# REUSABLE SCHRÖDINGER BRIDGES FROM COARSE ENDPOINT INTERACTIONS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Schrödinger bridges connect prescribed distributions through reference dynamics and provide a common framework for control, generation, and scientific inference. When many endpoint pairs share the same dynamics, can their bridges also share a smaller optimization problem? We show that fine endpoint information and the interaction governing endpoint pairing can be represented separately. Because the marginals are fixed, additive single-endpoint terms leave the optimizer unchanged. Approximating the remaining interaction on coarse states therefore gives a reusable solver that preserves the original fine marginals exactly. For positive finite references, a uniform log-interaction residual $\eta$ bounds each Kullback–Leibler (KL) direction by $\min\{2\eta, \eta^2/2\}$, with a sharp quadratic coefficient. A sparse-query converse then leads to a reference-only partition rule with an equal-budget approximation guarantee. We further decompose query error and quantify the dependence recovered by nested refinement, while fast mixing explains how coarse interactions arise dynamically. Our experiments verify the predicted error structure and demonstrate that reusing coarse endpoint interactions provides a valid and practical way to solve repeated bridge queries in molecular and stochastic-dynamics settings.

## 1 INTRODUCTION

Schrödinger bridges describe distribution-to-distribution evolution with the smallest change to a reference process, supporting stochastic control, generative modeling, and scientific inference (Schrödinger, 1931; Chen et al., 2016; Gushchin et al., 2024). Applications often reuse the same dynamics with different initial and final distributions. One would therefore expect much of the computation to be reusable. Yet each bridge must match both distributions at full resolution, even when a reduced model captures the relevant dynamics (Noé & Clementi, 2017; Schütte et al., 2023). Which information must enter the optimization, and which can simply be carried into the answer?

The answer follows from what the endpoint constraints already determine. Suppose an initial state receives an additional cost independent of its destination. Every feasible coupling places the same mass there and therefore incurs the same additional cost. The same reasoning applies at the final endpoint. Consequently, only costs depending jointly on the endpoints affect their pairing. This suggests a simple construction: group states with similar pairing preferences, optimize the masses exchanged between groups, and distribute those masses using the prescribed within-group laws. This preserves every prescribed fine marginal, including sparse and asymmetric endpoints.

We formalize this construction through the log-interaction left after fitting single-endpoint terms and coarse preferences. A small uniform residual guarantees accuracy for every admissible endpoint pair. Conversely, a two-point query detects interaction lost by merging distinct preferences. These results guide a reference-only partition competitive with the best equally sized representation. They connect bridge computation to finite-lag dependence studied systematically through transition manifolds, lumpability, and reaction coordinates (Bittracher et al., 2018; 2023; Zhang et al., 2025).

Application accuracy also depends on which interactions its endpoints emphasize. We therefore decompose query error into coarse-kernel mismatch and within-cell dependence, then quantify the dependence recovered by refinement. Fast mixing provides a dynamical example, producing coarse interior interactions and fine endpoint layers (Figure 1). Finally, molecular queries and complete-path comparisons connect this structure to storage and runtime. Using the same conditional sampler in exact and coarse methods makes the computational benefit directly measurable.



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

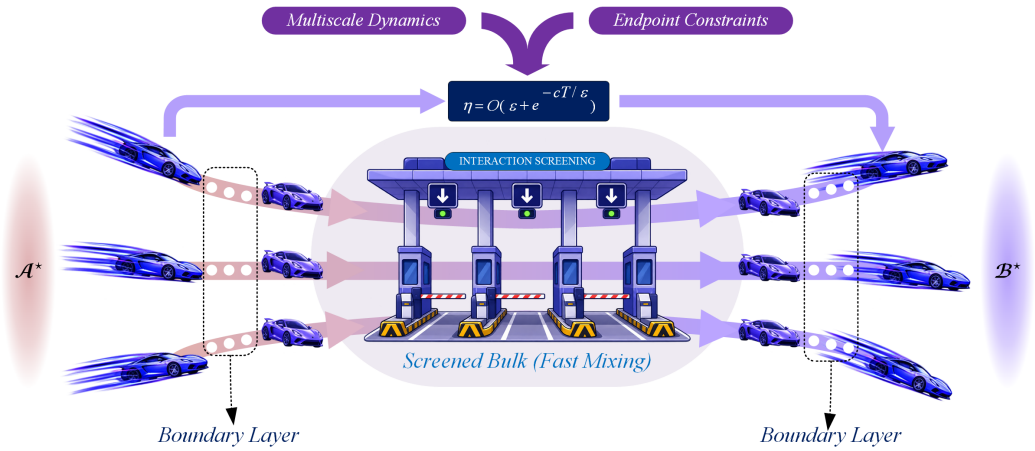


Figure 1: **Reusable Schrödinger bridges from coarse endpoint interactions.** The laws $P_0 = \mu$ and $P_T = \nu$ (shown as $\mathcal{A}^\star, \mathcal{B}^\star$) retain fine detail while their pairing is optimized coarsely. Fast mixing illustrates this separation through endpoint layers. Here $\eta$ is the interaction residual, $\varepsilon$ the time-scale ratio, $T$ the horizon, and $c > 0$ an effective mixing rate for within-fiber relaxation.

**Our contributions are fourfold.** (i) We characterize endpoint-uniform accuracy through sharp quadratic KL bounds and a sparse-query converse. (ii) We construct a reference-only partition competitive with the best worst-query KL among equally sized conditional lifts. (iii) We decompose actual-query error into kernel and partition terms, quantify nested refinement, and derive mixing-induced endpoint layers. (iv) We quantify compact serving through 100 molecular queries, component ablations, exact amortized baselines, and three independently trained learned path models.

## 2 Related Work

Classical Schrödinger bridge theory uses endpoint potentials to adapt a reference process to prescribed marginals (Fortet, 1940; Jamison, 1974; Léonard, 2012; Chen et al., 2016). Entropy projection supplies the variational geometry (Csiszár, 1975), and stability results quantify perturbations (Ghosal et al., 2022; Eckstein & Nutz, 2022; Nutz & Wiesel, 2023; Duan et al., 2026). Repeated queries raise a representation question: how much reference information is needed before the marginals are known? Our interaction criterion and converse answer it through bridge-query error.

Transition manifolds, lumpability, and molecular learning provide coordinates that retain important transition information (Bittracher et al., 2018; 2023; Schütte et al., 2023; Mardt et al., 2018; Zhang et al., 2025; Lu & Vanden-Eijnden, 2014; Hilder & Sharma, 2024). We ask when these coordinates also preserve endpoint interaction. Averaging and hypocoercivity provide a sufficient mechanism (Khasminskii, 1968; Legoll & Lelièvre, 2010; Hartmann et al., 2020; Dolbeault et al., 2015). Our fixed-horizon analysis studies short-time reference compression and complements long-horizon turnpike results (Clerc et al., 2023; Chiarini et al., 2022; Sanders et al., 2024).

Computational methods exploit several different sources of reuse. Kernel approximations, low-rank couplings, screening, and fused updates reduce the work within transport solvers (Solomon et al., 2015; Altschuler et al., 2019; Scetbon et al., 2021; Gasteiger et al., 2021; Alaya et al., 2019; Ye et al., 2026). Across tasks, meta-transport and neural operators predict useful initial potentials (Amos et al., 2023; Geuter et al., 2025). Our coarse representation reduces the interaction on which each query is solved, while its conditional lift preserves the fine marginals. Thus it addresses the representation cost of repeated queries and can be combined with improved scaling updates or learned initialization.

For full trajectories, Diffusion Schrödinger Bridge Matching (DSBM), Light and Optimal Schrödinger Bridge Matching (LightSB-M), and simulation-free score and flow matching ($SF^2M$) learn samplers (Shi et al., 2023; Gushchin et al., 2024; Tong et al., 2024). Quantized bridges also organize endpoints through anchors (Fuchs et al., 2026). We compare complete requests and isolate coarsening gains with a common conditional sampler. Compared with symmetry-based transport (Zhang et al.,

2026; Xu et al., 2026), our reusable structure concerns reference interaction and allows arbitrary admissible endpoint asymmetry through the original within-cell conditional laws.

## 3 Problem setup

Fix a horizon $T > 0$ and a Markov probability law $R$ on a standard Borel path space, with state coordinates $X_{[0,T]}$ in $\mathsf{X}$. Its endpoint joint and marginals are $r = R_{0T}$, $r_0 = R_0$, and $r_T = R_T$. Let $\mathcal{P}(\mathsf{X})$ denote probability laws on $\mathsf{X}$. For target endpoints $\mu, \nu \in \mathcal{P}(\mathsf{X})$, the dynamic Schrödinger bridge is

$$P^\star = \arg\min\{\mathrm{KL}(P\|R) : P_0 = \mu,\ P_T = \nu\}. \tag{1}$$

Relative entropy KL measures the information cost of changing $R$: (1) selects the least costly path law with the requested endpoints. Disintegrating $P$ and $R$ given $(X_0, X_T) = (x, y)$ (Léonard, 2014) reduces (1) to the endpoint problem

$$\gamma^\star = \arg\min_{\gamma\in\Pi(\mu,\nu)} \mathrm{KL}(\gamma\|r), \qquad P^\star = \int R^{xy}\,\gamma^\star(\,\mathrm{d}x\,\mathrm{d}y), \tag{2}$$

where $\Pi(\mu, \nu)$ is the set of probability couplings with marginals $\mu$ and $\nu$, and $R^{xy}$ is the conditional law of $R$ given $(X_0, X_T) = (x, y)$. We call $(\mu, \nu)$ $R$-*admissible* when $\Pi(\mu, \nu)$ contains a coupling of finite relative entropy with respect to $r$. This is the endpoint class used by the uniform coupling theorems. Thus each query changes the coupling and reuses the conditional paths $R^{xy}$.

To reduce the endpoint problem, fix a measurable coarse-graining map $q : \mathsf{X} \to \mathsf{B}$ into a standard Borel space. The fine states sharing $\beta \in \mathsf{B}$ form the fiber $\mathsf{F}_\beta = q^{-1}(\beta)$. Write $\mu^{\mathsf{B}} = q_{\#}\mu$ for the pushforward, defined by $(q_{\#}\mu)(E) = \mu(q^{-1}(E))$, and disintegrate $\mu(\,\mathrm{d}x) = \mu^{\mathsf{B}}(\,\mathrm{d}\beta)\mu_\beta(\,\mathrm{d}x)$, with analogous notation for $\nu$. For a coarse coupling $\Gamma \in \Pi(\mu^{\mathsf{B}}, \nu^{\mathsf{B}})$, define

$$\mathsf{L}_{\mu,\nu}\Gamma(\,\mathrm{d}x\,\mathrm{d}y) := \Gamma(\,\mathrm{d}\beta\,\mathrm{d}\beta')\mu_\beta(\,\mathrm{d}x)\nu_{\beta'}(\,\mathrm{d}y). \tag{3}$$

Operationally, one first samples $(\beta, \beta')$ from $\Gamma$ and then draws the fine endpoints from $\mu_\beta$ and $\nu_{\beta'}$. Integrating this conditional lift over either endpoint returns its prescribed marginal. It therefore preserves the full endpoint distributions, including their within-cell asymmetry.

## 4 Fixed marginals expose a coarse interaction

We now identify which part of the reference determines the coupling. Assume $r \sim r_0 \otimes r_T$, meaning that the two measures are mutually absolutely continuous, and define the endpoint density ratio and its logarithm by

$$\kappa_R(x, y) = \frac{\mathrm{d}r}{\mathrm{d}(r_0 \otimes r_T)}(x, y), \qquad \ell_R(x, y) = \log \kappa_R(x, y). \tag{4}$$

The ratio $\kappa_R(x, y)$ measures how much more likely $(x, y)$ is under the reference than under independent reference endpoints. Its logarithm $\ell_R$ is the reference log-interaction. For any $\gamma \in \Pi(\mu, \nu)$ and bounded $a, b$,

$$\int [a(x) + b(y)]\,\gamma(\,\mathrm{d}x\,\mathrm{d}y) = \int a\,\mathrm{d}\mu + \int b\,\mathrm{d}\nu, \tag{5}$$

so terms depending on one endpoint alone contribute the same constant to every feasible coupling. We measure the remaining two-endpoint dependence by

$$\eta_q(R) := \inf_{\substack{a,b:\mathsf{X}\to\mathbb{R}\\ c:\mathsf{B}\times\mathsf{B}\to\mathbb{R}}} \big\|\ell_R(x, y) - a(x) - b(y) - c(q(x), q(y))\big\|_\infty, \tag{6}$$

where the functions are bounded and measurable and the norm is the essential supremum under $r_0 \otimes r_T$. Thus $\eta_q$ measures the interaction left after fitting single-endpoint terms and coarse preferences. We work on domains with finite residual, which in this formulation also makes $\ell_R$ bounded. Positive finite kernels and positive continuous kernels on compact domains provide two important examples.

**Theorem 4.1** (Exact coarse-coordinate reduction). *Suppose $\ell_R(x,y) = a(x) + b(y) + c(q(x), q(y))$ almost everywhere for bounded measurable $a, b : \mathsf{X} \to \mathbb{R}$ and $c : \mathsf{B} \times \mathsf{B} \to \mathbb{R}$. For every $R$-admissible $(\mu, \nu)$, let*

$$\Gamma^\star = \arg \min_{\Gamma \in \Pi(\mu^{\mathsf{B}}, \nu^{\mathsf{B}})} \left\{ \mathrm{KL}(\Gamma \| \mu^{\mathsf{B}} \otimes \nu^{\mathsf{B}}) - \int c(\beta, \beta') \, \mathrm{d}\Gamma \right\}. \tag{7}$$

*The objective balances dependence against the reference preference $c$. Lifting $\Gamma^\star$ and retaining $R^{xy}$ gives the exact coupling and path laws. For finite cells, marginal constants drop out, so Sinkhorn directly scales $e^c$ without storing either removable fine-state endpoint potential.*

**Why it holds.** The entropy chain rule separates coarse dependence from within-cell dependence. For fixed coarse masses, the second term is minimized by the product of the prescribed conditionals. Together with the cancellation in (5), this gives the reduced problem and its exact lift (Theorem B.5).

The same construction remains accurate when the interaction is only approximately coarse. To quantify the error, fix bounded measurable $a, b : \mathsf{X} \to \mathbb{R}$ and $c : \mathsf{B} \times \mathsf{B} \to \mathbb{R}$, and define the fitted function and its uniform residual by

$$h(x,y) = a(x) + b(y) + c(q(x), q(y)), \qquad \eta(h; R) := \|\ell_R - h\|_\infty \le \eta, \tag{8}$$

where the norm has the same meaning as in (6). With the positive coarse kernel $K_{\mathsf{B}}(\beta, \beta') = e^{c(\beta,\beta')}$, let $\widehat{\Gamma}^\star$ solve (7) with $c$ and define its conditional lift $\widehat{\gamma}^\star = \mathsf{L}_{\mu,\nu} \widehat{\Gamma}^\star$.

**Theorem 4.2** (Uniform coupling and path approximation). *For every $R$-admissible endpoint pair, $\widehat{\gamma}^\star$ has marginals $(\mu, \nu)$ exactly and satisfies*

$$\mathrm{KL}(\widehat{\gamma}^\star \| \gamma^\star) \vee \mathrm{KL}(\gamma^\star \| \widehat{\gamma}^\star) \le 2\eta, \qquad \|\widehat{\gamma}^\star - \gamma^\star\|_{\mathrm{TV}} \le \sqrt{\eta}, \tag{9}$$

*where $\|p - q\|_{\mathrm{TV}} := \sup_E |p(E) - q(E)|$ denotes total-variation distance. Mixing the same conditional reference bridges, $\widehat{P} = \int R^{xy} \widehat{\gamma}^\star(\mathrm{d}x \, \mathrm{d}y)$, preserves both endpoints. Each path KL equals its corresponding coupling KL, so both path directions obey the same endpoint-uniform bound.*

Replacing the interaction changes each feasible objective by at most $\eta$. Entropy projection converts the resulting objective gap into KL, which transfers exactly to paths through their shared conditionals (Theorem B.7). Finite kernels admit a sharper accuracy bound and a converse.

### 4.1 Finite references: a criterion and a converse

Let $K \in \mathbb{R}^{N \times N}_{>0}$ be a positive reference kernel and $q : [N] \to [B]$ a surjective partition, where $1 \le B \le N$ and $[n] = \{1, \dots, n\}$. Define $\eta_q(K)$ by replacing $\ell_R$ with $\log K$ and the norm with $\max_{i,j \in [N]}$ in (6). Positive row or column scaling adds single-endpoint terms, so it leaves $\eta_q(K)$ unchanged. Endpoint probabilities may have zeros, in which case optimization is restricted to their active support.

**Theorem 4.3** (Finite-kernel accuracy and interaction criterion). *Let $a, b \in \mathbb{R}^N$ and $c \in \mathbb{R}^{B \times B}$ define $h_{ij} = a_i + b_j + c_{q(i),q(j)}$ with $\max_{i,j} |\log K_{ij} - h_{ij}| \le \eta$. For endpoint probabilities $\mu, \nu$ on $[N]$, let $\gamma^\star$ minimize $\sum_{ij} \gamma_{ij} \log(\gamma_{ij}/K_{ij})$ over $\Pi(\mu, \nu)$, and let $\widehat{\gamma}$ lift its corresponding coarse solution (7):*

$$\mathrm{KL}(\widehat{\gamma} \| \gamma^\star) \vee \mathrm{KL}(\gamma^\star \| \widehat{\gamma}) \le \min\{2\eta, \eta^2/2\}, \qquad \|\widehat{\gamma} - \gamma^\star\|_{\mathrm{TV}} \le \tanh(\eta/2). \tag{10}$$

*This holds for every endpoint probability pair, including sparse pairs. The quadratic coefficient $1/2$ is sharp as $\eta \to 0$. To state the converse, define the log cross-ratio variation*

$$D_q(K) := \max_{\substack{i \ne i', \, j \ne j' \in [N] \\ q(i) = q(i') \text{ or } q(j) = q(j')}} \left| \log \frac{K_{ij} K_{i'j'}}{K_{ij'} K_{i'j}} \right|. \tag{11}$$

*Then $D_q(K)/4 \le \eta_q(K) \le D_q(K)$, with the maximum zero if no rectangle is eligible. All indices range over $[N]$. For every positive coarse kernel, some endpoint pair supported on at most two states per side has $\mathrm{KL}(\widehat{\gamma} \| \gamma^\star) \ge \log \cosh(D_q(K)/4)$.*

Cross-ratios, which underlie Hilbert's projective metric (Bushell, 1973; Georgiou & Pavon, 2015), compare the preference for two possible endpoint pairings. A rectangle with two states in the same

cell therefore detects interaction lost by the partition. A two-point query makes that loss explicit, while residual tilting and marginal projection give the upper bound (Theorem B.15). The shared conditional reference bridges transfer both coupling bounds to the full path laws.

Let $\mathcal{R}_q(K)$ be the infimum, over positive coarse kernels chosen before queries, of the worst of the two coupling KL directions over all endpoint pairs. Corollary B.16 gives

$$\log\cosh\big(\eta_q(K)/4\big) \le \log\cosh\big(D_q(K)/4\big) \le \mathcal{R}_q(K) \le \min\{2\eta_q(K), \eta_q(K)^2/2\}. \tag{12}$$

Small residual therefore characterizes uniform accuracy up to dimension-independent constants. Refining a partition enlarges the interaction class and decreases the best residual (Proposition B.9). Additional cells can consequently retain more modes of endpoint dependence at the chosen horizon.

### 4.2 What controls the error of an actual query?

The uniform bound accounts for the most demanding endpoints. For an actual query, we can weight the interaction by its prescribed distributions. Fix $K$, $q : [N] \twoheadrightarrow [B]$, a positive coarse kernel $W$, and endpoints $\mu, \nu$. On active cells write $\mu_a(i) = \mu_i/(q_\#\mu)_a$ for $q(i) = a$, and similarly $\nu_b(j)$. Let $\widehat{\Gamma}$ scale $W$ to the coarse marginals and set $\widehat{\gamma} = \mathsf{L}_{\mu,\nu}\widehat{\Gamma}$. Define $e_{ij} = \log K_{ij} - \log W_{q(i),q(j)}$, $s_{ab} = \mu_a \otimes \nu_b$, and $\bar{e}_{ab} = \mathbb{E}_{s_{ab}} e$. Finally, let $\Gamma^\dagger$ scale $W_{ab}e^{\bar{e}_{ab}} = \exp(\mathbb{E}_{s_{ab}} \log K)$ to the same marginals. This corrected kernel depends only on $K$ and the query conditionals.

**Theorem 4.4** (Query error decomposition and bounds). *Put $F = \mathrm{KL}(\widehat{\gamma}\|\gamma^\star)$, $G = \mathrm{KL}(\gamma^\star\|\widehat{\gamma})$, and $C = \mathrm{KL}(\widehat{\Gamma}\|\Gamma^\dagger)$. Then*

$$F = C + \min_{\Lambda\in\Pi(q_\#\mu, q_\#\nu)} \mathrm{KL}(\mathsf{L}_{\mu,\nu}\Lambda\|\gamma^\star). \tag{13}$$

*Denote the minimum by $J$, attained at $\Gamma^\dagger$. For row and column functions $u_i, v_j$, set $g_{ij} = e_{ij} - u_i - v_j$ and $M_g(t) = \log \mathbb{E}_{\widehat{\gamma}} \exp\{t(g - \mathbb{E}_{\widehat{\gamma}} g)\}$. The forward bound uses $t = 1$, and the reverse bound holds for every $t > 1$:*

$$F \le M_g(1), \qquad G \le \frac{M_g(t) - tC}{t-1}. \tag{14}$$

The identity $F = C + J$ gives two distinct routes to improvement. Fitting a better coarse kernel reduces $C$, while resolving more within-cell dependence reduces the best achievable error $J$. In particular, for nested partitions $q_f, q_c$, the best lifted couplings satisfy $J_c = \mathrm{KL}(P_c^\dagger\|P_f^\dagger) + J_f$ (Corollary B.12). Refinement therefore recovers an exactly quantified amount of dependence. The moment bounds provide complementary query-level checks using $O(|I_0||I_T|)$ reference entries per pass on active supports $I_0, I_T$. Their evaluation is separate from online serving. Both results follow from the geometry of entropy projections (Csiszár, 1975; Dupuis et al., 2016).

## 5 Fast mixing as a sufficient mechanism

We next explain how a coarse interaction can arise from the dynamics itself. Consider a stationary reference on a compact smooth fiber bundle without boundary, with local state $x = (\beta, z)$, partition $q(x) = \beta$, and generator

$$L^\varepsilon = \varepsilon^{-1}A + S, \qquad (Ah)(\beta, z) = A_\beta h(\beta, \cdot)(z). \tag{15}$$

Here $\varepsilon > 0$ is the scale ratio, $A_\beta$ mixes within fibers, and $S$ acts at order one (Khasminskii, 1968; Legoll & Lelièvre, 2010; Hartmann et al., 2020). Assumptions (SF1)–(SF4) require elliptic reversible fibers, a uniform gap, forward/adjoint decay, controlled invariant conditionals, and positive-time smoothing. Appendix B.4 specifies the graph-norm and positivity conditions and the rate $c_{\mathrm{mix}} > 0$. Together, they yield the following interaction and boundary-layer estimates.

**Theorem 5.1** (Mixing-induced interaction and endpoint layers). *Under* (SF1)–(SF4)*, for fixed $0 < T_0 < T_1 < \infty$ and uniformly over $T \in [T_0, T_1]$, the stationary reference $R^\varepsilon$ satisfies*

$$\eta_q(R^\varepsilon) \le C(\varepsilon + e^{-c_{\mathrm{mix}}T/\varepsilon}). \tag{16}$$

*Let $m^\varepsilon$ be its invariant law and $\pi_\beta$ the invariant law of $A_\beta$. Fix $0 < \alpha < 1$ and $M \ge 1$. If $\mu = \rho_0 m^\varepsilon, \nu = \rho_T m^\varepsilon$ obey $e^{-M} \le \rho_0, \rho_T \le e^M$ and $\|\log\rho_0\|_{C^{2+\alpha}} \vee \|\log\rho_T\|_{C^{2+\alpha}} \le M$, then*

**Algorithm 1** Offline reference reduction and online bridge queries

**Require:** fixed reference operator or trajectories, with an optional physical coarse coordinate $q$

*Offline stage*

1: Estimate a positive kernel $K$ and choose $q$ by Theorem 6.1 or a reference-only physical/spectral coordinate.
2: Set $(K_{\mathsf{B}})_{ab} = K_{c_a,c_b}$ for selected centers, or use positive aggregation or log-interaction fitting for other partitions.
3: Store $q$, $K_{\mathsf{B}}$ and any required conditional-reference sampling objects.

*Online stage for an unseen admissible pair* $(\mu, \nu)$

4: Read each endpoint once, form $\mu^{\mathsf{B}}, \nu^{\mathsf{B}}$, and retain $\{\mu_\beta, \nu_{\beta'}\}$.
5: Run Sinkhorn scaling on $K_{\mathsf{B}}$ with marginals $(\mu^{\mathsf{B}}, \nu^{\mathsf{B}})$, obtaining $\widehat{\Gamma}$.
6: Form the implicit coupling $\widehat{\gamma} = \mathsf{L}_{\mu,\nu}\widehat{\Gamma}$ and sample endpoints hierarchically if requested.
7: **return** $\widehat{\gamma}$ and, on request, paths sampled by $(x, y) \sim \widehat{\gamma}$ followed by $R^{xy}$.

*the full bridge marginal $\mu_t^\varepsilon$ obeys*

$$\sup_\beta \|\mu_t^\varepsilon(\cdot \mid \beta) - \pi_\beta\|_{\mathrm{TV}} \le C_M\{\varepsilon + e^{-c_{\mathrm{mix}}t/\varepsilon} + e^{-c_{\mathrm{mix}}(T-t)/\varepsilon}\}. \tag{17}$$

*For small enough $\varepsilon$, $C, C_M$ are independent of $\varepsilon, T, t$ and of the selected endpoints within this class.*

In (16), mixing removes fast memory exponentially, while order-one coupling introduces an $O(\varepsilon)$ contribution. Fine endpoint detail is therefore concentrated in layers of width $(\varepsilon/c_{\mathrm{mix}})\log(1/\varepsilon)$. Appendix B.5 gives factor-gradient and fast-gap estimates, and Theorem B.21 covers finite states.

