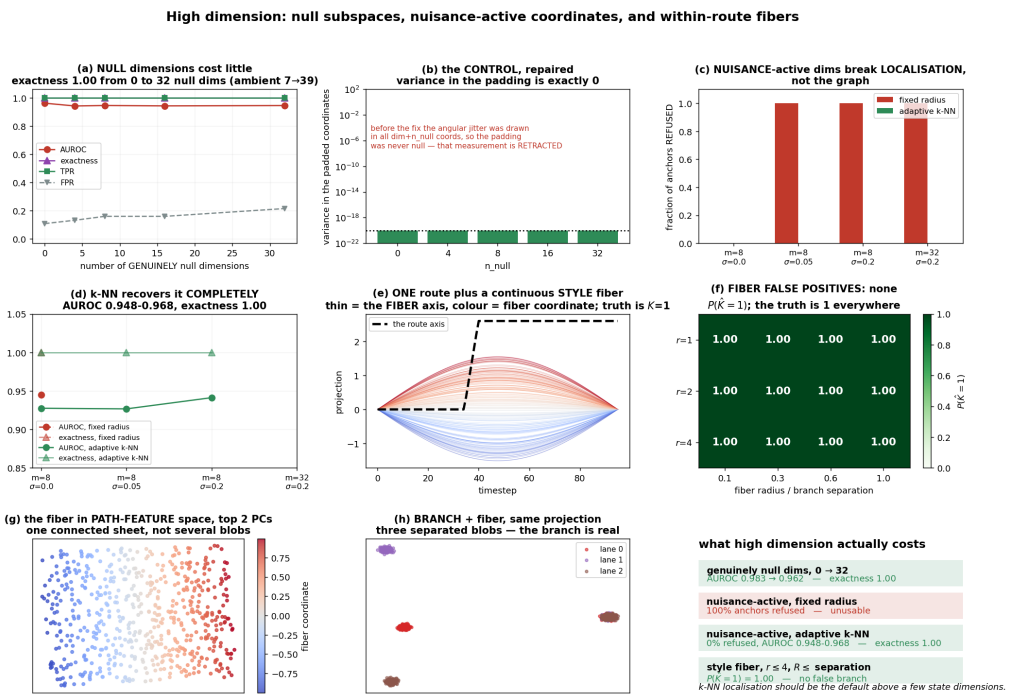

Figure 5: Different high-dimensional failure modes require different diagnoses. The repaired null control leaves padded coordinates constant. Nuisance-active context coordinates can empty a fixed-radius neighborhood, while adaptive localization reaches the graph. A within-route style fiber is not subdivided in the tested range.

path distances unchanged and costs nothing in exactness or recall: full-valency exactness and TPR are both 1.000 from 0 to 32 null dimensions. It is not free, however — the false-positive rate rises monotonically from 0.109 to 0.216 across that range, and we report it rather than calling null dimensions costless. Nuisance-active context coordinates instead concentrate distances and defeat the localizer before any path graph is formed: a fixed radius refuses every tested anchor once $\sigma \geq 0.05$. The adaptive k-nearest-neighbor branch reaches the graph in all four cases (0 refused) and holds exactness 1.000 with AUROC 0.927–0.941 at 8 nuisance dimensions. At 32 nuisance dimensions with $\sigma = 0.2$ it shows nothing at all: no anchor retains branchable truth, so AUROC, TPR and exactness are undefined there and only the vacuous FPR $= 0$ survives. We flag that cell rather than let 0 refusals read as success. A continuous within-route style fiber is never subdivided: $\mathbb{P}(\widehat{K} = 1) = 1.000$ at every one of the 12 combinations of fiber rank $r \in \{1, 2, 4\}$ and amplitude $R \in \{0.1, 0.3, 0.6, 1.0\}$. These are finite-range stress tests, not claims that nuisance can only harm localization or that fibers can never create false components.

Held-out paths. In-sample validation certifies an empirical partition on the paths used to propose it. We therefore freeze each accepted partition, assign an independent path sample by affinity mass, and recompute block mass and conductance. Each cell is now run over five geometry seeds, and the seed is the unit of uncertainty: redrawing the routes is independent in a way that anchors within one geometry are not. Survival ranges from 0.844 ± 0.038 in the crowded collinear six-way cell to 1.000 ± 0.000 in the orthogonal three- and four-way cells, over 412–490 accepted partitions and 1,142–1,190 scored anchors per cell. Coverage is 1.00 in every cell.

Held-out scoring lowers all-anchor agreement in two of the six cells: from 0.935 to 0.930 for the orthogonal six-way cell and from 0.915 to 0.893 for the orthogonal eight-way cell. It raises agreement in the two crowded collinear cells, from 0.875 to 0.894 and from 0.638 to 0.696. Thus held-out validation changes the decision rather than uniformly improving it.