## 6 An offline/online solver on coarse states

For a positive finite reference $K$, put $L_{ij} = \log K_{ij}$ and define

$$d_K(i,i') = \max\left\{\mathrm{osc}_j(L_{ij} - L_{i'j}),\ \mathrm{osc}_j(L_{ji} - L_{ji'})\right\}, \qquad \mathrm{osc}_j\, v_j = \max_j v_j - \min_j v_j. \tag{18}$$

Here $j \in [N]$. Removing endpoint scalings gives a pseudometric on incoming and outgoing pairing preferences. Its largest within-cell distance is $D_q(K)$, linking representation selection to the sparse-query converse and identifying fine distinctions needed for worst-query accuracy.

**Theorem 6.1** (Reference-only representation selection)**.** *Fix $1 \le B \le N$ and choose $B$ distinct centers $c_1, \dots, c_B$ by farthest-first traversal in $d_K$, starting arbitrarily. Assign states to nearest centers, with each center assigned to itself, giving $q_B : [N] \to [B]$. Put $r_B = \max_i \min_b d_K(i, c_b)$ and $(W_B)_{ab} = K_{c_a,c_b}$. For $W \in \mathbb{R}_{>0}^{B\times B}$, let $\mathcal{E}_K(q, W)$ be the supremum over all endpoint probabilities of the maximum of both KL directions between the full optimizer and the $(q, W)$ conditional lift. Set $\mathcal{R}_B^*(K) = \inf_{q:[N]\to[B],\, W>0} \mathcal{E}_K(q, W)$. Then*

$$\log\cosh(r_B/4) \le \mathcal{R}_B^*(K) \le \mathcal{E}_K(q_B, W_B) \le \min\{2r_B, r_B^2/2\}. \tag{19}$$

*Writing $h(u) = \operatorname{arcosh}(e^u)$ gives $\mathcal{E}_K(q_B, W_B) \le \min\{8h(\mathcal{R}_B^*), 8h(\mathcal{R}_B^*)^2\}$: the small-risk approximation factor is at most $16 + O(\mathcal{R}_B^*)$.*

Farthest-first traversal and Hilbert clustering are classical (Gonzalez, 1985; Nielsen & Sun, 2019). The interaction criterion connects their radius to optimal *bridge-query risk*, giving the equal-budget guarantee. Theorem B.19 adds covering-number bounds. Traversal costs $O(BN^2)$ offline and stores $O(N + B^2)$ entries. Spectral and physical coordinates provide application-specific alternatives.

Algorithm 1 separates reference processing from per-query scaling. Additive-potential fitting evaluates a finite kernel's residual, with a continuous extension under Appendix C.2's regularity conditions. Reference checks and query-moment evaluations are timed separately from online solves.

The converged lift has exact marginals, with numerical residual controlled by the stopping tolerance. Paths then follow $R^{xy}$, and conditional approximation contributes additively to path KL (Proposition B.13). Molecular paths below use the estimated finite-state reference.

**Corollary 6.2** (Per-query complexity)**.** *For $N$ fine states, $B$ coarse cells, and $I$ iterations, Algorithm 1 costs $O(N + IB^2)$ per implicit-coupling query. Endpoints are read once, and iterations use coarse masses. Path sampling and explicit output are accounted for in Appendix C.7.*

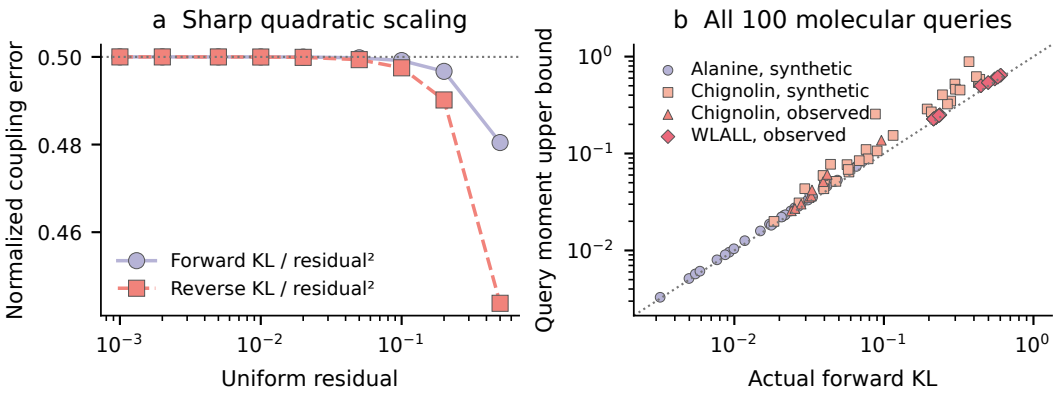


Figure 2: **Worst-query and actual-query guarantees.** (a) Both KL coefficients approach the sharp quadratic constant. (b) Product-centered moments bound forward KL for all 100 molecular queries. The dotted line denotes equality. These checks use the reference and returned coupling, with evaluation costs reported separately from online serving for each tested molecular query.

## 7 Experiments

We test interaction bounds, representation accuracy, and complete-path reuse. Appendix E collects supporting comparisons, and the code and data supplement provides full numerical records.

**Protocol and data.** All methods share a frozen reference and the same test endpoints. Symmetric KL $D_{\mathrm{sym}}$ averages both directions, with marginal $L^1$ tolerances $10^{-12}$ for dense targets and $10^{-10}$ for compared solvers (Cuturi, 2013). Online single-thread timings include endpoint processing. The 100 molecular queries comprise 30 synthetic pairs each for alanine and Chignolin, ten unsmoothed Chignolin occupancy pairs, and 30 observed WLALL pairs (mdshare, n.d.a; Lazzeri, 2025; mdshare, n.d.b). WLALL uses 15/5/5 training/validation/test trajectories, and observed marginals retain their exact zeros. Offline fitting and full path sampling are evaluated and timed separately.

### 7.1 From worst-query theory to actual-query error

On six irregular positive six-state references, we enumerate all $31 + 90$ partitions of sizes two and three, giving 726 cases. Exact residual programs bracket the equal-budget optimum, and two-point queries check the converse to $3.6 \times 10^{-13}$. A checkerboard family approaches $F/\eta^2 = G/\eta^2 = 1/2$, verifying the sharp coefficient (Figure 2a). Graded-rate, coupled, and kinetic chains then test resolution and endpoint layers. Appendices D.2 and F give the models and plots.

Molecular queries show why endpoint weighting is useful. On Chignolin's 30 synthetic queries, median forward KL is $0.0831$, compared with the product-centered bound $0.1082$. Two normalization passes tighten the median bound to $0.0838$, closely tracking the measured error. The reverse bound is $0.3637$, and all 100 queries satisfy the evaluated inequalities (Figure 2b). Table 7 and Appendix D.5 connect this application-scale accuracy to the exact query-error decomposition.

### 7.2 Frozen representations and component ablations

Next, reference-distance, spectral, geometric $k$-means, and farthest-first partitions are each paired with representative and aggregated kernels at five budgets. Positive nonnegative matrix factorization (NMF) provides a low-rank comparison with six ranks and three reference-only initializations (Lee & Seung, 2001). Figure 4 compares their accuracy, storage, and measured online cost.

On the 503-state WLALL reference, spectral aggregation with eight cells gives median symmetric KL $0.2268$ using $4.43$ KiB, compared with $0.2164$ and $62.88$ KiB for rank-eight NMF. Their median online times are $0.408$ and $0.588$ ms, while dense scaling takes $2.820$ ms. The coarse representation therefore offers substantially smaller storage at comparable median accuracy.

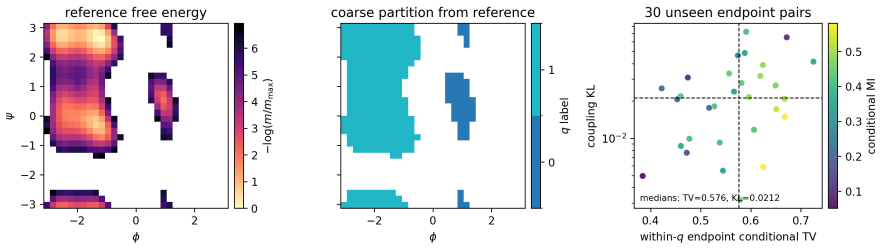


Figure 3: **Alanine-dipeptide reference trajectories.** A reference-only spectral rule selects $B = 2$ cells. Across 30 unseen synthetic pairs, median within-cell conditional TV is $0.576$, whereas median/maximum forward KL is $0.0212/0.0726$. The lift preserves every prescribed fine marginal.

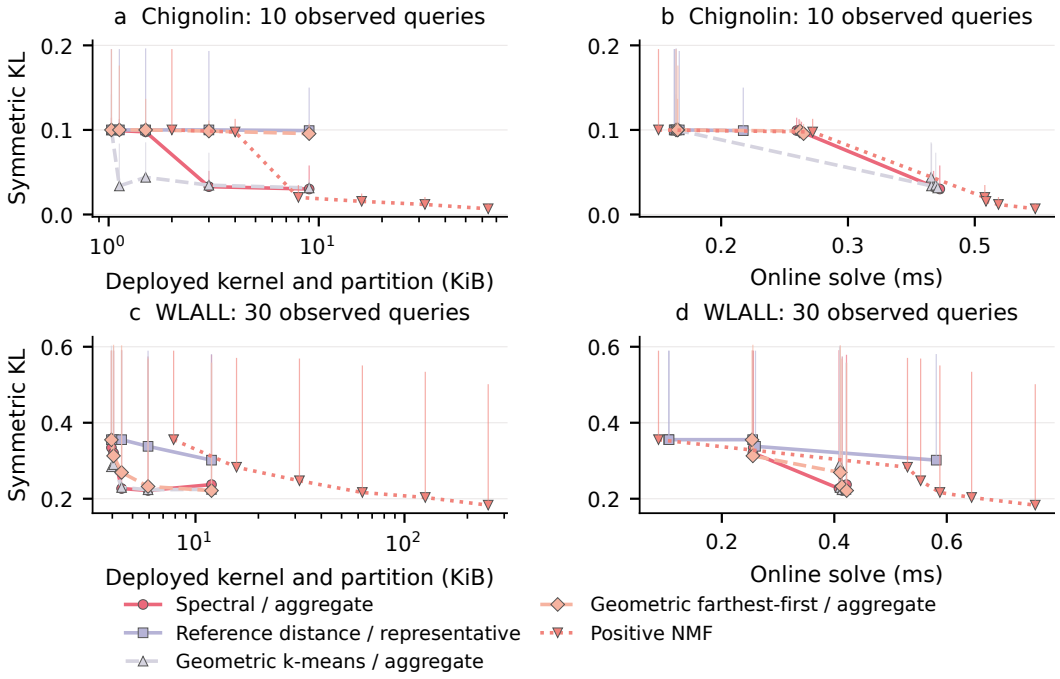


Figure 4: **Accuracy and deployment cost on observed molecular endpoints.** Points show median symmetric KL, with vertical marks extending to the 90th percentile. Each curve retains all tested budgets. Deployment storage counts kernels or factors together with partitions. Reference-distance and spectral representations are compared with geometric partitions and reference-selected NMF.

The ablations explain this accuracy. With partition and kernel fixed, using the requested endpoint conditionals reduces median coupling TV from $0.753$ to $0.058$ on alanine and from $0.632$ to $0.123$ on Chignolin, restoring the marginals. Independent-coupling controls and 135 lag–shrinkage settings test reference dependence through 10,692 supplementary sensitivity solves.

We also compare reuse through learned initialization. Meta-OT dual-objective and UNOT-MLP non-grid adapters (Amos et al., 2023; Geuter et al., 2025) train on independent endpoints and initialize exact fine-kernel refinement. On WLALL, their 2,000-update batch times are $0.679/0.686$ ms per query, compared with dense $0.918$ ms and hierarchical $0.720$ ms. Our $B = 8$ lift takes $0.040$ ms using $4.43$ KiB at median $D_{\mathrm{sym}} = 0.2268$. Tables 11–14 compare these exact-refinement and compact-representation tradeoffs across all four endpoint query collections.

Nested refinement then tests the resolution mechanism. Across 1,280 evaluations, the decomposition identities agree to $4.6 \times 10^{-11}$. On observed Chignolin queries, interaction-profile refinement from

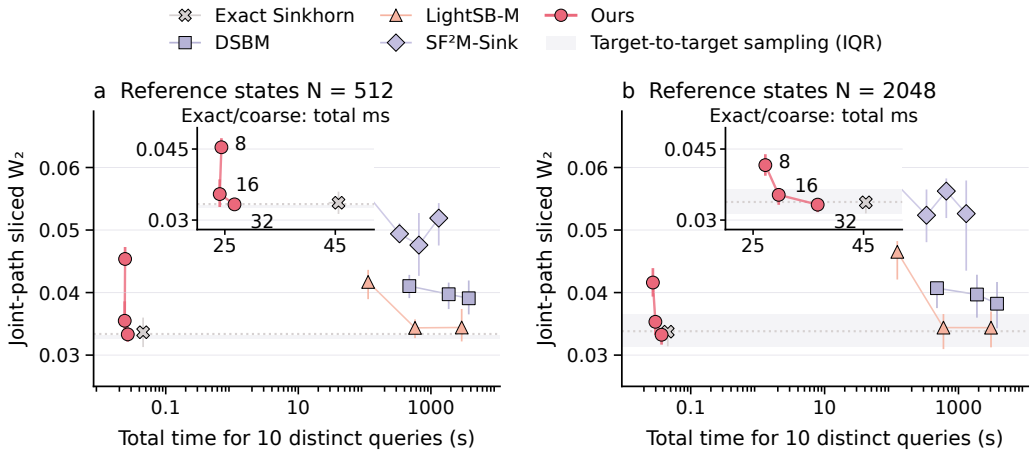


Figure 5: **Repeated complete path requests.** Ten queries per bank, three independent runs, and 2,048 five-time paths per query. All budgets are shown. Insets compare exact and coarse costs in milliseconds, with labels 8/16/32 denoting $B$. Intervals show run-to-run cost ranges and query interquartile ranges. Shading gives the independently evaluated target-to-target sampling floor.

$B = 16$ to 64 lowers median symmetric KL from 0.0329 to 0.0134. Appendix D.5 compares two reference-only hierarchies, complementing the fixed-budget kernel ablation.

### 7.3 Reference reuse for complete path queries

The remaining question is how coupling savings translate into full path requests. For the 30 observed WLALL queries, the coupling stage takes 1.197 ms in a batch, compared with 27.205 ms for dense scaling (22.7×). Serial medians are 0.408 and 2.820 ms (6.9×). Figure 6 in Appendix F includes offline processing, and the following experiments add the endpoint and conditional-path draws.

For finite-endpoint Brownian targets, each query draws 256 states per endpoint from banks of $N = 512/2048$ states. With $B = 32$, ten requests of 2,048 paths cost $0.027/0.037$ s including compilation, compared with exact Sinkhorn's $0.046/0.045$ s, giving $1.70\times$ / $1.24\times$ acceleration. Joint-path distances are $0.0333/0.0333$, near target-to-target levels $0.0334/0.0338$. The common conditional sampler isolates the coarsening gain. DSBM, LightSB-M, and SF$^2$M-Sink provide learned samplers through 180 fits (Shi et al., 2023; Gushchin et al., 2024; Tong et al., 2024). Figure 5 includes training and sampling, complemented by post-fit comparisons in Appendix E.6.

We conclude with four-step paths for 40 observed molecular queries. Their horizon-specific reference kernels give median symmetric path KL 0.000074 on Chignolin and 0.000193 on WLALL. Shared conditional laws make these exactly the corresponding coupling errors. Complete requests of 2,048 paths take $7.80/29.07$ ms, compared with dense $9.07/38.42$ ms ($1.16\times$ / $1.32\times$). Thus compression reduces complete service cost, with the gain determined by the common conditional simulation. Appendix D.8 provides same-horizon controls, with stage-wise measurements in the data supplement.

## 8 Conclusion

Reusable Schrödinger bridges share a coarse optimization across queries while preserving their fine endpoint laws. The key is to retain the reference interaction that determines pairing and carry the prescribed conditionals directly into the solution. Sharp bounds and a sparse-query converse characterize the required representation, while an exact decomposition identifies how kernel fitting and partition refinement improve individual queries. Fast mixing explains one dynamical source of coarse interaction. Together, molecular ablations and complete-path comparisons demonstrate how this separation supports compact representations and efficient reuse across changing endpoints.

## REFERENCES


Mokhtar Z. Alaya, Maxime Bérar, Gilles Gasso, and Alain Rakotomamonjy. Screening Sinkhorn algorithm for regularized optimal transport. In *Advances in Neural Information Processing Systems*, volume 32, pp. 12169–12179. Curran Associates, Inc., 2019. URL https://papers.neurips.cc/paper_files/paper/2019/hash/95688ba636a4720a85b3634acfec8cdd-Abstract.html.

Jason Altschuler, Francis Bach, Alessandro Rudi, and Jonathan Niles-Weed. Massively scalable Sinkhorn distances via the Nyström method. In *Advances in Neural Information Processing Systems*, volume 32, pp. 4427–4437. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/hash/f55cadb97eaff2ba1980e001b0bd9842-Abstract.html.

Brandon Amos, Giulia Luise, Samuel Cohen, and Ievgen Redko. Meta optimal transport. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 791–813. PMLR, 2023. URL https://proceedings.mlr.press/v202/amos23a.html.

Andreas Bittracher, Péter Koltai, Stefan Klus, Ralf Banisch, Michael Dellnitz, and Christof Schütte. Transition manifolds of complex metastable systems: Theory and data-driven computation of effective dynamics. *Journal of Nonlinear Science*, 28(2):471–512, 2018. doi: 10.1007/s00332-017-9415-0. URL https://doi.org/10.1007/s00332-017-9415-0.

Andreas Bittracher, Mattes Mollenhauer, Péter Koltai, and Christof Schütte. Optimal reaction coordinates: Variational characterization and sparse computation. *Multiscale Modeling & Simulation*, 21 (2):449–488, 2023. doi: 10.1137/21M1448367. URL https://doi.org/10.1137/21M1448367.

P. J. Bushell. On the projective contraction ratio for positive linear mappings. *Journal of the London Mathematical Society*, 6(2):256–258, 1973. doi: 10.1112/jlms/s2-6.2.256. URL https://doi.org/10.1112/jlms/s2-6.2.256.

Yongxin Chen, Tryphon T. Georgiou, and Michele Pavon. On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications*, 169(2):671–691, 2016. doi: 10.1007/s10957-015-0803-z. URL https://doi.org/10.1007/s10957-015-0803-z.

Alberto Chiarini, Giovanni Conforti, Giacomo Greco, and Zhenjie Ren. Entropic turnpike estimates for the kinetic Schrödinger problem. *Electronic Journal of Probability*, 27:1–32, 2022. doi: 10.1214/22-EJP850. URL https://doi.org/10.1214/22-EJP850. Article 131.

Gauthier Clerc, Giovanni Conforti, and Ivan Gentil. Long-time behaviour of entropic interpolations. *Potential Analysis*, 59(1):65–95, 2023. doi: 10.1007/s11118-021-09961-w. URL https://doi.org/10.1007/s11118-021-09961-w.

Imre Csiszár. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3(1):146–158, 1975. doi: 10.1214/aop/1176996454. URL https://doi.org/10.1214/aop/1176996454.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, volume 26, pp. 2292–2300. Curran Associates, Inc., 2013. URL https://proceedings.neurips.cc/paper/2013/hash/af21d0c97db2e27e13572cbf59eb343d-Abstract.html.

Jean Dolbeault, Clément Mouhot, and Christian Schmeiser. Hypocoercivity for linear kinetic equations conserving mass. *Transactions of the American Mathematical Society*, 367(6):3807–3828, 2015. doi: 10.1090/S0002-9947-2015-06012-7. URL https://doi.org/10.1090/S0002-9947-2015-06012-7.

Danny Duan, Hanbaek Lyu, and William Powell. Scaling limit of Sinkhorn-rescaled random matrices via stability of static Schrödinger bridges, 2026. URL https://arxiv.org/abs/2604.23461.

Paul Dupuis, Markos A. Katsoulakis, Yannis Pantazis, and Petr Plecháč. Path-space information bounds for uncertainty quantification and sensitivity analysis of stochastic dynamics. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):80–111, 2016. doi: 10.1137/15M1025645. URL https://doi.org/10.1137/15M1025645.

Stephan Eckstein and Marcel Nutz. Quantitative stability of regularized optimal transport and convergence of Sinkhorn's algorithm. *SIAM Journal on Mathematical Analysis*, 54(6):5922–5948, 2022. doi: 10.1137/21M145505X. URL https://doi.org/10.1137/21M145505X.

Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T. H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. POT: Python optimal transport. *Journal of Machine Learning Research*, 22(78): 1–8, 2021. URL https://jmlr.org/papers/v22/20-451.html.

Robert Fortet. Résolution d'un système d'équations de M. Schrödinger. *Journal de Mathématiques Pures et Appliquées*, 19(1–4):83–105, 1940. URL https://www.numdam.org/item/JMPA_1940_9_19_1-4_83_0/.

Tobias Fuchs, Florian Kalinke, and Nadja Klein. QDSB: Quantized diffusion Schrödinger bridges, 2026. URL https://arxiv.org/abs/2605.11983.

Johannes Gasteiger, Marten Lienen, and Stephan Günnemann. Scalable optimal transport in high dimensions for graph distances, embedding alignment, and more. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 5616–5627. PMLR, 2021. URL https://proceedings.mlr.press/v139/gasteiger21a.html.

Tryphon T. Georgiou and Michele Pavon. Positive contraction mappings for classical and quantum Schrödinger systems. *Journal of Mathematical Physics*, 56(3):033301, 2015. doi: 10.1063/1.4915289. URL https://doi.org/10.1063/1.4915289.

Jonathan Geuter, Gregor Kornhardt, Ingimar Tomasson, and Vaios Laschos. Universal neural optimal transport. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 19196–19232. PMLR, 2025. URL https://proceedings.mlr.press/v267/geuter25a.html.

Promit Ghosal, Marcel Nutz, and Espen Bernton. Stability of entropic optimal transport and Schrödinger bridges. *Journal of Functional Analysis*, 283(9):109622, 2022. doi: 10.1016/j.jfa.2022.109622. URL https://doi.org/10.1016/j.jfa.2022.109622.

Teofilo F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985. doi: 10.1016/0304-3975(85)90224-5. URL https://doi.org/10.1016/0304-3975(85)90224-5.

Konstantinos Gourgoulias, Markos A. Katsoulakis, Luc Rey-Bellet, and Jie Wang. How biased is your model? Concentration inequalities, information and model bias. *IEEE Transactions on Information Theory*, 66(5):3079–3097, 2020. doi: 10.1109/TIT.2020.2977067. URL https://doi.org/10.1109/TIT.2020.2977067.

Nikita Gushchin, Sergei Kholkin, Evgeny Burnaev, and Alexander Korotin. Light and optimal Schrödinger bridge matching. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 17100–17122. PMLR, 2024. URL https://proceedings.mlr.press/v235/gushchin24a.html.

Carsten Hartmann, Lara Neureither, and Upanshu Sharma. Coarse graining of nonreversible stochastic differential equations: Quantitative results and connections to averaging. *SIAM Journal on Mathematical Analysis*, 52(3):2689–2733, 2020. doi: 10.1137/19M1299852. URL https://doi.org/10.1137/19M1299852.

Bastian Hilder and Upanshu Sharma. Quantitative coarse-graining of Markov chains. *SIAM Journal on Mathematical Analysis*, 56(1):913–954, 2024. doi: 10.1137/22M1473996. URL https://doi.org/10.1137/22M1473996.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830. URL https://doi.org/10.1080/01621459.1963.10500830.

Richard Holley and Daniel W. Stroock. Logarithmic Sobolev inequalities and stochastic Ising models. *Journal of Statistical Physics*, 46(5–6):1159–1194, 1987. doi: 10.1007/BF01011161. URL https://doi.org/10.1007/BF01011161.

Benton Jamison. Reciprocal processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 30(1):65–86, 1974. doi: 10.1007/BF00532864. URL https://doi.org/10.1007/BF00532864.

R. Z. Khasminskii. On the principle of averaging the Itô stochastic differential equations. *Kybernetika*, 4(3):260–279, 1968. URL https://www.kybernetika.cz/content/1968/3/260.

Gianmarco Lazzeri. Simulation data and code for Optimal Rejection-Free Path Sampling. Zenodo dataset, 2025. URL https://doi.org/10.5281/zenodo.14922168. Version v1.

Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems*, volume 13, pp. 556–562. MIT Press, 2001. URL https://proceedings.neurips.cc/paper_files/paper/2000/hash/f9d1152547c0bde01830b7e8bd60024c-Abstract.html.

Frédéric Legoll and Tony Lelièvre. Effective dynamics using conditional expectations. *Nonlinearity*, 23(9):2131–2163, 2010. doi: 10.1088/0951-7715/23/9/006. URL https://doi.org/10.1088/0951-7715/23/9/006.

Christian Léonard. From the Schrödinger problem to the Monge–Kantorovich problem. *Journal of Functional Analysis*, 262(4):1879–1920, 2012. doi: 10.1016/j.jfa.2011.11.026. URL https://doi.org/10.1016/j.jfa.2011.11.026.

Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems*, 34(4):1533–1574, 2014. doi: 10.3934/dcds.2014.34.1533. URL https://www.aimsciences.org/article/doi/10.3934/dcds.2014.34.1533.

Antoine Liutkus, Umut Simsekli, Szymon Majewski, Alain Durmus, and Fabian-Robert Stöter. Sliced-Wasserstein flows: Nonparametric generative modeling via optimal transport and diffusions. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 4104–4113, 2019. URL https://proceedings.mlr.press/v97/liutkus19a.html.

Jianfeng Lu and Eric Vanden-Eijnden. Exact dynamical coarse-graining without time-scale separation. *The Journal of Chemical Physics*, 141(4):044109, 2014. doi: 10.1063/1.4890367. URL https://doi.org/10.1063/1.4890367.

Alessandra Lunardi. *Analytic Semigroups and Optimal Regularity in Parabolic Problems*, volume 16 of *Progress in Nonlinear Differential Equations and Their Applications*. Birkhäuser, Basel, 1995. doi: 10.1007/978-3-0348-0557-5. URL https://doi.org/10.1007/978-3-0348-0557-5.

Andreas Mardt, Luca Pasquali, Hao Wu, and Frank Noé. VAMPnets for deep learning of molecular kinetics. *Nature Communications*, 9(1):5, 2018. doi: 10.1038/s41467-017-02388-1. URL https://doi.org/10.1038/s41467-017-02388-1.

Robert T. McGibbon, Kyle A. Beauchamp, Matthew P. Harrigan, Christoph Klein, Jason M. Swails, Carlos X. Hernández, Christian R. Schwantes, Lee-Ping Wang, Thomas J. Lane, and Vijay S. Pande. MDTraj: A modern open library for the analysis of molecular dynamics trajectories. *Biophysical Journal*, 109(8):1528–1532, 2015. doi: 10.1016/j.bpj.2015.08.015.

mdshare. mdshare: Alanine dipeptide. Public molecular-dynamics data repository maintained by markovmodel, n.d.a. URL https://markovmodel.github.io/mdshare/ALA2/. Accessed 2026-08-09.

mdshare. mdshare: Pentapeptide. Public molecular-dynamics data repository maintained by markovmodel, n.d.b. URL https://markovmodel.github.io/mdshare/pentapeptide/. Accessed 2026-09-11.

Frank Nielsen and Ke Sun. Clustering in Hilbert's projective geometry: The case studies of the probability simplex and the elliptope of correlation matrices. In Frank Nielsen (ed.), *Geometric Structures of Information*, Signals and Communication Technology, pp. 297–331. Springer, Cham, 2019. doi: 10.1007/978-3-030-02520-5_11. URL https://doi.org/10.1007/978-3-030-02520-5_11.

Frank Noé and Cecilia Clementi. Collective variables for the study of long-time kinetics from molecular trajectories: Theory and methods. *Current Opinion in Structural Biology*, 43:141–147, 2017. doi: 10.1016/j.sbi.2017.02.006. URL https://doi.org/10.1016/j.sbi.2017.02.006.

Feliks Nüske, Hao Wu, Jan-Hendrik Prinz, Christoph Wehmeyer, Cecilia Clementi, and Frank Noé. Markov state models from short non-equilibrium simulations—analysis and correction of estimation bias. *The Journal of Chemical Physics*, 146(9):094104, 2017. doi: 10.1063/1.4976518. URL https://doi.org/10.1063/1.4976518.

Marcel Nutz and Johannes Wiesel. Stability of Schrödinger potentials and convergence of Sinkhorn's algorithm. *The Annals of Probability*, 51(2):699–722, 2023. doi: 10.1214/22-AOP1611. URL https://doi.org/10.1214/22-AOP1611.

Grigorios A. Pavliotis and Andrew M. Stuart. *Multiscale Methods: Averaging and Homogenization*, volume 53 of *Texts in Applied Mathematics*. Springer, New York, 2008. doi: 10.1007/978-0-387-73829-1. URL https://doi.org/10.1007/978-0-387-73829-1.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011. URL https://jmlr.org/papers/v12/pedregosa11a.html.

Julia Sanders, Marco Baldovin, and Paolo Muratore-Ginanneschi. Optimal control of underdamped systems: An analytic approach. *Journal of Statistical Physics*, 191(9):117, 2024. doi: 10.1007/s10955-024-03320-w. URL https://doi.org/10.1007/s10955-024-03320-w.

Meyer Scetbon, Marco Cuturi, and Gabriel Peyré. Low-rank Sinkhorn factorization. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9344–9354. PMLR, 2021. URL https://proceedings.mlr.press/v139/scetbon21a.html.

Martin K. Scherer, Benjamin Trendelkamp-Schroer, Fabian Paul, Guillermo Pérez-Hernández, Moritz Hoffmann, Nuria Plattner, Christoph Wehmeyer, Jan-Hendrik Prinz, and Frank Noé. PyEMMA 2: A software package for estimation, validation, and analysis of Markov models. *Journal of Chemical Theory and Computation*, 11(11):5525–5542, 2015. doi: 10.1021/acs.jctc.5b00743.

Bernhard Schmitzer. Stabilized sparse scaling algorithms for entropy regularized transport problems. *SIAM Journal on Scientific Computing*, 41(3):A1443–A1481, 2019. doi: 10.1137/16M1106018. URL https://doi.org/10.1137/16M1106018.

Erwin Schrödinger. Über die umkehrung der naturgesetze. *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-Mathematische Klasse*, pp. 144–153, 1931.

Christof Schütte, Stefan Klus, and Carsten Hartmann. Overcoming the timescale barrier in molecular dynamics: Transfer operators, variational principles and machine learning. *Acta Numerica*, 32:517–673, 2023. doi: 10.1017/S0962492923000016. URL https://doi.org/10.1017/S0962492923000016.

Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion Schrödinger bridge matching. In *Advances in Neural Information Processing Systems*, volume 36, pp. 62183–62223, 2023. doi: 10.52202/075280-2717. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/c428adf74782c2092d254329b6b02482-Abstract-Conference.html.

Justin Solomon, Fernando de Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas. Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics*, 34(4):66:1–66:11, 2015. doi: 10.1145/2766963. URL https://doi.org/10.1145/2766963.

Alexander Y. Tong, Nikolay Malkin, Kilian Fatras, Lazar Atanackovic, Yanlei Zhang, Guillaume Huguet, Guy Wolf, and Yoshua Bengio. Simulation-free Schrödinger bridges via score and flow matching. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 1279–1287, 2024. URL https://proceedings.mlr.press/v238/tong24a.html.

Christoph Wehmeyer and Frank Noé. Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *The Journal of Chemical Physics*, 148(24):241703, 2018. doi: 10.1063/1.5011399. URL https://doi.org/10.1063/1.5011399.

Yixian Xu, Yusong Wang, Shengjie Luo, Kaiyuan Gao, Tianyu He, Di He, and Chang Liu. Quotient-space diffusion models. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=3JPAkwSVc4. Oral presentation.

Felix X.-F. Ye, Xingjie Li, An Yu, Ming-Ching Chang, Linsong Chu, and Davis Wertheimer. FlashSinkhorn: IO-aware entropic optimal transport on GPU. In *Proceedings of the 43rd International Conference on Machine Learning*, volume 306 of *Proceedings of Machine Learning Research*. PMLR, 2026. URL https://arxiv.org/abs/2602.03067. Oral presentation.

Anru Zhang and Mengdi Wang. Spectral state compression of Markov processes. *IEEE Transactions on Information Theory*, 66(5):3202–3231, 2020. doi: 10.1109/TIT.2019.2956737. URL https://doi.org/10.1109/TIT.2019.2956737.

Jiechao Zhang, Huichun Zhang, Jian Sun, and Wei Zeng. Optimal transport with symmetry groups. In *Proceedings of the 43rd International Conference on Machine Learning*, volume 306 of *Proceedings of Machine Learning Research*, 2026. URL https://openreview.net/forum?id=C4JJrPSwpy.

Mingyuan Zhang, Zhicheng Zhang, Hao Wu, and Yong Wang. Flow matching for optimal reaction coordinates of biomolecular systems. *Journal of Chemical Theory and Computation*, 21(1):399–412, 2025. doi: 10.1021/acs.jctc.4c01139. URL https://doi.org/10.1021/acs.jctc.4c01139.

# Appendix

## Contents

**A Notation and conventions** . . . . . **16**

**B Proofs and theoretical details** . . . . . **16**

B.1 Reference endpoint dependence and fixed marginals . . . . . 17
B.2 Exact reduction and endpoint-uniform perturbation . . . . . 18
B.3 Directly verifiable finite-state case . . . . . 27
B.4 Slow–fast semigroup and kernel estimates . . . . . 31
B.5 Schrödinger factors and boundary layers . . . . . 35
B.6 Per-query complexity . . . . . 37

**C Algorithm and implementation details** . . . . . **38**

C.1 Positive coarse operators and residual fitting . . . . . 38
C.2 Continuous supremum certification . . . . . 38
C.3 Partition selection and heterogeneous mixing . . . . . 39
C.4 Coarse scaling and exact conditional lift . . . . . 40
C.5 Conditional reference-bridge sampling . . . . . 40
C.6 Verifying the slow–fast assumptions . . . . . 41
C.7 Complexity and solver composition . . . . . 41

**D Experimental models and protocols** . . . . . **41**

D.1 Common quantities and evaluation design . . . . . 41
D.2 Controlled references and finite-state theory . . . . . 42
D.3 Molecular references and held-out endpoints . . . . . 43
D.4 Reference representations and amortized competitors . . . . . 44
D.5 Query-dependent error and refinement . . . . . 45
D.6 Compatible Gaussian-reference comparisons . . . . . 45
D.7 Learned bridges and common path evaluation . . . . . 46
D.8 Finite-state conditional paths . . . . . 47

**E Supplementary results** . . . . . **47**

**F Supplementary figures** . . . . . **55**

**G Reproducibility, ethics, and AI use** . . . . . **59**

# A NOTATION AND CONVENTIONS

Table 1: Symbols used throughout the paper.

| Symbol | Meaning | Symbol | Meaning |
|---|---|---|---|
| $R$ | Markov reference path law | $R_{0T},\ R^{xy}$ | endpoint joint and conditional reference bridge |
| $r_0, r_T$ | reference endpoint marginals | $\kappa_R, \ell_R$ | endpoint density ratio and its logarithm |
| $q : \mathsf{X} \to \mathsf{B}$ | coarse-graining map | $\mathsf{F}_\beta$ | fiber $q^{-1}(\beta)$ |
| $\mathcal{G}_q, \eta_q$ | additive endpoint/coarse class and uniform residual | $\Pi(\mu, \nu)$ | couplings with marginals $\mu, \nu$ |
| $K$ | positive finite reference kernel | $D_q, \mathcal{R}_q$ | within-cell log cross-ratio and best worst-query KL |
| $d_K, r_B$ | bidirectional projective distance and covering radius | $\mathcal{E}_K, \mathcal{R}_B^*$ | representation risk and best risk with $B$ cells |
| $c_b, q_B, W_B$ | selected centers, nearest-center partition, and representative kernel | $\mathcal{N}_K, \mathcal{P}_K$ | reference covering and packing numbers |
| $\mathsf{E}, Q$ | within-cell averaging and deviation from its average | $\iota$ | lift of a coarse function to each fiber |
| $\mu, \nu$ | prescribed endpoint laws | $\mu^{\mathsf{B}}, \nu^{\mathsf{B}}$ | pushed-forward coarse marginals |
| $\gamma^\star, \Gamma^\star$ | full and coarse optimal couplings | $\mathsf{L}_{\mu,\nu}$ | conditional lift |
| $F, G$ | forward and reverse query KL | $C, J$ | coarse mismatch and best fixed-lift error |
| $g, M_g(t)$ | adjusted residual and centered log moment | $P^\dagger, \Gamma^\dagger$ | best query-specific fixed lift and its coarse coupling |
| $L^\varepsilon$ | fast–slow generator | $A, S$ | fast and order-one generator blocks |
| $P_t^\varepsilon, \bar{P}_t$ | full and effective semigroups | $\pi_\beta$ | frozen fast equilibrium |
| $m^\varepsilon, m^0$ | invariant law and leading product-form approximation | $\omega^\varepsilon$ | density $\mathrm{d}m^\varepsilon / \mathrm{d}m^0$ |
| $f_t^\varepsilon, g_t^\varepsilon$ | Schrödinger factors | $\mathcal{E}^\varepsilon_{K,\alpha}$ | regular positive endpoint class |
| $N, B$ | full-state count and coarse cardinality | $N_{\mathsf{B}}, N_{\mathsf{F}}$ | coarse count and maximum per-fiber state count |
| KL, TV | relative entropy and total variation | $T, \varepsilon$ | bridge horizon and fast scale |
| $T_0, T_1$ | fixed positive horizon bounds | $K$ or $M, \alpha$ | endpoint contrast bound and Hölder exponent |
| $\lambda_\beta, \lambda_\star$ | fiber gap and its infimum over $\beta$ | $c_{\mathrm{mix}}$ | positive fast-mixing decay rate |

Unless a result states otherwise, constants denoted by $C$, $C_K$, or $c_{\mathrm{mix}}$ may depend on the geometry, generator coefficients, analytic assumption bounds, $K$, $\alpha$, $T_0$, and $T_1$, but not on $0 < \varepsilon \le \varepsilon_0$, $t$, $T \in [T_0, T_1]$, $\beta$, or the selected endpoint pair. We write $\operatorname{osc}_E u := \sup_E u - \inf_E u$ and use $\|p - q\|_{\mathrm{TV}} = \sup_A |p(A) - q(A)| = \frac{1}{2}\|p - q\|_1$ when densities exist. Supremum-norm identities use the continuous kernel versions supplied by the smoothing assumptions. For finite-state references, the same notation denotes the maximum absolute entry over the full state space.

# B PROOFS AND THEORETICAL DETAILS

The proofs follow the information that a bridge query needs. We first remove endpoint-only terms, then identify the remaining coarse interaction and quantify the error of its conditional lift. Finite-state results turn this characterization into a converse and a reference-only construction. We then show how fast mixing produces a small interaction residual and how conditioning affects the resulting dynamics. The coupling results hold for all admissible endpoint pairs. The factor and drift estimates additionally use the regular positive endpoint class introduced in Definition B.29.

### B.1 Reference endpoint dependence and fixed marginals

Let $\mathsf{X}$ and $\mathsf{B}$ be standard Borel spaces, let $q : \mathsf{X} \to \mathsf{B}$ be a measurable coarse-graining map, and let $R$ be a Markov law on path space over $[0, T]$. Denote its endpoint joint and marginals by

$$r := R_{0T}, \qquad r_0 := R_0, \qquad r_T := R_T.$$

Throughout this subsection we assume $r \sim r_0 \otimes r_T$ and $\ell_R \in L^\infty(r_0 \otimes r_T)$. These assumptions are automatic for the finite-state strictly positive kernels used in the experiments and hold for a uniformly elliptic diffusion on a compact connected state space over any horizon in a compact subset of $(0, \infty)$, with its positive continuous transition density. Bounded interaction on the declared endpoint domain is the condition used below. Compactness provides one way to obtain it, and a specified bounded endpoint domain provides another for the Euclidean kernels used in finite-endpoint evaluations.

For a finite positive measure $s$ and a probability $\gamma$, define the generalized relative entropy

$$\mathcal{H}(\gamma \mid s) := \begin{cases} \displaystyle\int \log\left(\frac{\mathrm{d}\gamma}{\mathrm{d}s}\right) \mathrm{d}\gamma, & \gamma \ll s, \\ +\infty, & \text{otherwise.} \end{cases} \tag{20}$$

Multiplying $s$ by a positive constant adds the same constant to every objective value. Its total mass therefore does not affect the minimizer over a fixed feasible set. For probability measures we use $\|p - q\|_{\mathrm{TV}} := \sup_A |p(A) - q(A)| = \frac{1}{2}\|p - q\|_{L^1}$ whenever densities are available.

**Definition B.1** (Reference interaction)**.** *The interaction density and log-interaction of $R$ are*

$$\kappa_R(x, y) := \frac{\mathrm{d}r}{\mathrm{d}(r_0 \otimes r_T)}(x, y), \qquad \ell_R(x, y) := \log \kappa_R(x, y). \tag{21}$$

*If $R$ is stationary with invariant law $m$ and transition density $k_T(x, y)$ relative to $m$, then $\kappa_R(x, y) = k_T(x, y)$. In a finite state space, with transition matrix $K_T$ and invariant vector $m$, $\ell_R(i, j) = \log K_T(i, j) - \log m_j$ when matrix entries are interpreted relative to counting measure.*

The row and column parts of $\ell_R$ do not affect a Schrödinger minimization with fixed endpoint marginals. This motivates the following class of additive endpoint and coarse-coordinate functions.

**Definition B.2** (Additive/coarse class)**.** *Define the bounded additive endpoint/coarse class*

$$\mathcal{G}_q := \{(x, y) \mapsto a(x) + b(y) + c(q(x), q(y)) : a, b, c \text{ bounded and measurable}\}. \tag{22}$$

*The uniform coarse-interaction residual of the reference log-density is*

$$\eta_q(R) := \inf_{h \in \mathcal{G}_q} \|\ell_R - h\|_{L^\infty(r_0 \otimes r_T)}. \tag{23}$$

*For a particular fit $h(x, y) = a(x) + b(y) + c(q(x), q(y))$ we write*

$$\eta(h; R) := \|\ell_R - h\|_\infty. \tag{24}$$

*A numerical least-squares fit gives an upper bound on $\eta_q(R)$ once its supremum residual is bounded over the full domain. The fitting loss and the uniform error bound are distinct quantities.*

**Definition B.3** (Admissible endpoint pair)**.** *A pair $(\mu, \nu)$ is $R$-admissible if*

$$\mathfrak{C}_R(\mu, \nu) := \{\gamma \in \Pi(\mu, \nu) : \mathcal{H}(\gamma \mid r) < \infty\} \neq \varnothing. \tag{25}$$

*The collection of these pairs is denoted by $\mathfrak{A}_R$, and "all endpoints" below means all pairs in this class. Admissibility implies $\mu \ll r_0$ and $\nu \ll r_T$. On a positive finite reference, every probability pair is admissible, including pairs with exact zero masses. For an atomless reference marginal, an atomic target requires a different reference: specifying a finite endpoint joint together with its conditional paths defines such a probability law on the underlying continuous path space.*

**Unbounded endpoint interactions.** For a Brownian kernel on full Euclidean support, the log interaction differs from $x^\top y/(\sigma^2 T)$ by single-endpoint terms, where $\sigma > 0$ is the diffusion scale. A finite partition cannot uniformly remove this bilinear term: if distinct $x, x'$ share a cell, its log cross-ratio is $(x - x')^\top (y - y')/(\sigma^2 T)$, unbounded as $y, y'$ range over Euclidean space. A bounded residual would bound this cross-ratio by four times its supremum norm. For measurable partitions, a

positive-measure cell and Fubini's theorem supply such $x, x'$ outside the exceptional null set. Finite endpoint banks or a specified bounded endpoint domain give a finite interaction problem while leaving the original conditional reference paths unconstrained between the prescribed endpoints.

For $\mu \in \mathcal{P}(\mathsf{X})$ let $\mu^{\mathsf{B}} := q_{\#}\mu$ and fix a disintegration $\mu(\,\mathrm{d}x) = \mu^{\mathsf{B}}(\,\mathrm{d}\beta)\mu_\beta(\,\mathrm{d}x)$. Use analogous notation for $\nu$. Given $\Gamma \in \Pi(\mu^{\mathsf{B}}, \nu^{\mathsf{B}})$, define its endpoint- conditional lift by

$$\mathsf{L}_{\mu,\nu}\Gamma(\,\mathrm{d}x\;\mathrm{d}y) := \Gamma(\,\mathrm{d}\beta\;\mathrm{d}\beta')\mu_\beta(\,\mathrm{d}x)\nu_{\beta'}(\,\mathrm{d}y). \tag{26}$$

**Lemma B.4** (Conditional information decomposition)**.** *Let $\gamma \in \Pi(\mu,\nu)$ and $\Gamma = (q,q)_{\#}\gamma$. If $\gamma^{\beta,\beta'}$ denotes the conditional law of $(x,y)$ given $(q(x),q(y)) = (\beta,\beta')$, then*

$$\begin{aligned}\mathrm{KL}(\gamma\|\mu\otimes\nu) = {}& \mathrm{KL}(\Gamma\|\mu^{\mathsf{B}}\otimes\nu^{\mathsf{B}}) \\ &+ \int \mathrm{KL}\Big(\gamma^{\beta,\beta'}\Big\|\mu_\beta\otimes\nu_{\beta'}\Big)\,\Gamma(\,\mathrm{d}\beta\;\mathrm{d}\beta').\end{aligned} \tag{27}$$

*If $\mathrm{KL}(\Gamma\|\mu^{\mathsf{B}}\otimes\nu^{\mathsf{B}}) < \infty$, the unique entropy minimizer at fixed $\Gamma$ is $\mathsf{L}_{\mu,\nu}\Gamma$.*

*Proof.* Apply the chain rule for relative entropy to the maps $(x,y)\mapsto(q(x),q(y))$. The conditional law of $\mu\otimes\nu$ given $(\beta,\beta')$ is $\mu_\beta\otimes\nu_{\beta'}$. When the coarse term is finite, nonnegativity and strict convexity of conditional relative entropy give the last assertion. Finiteness of the coarse term ensures that the product conditionals attain a finite value, which is needed for this uniqueness claim. □

### B.2 Exact reduction and endpoint-uniform perturbation

The information decomposition makes exact reduction immediate: once the reference interaction depends only on coarse states, its objective cannot reward any additional dependence within a cell pair. After proving this case, we control the effect of a uniformly bounded interaction residual.

**Theorem B.5** (Exact coarse-coordinate reduction)**.** *Suppose that the reference interaction has the exact bounded additive endpoint/coarse decomposition*

$$\ell_R(x,y) = a(x) + b(y) + c(q(x),q(y)) \qquad (r_0\otimes r_T)\text{-a.e.} \tag{28}$$

*Let $(\mu,\nu)\in\mathfrak{A}_R$. Define the coarse problem*

$$\Gamma^\star := \arg\min_{\Gamma\in\Pi(\mu^{\mathsf{B}},\nu^{\mathsf{B}})}\left\{\mathrm{KL}(\Gamma\|\mu^{\mathsf{B}}\otimes\nu^{\mathsf{B}}) - \int c(\beta,\beta')\Gamma(\,\mathrm{d}\beta\;\mathrm{d}\beta')\right\}. \tag{29}$$

*Then the unique endpoint coupling of the full Schrödinger bridge is*

$$\gamma^\star = \mathsf{L}_{\mu,\nu}\Gamma^\star. \tag{30}$$

*In particular, endpoint laws may vary arbitrarily within fibers. The lift retains their conditional laws, including all prescribed asymmetry and correlations across the fine coordinates.*

*If $R^{xy}$ is a regular conditional bridge of the Markov reference, the full path-space optimizer is*

$$P^\star = \int R^{xy}\,\gamma^\star(\,\mathrm{d}x\;\mathrm{d}y). \tag{31}$$

*Thus the reduction changes neither the reference bridges nor the endpoint constraints.*

*Proof.* Choose compatible Polish realizations of the standard Borel spaces. The set $\Pi(\mu,\nu)$ is weakly compact, and relative entropy with respect to a fixed probability is weakly lower semicontinuous. Admissibility therefore gives an attained finite minimum, and strict convexity gives uniqueness. Data processing also gives $\mathrm{KL}(\mu\|r_0)+\mathrm{KL}(\nu\|r_T)<\infty$. The bounded decomposition functions also have finite integrals. We can therefore separate the fixed endpoint costs from the varying coupling terms.

For any feasible coupling, relative entropy decomposes as

$$\begin{aligned}\mathcal{H}(\gamma\mid r) = {}& \mathrm{KL}(\gamma\|r_0\otimes r_T) - \int \ell_R\;\mathrm{d}\gamma \\ = {}& \mathrm{KL}(\gamma\|\mu\otimes\nu) + \mathrm{KL}(\mu\|r_0) + \mathrm{KL}(\nu\|r_T) \\ & - \int a\;\mathrm{d}\mu - \int b\;\mathrm{d}\nu - \int c\;\mathrm{d}(q,q)_{\#}\gamma.\end{aligned} \tag{32}$$

Thus only mutual information and the $c$-interaction vary. Lemma B.4 minimizes the conditional information at fixed coarse coupling, yielding $\mathsf{L}_{\mu,\nu}\Gamma$ and the coarse problem (29). The attained full minimizer gives coarse attainment, and strict convexity gives uniqueness of both optimizers.