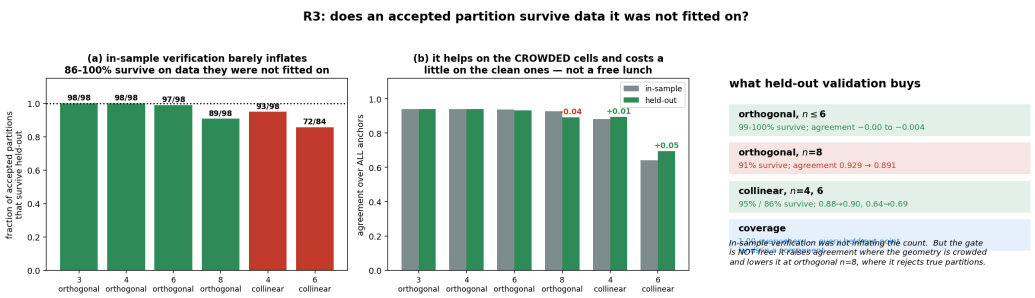

Figure 6: Held-out survival is high but not free. Bars on the left are conditioned on accepted training partitions and display their denominators. Agreement on the right is over every anchor, so a disappearing conditioning set cannot look successful.

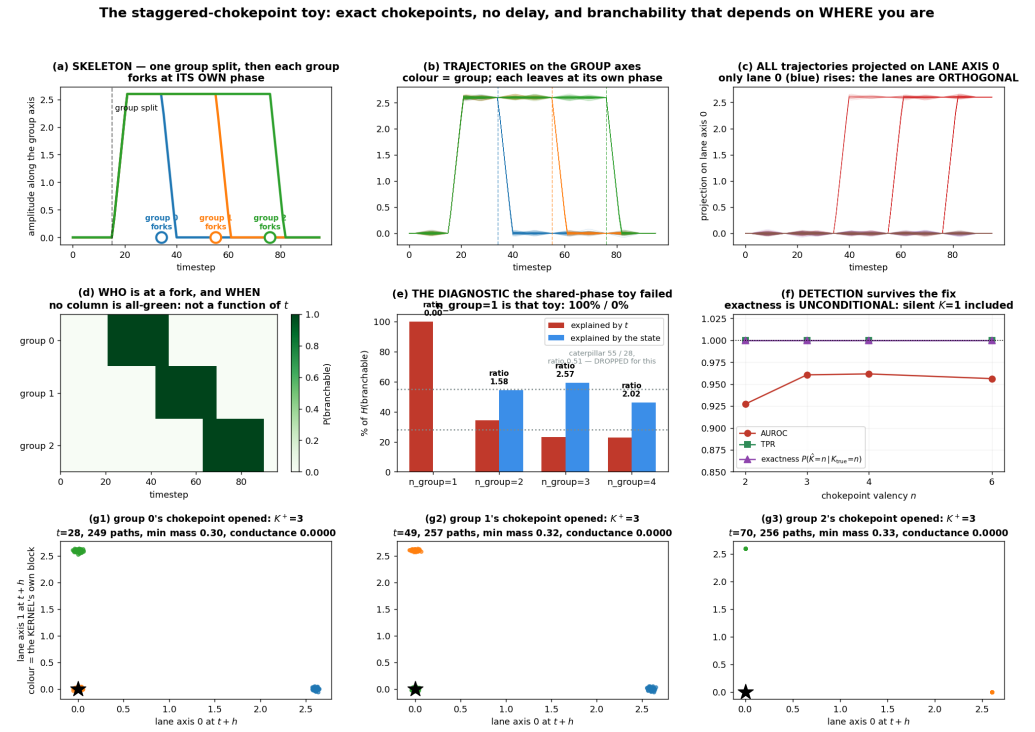

Figure 7: State-local staggered junctions. Lane groups reach exact chokepoints at different phases, so branchability is not a time-only label. The lower panels show the same all-anchor protocol and example accepted partitions. Conditional exactness is shown only where its conditioning set is nonempty; round-two all-anchor results are reported in the text.

The test uses new paths for the same geometry; it does not establish generalization to new geometry seeds.