For the path statement, disintegrate any feasible path law $P$ with endpoint coupling $\gamma$:

$$\mathrm{KL}(P\|R) = \mathrm{KL}(\gamma\|r) + \int \mathrm{KL}(P^{xy}\|R^{xy})\,\gamma(\mathrm{d}x\ \mathrm{d}y). \tag{33}$$

The conditional term is minimized uniquely by $P^{xy} = R^{xy}$, which proves (31). □

**Remark B.6** (Explicit endpoint scalings)**.** *Suppose, in addition, that* $r_0(\mathrm{d}x) = r_0^{\mathsf{B}}(\mathrm{d}\beta)\rho_\beta(\mathrm{d}x)$ *and* $r_T(\mathrm{d}y) = r_T^{\mathsf{B}}(\mathrm{d}\beta')\rho'_{\beta'}(\mathrm{d}y)$*. Whenever the endpoint conditionals have densities, define*

$$s_0(x) := \frac{\mathrm{d}\mu_{q(x)}}{\mathrm{d}\rho_{q(x)}}(x), \qquad s_T(y) := \frac{\mathrm{d}\nu_{q(y)}}{\mathrm{d}\rho'_{q(y)}}(y).$$

*Admissibility implies* $\mu \ll r_0$ *and* $\nu \ll r_T$*, so define*

$$u_0 := \frac{\mathrm{d}\mu}{\mathrm{d}r_0}, \qquad u_T := \frac{\mathrm{d}\nu}{\mathrm{d}r_T}.$$

*If* $\bar{f}$ *and* $\bar{g}$ *are coarse Schrödinger factors normalized so that*

$$\frac{\mathrm{d}\Gamma^\star}{\mathrm{d}(\mu^{\mathsf{B}} \otimes \nu^{\mathsf{B}})}(\beta,\beta') = \bar{f}(\beta)\bar{g}(\beta')e^{c(\beta,\beta')}, \tag{34}$$

*then one may choose full factors relative to* $r$ *in the form*

$$f(x) = u_0(x)e^{-a(x)}\bar{f}(q(x)), \qquad g(y) = u_T(y)e^{-b(y)}\bar{g}(q(y)). \tag{35}$$

*Equivalently,* $u_0(x) = s_0(x)\frac{\mathrm{d}\mu^{\mathsf{B}}}{\mathrm{d}r_0^{\mathsf{B}}}(q(x))$ *and analogously at time* $T$*. The full likelihood ratios consequently combine the coarse marginal ratios with the original within-fiber conditional densities. The resulting boundary scalings retain both kinds of prescribed endpoint detail.*

**Theorem B.7** (Uniform coupling and path approximation)**.** *Let* $h \in \mathcal{G}_q$*, written explicitly as* $h(x,y) = a(x) + b(y) + c(q(x), q(y))$*, satisfy*

$$\|\ell_R - h\|_\infty \le \eta < \infty, \tag{36}$$

*and define the finite positive endpoint measure*

$$\widehat{r}(\mathrm{d}x\ \mathrm{d}y) := e^{h(x,y)}r_0(\mathrm{d}x)r_T(\mathrm{d}y). \tag{37}$$

*For any* $(\mu,\nu) \in \mathfrak{A}_R$*, let* $\gamma^\star$ *minimize* $\mathcal{H}(\cdot \mid r)$ *and let* $\widehat{\gamma}^\star$ *minimize* $\mathcal{H}(\cdot \mid \widehat{r})$ *over* $\Pi(\mu,\nu)$*. Since* $\widehat{r}(\mathrm{d}x\ \mathrm{d}y) = e^{a(x)+b(y)+c(q(x),q(y))}r_0(\mathrm{d}x)r_T(\mathrm{d}y)$*, the same conditional- information argument as in Theorem B.5 computes* $\widehat{\gamma}^\star$ *by a coarse problem and the exact endpoint-conditional lift. Uniformly over all admissible endpoint pairs,*

$$\mathrm{KL}(\widehat{\gamma}^\star\|\gamma^\star) \le 2\eta, \qquad \mathrm{KL}(\gamma^\star\|\widehat{\gamma}^\star) \le 2\eta. \tag{38}$$

*Pinsker's inequality therefore gives*

$$\|\widehat{\gamma}^\star - \gamma^\star\|_{\mathrm{TV}} \le \sqrt{\eta}. \tag{39}$$

*Define a path law using the* original *reference bridges,*

$$P^\star := \int R^{xy}\,\gamma^\star(\mathrm{d}x\ \mathrm{d}y), \qquad \widehat{P} := \int R^{xy}\,\widehat{\gamma}^\star(\mathrm{d}x\ \mathrm{d}y). \tag{40}$$

*Then* $\widehat{P}_0 = \mu$*,* $\widehat{P}_T = \nu$ *and*

$$\mathrm{KL}(\widehat{P}\|P^\star) = \mathrm{KL}(\widehat{\gamma}^\star\|\gamma^\star) \le 2\eta, \qquad \mathrm{KL}(P^\star\|\widehat{P}) = \mathrm{KL}(\gamma^\star\|\widehat{\gamma}^\star) \le 2\eta. \tag{41}$$

*Thus the approximate solve preserves both endpoints exactly and uses the same conditional reference bridge as* $R$ *for every endpoint pair. Its error lies in the endpoint coupling.*

*Proof.* Write $F(\gamma) = \mathcal{H}(\gamma \mid r)$ and $\widehat{F}(\gamma) = \mathcal{H}(\gamma \mid \widehat{r})$. The uniform log-density bound makes the two references equivalent with bounded likelihood ratio. Thus they have the same nonempty finite-entropy feasible class, and the compactness/lower-semicontinuity argument used in Theorem B.5 gives both minimizers. Every finite-entropy feasible $\gamma$ also satisfies

$$|F(\gamma) - \widehat{F}(\gamma)| \le \eta. \tag{42}$$

Combining the two variational inequalities yields

$$F(\widehat{\gamma}^\star) \le \widehat{F}(\widehat{\gamma}^\star) + \eta \le \widehat{F}(\gamma^\star) + \eta \le F(\gamma^\star) + 2\eta. \tag{43}$$

The Schrödinger optimizer is the information projection onto the convex set of couplings with the prescribed marginals. The information-projection Pythagorean inequality (Csiszár, 1975) therefore gives, for every coupling with those marginals,

$$F(\gamma) - F(\gamma^\star) \ge \mathrm{KL}(\gamma \| \gamma^\star). \tag{44}$$

Combining (43) and (44) gives the first inequality in (38). Reversing the roles of $r$ and $\widehat{r}$ and using the Pythagorean inequality for $\widehat{\gamma}^\star$ gives the second on the same finite-entropy feasible class. Pinsker's inequality gives (39). Both $P^\star$ and $\widehat{P}$ use the same conditional law $R^{xy}$ given the endpoints. Their conditional relative entropies vanish, so the chain rule equates both path and coupling divergences in (41). □

**Corollary B.8** (Optimal uniform residual)**.** *For every $\delta > 0$, choose $h_\delta \in \mathcal{G}_q$ with $\|\ell_R - h_\delta\|_\infty \le \eta_q(R) + \delta$. The corresponding coarse lift satisfies Theorem B.7 with $\eta = \eta_q(R) + \delta$. If the infimum in Definition B.2 is attained, its conditional lift satisfies the bound $2\eta_q(R)$.*

**Proposition B.9** (Refinement cannot increase the best residual)**.** *Let $q_{\mathrm{f}} : \mathsf{X} \to \mathsf{B}_{\mathrm{f}}$ refine $q_{\mathrm{c}} : \mathsf{X} \to \mathsf{B}_{\mathrm{c}}$, in the sense that $q_{\mathrm{c}} = \phi \circ q_{\mathrm{f}}$ for a measurable map $\phi$. Then*

$$\eta_{q_{\mathrm{f}}}(R) \le \eta_{q_{\mathrm{c}}}(R). \tag{45}$$

*Thus a finer partition can represent every interaction available to the coarser one. Its additional computation and estimation costs purchase a weakly smaller best population residual.*

*Proof.* Since $q_{\mathrm{c}} = \phi \circ q_{\mathrm{f}}$, $\mathcal{G}_{q_{\mathrm{c}}} \subseteq \mathcal{G}_{q_{\mathrm{f}}}$. Taking infima over these nested classes proves the claim. □

**Proposition B.10** (Cellwise residual certificate)**.** *In the setting of Theorem B.7, replace (36) by a bounded nonnegative coarse envelope*

$$|\ell_R(x,y) - h(x,y)| \le \delta(q(x), q(y)). \tag{46}$$

*For coarse endpoint marginals $p_\mu = q_\# \mu$ and $p_\nu = q_\# \nu$, define*

$$U_\delta(p_\mu, p_\nu) := \sup_{\Gamma \in \Pi(p_\mu, p_\nu)} \int \delta \, \mathrm{d}\Gamma. \tag{47}$$

*Then both KL directions between the exact and reduced endpoint couplings are at most $2U_\delta(p_\mu, p_\nu)$. The same bounds hold for path laws reconstructed with the common conditional reference bridges. In particular, $U_\delta(p_\mu, p_\nu) \le \|\delta\|_\infty$ recovers Theorem B.7, while (47) gives a sharper, computable endpoint-weighted certificate when mixing quality varies across coarse-cell pairs.*

*Proof.* Write $e = \ell_R - h$. The two objectives differ by $-\int e \, \mathrm{d}\gamma$. Optimality of their minimizers and the Pythagorean inequality used in Theorem B.7 give

$$\mathrm{KL}(\widehat{\gamma}^\star \| \gamma^\star) \le \int |e| \, \mathrm{d}\widehat{\gamma}^\star + \int |e| \, \mathrm{d}\gamma^\star. \tag{48}$$

Both coarse pushforwards belong to $\Pi(p_\mu, p_\nu)$, so each integral is at most $U_\delta(p_\mu, p_\nu)$. Reversing the two reference measures proves the other KL direction. The path statements follow from the conditional entropy chain rule because the conditional reference bridge is shared. □

### B.2.1 Query-dependent error and fixed-partition approximation

Uniform error measures the hardest endpoint pair. For a given query, the same representation can be much more accurate because its marginals place little mass on difficult cell pairs. We now separate the error caused by the deployed coarse kernel from the dependence that no fixed-conditional lift can represent. The proof combines the Pythagorean identity for information projections (Csiszár, 1975) with the variational formula for exponential moments (Dupuis et al., 2016), giving query-weighted bounds in the spirit of Gourgoulias et al. (2020), evaluated under the available coarse lift.

**Proposition B.11** (Query-dependent error decomposition and bounds)**.** *Let $N \geq 1$, $1 \leq B \leq N$, and $K \in \mathbb{R}_{>0}^{N\times N}$. Fix a surjective partition $q : [N] \to [B]$, a coarse kernel $W \in \mathbb{R}_{>0}^{B\times B}$, and probability vectors $\mu, \nu$. Write $I_0 = \operatorname{supp}\mu$, $I_T = \operatorname{supp}\nu$, $p = q_{\#}\mu$, and $r = q_{\#}\nu$, and omit zero-mass coarse rows and columns. For $a \in q(I_0)$ and $b \in q(I_T)$, let $\mu_a(i) = \mu_i/p_a$ on $q^{-1}(a) \cap I_0$, $\nu_b(j) = \nu_j/r_b$ on $q^{-1}(b) \cap I_T$, and $s_{ab}(i,j) = \mu_a(i)\nu_b(j)$ on this cell pair. Let $\widehat{\Gamma}$ be the scaling of $W$ to $p, r$, let $P = \mathsf{L}_{\mu,\nu}\widehat{\Gamma}$, and let $Q$ be the optimizer for $K$ with marginals $\mu, \nu$. Define*

$$e_{ij} = \log K_{ij} - \log W_{q(i),q(j)}, \qquad \bar{e}_{ab} = \sum_{i\in q^{-1}(a)\cap I_0} \sum_{j\in q^{-1}(b)\cap I_T} s_{ab}(i,j)e_{ij}. \tag{49}$$

*Let $\Gamma^\dagger$ be the scaling of $W_{ab}\exp(\bar{e}_{ab})$ to $p, r$, and set $P^\dagger = \mathsf{L}_{\mu,\nu}\Gamma^\dagger$. The identity*

$$W_{ab}\exp(\bar{e}_{ab}) = \exp\left(\sum_{i,j} s_{ab}(i,j)\log K_{ij}\right)$$

*shows that this corrected kernel, $P^\dagger$, and its representation error are independent of the deployed $W$. Then*

$$\begin{aligned} F &:= \mathrm{KL}(P\|Q) = C + J, \qquad C := \mathrm{KL}(\widehat{\Gamma}\|\Gamma^\dagger), \\ J &:= \mathrm{KL}(P^\dagger\|Q) = \min_{\Lambda\in\Pi(p,r)} \mathrm{KL}(\mathsf{L}_{\mu,\nu}\Lambda\|Q). \end{aligned} \tag{50}$$

*The minimizing $\Lambda$ is unique. Since $J \geq 0$, we have $C \leq F$. Here $J$ measures the best forward KL within the specified class of endpoint-conditional lifts. For finite functions $u : I_0 \to \mathbb{R}$ and $v : I_T \to \mathbb{R}$, put*

$$g_{ij} = e_{ij} - u_i - v_j, \qquad M_g(t) = \log \mathbb{E}_P \exp\{t(g - \mathbb{E}_P g)\}, \qquad G := \mathrm{KL}(Q\|P). \tag{51}$$

*The forward bound holds at $t = 1$. For every $t > 1$, the two further inequalities below control reverse error:*

$$F \leq M_g(1), \qquad tF + (t-1)G \leq M_g(t), \qquad G \leq \frac{M_g(t) - tC}{t-1}. \tag{52}$$

*All results use the active rectangle $I_0 \times I_T$ and allow arbitrary positive masses on that support.*

*Proof.* For a positive reference matrix $A$, write $\mathcal{F}_A(Z) = \sum_{ij} Z_{ij}\log(Z_{ij}/A_{ij})$. If $Z_A$ minimizes this objective over fixed marginals, positivity gives $\log((Z_A)_{ij}/A_{ij}) = \alpha_i + \beta_j$ on the active rectangle. Consequently, every feasible $Z$ satisfies

$$\mathcal{F}_A(Z) - \mathcal{F}_A(Z_A) = \mathrm{KL}(Z\|Z_A), \tag{53}$$

because subtracting the right side leaves $\sum_{ij}(Z - Z_A)_{ij}(\alpha_i + \beta_j) = 0$. The same equality holds on the active coarse rectangle, whose constraints are affine and optimizers positive.

Substitute $Z_{ij} = \Lambda_{ab}s_{ab}(i,j)$, where $a = q(i)$ and $b = q(j)$, into the fine objective. The conditional entropy terms separate into a function of $a$ and a function of $b$, so their expectation is constant for $\Lambda \in \Pi(p, r)$. Therefore

$$\mathcal{F}_K(\mathsf{L}_{\mu,\nu}\Lambda) = \sum_{ab} \Lambda_{ab}\log\frac{\Lambda_{ab}}{W_{ab}\exp(\bar{e}_{ab})} + c_{\mu,\nu}, \tag{54}$$

where $c_{\mu,\nu} = \sum_i \mu_i \log \mu_{q(i)}(i) + \sum_j \nu_j \log \nu_{q(j)}(j)$ does not depend on $\Lambda$. Strict convexity identifies the unique restricted minimizer as $P^\dagger$. Applying (53) to the fine and coarse problems gives

$$\begin{aligned} \mathrm{KL}(P\|Q) &= \mathcal{F}_K(P) - \mathcal{F}_K(P^\dagger) + \mathcal{F}_K(P^\dagger) - \mathcal{F}_K(Q) \\ &= \mathrm{KL}(\widehat{\Gamma}\|\Gamma^\dagger) + \mathrm{KL}(P^\dagger\|Q). \end{aligned} \tag{55}$$

The fine identity identifies $J$ in (50) as the minimum forward KL to $Q$ among fixed-conditional lifts.

For the moment bound, set $Z_g = \mathbb{E}_P e^g$ and $V_{ij} = P_{ij} e^{g_{ij}}/Z_g$. Since $P$ is a positive row and column scaling of $W_{q(i),q(j)}$, the matrix $V$ is a positive row and column scaling of $K$. Its fixed-marginal information projection is consequently $Q$. Equation (53) implies

$$M_g(1) = \mathrm{KL}(P\|V) = \mathrm{KL}(P\|Q) + \mathrm{KL}(Q\|V) \geq F. \tag{56}$$

Moreover, $\log(Q_{ij}/P_{ij})$ differs from $g_{ij}$ by a row plus column function. Equal marginals therefore give the exact identity

$$F + G = \sum_{ij} (Q_{ij} - P_{ij}) g_{ij} = \mathbb{E}_Q g - \mathbb{E}_P g. \tag{57}$$

The entropy variational formula $\log \mathbb{E}_P e^f = \sup_{Z \ll P} \{\mathbb{E}_Z f - \mathrm{KL}(Z\|P)\}$, applied to $f = t(g - \mathbb{E}_P g)$ and evaluated at $Z = Q$, yields

$$M_g(t) \geq t(\mathbb{E}_Q g - \mathbb{E}_P g) - G = tF + (t-1)G.$$

Since $t > 1$ and $F \geq C$, rearrangement gives the last two inequalities in (52). They control reverse KL through moments under $P$, with $t > 1$ accounting for the change of measure. □

**Corollary B.12** (Exact representation-error reduction under refinement)**.** *Fix a positive finite kernel $K$, endpoint probabilities $\mu, \nu$, and their exact coupling $Q$. For a partition $q$, define the conditional-lift family and its best forward-KL approximation by*

$$\mathcal{M}_q(\mu,\nu) := \{\mathsf{L}^q_{\mu,\nu}\Gamma : \Gamma \in \Pi(q_\#\mu, q_\#\nu)\},$$
$$P^\dagger_q := \arg\min_{P \in \mathcal{M}_q(\mu,\nu)} \mathrm{KL}(P\|Q), \tag{58}$$

*where the superscript on $\mathsf{L}^q$ identifies the partition used for conditioning, and put $J_q := \mathrm{KL}(P^\dagger_q\|Q)$. If $q_\mathrm{f}$ refines $q_\mathrm{c}$, so that $q_\mathrm{c} = \phi \circ q_\mathrm{f}$ for a map $\phi$, then*

$$J_{q_\mathrm{c}} = \mathrm{KL}(P^\dagger_{q_\mathrm{c}}\|P^\dagger_{q_\mathrm{f}}) + J_{q_\mathrm{f}}. \tag{59}$$

*Consequently, for any nested sequence $q_0, \ldots, q_L$ with $q_{\ell+1}$ refining $q_\ell$,*

$$J_{q_0} - J_{q_L} = \sum_{\ell=0}^{L-1} \mathrm{KL}(P^\dagger_{q_\ell}\|P^\dagger_{q_{\ell+1}}) \geq 0. \tag{60}$$

*At singleton resolution $J_q = 0$. The identities also hold for sparse endpoints after restricting the coupling and all conditional laws to their positive support, with no lower bound on nonzero masses.*

*Proof.* Let $p_u, r_v$ be the endpoint masses of fine cells $u, v$, and let $p^\mathrm{c}_a, r^\mathrm{c}_b$ be their coarse-cell masses. For any $\Gamma^\mathrm{c} \in \Pi((q_\mathrm{c})_\#\mu, (q_\mathrm{c})_\#\nu)$, define on positive-mass fine cells

$$\Gamma^\mathrm{f}_{uv} := \Gamma^\mathrm{c}_{\phi(u),\phi(v)} \frac{p_u}{p^\mathrm{c}_{\phi(u)}} \frac{r_v}{r^\mathrm{c}_{\phi(v)}}.$$

Summing over $v$ or $u$ gives the fine-cell marginals $p_u, r_v$. Substituting the endpoint conditionals into the two lifts gives $\mathsf{L}^{q_\mathrm{f}}_{\mu,\nu}\Gamma^\mathrm{f} = \mathsf{L}^{q_\mathrm{c}}_{\mu,\nu}\Gamma^\mathrm{c}$. Thus $\mathcal{M}_{q_\mathrm{c}}(\mu,\nu) \subseteq \mathcal{M}_{q_\mathrm{f}}(\mu,\nu)$.

For any $P = \mathsf{L}^{q_\mathrm{f}}_{\mu,\nu}\Lambda$ in the finer family, the restricted objective calculation (54) and the affine scaling identity (53) give

$$\mathrm{KL}(P\|Q) - \mathrm{KL}(P^\dagger_{q_\mathrm{f}}\|Q) = \mathrm{KL}(\Lambda\|\Gamma^\dagger_{q_\mathrm{f}}) = \mathrm{KL}(P\|P^\dagger_{q_\mathrm{f}}),$$

where $\Gamma^\dagger_{q_\mathrm{f}}$ is the coarse coupling whose lift is $P^\dagger_{q_\mathrm{f}}$. The second equality uses the entropy chain rule: both lifted measures have the same conditional product law in each fine cell pair. Taking $P = P^\dagger_{q_\mathrm{c}}$ proves (59). Summing the identity gives (60). For singleton cells, $Q$ belongs to the lift family, so $J_q = 0$. □

The identity gives two complementary ways to improve a representation. Changing the deployed coarse kernel addresses $C$, while subdividing cells enlarges the lift family and reduces its optimal error $J$ by the displayed KL term. All partitions can be constructed before queries, retaining earlier cell boundaries. The deployed error also includes the separately fitted coarse-kernel term $C$.

**Gauge invariance and evaluation cost.** Changing $e$ by $u_i + v_j$ changes $\bar{e}_{ab}$ by $\mathbb{E}_{\mu_a} u + \mathbb{E}_{\nu_b} v$, a coarse row plus column function. Thus $\Gamma^\dagger$, $C$, and $J$ are unchanged. One may equivalently scale $\widehat{\Gamma}_{ab} \exp(\bar{e}_{ab})$, since $\widehat{\Gamma}/W$ is itself a coarse row factor times a coarse column factor. Computing the conditional averages and a log moment generally requires $O(|I_0||I_T|)$ reference accesses and arithmetic, followed by a coarse solve for $\Gamma^\dagger$. Blockwise accumulation avoids an additional dense matrix but does not remove this arithmetic. The resulting query-dependent projection explains the error of the unchanged reference-only representation. Its dense scans belong to this analysis, separately from the online solve with the deployed reference-only representation.

The audit compares the frozen reference-only gauge, independent-product centering, and one or two fixed alternating row and column normalizations. For a current residual $g$ and the same returned coupling $P$, a row normalization subtracts $\log \sum_j P_{ij} e^{g_{ij}} - \log \mu_i$ from row $i$. A column normalization analogously subtracts $\log \sum_i P_{ij} e^{g_{ij}} - \log \nu_j$ after the row update. These additive gauges leave $P$ unchanged. Their dense-pass and moment-evaluation costs are measured separately from online solution time. Taking the minimum over tested gauges and reverse exponents $t > 1$ preserves these bounds in exact arithmetic. Evaluations use the stated numerical precision and solver tolerances.

The relation to exact scaling also identifies the cost of optimizing the gauge without restricting the number of passes:

$$\inf_{u,v} M_{e-u-v}(1) = F. \tag{61}$$

Indeed, the convex objective has derivatives $\mu_i - \sum_j V_{ij}$ and $\nu_j - \sum_i V_{ij}$. At a minimizer, the tilted matrix has the prescribed marginals and equals $Q$, making the last KL term in (56) vanish. Positive finite matrix scaling supplies such a minimizer. Fully optimizing the gauge solves the full-state problem. A fixed number of normalizations gives an intermediate bound at its stated evaluation cost.

**Proposition B.13** (Approximate conditional-bridge sampling)**.** *Under Theorem B.7, let $\widetilde{R}^{xy}$ be a measurable kernel on paths from $x$ to $y$, and define the sampled path law by*

$$\widehat{P}_{\mathrm{sam}} := \int \widetilde{R}^{xy}\, \widehat{\gamma}^\star(\,\mathrm{d}x\ \mathrm{d}y). \tag{62}$$

*If the following bounds hold for $\widehat{\gamma}^\star$-almost every $(x, y)$ in the forward direction and for $\gamma^\star$-almost every $(x, y)$ in the reverse direction,*

$$\mathrm{KL}(\widetilde{R}^{xy} \| R^{xy}) \le \delta_+, \qquad \mathrm{KL}(R^{xy} \| \widetilde{R}^{xy}) \le \delta_-, \tag{63}$$

*then $\widehat{P}_{\mathrm{sam}}$ has endpoints $(\mu, \nu)$ and*

$$\mathrm{KL}(\widehat{P}_{\mathrm{sam}} \| P^\star) \le 2\eta + \delta_+, \qquad \mathrm{KL}(P^\star \| \widehat{P}_{\mathrm{sam}}) \le 2\eta + \delta_-. \tag{64}$$

*More generally, the uniform $\delta_\pm$ can be replaced by their integrals against $\widehat{\gamma}^\star$ and $\gamma^\star$, respectively.*

*Proof.* The endpoint support of $\widetilde{R}^{xy}$ preserves the endpoint coupling $\widehat{\gamma}^\star$, hence the prescribed marginals. Disintegrating both path laws by their endpoints gives the exact identities

$$\mathrm{KL}(\widehat{P}_{\mathrm{sam}} \| P^\star) = \mathrm{KL}(\widehat{\gamma}^\star \| \gamma^\star) + \int \mathrm{KL}(\widetilde{R}^{xy} \| R^{xy})\, \widehat{\gamma}^\star(\,\mathrm{d}x\ \mathrm{d}y), \tag{65}$$

$$\mathrm{KL}(P^\star \| \widehat{P}_{\mathrm{sam}}) = \mathrm{KL}(\gamma^\star \| \widehat{\gamma}^\star) + \int \mathrm{KL}(R^{xy} \| \widetilde{R}^{xy})\, \gamma^\star(\,\mathrm{d}x\ \mathrm{d}y). \tag{66}$$

Applying Theorem B.7 and the two bounds in (63) completes both directional KL estimates. □

### B.2.2 Finite kernels: a constructive criterion and a query converse

The query analysis describes a chosen endpoint pair. We next return to uniform accuracy and ask which reference distinctions some query can detect. On a finite state space, log cross-ratios answer this question and also construct a positive coarse kernel. Let $N \ge 1$, write $[N] = \{1, \dots, N\}$, and let $K \in \mathbb{R}^{N \times N}_{>0}$ be strictly positive. It may be a nonreversible transition matrix or a finite positive reference matrix. Let $1 \le B \le N$ and let $q : [N] \to [B]$ be a surjective partition. Put $L_{ij} = \log K_{ij}$ and define

$$\eta_q(K) := \min_{a,b \in \mathbb{R}^N,\, C \in \mathbb{R}^{B \times B}} \max_{i,j \in [N]} |L_{ij} - a_i - b_j - C_{q(i),q(j)}|. \tag{67}$$

The minimum exists because the fitted matrices form a closed linear subspace in finite dimensions. Positive row and column scaling of $K$ leaves the residual unchanged, so it equals that of the normalized reference interaction, whose logarithm differs only by endpoint terms.

For probability vectors $\mu, \nu$ on $[N]$, define

$$\gamma_K^{\mu,\nu} := \operatorname*{arg\,min}_{\gamma \in \Pi(\mu,\nu)} \sum_{i,j} \gamma_{ij} \log \frac{\gamma_{ij}}{K_{ij}}. \tag{68}$$

Normalizing $K$ changes this objective by a constant. For a positive coarse matrix $W \in \mathbb{R}_{>0}^{B\times B}$, let $\Gamma_W^{\mu,\nu}$ be the analogous optimizer with reference $W$ and marginals $q_\#\mu, q_\#\nu$, and write $\widehat{\gamma}_W^{\mu,\nu} := \mathsf{L}_{\mu,\nu}\Gamma_W^{\mu,\nu}$. The conditional-information decomposition identifies it as the full-state optimizer for $W_{q(i),q(j)}$. For sparse endpoints, restrict all arguments to $\operatorname{supp}(\mu) \times \operatorname{supp}(\nu)$, where the optimizers are positive. Zero-mass rows and columns need neither entropy terms nor conditional ratios.

**Theorem B.14** (Constructive finite-kernel criterion)**.** *For distinct $i, i'$ and distinct $j, j'$, define the log cross-ratio $\delta_{ii';jj'} := L_{ij} + L_{i'j'} - L_{ij'} - L_{i'j}$. Let*

$$D_q(K) := \max_{\substack{i\neq i',\ j\neq j' \\ q(i)=q(i') \text{ or } q(j)=q(j')}} \left|\log \frac{K_{ij}K_{i'j'}}{K_{ij'}K_{i'j}}\right|, \qquad \frac{D_q(K)}{4} \le \eta_q(K) \le D_q(K), \tag{69}$$

*where the maximum of an empty set is defined as zero. Thus exact reduction is equivalent to vanishing log cross-ratios whenever either pair lies in one coarse cell. The criterion allows arbitrary row and column factors and imposes no symmetry on the kernel or the endpoint laws.*

*Choose row representatives $i_\beta \in q^{-1}(\beta)$ and column representatives $j_\kappa \in q^{-1}(\kappa)$ for $\beta, \kappa \in [B]$. For a finite real-valued family $v$, write $\operatorname{mid}(v) := (\max v + \min v)/2$ and $\operatorname{osc}(v) := \max v - \min v$. Define*

$$\begin{aligned} r_i(j) &:= L_{ij} - L_{i_{q(i)},j}, & a_i &:= \operatorname{mid}_{j\in[N]} r_i(j), \\ s_j(\beta) &:= L_{i_\beta,j} - L_{i_\beta,j_{q(j)}}, & b_j &:= \operatorname{mid}_{\beta\in[B]} s_j(\beta), \\ C_{\beta\kappa} &:= L_{i_\beta,j_\kappa}, & W_{\beta\kappa} &:= e^{C_{\beta\kappa}} = K_{i_\beta,j_\kappa}. \end{aligned} \tag{70}$$

*With $d_r := \max_i \operatorname{osc}_j r_i(j)$, $d_c := \max_j \operatorname{osc}_\beta s_j(\beta)$, and $A_q := \max\{d_r, d_c\}$, this positive coarse matrix satisfies*

$$\|L - a - b - C \circ (q,q)\|_\infty \le \frac{d_r + d_c}{2} \le A_q \le D_q(K), \qquad \frac{A_q}{4} \le \eta_q(K) \le A_q. \tag{71}$$

*Given the full log kernel, the construction and $A_q$ require $O(N^2)$ arithmetic. For $A_q > 0$, a maximizing row or column range identifies an eligible rectangle with absolute log cross-ratio $A_q$.*

*Proof.* For any fitted matrix $a_i + b_j + C_{q(i),q(j)}$, the four terms of an eligible rectangle cancel. Its log cross-ratio therefore equals the corresponding four-term difference of the residual, with absolute value at most four times the residual norm. Taking the infimum gives $D_q(K) \le 4\eta_q(K)$.

For $\beta = q(i)$ and $\kappa = q(j)$, the construction gives the exact identity

$$L_{ij} - a_i - b_j - C_{\beta\kappa} = [r_i(j) - a_i] + [s_j(\beta) - b_j]. \tag{72}$$

Each centered range is bounded by half its oscillation. Differences of $r_i$ are log cross-ratios with rows $i, i_\beta$ in the same cell, while differences of $s_j$ are log cross-ratios with columns $j, j_\kappa$ in the same cell. Degenerate rectangles contribute zero. Hence $d_r, d_c \le D_q(K)$, which proves the upper bounds, while $A_q \le D_q(K) \le 4\eta_q(K)$ gives the remaining lower bound. Evaluating all $r_i$ uses $N^2$ entries and all $s_j$ uses $BN \le N^2$ entries. Nonzero extrema identify the asserted rectangle without computing $D_q(K)$. Its exhaustive evaluation uses $O(N^3)$ row-pair and column-pair range scans. □

**Theorem B.15** (Quadratic finite-kernel perturbation)**.** *Suppose $h_{ij} = a_i + b_j + C_{q(i),q(j)}$ and $\|\log K - h\|_\infty \le \eta$ for $\eta \ge 0$. Set $W = e^C$ entrywise. For every pair of endpoint probabilities, let $P = \widehat{\gamma}_W^{\mu,\nu}$ and $Q = \gamma_K^{\mu,\nu}$. Then*

$$\mathrm{KL}(P\|Q) \vee \mathrm{KL}(Q\|P) \le \min\{2\eta, \eta^2/2\}, \qquad \|P - Q\|_{\mathrm{TV}} \le \tanh(\eta/2). \tag{73}$$

*Moreover, their sum $J := \mathrm{KL}(P\|Q) + \mathrm{KL}(Q\|P)$ obeys $J \le 2\eta \tanh(\eta/2) \le \eta^2$. The coefficient $1/2$ in the individual quadratic bounds is asymptotically sharp. The entropy chain rule gives the same directional KL bounds for path laws sharing the original conditional reference bridges.*

*Proof.* Work on the common active support rectangle and put $e = \log K - h$. Diagonal scaling gives $P_{ij} = u_i e^{h_{ij}} v_j$ for positive scalings $u, v$. Let $Z = \sum_{i,j} P_{ij} e^{e_{ij}}$ and define the probability $V_{ij} = P_{ij} e^{e_{ij}}/Z$. Since $V$ differs from $K$ only by positive row and column factors, $Q$ is its relative-entropy projection onto the same marginal constraints. In particular, $\log(Q_{ij}/V_{ij})$ is a row function plus a column function. Pairing it with $P - Q$ gives zero and proves

$$\mathrm{KL}(P\|V) = \mathrm{KL}(P\|Q) + \mathrm{KL}(Q\|V). \tag{74}$$

The last term is nonnegative, so

$$\mathrm{KL}(P\|Q) \le \log \mathbb{E}_P e^e - \mathbb{E}_P e \le \frac{(\max e - \min e)^2}{8} \le \frac{\eta^2}{2}. \tag{75}$$

Hoeffding's log-moment estimate (Hoeffding, 1963) gives the middle inequality: for a bounded random variable $X$, let $\psi(s) = \log \mathbb{E} e^{s(X - \mathbb{E}X)}$. Then $\psi(0) = \psi'(0) = 0$ and $\psi''(s)$ is the variance of $X$ under its exponential tilt. On an interval of length $d$, variance is at most $d^2/4$ under every probability law. Integrating this bound twice on $[0, 1]$ gives $\psi(1) \le d^2/8$. Swapping $K$ and $e^h$, and replacing $e$ by $-e$, proves the other KL direction. The previous perturbation bound also supplies $2\eta$.

For the total-variation estimate, diagonal scaling also gives $\log(Q_{ij}/P_{ij}) = e_{ij} + \alpha_i + \beta_j$ for some real vectors $\alpha, \beta$. Marginal cancellation yields

$$J = \sum_{i,j} (Q_{ij} - P_{ij}) e_{ij} \le 2\eta t, \qquad t := \|P - Q\|_{\mathrm{TV}}. \tag{76}$$

If $t > 0$, let $E = \{(i, j) : Q_{ij} \ge P_{ij}\}$ and put $a = Q(E)$, $b = P(E)$, so $a - b = t$. Applying the log-sum inequality to $E$ and its complement in both directions gives

$$J \ge t \log \frac{a(1-b)}{b(1-a)} \ge 2t \log \frac{1+t}{1-t} = 4t \operatorname{artanh}(t). \tag{77}$$

For the second inequality, use $b(1-a) = b(1-b-t) \le (1-t)^2/4$ and $a(1-b) = b(1-a) + t$. Combining the bounds and cancelling $t$ gives $t \le \tanh(\eta/2)$, which also holds when $t = 0$. Substitution into (76) gives the bound on $J$. The next proof gives a sharp two-state example. □

**Corollary B.16** (Endpoint-query converse and minimax error)**.** *Define the best endpoint-uniform error among frozen positive coarse kernels by*

$$\mathcal{R}_q(K) := \inf_{W \in \mathbb{R}_{>0}^{B \times B}} \sup_{\mu,\nu \in \mathcal{P}([N])} \max\left\{\mathrm{KL}(\widehat{\gamma}_W^{\mu,\nu} \| \gamma_K^{\mu,\nu}), \mathrm{KL}(\gamma_K^{\mu,\nu} \| \widehat{\gamma}_W^{\mu,\nu})\right\}. \tag{78}$$

*Then*

$$\log\cosh \frac{\eta_q(K)}{4} \le \log\cosh \frac{D_q(K)}{4} \le \mathcal{R}_q(K) \le \min\{2\eta_q(K), \eta_q(K)^2/2\}. \tag{79}$$

*The lower bound is witnessed by an endpoint pair supported on at most two states on each side. It applies even to endpoint-dependent algorithms whose outputs always have the form* $\mathsf{L}_{\mu,\nu}\Gamma$*. The anchored quantity gives the directly witnessed lower bound* $\log\cosh(A_q/4)$*. Consequently, universal exact conditional lifting is possible exactly when* $\eta_q(K) = 0$*, and its best uniform KL error is quadratic in a small residual, up to constants independent of* $N$ *and* $B$*, even for sparse queries.*

*Proof.* Take an eligible nondegenerate rectangle with log cross-ratio $\delta$ and choose $\mu = (\delta_i + \delta_{i'})/2$ and $\nu = (\delta_j + \delta_{j'})/2$, where $\delta_i$ denotes the unit point mass at state $i$. At least one coarse marginal is a point mass, so the coarse coupling is fixed by its other marginal. Its exact conditional lift is therefore $P = \mu \otimes \nu$, independently of the coarse kernel or the algorithm choosing the coarse coupling. The full optimizer on this support is

$$Q = \frac{1}{2}\begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}, \qquad p = (1 + e^{-\delta/2})^{-1}. \tag{80}$$

Equal marginals force this matrix form, and diagonal scaling preserves its odds ratio, which equals $e^\delta$. Writing $u = \delta/4$, direct substitution gives

$$\mathrm{KL}(P\|Q) = \log\cosh u, \qquad \mathrm{KL}(Q\|P) = u \tanh u - \log\cosh u. \tag{81}$$

Choosing a maximizing rectangle gives the lower bound when $D_q(K) > 0$. For $D_q(K) = 0$, nonnegativity gives the same conclusion. The construction at an anchored maximizer proves the $A_q$ bound. All these sparse endpoint pairs are admissible because the finite kernel is positive.

Use a minimizer of (67) and apply Theorem B.15 to $W = e^C$ for the upper bound. The inequality $\eta_q(K) \le D_q(K)$ gives the first term in (79). The expansion $\log\cosh x = x^2/2 + O(x^4)$ at zero proves the dimension-free quadratic comparison, and Theorem B.14 characterizes exact lifting.

Finally, take $N = 2$, one coarse cell, and $\log K = \left(\begin{smallmatrix} \eta & -\eta \\ -\eta & \eta \end{smallmatrix}\right)$. Its cross-ratio is $4\eta$, and the zero fit has residual $\eta$, so (69) implies $\eta_q(K) = \eta$. The half–half query has $\mathrm{KL}(P\|Q) = \log\cosh\eta = \eta^2/2 + O(\eta^4)$, proving sharpness of the leading coefficient in Theorem B.15. Row normalization makes this example a Markov kernel without changing its residual, odds ratios, or transport optimizers. □

**Remark B.17** (Relative kernel approximation)**.** *If $a, b$ and a positive coarse kernel $W$ satisfy $|K_{ij}/(e^{a_i+b_j}W_{q(i),q(j)})-1| \le \rho < 1$ uniformly, with $\rho \ge 0$, their log residual is at most $-\log(1-\rho)$. Theorem B.15 therefore bounds each coupling and shared- conditional path KL by $\min\{-2\log(1-\rho), \frac{1}{2}\log^2(1-\rho)\}$. The condition is a relative approximation on every entry, which directly supplies the logarithmic control used by the theorem for all prescribed endpoint pairs.*

### B.2.3 Reference-only selection of the coarse states

The converse also suggests how to choose a partition. States should share a cell when their incoming and outgoing pairing preferences differ little after endpoint scalings are removed. The resulting distance converts the reference-only clustering problem into a guarantee on bridge-query error.

Let $K \in \mathbb{R}^{N\times N}_{>0}$, $L_{ij} = \log K_{ij}$, and $[N] = \{1, \dots, N\}$. For a finite family $v$, write $\mathrm{osc}(v) = \max v - \min v$ and $\mathrm{mid}(v) = (\max v + \min v)/2$. Define

$$d_K(i, i') = \max\left\{\mathrm{osc}_{j\in[N]}(L_{ij} - L_{i'j}), \mathrm{osc}_{j\in[N]}(L_{ji} - L_{ji'})\right\}. \tag{82}$$

Each term is the Hilbert projective distance between the corresponding positive rows or columns. Their maximum treats the two endpoints on equal footing, including for a nonreversible kernel.

For a surjective $q : [N] \to [B]$ and $W \in \mathbb{R}^{B\times B}_{>0}$, use the full optimizer $\gamma_K^{\mu,\nu}$ and conditional lift $\widehat{\gamma}_{q,W}^{\mu,\nu}$ of the coarse optimizer. Define

$$\begin{aligned} \mathcal{E}_K(q, W) &= \sup_{\mu,\nu\in\mathcal{P}([N])} \max\left\{\mathrm{KL}(\widehat{\gamma}_{q,W}^{\mu,\nu}\|\gamma_K^{\mu,\nu}), \mathrm{KL}(\gamma_K^{\mu,\nu}\|\widehat{\gamma}_{q,W}^{\mu,\nu})\right\}, \\ \mathcal{R}_B^*(K) &= \inf_{q:[N]\to[B]} \inf_{W\in\mathbb{R}^{B\times B}_{>0}} \mathcal{E}_K(q, W). \end{aligned} \tag{83}$$

The supremum includes probability vectors with zero entries and sparse support. As before, each coupling comparison is restricted to its active support rectangle, and zero-mass rows and columns contribute zero to KL. The positive full kernel ensures that all these optimizers are well defined.

**Lemma B.18** (Projective distance and partition diameter)**.** *The function $d_K$ is a pseudometric on $[N]$, invariant under arbitrary positive row and column scalings of $K$. For every surjective partition,*

$$D_q(K) = \max_{b\in[B]} \max_{i,i'\in q^{-1}(b)} d_K(i, i'). \tag{84}$$

*For distinct centers $c_1, \dots, c_B$ and $q$ with $q(c_b) = b$ and $d_K(i, c_{q(i)}) \le r$ for every $i$, the kernel $W_{ab} = K_{c_a,c_b}$ has a row–column fit with residual at most $r$, giving $\mathcal{E}_K(q, W) \le \min\{2r, r^2/2\}$.*

*Proof.* Oscillation is unchanged by a sign or by adding a constant, and $\mathrm{osc}(u+v) \le \mathrm{osc}(u) + \mathrm{osc}(v)$. These identities prove symmetry, the triangle inequality, and invariance under row and column scaling. Nonnegativity and $d_K(i, i) = 0$ follow directly. For fixed rows $i, i'$, their oscillation is

$$\max_{j,j'} |L_{ij} + L_{i'j'} - L_{ij'} - L_{i'j}|.$$

Within-cell row and column maxima give (84), with zero for degenerate rectangles and singletons.

Use $i_b = j_b = c_b$ in (70). Its row oscillations satisfy $d_r \le \max_i d_K(i, c_{q(i)}) \le r$. For every column $j$,

$$\mathrm{osc}_{b\in[B]}(L_{c_b,j} - L_{c_b,c_{q(j)}}) \le \mathrm{osc}_{i\in[N]}(L_{ij} - L_{i,c_{q(j)}}) \le d_K(j, c_{q(j)}) \le r.$$

Thus $d_c \le r$ as well. The anchored residual is bounded by $(d_r + d_c)/2 \le r$, and Theorem B.15 gives the risk bound for this particular $W$, without optimizing or changing the centers. □

**Theorem B.19** (Uniform-risk approximation at a fixed coarse-state budget)**.** *Fix $1 \le B \le N$. Choose $c_1 \in [N]$ and, for $b < B$, choose a distinct unselected state maximizing $\min_{a \le b} d_K(i, c_a)$ as $c_{b+1}$. Let $q_B$ assign each state to a nearest center, require $q_B(c_b) = b$, and break remaining ties arbitrarily. Put*

$$r_B = \max_{i \in [N]} \min_{b \in [B]} d_K(i, c_b), \qquad (W_B)_{ab} = K_{c_a, c_b}.$$

*Then*

$$\log\cosh(r_B/4) \le \mathcal{R}_B^*(K) \le \mathcal{E}_K(q_B, W_B) \le \min\{2r_B, r_B^2/2\}. \tag{85}$$

*For $h(u) = \operatorname{arcosh}(e^u)$, $u \ge 0$, this implies*

$$\mathcal{E}_K(q_B, W_B) \le \min\{8h(\mathcal{R}_B^*(K)),\ 8h(\mathcal{R}_B^*(K))^2\}. \tag{86}$$

*In particular, the upper approximation factor is $16 + O(\mathcal{R}_B^*(K))$ in the small-risk regime. If $\mathcal{R}_B^*(K) = 0$, the selected representation is exact for every endpoint pair, including sparse pairs.*

*Proof.* The upper bound follows from Lemma B.18. For the lower bound, it suffices to show $r_B \le D_q(K)$ for every surjective $B$-cell partition $q$. If $r_B = 0$, this is immediate. Otherwise $B < N$, so choose an additional unselected state $c_{B+1}$ attaining the final radius $r_B$. Greedy covering radii are nonincreasing, so for $1 \le a < b \le B+1$ the selection rule gives $d_K(c_a, c_b) \ge r_B$. Any partition into $B$ cells places two of these $B+1$ states in the same cell. Equation (84) therefore gives $D_q(K) \ge r_B$. The endpoint-query converse now yields

$$\mathcal{E}_K(q, W) \ge \log\cosh(D_q(K)/4) \ge \log\cosh(r_B/4)$$

for all $q$ and positive $W$. Taking infima gives the lower bound without an attainment assumption.

The function $x \mapsto \log\cosh(x/4)$ is increasing on $[0, \infty)$. Its inversion gives $r_B \le 4h(\mathcal{R}_B^*(K))$, which proves (86). The expansion $h(u)^2 = 2u + \frac{2}{3}u^2 + O(u^3)$ at zero gives the stated factor. If the optimal risk is zero, this same inversion forces $r_B = 0$, and the constructive upper bound forces exactness. In particular, when $B = N$, every state is a center and $r_B = 0$. For distinct states at zero pseudometric distance, self-assignment keeps the partition surjective despite nearest-center ties. □

**Corollary B.20** (Coarse-state budget and reference covering numbers)**.** *For $s \ge 0$, let $\mathcal{N}_K(s)$ be the smallest number of closed $d_K$-balls of radius $s$, centered in $[N]$, covering $[N]$. Let $\mathcal{P}_K(s)$ be the maximum cardinality of a subset whose distinct points have pairwise distance strictly greater than $s$. For $\delta \ge 0$, let $B_\delta(K)$ be the smallest $B$ for which some surjective partition $q : [N] \to [B]$ and positive coarse kernel $W$ satisfy $\mathcal{E}_K(q, W) \le \delta$. Set $h(\delta) = \operatorname{arcosh}(e^\delta)$ and $s_\delta = \max\{\delta/2, \sqrt{2\delta}\}$. Then*

$$\mathcal{P}_K(4h(\delta)) \le B_\delta(K) \le \mathcal{N}_K(s_\delta). \tag{87}$$

*At zero error, $B_0(K)$ is exactly the number of equivalence classes of the relation $d_K(i, i') = 0$.*

*Proof.* Any representation of risk at most $\delta$ has $D_q(K) \le 4h(\delta)$ by the endpoint-query converse. No two states of a $4h(\delta)$-packing can then belong to the same cell, proving the lower bound. Conversely, choose a minimum cover of radius $s_\delta$, assign states to nearest centers, and assign each center to itself. Lemma B.18 gives risk at most $\min\{2s_\delta, s_\delta^2/2\} \le \delta$, proving the upper bound. At $\delta = 0$, both packing and covering numbers count zero-distance equivalence classes of reference states. □

### B.3 A DIRECTLY VERIFIABLE FINITE-STATE CASE

So far the interaction residual has been specified directly from the reference kernel. A slow–fast chain provides a concrete mechanism that makes it small: rapid within-cell mixing removes initial detail before the coarse state changes appreciably. In finite dimensions, matrix gaps, invariant vectors, and positive kernels make each step of this argument quantitatively explicit.

**Theorem B.21** (Finite-state slow–fast approximation)**.** *Let $\mathsf{X} = \{(\beta, z) : \beta \in \mathsf{B}, z \in \mathsf{F}_\beta\}$ be finite. Define its slow–fast generator by*

$$L^\varepsilon = \varepsilon^{-1} A + S, \qquad A = \operatorname{diag}(A_\beta)_{\beta \in \mathsf{B}}, \tag{88}$$

*Write $P_t^\varepsilon := e^{tL^\varepsilon}$. Every $A_\beta$ is an irreducible conservative generator with invariant probability $\pi_\beta$ and $S$ is a conservative generator. Define the fiber-average and coarse-function lift by $(\mathsf{E}h)(\beta) := \sum_{z\in\mathsf{F}_\beta} h(\beta,z)\pi_\beta(z)$ and $(\iota v)(\beta,z) := v(\beta)$. Here $\mathsf{E}$ is conditional expectation, distinct from the fast generator $A$. Put $Q := I - \iota\mathsf{E}$, $\bar{L} := \mathsf{E}S\iota$, and $\bar{P}_t := e^{t\bar{L}}$. Assume that*

$$\max_\beta \|e^{tA_\beta}(I - \mathbf{1}\pi_\beta^\top)\|_{\infty\to\infty} \le C_0 e^{-\lambda t}, \qquad t \ge 0, \tag{89}$$

*that $\|S\|_{\infty\to\infty} \le M_S$, and that $\bar{L}$ is irreducible. Then $L^\varepsilon$ is irreducible, and hence has a unique strictly positive invariant law $m^\varepsilon$, for every $\varepsilon > 0$. Moreover, for every $0 < T_0 < T_1 < \infty$, there exist $\varepsilon_0, C, c_{\mathrm{mix}} > 0$ such that the following holds. Let $R^\varepsilon$ denote the path law generated by $L^\varepsilon$ and initialized from $m^\varepsilon$, and let $q(\beta,z) := \beta$. The stationary reference obeys, simultaneously for $0 < \varepsilon \le \varepsilon_0$ and $T \in [T_0, T_1]$,*

$$\eta_q(R^\varepsilon) \le C\left(\varepsilon + e^{-c_{\mathrm{mix}}T/\varepsilon}\right), \tag{90}$$

*The constants depend only on the displayed finite generators and projections, $C_0, \lambda, M_S, T_0, T_1$, and not on the endpoint pair, $T$, or $\varepsilon$ in the indicated ranges. More explicitly, let $\bar{m}$ be invariant for $\bar{L}$ and define*

$$c_T(\beta,\beta') := \log \frac{\bar{P}_T(\beta,\beta')}{\bar{m}(\beta')}, \qquad \bar{P}_T = e^{T\bar{L}}, \tag{91}$$

*then, for every $R^\varepsilon$-admissible endpoint pair, the endpoint- conditional lift of the coarse problem with interaction $c_T$ has the exact endpoint marginals. Both KL directions to the full optimizer are at most $2C(\varepsilon + e^{-c_{\mathrm{mix}}T/\varepsilon})$, also for the path laws using the original conditional reference bridges.*

*Proof.* Put $\mathsf{P}_0 := \iota\mathsf{E}$. Then $\mathsf{E}\iota = I$, $A\mathsf{P}_0 = \mathsf{P}_0 A = 0$, and $AQ = QA = A$. The matrix $\bar{L}$ is a conservative coarse generator: for $\beta \ne \beta'$ its entry is the $\pi_\beta$-average of the $S$-rates from fiber $\beta$ to fiber $\beta'$. Every positive edge of $\bar{L}$ is therefore realized by an edge of $S$. Combining these edges with irreducibility of every $A_\beta$ shows that $L^\varepsilon$ is irreducible for every $\varepsilon > 0$ on the full state space.

On $\operatorname{Ran} Q$, the restriction $A_Q$ is exponentially stable by (89). Duhamel's formula for the perturbation of $\varepsilon^{-1}A_Q$ by the bounded matrix $QSQ$ gives the evolution $E_Q^\varepsilon(t)$ on $\operatorname{Ran} Q$ and the bound

$$\|E_Q^\varepsilon(t)\|_{\infty\to\infty} \le C_0 \exp\left[-\left(\frac{\lambda}{\varepsilon} - C_0\|QSQ\|_{\infty\to\infty}\right)t\right], \tag{92}$$

Thus, after reducing $\varepsilon_0$ if needed, for some $c_{\mathrm{mix}} > 0$,

$$\|E_Q^\varepsilon(t-s)\|_{\infty\to\infty} \le Ce^{-c_{\mathrm{mix}}(t-s)/\varepsilon}. \tag{93}$$

For $u(t) = P_t^\varepsilon f$, set $p(t) = \mathsf{E}u(t)$ and $v(t) = Qu(t)$. The two projected Kolmogorov equations are

$$\dot{p} = \bar{L}p + \mathsf{E}Sv, \qquad \dot{v} = (\varepsilon^{-1}A_Q + QSQ)v + QS\iota p. \tag{94}$$

Both $P_t^\varepsilon$ and $\bar{P}_t$ are contractions in supremum norm. Variation of constants in the second equation, followed by the first, hence yields

$$\|QP_t^\varepsilon\|_{\infty\to\infty} \le C(e^{-c_{\mathrm{mix}}t/\varepsilon} + \varepsilon), \tag{95}$$

$$\|\mathsf{E}P_t^\varepsilon - \bar{P}_t\mathsf{E}\|_{\infty\to\infty} \le C\varepsilon \tag{96}$$

uniformly for $0 \le t \le T_1$. Since $P_t^\varepsilon = \iota\mathsf{E}P_t^\varepsilon + QP_t^\varepsilon$, this implies, entry by entry,

$$P_T^\varepsilon((\beta,z),(\beta',z')) = \bar{P}_T(\beta,\beta')\pi_{\beta'}(z') + O(\varepsilon + e^{-c_{\mathrm{mix}}T/\varepsilon}). \tag{97}$$

For the stationary expansion, regard invariant laws as row vectors and set $\bar{m}^\varepsilon := m^\varepsilon\iota$ and $\rho^\varepsilon := m^\varepsilon Q$. Stationarity gives

$$\rho^\varepsilon A = -\varepsilon m^\varepsilon S, \qquad m^\varepsilon S\iota = 0. \tag{98}$$

The second identity shows that the right-hand side of the first lies in the left subspace on which the fast modes decay. On that subspace, right multiplication by $A$ has a bounded inverse (equivalently, integrate the transpose of the fast decay). Consequently $\|\rho^\varepsilon\|_1 \le C\varepsilon$. Decomposing $m^\varepsilon = \bar{m}^\varepsilon\mathsf{E} + \rho^\varepsilon$

in the second identity then gives $\bar{m}^\varepsilon \bar{L} = O(\varepsilon)$. Because $\bar{L}$ is irreducible, right multiplication by $\bar{L}$ is invertible on zero-mass row vectors. Normalization therefore implies $\bar{m}^\varepsilon = \bar{m} + O(\varepsilon)$ and

$$m^\varepsilon(\beta, z) = \bar{m}(\beta)\pi_\beta(z) + O(\varepsilon), \tag{99}$$

where $\bar{m}$ is the strictly positive invariant law of $\bar{L}$. Irreducibility also gives $\bar{P}_T(\beta, \beta') > 0$ for every $T > 0$. Continuity in $T$ and compactness of $[T_0, T_1]$, together with strict positivity of the finitely many $\pi_\beta$ and $\bar{m}$, give a common positive lower bound for all leading entries in (97)–(99). For sufficiently small $\varepsilon$, the exact entries lie in one fixed positive compact interval as well. Since the stationary interaction is

$$\kappa_{R^\varepsilon}(x, y) = \frac{P_T^\varepsilon(x, y)}{m^\varepsilon(y)}, \tag{100}$$

the ratio map and logarithm are uniformly Lipschitz on these intervals, and

$$\log \kappa_{R^\varepsilon}((\beta, z), (\beta', z')) = \log \frac{\bar{P}_T(\beta, \beta')}{\bar{m}(\beta')} + O(\varepsilon + e^{-c_{\mathrm{mix}} T/\varepsilon}) \tag{101}$$

uniformly in $T \in [T_0, T_1]$. The leading term is exactly the logarithm of the strictly positive coarse kernel in (91), proving (90). Both coupling conclusions and both path conclusions now follow from Theorem B.7 and its path identities for shared conditional reference bridges in (41). □

**Corollary B.22** (Computable finite-state certificate)**.** *Under Theorem B.21, define for any fixed $\varepsilon, T$ in its range*

$$\Delta_T^\varepsilon := \max_{x=(\beta,z),\, y=(\beta',z')} \left| \log \frac{P_T^\varepsilon(x, y)}{m^\varepsilon(y)} - \log \frac{\bar{P}_T(\beta, \beta')}{\bar{m}(\beta')} \right|. \tag{102}$$

*Then $\eta_q(R^\varepsilon) \le \Delta_T^\varepsilon$, and all four coupling and reconstructed-path KL directions in Theorem B.21 are at most $\min\{2\Delta_T^\varepsilon, (\Delta_T^\varepsilon)^2/2\}$. Their coupling total variation is at most $\tanh(\Delta_T^\varepsilon/2)$. This is an exact finite-matrix certificate: $P_T^\varepsilon$, $m^\varepsilon$, $\bar{P}_T$, and $\bar{m}$ are obtained by matrix exponentiation and invariant-vector solves. The resulting bound evaluates the specified reference matrices directly, giving a numerical counterpart to the uniform asymptotic estimate at the chosen horizon and scale.*

*Proof.* Use the coarse interaction in (91), set both endpoint-only terms to zero, and take the displayed maximum as a witness in Definition B.2. The KL and total-variation conclusions follow from Theorem B.15, with path KL transferred by the shared-conditional entropy identity (41). □

For an a priori envelope, the proof is also quantitative. Taking

$$\varepsilon_0 \le \frac{\lambda}{2C_0 \|QSQ\|_{\infty\to\infty}} \tag{103}$$

(with the right-hand side interpreted as infinity when $QSQ = 0$) permits $c_{\mathrm{mix}} = \lambda/2$ in (93). The continuously injected fast component is then bounded by $2C_0\varepsilon\|QS\iota\|/\lambda$. The effective-semigroup coefficient is obtained by multiplying its time integral by $\|\mathsf{E}SQ\|_{\infty\to\infty}$. The stationary coefficient uses the inverse norms of $A$ on the fast zero-mean subspace and the group inverse of $\bar{L}$ on coarse zero-mass vectors. Finally, converting to a log bound uses the reciprocal of

$$\min_{\substack{T\in[T_0,T_1]\\ \beta,\beta',z'}} \bar{P}_T(\beta, \beta')\pi_{\beta'}(z') \wedge \bar{m}(\beta')\pi_{\beta'}(z'). \tag{104}$$

These reference matrix norms, group inverses, and finite minima give an asymptotic prefactor before any query. Equation (102) evaluates the uniform discrepancy at the chosen horizon and scale.