D Additional controlled visualizations

D.1 Detecting branches versus creating them

A growing line of work makes trajectory branching something a method produces, and it is worth saying precisely how that differs from what we do. Branching RL treats multiplicity as a property of the environment: an agent takes several base actions at a state and the trajectory is a tree by construction [17]. Lookahead tree rollouts branch at high-uncertainty generation steps and prune by simulated similarity, and policy-aware dynamic branching

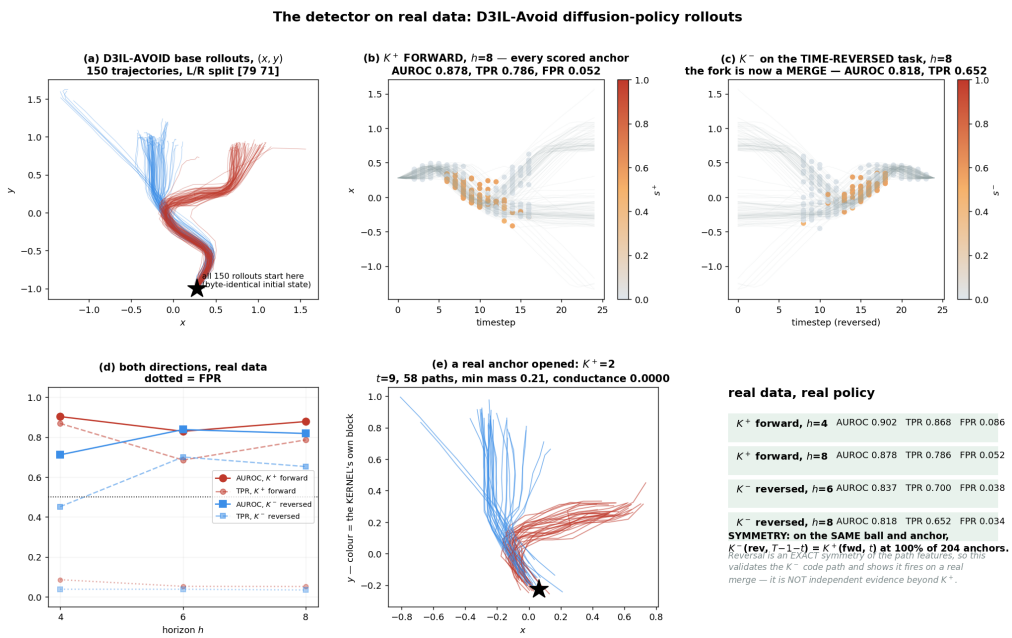

Figure 8: Real forward split localization and an exact reversal check. Forward scores concentrate around the visible route separation. The reversed run turns that fork into a merge only by transforming the same data, so it validates implementation symmetry and is not independent merge evidence.

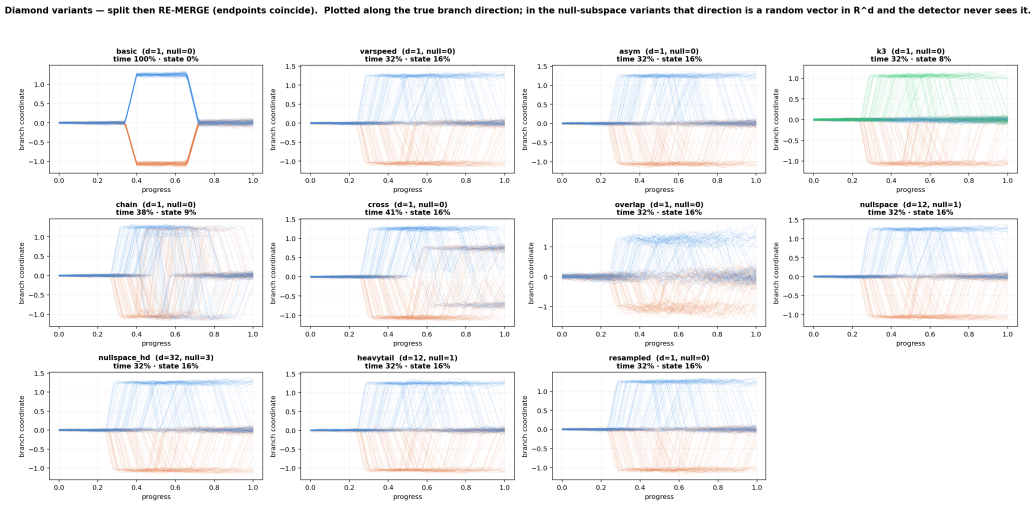

Figure 9: Randomized split–remerge geometries used to instantiate endpoint non-identification. The serialized detector comparison contains 12 trials.

spends a sampling budget at states the agent's own signals mark as critical [18, 19]; both decide where to spend exploration online, given the ability to re-sample. Generative trajectory policies model multimodal offline behaviour as a whole [20], and skill-step abstraction branches logged trajectories in a learned latent space for goal-conditioned control [21]; both represent or exploit multiplicity without localizing the contexts at which it appears or reporting a directed count.