**Corollary B.23** (Finite-state Schrödinger boundary layers)**.** *Under the hypotheses of Theorem B.21, let $0 \le K < \infty$ and, for every $T \in [T_0, T_1]$, consider endpoint probabilities*

$$\mu_0 = \rho_0 m^\varepsilon, \qquad \mu_T = \rho_T m^\varepsilon, \qquad e^{-K} \le \rho_0, \rho_T \le e^K. \tag{105}$$

*Both likelihoods have $m^\varepsilon$-weighted mean one, so $\mu_0$ and $\mu_T$ are probability distributions. On this fixed finite state space, no fiber invariance or additional regularity is imposed on either likelihood. Let $(P_t^\varepsilon)^{\dagger,\varepsilon}$ denote the adjoint in $\ell^2(m^\varepsilon)$. Choose the positive Schrödinger factors with normalization*

$$f_0^\varepsilon P_T^\varepsilon g_T^\varepsilon = \rho_0, \qquad g_T^\varepsilon (P_T^\varepsilon)^{\dagger,\varepsilon} f_0^\varepsilon = \rho_T, \qquad \sum_x m^\varepsilon(x) g_T^\varepsilon(x) = 1. \tag{106}$$

*Define the time-dependent Schrödinger factors by*

$$f_t^\varepsilon := (P_t^\varepsilon)^{\dagger,\varepsilon} f_0^\varepsilon, \qquad g_t^\varepsilon := P_{T-t}^\varepsilon g_T^\varepsilon. \tag{107}$$

*There are $\varepsilon_0, C_K, c_{\mathrm{mix}} > 0$, uniform over the endpoints in (105) and over $T \in [T_0, T_1]$, such that the following statements hold for $0 < \varepsilon \le \varepsilon_0$ and $0 \le t \le T$. Define the exact bridge marginal at time $t$ by $\mu_t^\varepsilon(\mathrm{d}x) := f_t^\varepsilon(x) g_t^\varepsilon(x) m^\varepsilon(\mathrm{d}x)$ and set*

$$\Delta_t^\varepsilon := \varepsilon + e^{-c_{\mathrm{mix}} t/\varepsilon} + e^{-c_{\mathrm{mix}}(T-t)/\varepsilon}. \tag{108}$$

$$\sup_\beta \mathrm{osc}_{z \in \mathsf{F}_\beta} \log f_t^\varepsilon(\beta, z) \le C_K \left(\varepsilon + e^{-c_{\mathrm{mix}} t/\varepsilon}\right), \tag{109}$$

$$\sup_\beta \mathrm{osc}_{z \in \mathsf{F}_\beta} \log g_t^\varepsilon(\beta, z) \le C_K \left(\varepsilon + e^{-c_{\mathrm{mix}}(T-t)/\varepsilon}\right). \tag{110}$$

*If $\mu_t^\varepsilon(\cdot \mid \beta)$ is the bridge conditional law in fiber $\beta$, then*

$$\sup_\beta \left\| \log \frac{\mathrm{d}\mu_t^\varepsilon(\cdot \mid \beta)}{\mathrm{d}\pi_\beta} \right\|_\infty \le C_K \Delta_t^\varepsilon, \tag{111}$$

$$\sup_\beta \|\mu_t^\varepsilon(\cdot \mid \beta) - \pi_\beta\|_{\mathrm{TV}} \le C_K \Delta_t^\varepsilon. \tag{112}$$

*Choose the explicit boundary-layer width*

$$\tau_\varepsilon := \frac{\varepsilon}{c_{\mathrm{mix}}} \log \frac{1}{\varepsilon}, \tag{113}$$

*For sufficiently small $\varepsilon$, both factor oscillations and both conditional-fiber errors are $O_K(\varepsilon)$ uniformly on $t \in [\tau_\varepsilon, T - \tau_\varepsilon]$. Thus bounded-contrast endpoint fiber structure is confined to two layers of width $O(\varepsilon \log(1/\varepsilon))$, with equilibrium conditionals in the bulk up to the stated error.*

*Proof.* Write $k_T^{\varepsilon,\mathrm{stat}}(x, y) := P_T^\varepsilon(x, y)/m^\varepsilon(y)$. Equations (97) and (99), together with positivity of $\bar{P}_T$ uniformly on $[T_0, T_1]$, give constants independent of $T$ and small $\varepsilon$ such that

$$0 < \underline{k} \le k_T^{\varepsilon,\mathrm{stat}}(x, y) \le \overline{k} < \infty. \tag{114}$$

Strict positivity gives the finite Schrödinger factors, unique up to their common reciprocal scaling. Under the normalization in (106),

$$\underline{k} \le P_T^\varepsilon g_T^\varepsilon \le \overline{k}. \tag{115}$$

The first Schrödinger equation and the endpoint likelihood bounds therefore give $C_K^{-1} \le f_0^\varepsilon \le C_K$, and hence $C_K^{-1} \le \sum_x m^\varepsilon(x) f_0^\varepsilon(x) \le C_K$. Since

$$(P_T^\varepsilon)^{\dagger,\varepsilon} f_0^\varepsilon(y) = \sum_x k_T^{\varepsilon,\mathrm{stat}}(x, y) f_0^\varepsilon(x) m^\varepsilon(x), \tag{116}$$

the same kernel bounds give

$$\underline{k} \sum_x m^\varepsilon(x) f_0^\varepsilon(x) \le (P_T^\varepsilon)^{\dagger,\varepsilon} f_0^\varepsilon(y) \le \overline{k} \sum_x m^\varepsilon(x) f_0^\varepsilon(x). \tag{117}$$

The second Schrödinger equation bounds $g_T^\varepsilon$ above and away from zero. Hence, uniformly over all asserted ranges,

$$C_K^{-1} \le f_0^\varepsilon, g_T^\varepsilon \le C_K \tag{118}$$

For positive $h$ satisfying (118), Markov contraction and (95) give, uniformly for $0 \le s \le T_1$,

$$\sup_\beta \mathrm{osc}_{\mathsf{F}_\beta} \log P_s^\varepsilon h \le C_K(\varepsilon + e^{-c_{\mathrm{mix}} s/\varepsilon}). \tag{119}$$

Here the fiber decomposition controls oscillation, while positivity transfers the bound to logarithms.

The adjoint estimate follows from the same expansions without assuming reversibility. Here we use the additive kernel reduction established in the proof of Theorem B.21, which is uniform for

$0 \le s \le T_1$. Strict positivity is required at the full horizon $T \ge T_0$. Writing $\bar{P}_s^{\dagger,\bar{m}}$ for the adjoint of $\bar{P}_s$ in $\ell^2(\bar{m})$ gives

$$(P_s^\varepsilon)^{\dagger,\varepsilon} h(y) = \frac{1}{m^\varepsilon(y)} \sum_x m^\varepsilon(x) P_s^\varepsilon(x,y) h(x) \tag{120}$$

has the following expansion. Write $y = (\beta', z')$, $\delta_s^\varepsilon := \varepsilon + e^{-c_{\mathrm{mix}} s/\varepsilon}$, and $\bar{h}(\beta) := (\mathsf{E}h)(\beta)$. The kernel and invariant-law expansions give uniformly for $0 \le s \le T_1$,

$$\begin{aligned} &\sum_x m^\varepsilon(x) P_s^\varepsilon(x,y) h(x) \\ &\quad = \pi_{\beta'}(z') \sum_\beta \bar{m}(\beta) \bar{P}_s(\beta,\beta') \bar{h}(\beta) + O(\delta_s^\varepsilon) \|h\|_\infty. \end{aligned} \tag{121}$$

Since $m^\varepsilon(y) = \bar{m}(\beta')\pi_{\beta'}(z') + O(\varepsilon)$, uniform positivity of the leading masses gives the expansion

$$\begin{aligned} (P_s^\varepsilon)^{\dagger,\varepsilon} h(y) = \frac{1}{\bar{m}(\beta')} &\sum_\beta \bar{m}(\beta) \bar{P}_s(\beta,\beta') \bar{h}(\beta) \\ &+ O(\delta_s^\varepsilon) \|h\|_\infty. \end{aligned} \tag{122}$$

The leading term is precisely $(\iota \bar{P}_s^{\dagger,\bar{m}} \mathsf{E}h)(y)$, proving

$$\left\| (P_s^\varepsilon)^{\dagger,\varepsilon} h - \iota \bar{P}_s^{\dagger,\bar{m}} \mathsf{E}h \right\|_\infty \le C(\varepsilon + e^{-c_{\mathrm{mix}} s/\varepsilon}) \|h\|_\infty. \tag{123}$$

The adjoint is Markov because $m^\varepsilon$ is invariant, so it also preserves the lower bound on $h$. The logarithmic version of (123), applied to $f_0^\varepsilon$, and (119), applied to $g_T^\varepsilon$, prove (109)– (110).

It remains to identify the correct fiber equilibrium. Summing (99) over a fiber first gives $m_{\mathsf{B}}^\varepsilon(\beta) = \bar{m}(\beta) + O(\varepsilon)$, where $m_{\mathsf{B}}^\varepsilon(\beta) := \sum_{z \in \mathsf{F}_\beta} m^\varepsilon(\beta, z)$, and then $m_\beta^\varepsilon(z) = \pi_\beta(z) + O(\varepsilon)$. Dividing entrywise and using $\min_{\beta,z} \pi_\beta(z) > 0$ on the fixed finite state space gives

$$\sup_{\beta,z} \left| \log \frac{m_\beta^\varepsilon(z)}{\pi_\beta(z)} \right| \le C\varepsilon, \qquad m_\beta^\varepsilon(z) := \frac{m^\varepsilon(\beta,z)}{\sum_{z'} m^\varepsilon(\beta,z')}. \tag{124}$$

The bridge time marginal has density $f_t^\varepsilon g_t^\varepsilon$ relative to $m^\varepsilon$. Consequently, inside fiber $\beta$ its density relative to $\pi_\beta$ is the normalization of

$$z \longmapsto f_t^\varepsilon(\beta,z) g_t^\varepsilon(\beta,z) \frac{m_\beta^\varepsilon(z)}{\pi_\beta(z)}. \tag{125}$$

To bound this normalized density, set

$$a_\beta(z) := \log f_t^\varepsilon(\beta,z) + \log g_t^\varepsilon(\beta,z) + \log \frac{m_\beta^\varepsilon(z)}{\pi_\beta(z)}, \tag{126}$$

The elementary log-sum-exp inequality gives

$$\min_z a_\beta(z) \le \log \sum_z \pi_\beta(z) e^{a_\beta(z)} \le \max_z a_\beta(z). \tag{127}$$

Thus the normalized log-density is bounded by $\mathrm{osc}_z\, a_\beta$, and (111) follows by adding the two factor oscillation bounds and (124). If $\delta_{t,\beta}$ denotes its supremum norm, then on its uniformly bounded range, $\mathrm{TV} \le \frac{1}{2}(e^{\delta_{t,\beta}} - 1) \le C_K \delta_{t,\beta}$, proving (112). After replacing $c_{\mathrm{mix}}$ by the minimum of the forward and stationary-adjoint decay rates, decrease $\varepsilon_0$ so that $\varepsilon_0 < 1$ and $2(\varepsilon/c_{\mathrm{mix}}) \log(1/\varepsilon) \le T_0$ for every $0 < \varepsilon \le \varepsilon_0$. Then $e^{-c_{\mathrm{mix}} \tau_\varepsilon/\varepsilon} = \varepsilon$, giving the bulk bound uniformly in $T \in [T_0, T_1]$. □

### B.4 Slow–fast references screen interaction

The finite-state proof used two operations: projecting onto fiber averages and integrating the decay of the remaining modes. We now carry out the same argument for smooth diffusions. The analytic assumptions specify the norms in which these operations and the conversion to kernel bounds are uniform. Fix $\varepsilon_0 > 0$, $0 < \alpha < 1$, and $0 < T_0 < T_1 < \infty$. Let $\mathsf{X}$ be a compact smooth fiber bundle without boundary over a compact base $\mathsf{B}$, with compact connected fibers $\mathsf{F}_\beta$. We use $x = (\beta, z)$

locally and take $q(x) = \beta$. Let $\pi_\beta$ be the smooth positive fiber probabilities specified in (SF1), and define the fiber average, the lift of a coarse function, and the complementary projection by

$$(\mathsf{E}h)(\beta) := \int_{\mathsf{F}_\beta} h(\beta, z)\pi_\beta(\,\mathrm{d}z), \qquad (\iota v)(\beta, z) := v(\beta), \qquad Q := I - \iota\mathsf{E}. \tag{128}$$

For a function $u$ on one fiber, write $\mathsf{E}_\beta u := (\int u\,\mathrm{d}\pi_\beta)\mathbf{1}$. Consider

$$L^\varepsilon = \varepsilon^{-1}A + S, \qquad 0 < \varepsilon \le \varepsilon_0, \tag{129}$$

where $(Ah)(\beta, z) = A_\beta h(\beta, \cdot)(z)$ is vertical and $S$ contains the slow generator and all order-one slow–fast interaction. Set $P_t^\varepsilon := e^{tL^\varepsilon}$ and define

$$\bar{L} := \mathsf{E}S\iota, \qquad \bar{P}_t := e^{t\bar{L}}. \tag{130}$$

Introduce a smooth positive probability $\bar{m}$ on $\mathsf{B}$ and the comparison probability

$$m^0(\,\mathrm{d}\beta\;\mathrm{d}z) := \bar{m}(\,\mathrm{d}\beta)\pi_\beta(\,\mathrm{d}z). \tag{131}$$

Superscript $^{\dagger,0}$ denotes the adjoint in $L^2(m^0)$, while $\bar{P}_t^\dagger$ and $\bar{L}^\dagger$ denote adjoints in $L^2(\bar{m})$. The vertical norm $\|h\|_{C^1_F} := \|h\|_\infty + \|\nabla_F h\|_\infty$ controls a function and its first derivatives within each fiber. The conditions below apply throughout to the forward and adjoint evolutions.

**(SF1) Uniform fast equilibrium and gap.** For every $\beta$, $A_\beta$ is a smooth uniformly elliptic conservative generator on $\mathsf{F}_\beta$, reversible with respect to a smooth positive probability $\pi_\beta$. Let $\lambda_\beta$ denote the largest constant for which, for every $u$ satisfying $\int u\,\mathrm{d}\pi_\beta = 0$,

$$\langle -A_\beta u, u\rangle_{L^2(\pi_\beta)} \ge \lambda_\beta \|u\|^2_{L^2(\pi_\beta)}, \qquad \lambda_\star := \inf_{\beta\in\mathsf{B}} \lambda_\beta > 0. \tag{132}$$

The condition $\lambda_\star > 0$ is the uniform fiber-gap assumption. Uniform Schauder estimates imply, for some $c_0, C_0 > 0$,

$$\|e^{sA_\beta}(I - \mathsf{E}_\beta)u\|_{C^1} \le C_0(1 + s^{-1/2})e^{-c_0 s}\|u\|_{C^\alpha}. \tag{133}$$

**(SF2) Order-one coupling and fast-mode stability.** The coefficients of $S$ are uniformly bounded in $C^{3+\alpha}$. The maps $QS\iota$, $QSQ$, $\mathsf{E}SQ$ and their $m^0$-adjoints are uniformly bounded between the anisotropic Hölder graph spaces below. All cross-diffusion coefficients included in $S$ are of order one. This interaction may continually inject fast modes and may fail to commute with $A$ or $\mathsf{E}$.

More precisely, let $\mathcal{X}$ be an anisotropic Hölder graph space with continuous embeddings

$$C^{2+\alpha}(\mathsf{X}) \hookrightarrow \mathcal{X} \hookrightarrow C^1(\mathsf{X}), \tag{134}$$

where the graph norm controls the slow derivatives on which $S$ acts. The last embedding includes both vertical and horizontal first derivatives for a fixed smooth bundle metric. Its vertical part suffices for interaction screening, while its horizontal part controls the order-one cross drift in Theorem B.33. $Q$ and $\iota\mathsf{E}$ are assumed bounded on $\mathcal{X}$. Let $V_Q^\varepsilon(t, s)$ and $V_Q^{\varepsilon,\dagger,0}(t, s)$ be the strongly continuous evolutions on $\operatorname{Ran} Q$ generated by $\varepsilon^{-1}A + QSQ$ and its $m^0$-adjoint, respectively. Both are assumed well posed on $\mathcal{X}$. For constants $C, c_{\mathrm{mix}} > 0$ independent of $s, t$, and $\varepsilon$, we impose the fast-mode decay estimate

$$\|V_Q^\varepsilon(t, s)\|_{\mathcal{X}\to\mathcal{X}} + \|V_Q^{\varepsilon,\dagger,0}(t, s)\|_{\mathcal{X}\to\mathcal{X}} \le Ce^{-c_{\mathrm{mix}}(t-s)/\varepsilon}, \qquad 0 \le s \le t \le T_1, \tag{135}$$

The fast-mode estimate is supplemented by

$$\|QS\iota v\|_{\mathcal{X}} \le C\|v\|_{C^{2+\alpha}(\mathsf{B})}, \qquad \|\mathsf{E}SQw\|_{C^1(\mathsf{B})} \le C\|w\|_{\mathcal{X}}, \tag{136}$$

$$\sup_{0\le t\le T_1} \|\mathsf{E}P_t^\varepsilon h\|_{C^{2+\alpha}(\mathsf{B})} \le C\|h\|_{C^{2+\alpha}(\mathsf{X})}. \tag{137}$$

The mapping and regularity-propagation bounds in (136)–(137) also hold for the $m^0$-adjoint operators. These graph norms control unbounded coupling through the relevant slow derivatives.

**(SF3) Invariant-law correction and effective dynamics.** The invariant probability of $L^\varepsilon$ may depend on $\varepsilon$. We compare it with $m^0$ through

$$m^\varepsilon = \omega^\varepsilon m^0, \qquad C^{-1} \le \omega^\varepsilon \le C, \qquad \|\log \omega^\varepsilon\|_{C^{2+\alpha}} \le C. \tag{138}$$

Writing $m^\varepsilon(\,\mathrm{d}\beta\;\mathrm{d}z) = \bar{m}^\varepsilon(\,\mathrm{d}\beta) m^\varepsilon_\beta(\,\mathrm{d}z)$, assume the first-order stationary expansion

$$\sup_\beta \left\| \log \frac{\mathrm{d}m^\varepsilon_\beta}{\mathrm{d}\pi_\beta} \right\|_{C^{2+\alpha}(\mathsf{F}_\beta)} \le C\varepsilon. \tag{139}$$

Thus the invariant conditional may differ from $\pi_\beta$ by an order-$\varepsilon$ correction. The adjoint of $L^\varepsilon$ in $L^2(m^0)$ has the same slow–fast analytic structure and satisfies the adjoint conditions in (SF2). The operator $\bar{L}$ in (130) is required to be a smooth uniformly elliptic Markov generator on $\mathsf{B}$ with invariant law $\bar{m}$. Since $\mathsf{E}$ and $\iota$ are adjoint in $L^2(m^0)$, the effective adjoint satisfies

$$\mathsf{E} S^{\dagger,0} \iota = \bar{L}^\dagger \tag{140}$$

The right-hand adjoint uses $L^2(\bar{m})$, the same measure as averaging and stationary reversal.

**(SF4) Uniform positive-time smoothing.** For every $0 < t_0 < t_1 < \infty$, each of $P^\varepsilon_t$, $(P^\varepsilon_t)^{\dagger,0}$, $\bar{P}_t$, and $\bar{P}^\dagger_t$ satisfies a uniform $L^1 \to C^{2+\alpha}$ smoothing bound for $t \in [t_0, t_1]$. Their heat kernels relative to $m^0$ and $\bar{m}$ are strictly positive. These parameter-uniform smoothing hypotheses complement the fast-mode estimates by controlling transition densities pointwise. For compact uniformly elliptic models, their verification uses parameter-dependent Schauder and heat-kernel estimates.

**Lemma B.24** (Fast-mode decay under slow–fast coupling)**.** *For $h \in C^{2+\alpha}(\mathsf{X})$ and $0 < t \le T_1$, assumptions* (SF1)–(SF4) *imply*

$$\|QP^\varepsilon_t h\|_{\mathcal{X}} \le C\left(e^{-c_{\mathrm{mix}} t/\varepsilon}\|Qh\|_{\mathcal{X}} + \varepsilon\|h\|_{C^{2+\alpha}}\right). \tag{141}$$

*The same estimate holds for the $m^0$-adjoint semigroup. By (134), it controls the supremum norm and first derivatives, including the vertical variation relevant to mixing. The two terms respectively measure decaying initial detail and the fast variation continually supplied by the slow coupling.*

*Proof.* Set $u_t = P^\varepsilon_t h$, $u^0_t = \iota \mathsf{E} u_t$, and $u^\perp_t = Q u_t$. Since $A\iota = 0$ and $\mathsf{E}A = 0$, projecting the Kolmogorov equation onto the fast modes yields

$$\partial_t u^\perp_t = \varepsilon^{-1} A u^\perp_t + QSQ u^\perp_t + QS\iota \mathsf{E} u_t. \tag{142}$$

The fast-mode decay estimate (135) and variation of constants give

$$u^\perp_t = V^\varepsilon_Q(t,0)Qh + \int_0^t V^\varepsilon_Q(t,s) QS\iota \mathsf{E} u_s \,\mathrm{d}s. \tag{143}$$

The mapping and propagation bounds (136)– (137), together with

$$\int_0^t e^{-c_{\mathrm{mix}}(t-s)/\varepsilon}\,\mathrm{d}s \le \frac{\varepsilon}{c_{\mathrm{mix}}}, \tag{144}$$

give the $\mathcal{X}$ estimate and, by the same argument, its $m^0$-adjoint version. When $QS\iota \neq 0$, the quasi-steady balance $\varepsilon^{-1}Au^\perp + QSu^0 = 0$ gives $u^\perp = -\varepsilon A^{-1} QSu^0$, where $A^{-1}$ acts on zero-mean fast modes. Slow forcing thus sustains an order-$\varepsilon$ residual after the initial memory decays. □

**Lemma B.25** (Effective semigroup estimate)**.** *For $h \in C^{2+\alpha}(\mathsf{X})$ and $0 < t \le T_1$,*

$$\|\mathsf{E}P^\varepsilon_t h - \bar{P}_t \mathsf{E} h\|_{C^1(\mathsf{B})} \le C\varepsilon \|h\|_{C^{2+\alpha}}, \tag{145}$$

$$\|P^\varepsilon_t h - \iota \bar{P}_t \mathsf{E} h\|_{C^1_F} \le C\left(\varepsilon + e^{-c_{\mathrm{mix}} t/\varepsilon}\right)\|h\|_{C^{2+\alpha}}. \tag{146}$$

*The same statements hold with $(P^\varepsilon_t)^{\dagger,0}$ compared to $\iota \bar{P}^\dagger_t \mathsf{E}$, with constants uniform in $t$ and $\varepsilon$.*

*Proof.* Projecting the Kolmogorov equation gives

$$\partial_t \mathsf{E} u_t = \bar{L} \mathsf{E} u_t + \mathsf{E} SQ u_t. \tag{147}$$

Duhamel's formula for $\bar{P}_t$ and Lemma B.24 show that the integral of $\mathsf{E}SQu_t$ is $O(\varepsilon)$: the initial fast layer integrates to $O(\varepsilon)$ and the continuously injected component is itself $O(\varepsilon)$. This proves (145). Combining the averaged error with the complementary estimate (141) then gives (146). □

Write $k_t^{\varepsilon,0}(x,y)$ for the density of $P_t^\varepsilon$ relative to $m^0$, and $\bar{k}_t(\beta,\beta')$ for that of $\bar{P}_t$ relative to $\bar{m}$.

**Lemma B.26** (Positive fixed-horizon kernel expansion)**.** *Fix $0 < T_0 < T_1 < \infty$. Under* (SF1)–(SF4)*, there are $C, c_{\mathrm{mix}}, \varepsilon_0 > 0$ such that, for $T \in [T_0, T_1]$ and $0 < \varepsilon \le \varepsilon_0$,*

$$\sup_{x=(\beta,z),\, y=(\beta',z')} \left| k_T^{\varepsilon,0}(x,y) - \bar{k}_T(\beta,\beta') \right| \le C\left(\varepsilon + e^{-c_{\mathrm{mix}}T/\varepsilon}\right). \tag{148}$$

*Moreover, after possibly decreasing $\varepsilon_0$,*

$$\sup_{x,y} \left| \log \frac{k_T^{\varepsilon,0}(x,y)}{\bar{k}_T(q(x),q(y))} \right| \le C\left(\varepsilon + e^{-c_{\mathrm{mix}}T/\varepsilon}\right). \tag{149}$$

*Both kernels in* (149) *are strictly positive Markov kernels, so their logarithmic ratio is well defined.*

*Proof.* Put $J_t := \iota \bar{P}_t \mathsf{E}$ and split the horizon into three equal intervals of length $a = T/3 \ge T_0/3$. The following identity telescopes the three factors:

$$(P_a^\varepsilon)^3 - J_a^3 = (P_a^\varepsilon - J_a)(P_a^\varepsilon)^2 + J_a(P_a^\varepsilon - J_a)P_a^\varepsilon + J_a^2(P_a^\varepsilon - J_a). \tag{150}$$

For the first two terms, the right positive-time factors map $L^1$ into $C^{2+\alpha}$, so Lemma B.25 applies to the difference factor. For the last term, apply the same argument after taking the $m^0$-adjoint. The left positive-time factors then provide the required smoothing. Let $c_Q > 0$ denote a decay rate furnished by Lemma B.25 before this time splitting. Each difference factor is estimated over $a = T/3$, giving $C(\varepsilon + e^{-c_Q T/(3\varepsilon)})$. Set $c_{\mathrm{mix}} = c_Q/3$ for the kernel estimate and, if necessary, decrease the common rate in subsequent statements to this value. Earlier bounds remain valid with this common, possibly smaller decay rate. The forward and adjoint bounds in (SF4) therefore give

$$\|P_T^\varepsilon - J_T\|_{L^1(m^0)\to L^\infty(m^0)} \le C\left(\varepsilon + e^{-c_{\mathrm{mix}}T/\varepsilon}\right). \tag{151}$$

This operator norm is the essential-supremum kernel bound. The continuous kernel versions supplied by (SF4) upgrade it to the pointwise bound (148). Compactness, uniform ellipticity of $L$, and $T \in [T_0, T_1]$ imply

$$\inf_{T\in[T_0,T_1],\, \beta,\beta'} \bar{k}_T(\beta,\beta') =: \underline{k} > 0. \tag{152}$$

For small $\varepsilon$, the right-hand side of (148) is at most $\underline{k}/2$. The mean-value theorem for $\log$ then gives (149). Thus positive-time smoothing and the coarse kernel lower bound together convert semigroup convergence into the relative estimate needed by the bridge approximation theorem. □

**Theorem B.27** (Slow–fast interaction screening)**.** *Under* (SF1)–(SF4)*, there exist $C, c_{\mathrm{mix}}, \varepsilon_1 > 0$ such that the following holds for every $0 < \varepsilon \le \varepsilon_1$ and $T \in [T_0, T_1]$. Let $R^\varepsilon$ be the Markov path law generated by $L^\varepsilon$, initialized from its invariant probability $m^\varepsilon$, and observed over horizon $T$. Its uniform coarse-interaction residual obeys*

$$\eta_q(R^\varepsilon) \le C\left(\varepsilon + e^{-c_{\mathrm{mix}}T/\varepsilon}\right). \tag{153}$$

*The coupling $S$ produces the persistent term, and initial fast modes decay exponentially.*

*Proof.* Let $k_T^{\varepsilon,\mathrm{stat}}$ be the transition density relative to the actual invariant law $m^\varepsilon$. By (138),

$$k_T^{\varepsilon,\mathrm{stat}}(x,y) = \frac{k_T^{\varepsilon,0}(x,y)}{\omega^\varepsilon(y)}. \tag{154}$$

Stationarity gives $\kappa_{R^\varepsilon} = k_T^{\varepsilon,\mathrm{stat}}$. Set $a^\varepsilon = 0$, $b^\varepsilon(y) = -\log \omega^\varepsilon(y)$, and

$$c_T(\beta,\beta') := \log \bar{k}_T(\beta,\beta'). \tag{155}$$

The column term absorbs the stationary correction exactly. The remaining residual is $\log k_T^{\varepsilon,0}(x,y) - \log \bar{k}_T(q(x),q(y))$, so the bound is Lemma B.26, followed by the infimum in Definition B.2. □

**Corollary B.28** (All-admissible-endpoint slow–fast reduction)**.** *Under the hypotheses of Theorem B.27, for every $(\mu,\nu)\in\mathfrak{A}_{R^\varepsilon}$, define the exact endpoint optimizer and exact bridge path law by*

$$\gamma^{\varepsilon,\star}:=\arg\min_{\gamma\in\Pi(\mu,\nu)}\mathcal{H}(\gamma\mid R^\varepsilon_{0T}),\qquad P^{\varepsilon,\star}:=\int R^{\varepsilon,xy}\gamma^{\varepsilon,\star}(\,\mathrm{d}x\,\mathrm{d}y),$$

*where $R^{\varepsilon,xy}$ is the conditional law of $R^\varepsilon$ given its endpoints $(x,y)$. Let $\widetilde{\gamma}^\star$ be the exact endpoint-conditional lift of the coarse problem with interaction $c_T$ from (155). Then*

$$\mathrm{KL}(\widetilde{\gamma}^\star\|\gamma^{\varepsilon,\star})\vee\mathrm{KL}(\gamma^{\varepsilon,\star}\|\widetilde{\gamma}^\star)\le 2C\left(\varepsilon+e^{-c_{\mathrm{mix}}T/\varepsilon}\right). \tag{156}$$

*The path law reconstructed with the original conditional reference bridges*

$$\widetilde{P}^\varepsilon:=\int R^{\varepsilon,xy}\widetilde{\gamma}^\star(\,\mathrm{d}x\,\mathrm{d}y) \tag{157}$$

*has the exact prescribed endpoints and obeys*

$$\mathrm{KL}(\widetilde{P}^\varepsilon\|P^{\varepsilon,\star})\ \vee\ \mathrm{KL}(P^{\varepsilon,\star}\|\widetilde{P}^\varepsilon)\le 2C\left(\varepsilon+e^{-c_{\mathrm{mix}}T/\varepsilon}\right). \tag{158}$$

*The error bound is uniform over admissible endpoints because the relative interaction estimate holds over the entire state space, including regions on which a query may concentrate its probability mass.*

*Proof.* Apply Theorem B.7 with the interaction bound from Theorem B.27. The lift preserves marginals, and shared conditional reference bridges give equal path and coupling divergences. □

### B.5 Schrödinger factors and conditional fast dynamics

The interaction estimate explains why a coarse solve approximates the endpoint coupling. To locate the fine endpoint detail along the path, we now study the two Schrödinger factors. Each evolves under the reference semigroup from one end of the time interval, so fast-mode decay produces two boundary layers. Uniform derivatives additionally require positive, regular endpoint likelihoods.

**Definition B.29** (Regular positive endpoint class)**.** *For $K\ge 1$ and $0<\alpha<1$, define*

$$\mathcal{E}^\varepsilon_{K,\alpha}:=\left\{\mu\in\mathcal{P}(\mathsf{X}):\rho:=\frac{\mathrm{d}\mu}{\mathrm{d}m^\varepsilon},\quad e^{-K}\le\rho\le e^K,\quad\|\log\rho\|_{C^{2+\alpha}(\mathsf{X})}\le K\right\}. \tag{159}$$

*The class allows nonconstant fiber densities, including asymmetric mixtures, smooth localized bumps, and multimodal slow–fast correlations within the stated contrast and regularity bounds.*

For $T\in[T_0,T_1]$ and $\mu_0,\mu_T\in\mathcal{E}^\varepsilon_{K,\alpha}$, write $\rho_0:=\mathrm{d}\mu_0/\,\mathrm{d}m^\varepsilon$ and $\rho_T:=\mathrm{d}\mu_T/\,\mathrm{d}m^\varepsilon$. From this point on, a superscript $^{\dagger,\varepsilon}$ denotes the stationary adjoint in $L^2(m^\varepsilon)$. It is related to the $m^0$-adjoint by

$$(P^\varepsilon_t)^{\dagger,\varepsilon}h=(\omega^\varepsilon)^{-1}(P^\varepsilon_t)^{\dagger,0}(\omega^\varepsilon h),\qquad(P^\varepsilon_t)^{\dagger,0}\omega^\varepsilon=\omega^\varepsilon. \tag{160}$$

Normalize the Schrödinger factors $(f^\varepsilon_0,g^\varepsilon_T)$ by $\int g^\varepsilon_T\,\mathrm{d}m^\varepsilon=1$ and write the Schrödinger system as

$$f^\varepsilon_0P^\varepsilon_Tg^\varepsilon_T=\rho_0,\qquad g^\varepsilon_T(P^\varepsilon_T)^{\dagger,\varepsilon}f^\varepsilon_0=\rho_T. \tag{161}$$

The time-$t$ bridge density relative to $m^\varepsilon$ is $\rho^\varepsilon_t=f^\varepsilon_tg^\varepsilon_t$, where

$$f^\varepsilon_t:=(P^\varepsilon_t)^{\dagger,\varepsilon}f^\varepsilon_0,\qquad g^\varepsilon_t:=P^\varepsilon_{T-t}g^\varepsilon_T. \tag{162}$$

**Lemma B.30** (Uniform Schrödinger-factor bounds)**.** *After decreasing $\varepsilon_0$ as in Lemma B.26, for every $T\in[T_0,T_1]$, $0<\varepsilon\le\varepsilon_0$, and $\mu_0,\mu_T\in\mathcal{E}^\varepsilon_{K,\alpha}$, assumptions* (SF1)–(SF4) *imply that a constant $C_K<\infty$, independent of $T$, $\varepsilon$, and the particular endpoints, satisfies*

$$C_K^{-1}\le f^\varepsilon_0,g^\varepsilon_T\le C_K,\qquad\|f^\varepsilon_0\|_{C^{2+\alpha}}+\|g^\varepsilon_T\|_{C^{2+\alpha}}\le C_K. \tag{163}$$

*Proof.* By Lemma B.26, (154), and the uniform bounds on $\omega^\varepsilon$, the transition density relative to $m^\varepsilon$ has positive lower and upper bounds $0<\underline{k}\le k^{\varepsilon,\mathrm{stat}}_T\le\overline{k}$ independent of $\varepsilon$. Since $\int g^\varepsilon_T\,\mathrm{d}m^\varepsilon=1$,

$$\underline{k}\le P^\varepsilon_Tg^\varepsilon_T\le\overline{k}. \tag{164}$$

The first equation in (161) and $e^{-K}\le\rho_0\le e^K$ bound $f^\varepsilon_0$ above and below. Its integral is therefore bounded above and below. Applying the same kernel bounds to $(P^\varepsilon_T)^{\dagger,\varepsilon}f^\varepsilon_0$ and using the second Schrödinger equation bounds $g^\varepsilon_T$. Uniform positive-time Schauder estimates and the conjugacy (160) bound the derivatives of the two denominators in (161). Their positive lower bounds permit division in $C^{2+\alpha}$. Combining these estimates with the endpoint regularity gives the bounds (163). □

**Lemma B.31** (Uniform conditional log-density estimate)**.** *Fix $H \geq 1$ and let $h \in C^{2+\alpha}(\mathsf{X})$ satisfy $H^{-1} \leq h \leq H$ and $\|h\|_{C^{2+\alpha}} \leq H$. Then there is a constant $C_H$, independent of $h, s$, and $\varepsilon$, such that, for $0 \leq s \leq T_1$,*

$$\sup_{\beta} \operatorname{osc}_{z \in \mathsf{F}_\beta} \log P_s^\varepsilon h(\beta, z) + \|\nabla_F \log P_s^\varepsilon h\|_\infty \leq C_H \left(\varepsilon + e^{-c_{\mathrm{mix}} s/\varepsilon}\right). \tag{165}$$

*The same estimate holds with $P_s^\varepsilon$ replaced by the stationary adjoint $(P_s^\varepsilon)^{\dagger,\varepsilon}$.*

*Proof.* The Markov property gives $H^{-1} \leq P_s^\varepsilon h \leq H$. By Lemma B.24, its $Q$-component and vertical gradient are bounded, for $s > 0$, by $C_H(\varepsilon + e^{-c_{\mathrm{mix}} s/\varepsilon})$. On an interval bounded away from zero in value, the logarithm is Lipschitz and $\nabla_F \log u = (\nabla_F u)/u$. These observations prove the estimate. The case $s = 0$ follows directly from the hypotheses on $h$ after increasing $C_H$.

For the stationary adjoint, set $v_s = (P_s^\varepsilon)^{\dagger,0}(\omega^\varepsilon h)$. Positivity and (160) give $H^{-1}\omega^\varepsilon \leq v_s \leq H\omega^\varepsilon$. Apply the $m^0$-adjoint Lemma B.24 to $\omega^\varepsilon h$, divide by $\omega^\varepsilon$, and use (139) fiberwise. □

**Theorem B.32** (Two boundary layers for Schrödinger factors)**.** *Fix $K \geq 1$ and assume* (SF1)–(SF4)*. There are constants $C_K, c_{\mathrm{mix}} > 0$, independent of the endpoint pair, $t$, $T$, and $\varepsilon$, such that for every $T \in [T_0, T_1]$, $0 < \varepsilon \leq \varepsilon_0$, $\mu_0, \mu_T \in \mathcal{E}^\varepsilon_{K,\alpha}$, and $0 \leq t \leq T$,*

$$\sup_{\beta} \operatorname{osc}_{\mathsf{F}_\beta} \log f_t^\varepsilon + \|\nabla_F \log f_t^\varepsilon\|_\infty \leq C_K \left(\varepsilon + e^{-c_{\mathrm{mix}} t/\varepsilon}\right), \tag{166}$$

$$\sup_{\beta} \operatorname{osc}_{\mathsf{F}_\beta} \log g_t^\varepsilon + \|\nabla_F \log g_t^\varepsilon\|_\infty \leq C_K \left(\varepsilon + e^{-c_{\mathrm{mix}} (T-t)/\varepsilon}\right). \tag{167}$$

*The forward factor has a left boundary layer, and the backward factor has a right boundary layer. Between them, mixing reduces the variation continually injected by order-one coupling to $O(\varepsilon)$.*

*Let $\mu_t^\varepsilon(\,\mathrm{d}z \mid \beta)$ denote the conditional fiber law of the exact Schrödinger bridge at time $t$. Then*

$$\sup_{\beta} \left\| \log \frac{\mathrm{d}\mu_t^\varepsilon(\cdot \mid \beta)}{\mathrm{d}\pi_\beta} \right\|_\infty \leq C_K \left(\varepsilon + e^{-c_{\mathrm{mix}} t/\varepsilon} + e^{-c_{\mathrm{mix}} (T-t)/\varepsilon}\right), \tag{168}$$

*which also implies the total-variation bound*

$$\sup_{\beta} \|\mu_t^\varepsilon(\cdot \mid \beta) - \pi_\beta\|_{\mathrm{TV}} \leq C_K \left(\varepsilon + e^{-c_{\mathrm{mix}} t/\varepsilon} + e^{-c_{\mathrm{mix}} (T-t)/\varepsilon}\right). \tag{169}$$

*Choose the explicit boundary-layer width*

$$\tau_\varepsilon := \frac{\varepsilon}{c_{\mathrm{mix}}} \log \frac{1}{\varepsilon}, \tag{170}$$

*Then, for all sufficiently small $\varepsilon$ such that $2\tau_\varepsilon \leq T_0$,*

$$\sup_{t \in [\tau_\varepsilon, T - \tau_\varepsilon]} \sup_{\beta} \|\mu_t^\varepsilon(\cdot \mid \beta) - \pi_\beta\|_{\mathrm{TV}} \leq C_K \varepsilon. \tag{171}$$

*Proof.* Applying Lemmas B.30 and B.31 to the two factor evolutions gives (166)–(167).

Relative to the actual invariant conditional $m_\beta^\varepsilon$, the unnormalized conditional fiber density at time $t$ is $f_t^\varepsilon(\beta, z) g_t^\varepsilon(\beta, z)$. Its fiber log-oscillation is at most the sum of the two factor log-oscillations. Normalization over the fiber changes the log-density by a fiber-constant term lying between its minimum and maximum. Finally, (139) changes the comparison measure from $m_\beta^\varepsilon$ to $\pi_\beta$ at an additional $O(\varepsilon)$ cost. This proves (168). The elementary inequality $\|p - q\|_{\mathrm{TV}} \leq (e^\delta - 1)/2$ when $\|\log(\,\mathrm{d}p/\,\mathrm{d}q)\|_\infty \leq \delta$ gives (169). With the width in (170), both exponential terms are at most $\varepsilon$ throughout the interior interval, which proves the uniform bulk estimate (171). □

**Theorem B.33** (Spectral gap of the frozen conditioned fast dynamics)**.** *Under the hypotheses of Theorem B.32, assume in addition that the diffusion covariance has vertical block $\varepsilon^{-1} a_F$, where $a_F$ is the vertical diffusion covariance of $A_\beta$, and all slow–vertical cross blocks are uniformly $O(1)$ in $C^1$. The forward Schrödinger bridge has generator*

$$L_t^{\varepsilon,\star} \varphi = L^\varepsilon \varphi + 2\Gamma_{L^\varepsilon}(\log g_t^\varepsilon, \varphi), \tag{172}$$

*where the carré du champ uses the standard convention* $\Gamma_{L^\varepsilon}(u,v) = \frac{1}{2}[L^\varepsilon(uv) - uL^\varepsilon v - vL^\varepsilon u]$. *Its additional vertical drift is the vector field* $u^\varepsilon_{F,t}$ *defined by writing the vertical first-order part of* $2\Gamma_{L^\varepsilon}(\log g^\varepsilon_t, \varphi)$ *as* $\langle u^\varepsilon_{F,t}, \nabla_F \varphi\rangle$. *It satisfies*

$$\varepsilon \|u^\varepsilon_{F,t}\|_\infty \le C_K \left(\varepsilon + e^{-c_{\mathrm{mix}}(T-t)/\varepsilon}\right). \tag{173}$$

*Thus the bulk control has order one, below the* $O(\varepsilon^{-1})$ *scale of the reference fast dynamics.*

*More precisely, freeze* $(t,\beta)$ *and define the conservative vertical operator*

$$L^{\varepsilon,\star}_{F,t,\beta} := \varepsilon^{-1}\left[A_\beta + 2\Gamma_{A_\beta}(\log g^\varepsilon_t(\beta,\cdot),\cdot)\right]. \tag{174}$$

*Here* $\Gamma_{A_\beta}(u,v) := \frac{1}{2}[A_\beta(uv) - uA_\beta v - vA_\beta u]$. *Its reversible invariant probability is*

$$\pi^{\varepsilon,\star}_{t,\beta}(\mathrm{d}z) := \frac{(g^\varepsilon_t(\beta,z))^2}{\int (g^\varepsilon_t(\beta,z'))^2 \pi_\beta(\mathrm{d}z')}\pi_\beta(\mathrm{d}z). \tag{175}$$

*For* $0 \le t \le T - \tau_\varepsilon$, *its* $L^2$ *Poincaré spectral gap with respect to* (175) *satisfies the uniform lower bound*

$$\mathrm{gap}(L^{\varepsilon,\star}_{F,t,\beta}) \ge \frac{\lambda_\beta}{\varepsilon} - C_K \ge \frac{\lambda_\star}{\varepsilon} - C_K. \tag{176}$$

*This gap describes the reversible frozen vertical operator* (174). *The full* $Q$*-block may be nonreversible, and its corresponding evolution is controlled by* (135). *The corresponding reverse-time estimates use* $f^\varepsilon_t$, *the left boundary layer, and the stationary correction in* (178).

*Proof.* The leading vertical part of $2\Gamma_{L^\varepsilon}$ is $\varepsilon^{-1}\langle a_F \nabla_F \log g^\varepsilon_t, \nabla_F\varphi\rangle$. Equation (173) follows from (167). For the cross terms, decompose $P^\varepsilon_s h = \iota \mathsf{E} P^\varepsilon_s h + Q P^\varepsilon_s h$. Equation (137) bounds the first term in $C^1$, while Lemma B.24 and the full $C^1$ embedding (134) bound the second. Applying these estimates to $h = g^\varepsilon_T$, whose regularity and positive lower bound follow from Lemma B.30, gives $\|\nabla_B \log g^\varepsilon_t\|_\infty \le C_K$. Here $\nabla_B$ denotes the horizontal gradient for the fixed bundle metric. The order-one cross covariance therefore contributes $O(\varepsilon)$ to the normalized drift.

For frozen $(t,\beta)$, direct substitution shows that (175) is invariant and reversible. Because (167) gives $\mathrm{osc}_{\mathsf{F}_\beta} \log g^\varepsilon_t = O(\varepsilon)$ in the bulk, the Holley–Stroock perturbation bound (Holley & Stroock, 1987) gives the unscaled gap

$$\mathrm{gap}\left(A_\beta + 2\Gamma_{A_\beta}(\log g^\varepsilon_t,\cdot)\right) \ge \lambda_\beta e^{-2\,\mathrm{osc}_{\mathsf{F}_\beta}\log g^\varepsilon_t} \ge \lambda_\beta e^{-C_K\varepsilon}. \tag{177}$$

Smoothness and uniform ellipticity on the compact bundle also give $\sup_\beta \lambda_\beta < \infty$, for example by applying the Rayleigh quotient to a finite family of nonconstant local test functions. Restoring the $\varepsilon^{-1}$ time scale and using $e^{-C_K\varepsilon} \ge 1 - C_K\varepsilon$ therefore gives (176). The resulting bound separates the order-one bulk correction from the order-$\varepsilon^{-1}$ control allowed inside the terminal layer.

For the reverse-time statement, write $\theta^\varepsilon_\beta = \mathrm{d}m^\varepsilon_\beta / \mathrm{d}\pi_\beta$. The conservative frozen vertical drift–diffusion part of the stationary reversal, followed by its $f^\varepsilon_t$-transform, is

$$\varepsilon^{-1}\left[A_\beta + 2\Gamma_{A_\beta}(\log(\theta^\varepsilon_\beta f^\varepsilon_t),\cdot)\right]. \tag{178}$$

Equations (139) and (166) make the fiber oscillation of its log potential $O(\varepsilon)$ outside the left layer. Holley–Stroock gives the reverse frozen gap, excluding order-one cross and zeroth-order terms. □

**Remark B.34** (Scope of the endpoint-uniform factor estimates)**.** *Theorems B.5, B.7, and Corollary B.28 concern endpoint couplings and apply to all admissible endpoint pairs because their premise is a uniform relative interaction bound. Theorems B.32 and B.33 additionally control* $\nabla_F \log f^\varepsilon_t$ *and* $\nabla_F \log g^\varepsilon_t$. *The positive likelihood-ratio and regularity bounds in Definition B.29 are exactly what makes these factor and drift estimates uniform over the stated endpoint class.*

### B.6 Complexity consequence

**Corollary B.35** (Per-query complexity)**.** *Suppose a discretization contains* $N_{\mathsf{B}}$ *coarse cells and at most* $N_{\mathsf{F}}$ *states in each fiber. Let* $C_{\mathsf{B}}(N_{\mathsf{B}})$ *be the cost of one application of the positive reduced kernel and let* $I_{\mathsf{B}}$ *be the number of coarse-state Sinkhorn iterations. Let* $I_{\mathsf{X}}$ *denote the number of iterations*

*used by a full-state Sinkhorn solve. The reduced coupling from Corollary B.28 has, for every new endpoint pair, an exact implicit representation with operation count*

$$O(N_{\mathsf{B}}N_{\mathsf{F}}) + O(I_{\mathsf{B}}C_{\mathsf{B}}(N_{\mathsf{B}})) \tag{179}$$

*operations, apart from the cost of drawing requested original-reference conditional bridges. The first term reads the endpoints, forms their coarse marginals, and stores their exact fiber conditionals. The fiber dimension does not occur inside the iterative coarse-state solve. In contrast, an unstructured full dense Sinkhorn solve costs $O(I_{\mathsf{X}}N_{\mathsf{B}}^2N_{\mathsf{F}}^2)$. Materializing every entry of the lifted $N_{\mathsf{B}}N_{\mathsf{F}}$-by-$N_{\mathsf{B}}N_{\mathsf{F}}$ coupling is not included, since that output alone has quadratic size in the number of full states.*

*For a fixed generator, $q$, $\pi_\beta$, $\bar{L}$, and the positive reduced kernel are computed once from reference data. Queries reuse them, forming only their own coarse masses and exact fiber conditionals.*

**Remark B.36** (Positivity and scope)**.** *The reduction uses the conditional expectation $\mathsf{E}$ and the Markov kernel $\bar{P}_T(\beta, \mathrm{d}\beta')\pi_{\beta'}(\mathrm{d}z')$, each of which preserves positivity and mass. The generator-level rate applies under the stated compactness, uniform fast-mixing, and order-one coupling assumptions. The perturbation theorem needs only a bounded uniform interaction fit. Regularity of the endpoint likelihoods is required for the factor and drift estimates, whereas the coupling approximation covers every admissible endpoint pair, with arbitrary within-fiber asymmetry.*

# C Algorithm and complexity details

The reduction leaves three implementation choices: constructing a coarse reference, bounding its interaction residual, and sampling a path from the resulting coupling. We describe them in order, separating reference preprocessing from work tied to the requested endpoints.

## C.1 Discrete endpoint dependence and additive-potential fitting

For a strictly positive stationary transition matrix $K_T$ on $N$ states with invariant vector $m$, its endpoint joint is $J_{ij} = m_i K_T(i,j)$ for $i, j \in \{1, \dots, N\}$. The discrete reference log-interaction is therefore

$$\ell_{ij} = \log K_T(i,j) - \log m_j = \log J_{ij} - \log m_i - \log m_j. \tag{180}$$

Given coarse labels $q_i \in \{1, \dots, N_{\mathsf{B}}\}$, fit row effects $a_i$, column effects $b_j$, and coarse-pair effects $c_{q_iq_j}$, with residual $e_{ij}$ defined by

$$\ell_{ij} = a_i + b_j + c_{q_iq_j} + e_{ij} \tag{181}$$

The fit alternates row, column, and block residual means, centering the nonidentifiable additive terms after every sweep. Exponentiating the coarse effects gives a strictly positive reduced kernel $K_{\mathsf{B}} = \exp(c)$. The squared-error objective chooses the fit, and exhaustive evaluation of $\|e\|_\infty$ then gives its finite-state uniform bound in exact arithmetic. Thus the two computations play different roles. We store $q$, $K_{\mathsf{B}}$, and the evaluated bound from the reference alone for reuse across queries.