Our setting is the complement: the rollouts are fixed and already collected, no re-sampling is possible, and the question is where — and in which direction — they divide. That output is in principle the input those methods need, a set of candidate expansion points with a

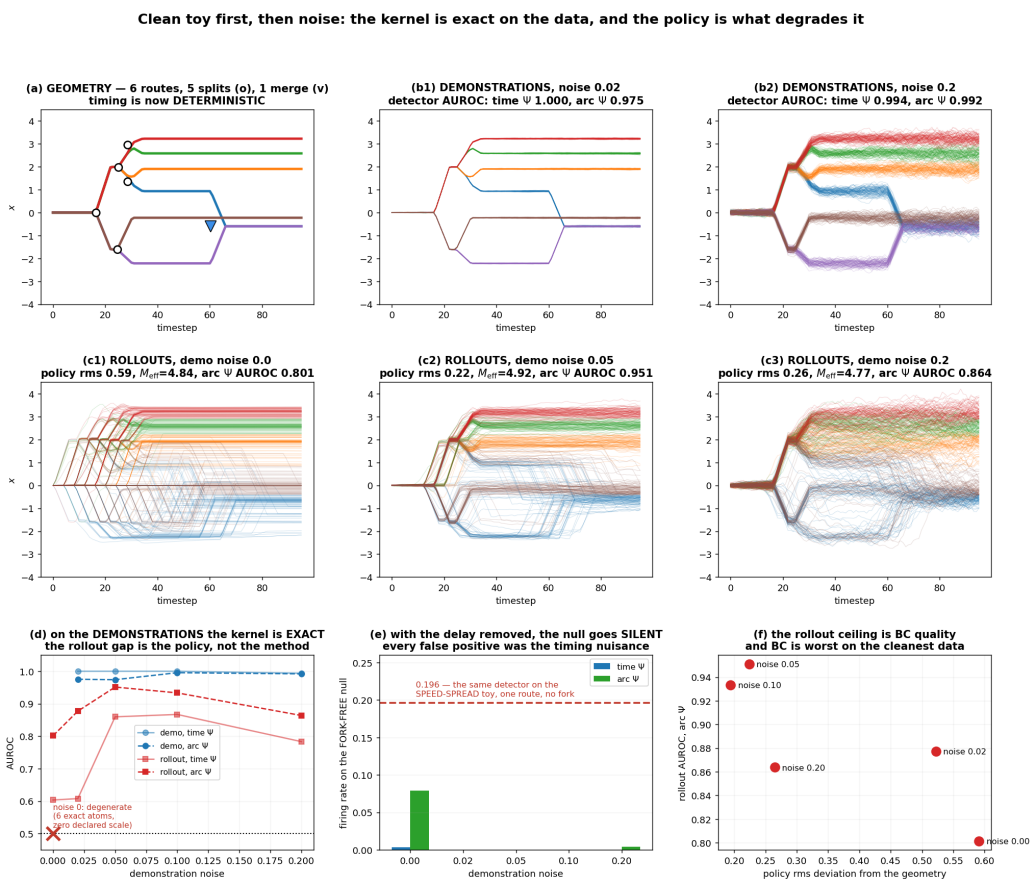

Figure 10: Forward and reverse performance across the controlled noise ladder, with a matched fork-free null. Main-text values use a single explicitly sourced rung rather than a range that silently drops the hardest rung.

certified local valency, and Section 7.1 is one instance of feeding a detected partition back into training. We have not tested BLPTK as an expansion-point proposer for any of them, so the connection is stated as a relationship between problems, not a demonstrated capability.

D.2 Scope of the controlled suites, and two provenance limits

Grader resolution, anchor reuse, and reproducibility. The synthetic grader's truth is the noiseless single-linkage count at an absolute separation threshold while the estimator is given a declared scale from a data quantile, so where routes separate by more than one and less than the other the two disagree by construction; that mismatch, not a detector failure, accounts for the entire residual error in the orthogonal arm (Table 2). Adaptive context localization also turns a radius declaration into a neighbor-count declaration. Anchors reuse rollouts, so anchor-level uncertainty is anti-conservative, and held-out paths test partition survival within a geometry rather than transfer to unseen geometries. Finally, the real-data driver has since been revised: it no longer regenerates the stored avoid_detect.json schema, so that artifact is not currently reproducible end to end from the script in the repository, and freezing a regenerating commit is outstanding work.

The controlled suites divide the labour and no single one establishes the whole claim. The valency, affinity-scale and sample-size sweeps of Appendix C.18 use a shared-phase geometry in which a clock already separates branchable contexts, so they measure counting and resolvability, not localization; localization rests on the staggered family of Section 5, on the within-slice real-data measurement of Section 6, and on the arm of Section 7. Anchors reuse

rollouts, so anchor-level uncertainty is anti-conservative, and held-out paths test partition survival within a geometry rather than transfer to unseen geometries.

Two provenance limits should also be stated plainly. The certificate is a deterministic feasibility statement about the exact weighted graph that was validated — including any Monte Carlo resampling of the local ball — not a statistical confidence bound on a population quantity. And the real-data driver has since been revised: it no longer regenerates the stored avoid_detect.json schema, so that artifact is not currently reproducible end to end from the script in the repository, and freezing a regenerating commit is outstanding work.