When only reference trajectories are available, estimate a positive regularized lagged joint $\widehat{J}$. Summing its fine entries over both coarse cells gives the directly aggregated joint

$$\widehat{J}_{\mathsf{B}}(\beta, \beta') = \sum_{i:q_i=\beta} \sum_{j:q_j=\beta'} \widehat{J}_{ij} \tag{182}$$

and preserves positivity and total mass. Normalizing by the two coarse marginals gives the positive kernel used in the molecular experiments. Because this normalization changes only row and column factors, scaling $\widehat{J}_{\mathsf{B}}$ directly gives the same coupling for fixed target marginals. We therefore obtain two reference-only constructions: additive-potential fitting approximates the log interaction, with its uniform residual evaluated afterward, while aggregation preserves the coarse reference joint.

## C.2 Certifying a continuous supremum residual

For a continuous kernel, a covering argument connects finitely many residual evaluations to the full-domain bound. Let $\mathsf{X}$ be compact with metric $d$, equip $\mathsf{X}^2$ with the sum metric, and define $h(x,y) = a(x) + b(y) + c(q(x), q(y))$ and $e = \ell_R - h$. For a finite partition $q : \mathsf{X} \to [B]$, write $D_{ab} = q^{-1}(a) \times q^{-1}(b)$ for each nonempty cell pair. Since the coarse term can jump at cell interfaces, both regularity and coverage are imposed within each $D_{ab}$.

**Proposition C.1** (Covering certificate). *Fix $0 < \delta_{\mathrm{conf}} < 1$. For each nonempty $D_{ab}$, suppose a representative of $e$ is $L_{ab}$-Lipschitz on $D_{ab}$ in the restricted sum metric, where $L_{ab} \geq 0$ is finite. Let $\mathcal{Z}_{ab} \subset D_{ab}$ be a nonempty finite $r_{ab}$-net of that cell pair, with $r_{ab} \geq 0$: every point in $D_{ab}$ has a point of $\mathcal{Z}_{ab}$ within distance $r_{ab}$. Define $\mathcal{Z} = \bigcup_{a,b} \mathcal{Z}_{ab}$, and assume that finite $\xi_{ab} \geq 0$ and evaluated residuals $\widehat{e}$ satisfy the simultaneous confidence statement below. All maxima over $(a, b)$ omit empty cell pairs:*

$$\Pr\left\{ \max_{a,b} \max_{w \in \mathcal{Z}_{ab}} (|\widehat{e}(w) - e(w)| - \xi_{ab}) \leq 0 \right\} \geq 1 - \delta_{\mathrm{conf}}. \tag{183}$$

*Then, with probability at least $1 - \delta_{\mathrm{conf}}$, the full uniform residual obeys*

$$\eta(h; R) \leq \max_{a,b} \left\{ \max_{w \in \mathcal{Z}_{ab}} |\widehat{e}(w)| + \xi_{ab} + L_{ab} r_{ab} \right\}. \tag{184}$$

*Proof.* For every $(x, y) \in D_{ab}$ choose $w \in \mathcal{Z}_{ab}$ within distance $r_{ab}$, so the point and its witness belong to the same cell pair. The triangle inequality, the Lipschitz bound, and (183) give $|e(x, y)| \leq |\widehat{e}(w)| + \xi_{ab} + L_{ab} r_{ab}$. Taking the supremum in each cell pair and then the finite maximum proves the claim. A representative defined outside a null set suffices when the coverage and regularity assumptions are imposed there, since the claimed norm is an essential supremum. □

Here the probability is over evaluation randomness, conditional on the fitted map, potentials, and net. If these objects are data-dependent, the evaluation sample must be independent, or the confidence statement must hold uniformly over their possible outputs. For example, let $n_{\mathrm{net}} = |\mathcal{Z}|$ and suppose the evaluation errors are conditionally centered sub-Gaussian with proxy variance $\sigma^2/n$, where $\sigma \geq 0$ and $n \geq 1$ is the evaluation sample size. A union bound over all cell-pair net points gives the simultaneous statement by setting every $\xi_{ab}$ equal to

$$\xi_{\mathrm{conf}} = \sigma \sqrt{\frac{2 \log(2 n_{\mathrm{net}} / \delta_{\mathrm{conf}})}{n}}. \tag{185}$$

In a model with known coefficients, within-cell derivative or interval bounds on the transition log-density and fitted potentials provide $L_{ab}$. Numerical partial-differential-equation error estimates or simultaneous concentration on an independently evaluated net provide $\xi_{ab}$. Uniform bounds $L_{ab} \leq L_e$, $r_{ab} \leq r_{\mathrm{net}}$, and $\xi_{ab} \leq \xi_{\mathrm{conf}}$ recover the simpler expression $\max_{w \in \mathcal{Z}} |\widehat{e}(w)| + \xi_{\mathrm{conf}} + L_e r_{\mathrm{net}}$. In a finite state space, exhaustive evaluation takes $r_{ab} = 0$ and gives the matrix residual in exact arithmetic. More generally, evaluation error is included in $\xi_{ab}$. Thus the perturbation theorem can use a covering bound, an exhaustive finite maximum, or the analytic estimate of Theorem 5.1. A held-out maximum without coverage and regularity information describes the evaluated points only.

### C.3 Learning a coarse partition from reference trajectories

Form the normalized lagged joint from $d_0 = \widehat{J}\mathbf{1}$ and $d_T = \widehat{J}^\top \mathbf{1}$:

$$M = \mathrm{diag}(d_0)^{-1/2} \widehat{J}\, \mathrm{diag}(d_T)^{-1/2}. \tag{186}$$

For a nonreversible reference, left and right singular coordinates of $M$ describe complementary kinetic information. We therefore concatenate the nonconstant leading coordinates, weight states by their stationary mass, and cluster them to obtain $q$. If $\sigma_1 \geq \sigma_2 \geq \cdots$ are the singular values and $\mathcal{K}$ is the fixed set of candidate coarse cardinalities, the spectral-ratio rule chooses

$$N_{\mathsf{B}} = \arg\max_{k \in \mathcal{K}} \frac{\sigma_k}{\sigma_{k+1}}, \qquad \sigma_{k+1}/\sigma_1 > \tau_{\mathrm{spec}}, \tag{187}$$

where $\tau_{\mathrm{spec}} > 0$ is a fixed numerical floor restricting the eligible candidates. The singular vectors determine the coarse coordinate, whereas the subsequent positive aggregation or additive-potential fit determines its interaction kernel. Sinkhorn therefore receives the positive table (182) or $\exp(c)$, independently of the signs of the spectral coordinates used to construct the partition.

**Choice of $N_{\mathsf{B}}$ and nonuniform mixing.** Proposition B.9 gives the population bias direction: a refined partition has a weakly smaller best residual. When a reference residual curve is available, a principled rule chooses the smallest $N_{\mathsf{B}}$ after its certified or held-out value stops decreasing materially. Too few cells leave slow interaction inside fibers. Additional cells increase coarse solve cost and estimation variance without a population-residual penalty. This criterion uses no endpoint tasks.

With variable fiber gaps, the endpoint-uniform rate uses $\lambda_\star = \inf_\beta \lambda_\beta$, so one small local gap controls the global bound. When its eigenfunction identifies a coherent slow mode, promote that mode to the coarse coordinate by refining its cell. This refinement assigns the slowly relaxing variation to the represented coordinate. For finite partitions, we can also compute cell-pair residual envelopes $\delta_{\beta\beta'}$. Proposition B.10 then bounds each KL direction by $2U_\delta(q_\#\mu, q_\#\nu)$ through a coarse optimal-transport linear program weighted by the prescribed endpoint marginals.

### C.4 Coarse-state scaling and exact conditional lift

For a positive reduced reference matrix $W$ and coarse marginals $p_0 = \mu^{\mathsf{B}}$, $p_T = \nu^{\mathsf{B}}$, initialize positive scaling vectors $u, v$ and alternate elementwise divisions in the Sinkhorn updates

$$u \leftarrow p_0/(Wv), \qquad v \leftarrow p_T/(W^\top u), \qquad \widehat{\Gamma} = \operatorname{diag}(u) W \operatorname{diag}(v), \tag{188}$$

with logarithmic stabilization when needed. The objective relative to $p_0 \otimes p_T$ differs from $\sum_{ab} \Gamma_{ab} \log(\Gamma_{ab}/W_{ab})$ only by the fixed constant $-\sum_a (p_0)_a \log(p_0)_a - \sum_b (p_T)_b \log(p_T)_b$. Positive row and column factors are likewise absorbed into $u, v$, so the additive endpoint potentials need not be stored or evaluated online. In a finite partition, the lifted matrix is

$$\widehat{\gamma}_{ij} = \widehat{\Gamma}_{q_i q_j} \frac{\mu_i}{\mu^{\mathsf{B}}_{q_i}} \frac{\nu_j}{\nu^{\mathsf{B}}_{q_j}}, \tag{189}$$

with the usual zero-mass convention. Summing (189) over $j$ or $i$ gives $\mu_i$ or $\nu_j$ at the exact scaling solution. At a finite stopping tolerance, the fine and coarse $L^1$ marginal residuals still agree because each endpoint conditional sums to one. The implicit representation stores $\widehat{\Gamma}$, the endpoints, their coarse masses, and $q$. To sample, first draw $(\beta, \beta') \sim \widehat{\Gamma}$, then independently draw $i \sim \mu(\cdot \mid \beta)$ and $j \sim \nu(\cdot \mid \beta')$. A path request continues with the conditional reference bridge $R^{ij}$.

### C.5 Conditional reference-bridge sampling

For a finite Markov generator $L$, take a uniformization rate $\Lambda > 0$ with $\Lambda \geq \max_i(-L_{ii})$ and set the stochastic matrix $U = I + L/\Lambda$. Given endpoints $i, j$ and horizon $T$, the event count $M$ has probabilities proportional to $e^{-\Lambda T}(\Lambda T)^m (U^m)_{ij}/m!$ for integers $m \geq 0$. Given $M = m$, at event $k$ in state $x$, sample the next state $y$ with probability $U_{xy}(U^{m-k-1})_{yj}/(U^{m-k})_{xj}$ for $k = 0, \ldots, m-1$, whenever the denominator is positive. Sorted independent uniform event times in $[0, T]$ complete the bridge. The construction is exact if the event count is sampled without truncation.

For the molecular references, a one-step transition matrix defines a discrete Markov path law. Appendix D.8 samples its multi-step bridges using consistent forward–backward powers. Continuous diffusion references use their corresponding conditioned simulators. In each case, endpoint scaling selects the coupling, and conditional simulation supplies the path between the sampled endpoints.

If a supplied conditional sampler $\widetilde{R}^{xy}$ has the exact endpoints and conditional KL errors bounded by $\delta_+, \delta_-$ in the two directions, Proposition B.13 adds these errors to the coupling bounds. Thus the general path bounds are $2\eta + \delta_\pm$. For a finite endpoint kernel they improve to $\min\{2\eta, \eta^2/2\} + \delta_\pm$ by Theorem B.15. This separates the coupling approximation from conditional simulation error.

**Changing the horizon.**

**Proposition C.2** (A common partition across horizons)**.** *Let $K \in \mathbb{R}^{N\times N}_{>0}$, $q : [N] \twoheadrightarrow [B]$, and $Q_{ib} = \mathbf{1}_{\{q(i)=b\}}$. Suppose $H = D_a Q W Q^\top D_b$ satisfies $\|\log K - \log H\|_\infty \leq \eta$, where $D_a = \operatorname{diag}(e^{a_1}, \ldots, e^{a_N})$, $D_b = \operatorname{diag}(e^{b_1}, \ldots, e^{b_N})$, and $W > 0$. For each integer $m \geq 2$, define*

$$C_m = W Q^\top D_b K^{m-2} D_a Q W, \qquad H_m = D_a Q C_m Q^\top D_b.$$

*Then $C_m > 0$, $\|\log K^m - \log H_m\|_\infty \leq 2\eta$, and hence $\eta_q(K^m) \leq 2\eta_q(K)$. For every endpoint pair, the lift using $q, C_m$ has both coupling KL directions at most $\min\{4\eta, 2\eta^2\}$.*

*Proof.* The factorization gives $H_m = HK^{m-2}H$. Since multiplication by a nonnegative matrix preserves entrywise inequalities, $e^{-\eta}H \le K \le e^{\eta}H$ implies $e^{-2\eta}H_m \le K^m \le e^{2\eta}H_m$. The middle coarse matrix is positive: for $m = 2$, $Q^\top D_b D_a Q$ is positive diagonal and $W > 0$, for $m > 2$, the intermediate full matrix is positive. Taking logarithms proves the residual bound. Applying it to a minimizing fit for $K$ proves the assertion for $\eta_q$, and Theorem B.15 gives the KL estimates. □

Proposition C.2 approximates the two boundary factors and retains the exact interior evolution. Consequently, its log-error does not accumulate over the intervening steps. This extends the exact low-rank closure under powers (Zhang & Wang, 2020) to a uniform log-error bound. The same partition serves all horizons, while constructing each $C_m$ uses lag-specific full-state preprocessing. Direct aggregation of a separately estimated lagged joint gives another offline construction.

The representation is indexed by its reference lag $T$. If a new query changes $T$, one may retain the partition $q$ and construct a positive coarse kernel for that lag before solving. A collection of preselected lags can share $q$ and store one coarse matrix per lag. For two positive finite kernels $K_T, K_S$, if $\inf_{a,b} \| \log K_T - \log K_S - a - b\|_\infty \le d$ and a coarse kernel fits $K_S$ with residual $\eta_S$, the same coarse kernel fits $K_T$ with residual at most $\eta_S + d$, by adding those row and column terms. The finite KL bound therefore uses $\min\{2(\eta_S + d), (\eta_S + d)^2/2\}$. Thus a nearby lag can reuse a coarse kernel with an enlarged interaction bound, while a precomputed kernel handles its lag.

### C.6 Verifying the slow–fast assumptions

Each analytic assumption supplies a specific step of the proof, following the averaging and semigroup framework of Pavliotis & Stuart (2008); Lunardi (1995). For (SF1), compute the fiber invariant law and a uniform Poincaré bound, then establish the stated $C^1$ mixing estimate. For (SF2), control both forward and adjoint $Q$ evolutions and the couplings $QS\iota$ and $\mathsf{E}SQ$ in the declared graph norm. Next, solve the effective invariant problem and bound the conditional density correction for (SF3). Finally, uniform heat-kernel or Schauder bounds give the (SF4) smoothing estimate (Lunardi, 1995).

Finite matrices make this route particularly explicit. Theorem B.21 allows nonreversible fast blocks and derives the required estimates from their mixing bounds. Corollary B.22 then converts the transition matrix directly into an endpoint-uniform interaction bound. For the continuum frozen-gap result, reversibility additionally enables the Holley–Stroock comparison. Whenever a residual is established by another route, Theorem B.7 applies directly to each such reference bound.

### C.7 Complexity accounting

Let $N_\beta$ count states in cell $\beta$, with $N = \sum_{\beta=1}^{N_{\mathsf{B}}} N_\beta$ full states. Reading and aggregating dense endpoints costs $\Theta(N)$, followed by $I_{\mathsf{B}} C_{\mathsf{B}}(N_{\mathsf{B}})$ work for coarse scaling. The deployed representation—operator, labels, and scaling state, excluding the two common input endpoints—requires

$$O(N_{\mathsf{B}}^2 + N) \tag{190}$$

for a dense reduced kernel, compared with $O(N^2)$ for a dense full kernel. Every method also reads the same two $\Theta(N)$ endpoint inputs. The conditional ratios in (189) can be evaluated lazily or formed in place, so they need no second copy. An exact positive factorization of width $r$ uses $O(rN)$ kernel memory and work per product, providing the structure-aware comparator in the controlled benchmark. These counts describe implicit couplings. A requested full matrix adds quadratic output cost.

The abstract cost $C_{\mathsf{B}}(N_{\mathsf{B}})$ also allows different implementations of the reduced kernel product. Dense or distributed products and exact positive factors preserve the scaling problem, as does a FlashSinkhorn implementation when the coarse costs have its required squared-Euclidean form (Ye et al., 2026). Such choices change product cost and numerical stopping error. If the coarse kernel itself is approximated, its additional log-interaction error is included in the perturbation bound.

## D Experimental models and protocols

### D.1 Common quantities and evaluation design

The experiments address three questions in sequence: whether reference interaction controls approximation error, how reference-only representations compare at a fixed deployment budget, and

how these savings translate into complete path requests. Every comparison fixes the reference and evaluation queries before fitting its representations. Consequently, changes in accuracy reflect the representation or solver being tested, under common endpoint preparation and stopping rules.

For a returned coupling $\widehat{\gamma}$ and the dense optimum $\gamma^\star$, we measure

$$D_+ = \mathrm{KL}(\widehat{\gamma}\|\gamma^\star), \qquad D_- = \mathrm{KL}(\gamma^\star\|\widehat{\gamma}), \qquad D_{\mathrm{sym}} = \tfrac{1}{2}(D_+ + D_-), \qquad \mathrm{TV}_{\mathrm{cpl}} = \tfrac{1}{2}\|\widehat{\gamma} - \gamma^\star\|_1. \tag{191}$$

Thus $D_{\mathrm{sym}}$ gives equal weight to both directions of approximation. Feasibility is measured separately by

$$r_{\mathrm{marg}} = \max\{\|\widehat{\gamma}\mathbf{1} - \mu\|_1, \|\widehat{\gamma}^\top\mathbf{1} - \nu\|_1\}. \tag{192}$$

Scaling is restricted to the positive endpoint supports, and zero masses are restored exactly. The controlled chain uses tolerance $10^{-11}$, kinetic and molecular problems use $10^{-10}$, and the exact-factorization scaling benchmark uses $10^{-9}$. Molecular accuracy references additionally use $10^{-12}$. All methods within a timing comparison share a tolerance, use double precision, and stop on $r_{\mathrm{marg}}$. Controlled and kinetic problems use log-domain scaling, while positive-operator implementations rebalance their logarithmic scaling factors during iteration to control numerical range.

Implicit-coupling timings include endpoint processing and all scaling iterations, with dense coupling materialization excluded for every method. Complete-path timings additionally include sampler preparation and all requested paths. Reference fitting is charged once in repeated-query totals. Deployment storage counts the representation required for a new query, and query workspaces and shared conditional-path arrays are accounted for separately. Processor comparisons use one numerical-library thread. Five repetitions are used for the original molecular comparison and seven for the additional amortized comparison, after an unrecorded warm-up. Neural sampling uses three warm-ups followed by seven synchronized repetitions. This common accounting separates reference reuse from repeated sampling of an already fitted bridge at fixed endpoint constraints.

### D.2 Controlled references and finite-state theory

We first use a periodic $16 \times 24$ chain to connect the finite-state theorem with measurable interaction and boundary layers. Let $\beta$ and $z$ index the coarse and fiber grids, with $\theta_\beta = 2\pi\beta/16$, $\varphi_z = 2\pi z/24$, and $\Delta z = 2\pi/24$. Define

$$U_{\mathsf{B}}(\beta) = 1.5(1 - \cos 2\theta_\beta), \qquad U_{\mathsf{F}}(\beta, z) = 0.6\{1 - \cos(\varphi_z - 0.65\sin\theta_\beta)\}. \tag{193}$$

The fiber rate from $z$ to $z \pm 1$ and coarse rate from $\beta$ to $\beta + d$, $d \in \{-1, 1\}$, are respectively

$$\frac{1}{\varepsilon\Delta z^2}\exp\left[-\frac{U_{\mathsf{F}}(\beta, z \pm 1) - U_{\mathsf{F}}(\beta, z)}{2}\right], \tag{194}$$

$$0.15\exp\left[-\frac{U_{\mathsf{B}}(\beta + d) - U_{\mathsf{B}}(\beta)}{2} + d\{0.45\sin(\varphi_z - 0.65\sin\theta_\beta) + 0.2\}\right]. \tag{195}$$

The fiber equilibrium therefore varies across coarse states, and the coarse motion depends on the current fiber phase. The constant drive breaks reversibility. At $T = 1$, we evaluate $\varepsilon \in \{1, 0.5, 0.25, 0.125, 0.0625\}$ with five independent endpoint pairs per scale. Endpoints exponentially tilt the invariant distribution by bounded coarse, fiber, and mixed harmonics at strength 2.5. All $N^2$ reference entries enter the residual calculation. The minimum frozen-block gap is 1.052878, and $\|S\|_\infty = 1.101230$, so this example directly tests Theorem B.21.

The kinetic experiment then asks whether the same interaction picture extends to a singularly coupled reference. On a periodic 20-point position grid and a reflecting 17-point velocity grid in $[-3.5, 3.5]$, use

$$L^\varepsilon = \varepsilon^{-1}L_{\mathrm{OU}} + \varepsilon^{-1/2}0.65v\nabla_x + \varepsilon^{-1/2}F(x)\nabla_v + 0.01\Delta_x, \qquad F(x) = -0.35\,\partial_x[0.8(1-\cos x)]. \tag{196}$$

Here $L_{\mathrm{OU}}$ is the Ornstein–Uhlenbeck velocity generator with unit temperature and friction. Birth–death velocity jumps and upwind transport give nonnegative rates. Five endpoint pairs per scale use opposing velocity biases and strength-4 mixed-coordinate tilts. Since the cross terms have order $\varepsilon^{-1/2}$, the measured finite-state residual supplies the perturbation guarantee for this experiment. At 61 equally spaced times we compute

$$\sum_x p_t(x)\,\mathrm{TV}\{p_t(\cdot \mid x), m(\cdot \mid x)\}, \tag{197}$$

which measures fiber detail relative to the stationary conditional. Boundary width is the first distance from an endpoint at which this quantity falls below its interior minimum plus $1/e$ of its endpoint excess. A separate $\varepsilon = 0.125$ experiment holds both position marginals fixed and varies only conditional-velocity tilt strength over $0, 1, 2, 4, 8, 12$, using 20 independent pairs per strength.

To isolate arithmetic scaling, we also construct an exact positive factorization on $24 \times N_{\mathsf{F}}$ states. With $\theta_\beta = 2\pi\beta/24$ and $a_k = 2\pi(k + 0.37)/8$, set

$$A_{\beta k} = \exp\{0.75\cos(\theta_\beta - a_k) + 0.18\sin(2\theta_\beta + a_k)\},$$
$$B_{k\beta'} = \exp\{0.68\cos(a_k - \theta_{\beta'} + 0.31) + 0.16\sin(a_k + 2\theta_{\beta'})\}. \tag{198}$$

Normalize $W = AB/\operatorname{mean}(AB)$ and form $K_{ij} = g_i W_{q(i),q(j)} h_j$ with positive endpoint factors. Dense, supplied-rank-eight, and coarse scaling then solve the same problem. Nine fiber sizes from 4 to 512 give up to 12,288 states, with eight independent endpoint pairs per size. Dense verification through $N = 3072$ gives coupling TV below $2.4 \times 10^{-13}$, while the exact factorization identifies the common operator at larger sizes without materializing a quadratic endpoint coupling.

Finally, six positive six-state references test the minimax brackets without a prescribed partition. An irregular reversible ring, a directed cycle, and a dense nonreversible generator are each evaluated at horizons 0.6 and 1.4. We enumerate all 31 two-cell and 90 three-cell partitions, compute $D_q$, and solve the Chebyshev problem for $\eta_q$. Sparse two-point queries independently check the cross-ratio formulas. To study gradual relaxation spectra, six binary coordinates have flip rates $a_l(\alpha) = a_l(0)^{1-\alpha} a_l(1)^\alpha$, where $a(0) = (0.06, 0.06, 0.06, 3, 3, 3)$ and $a(1)$ is geometrically spaced from 0.06 to 3. We use $T = 1$, five values of $\alpha$ from zero to one, and budgets $B \in \{2, 4, 8, 16, 32\}$. These references connect the exact inequalities with both separated and graded relaxation rates in the same finite state space.

### D.3 Molecular references and held-out endpoints

We next freeze references estimated from three public molecular systems. If $C$ is the retained transition-count matrix, the reversible empirical joint is $J^{\mathrm{emp}} = (C + C^\top)/(2\sum_{ij} C_{ij})$, with marginal $m = J^{\mathrm{emp}}\mathbf{1}$. Product shrinkage gives $J_\alpha = (1-\alpha)J^{\mathrm{emp}} + \alpha m m^\top$. Validation selects $\alpha$ by directional transition negative log likelihood under $P_\alpha(i,j) = J_\alpha(i,j)/m_i$. This uses transition probabilities on a fixed state alphabet. The corresponding bridge benchmarks subsequently condition this estimated reference on new endpoint distributions drawn after reference selection.

**Alanine dipeptide.** Three 250-ns trajectories contain 250,000 backbone-dihedral frames each at 1-ps spacing (Nüske et al., 2017; Wehmeyer & Noé, 2018; mdshare, n.d.a). A periodic $30 \times 30$ grid retains 311 states with at least 20 training frames and positive lagged activity. The first two trajectories define support, and the third selects shrinkage at a 250-ps lag. Pooling the trajectories after selection gives $\alpha = 0.3$. The reference-only adjacent singular-value ratio selects $B = 2$. Thirty independent endpoint pairs are then generated by smooth Fourier tilts of the invariant law. Each tilt is reference-centered, normalized to unit reference variance, clipped to $[-3, 3]$, and exponentiated at strength two. Therefore no evaluation endpoint enters reference estimation or partition selection.

**Chignolin.** We use the long equilibrium trajectories `run0`, `run2`, and `run3` in the public archive of Lazzeri (2025), sampled at 100-ps spacing and simulated at 340 K. Every fifth frame is retained, and backbone torsions give 36 sine–cosine features. The first trajectory fits 128 microstates, the second selects shrinkage and coarse cardinality, and their pooled counts fit the frozen reference at a 5-ns lag. Transition validation selects $\alpha = 0.1$ and $B = 16$, using the smallest cardinality within 0.02 nats of the best candidate. The third trajectory supplies the held-out likelihood and observed endpoints. Thirty synthetic pairs use strength-two nonlinear projections of the torsion features. Separately, 20 consecutive windows of the 52,265 retained test frames give ten adjacent-window occupancy pairs, with empirical zero masses unchanged throughout scaling and reconstruction.

**WLALL pentapeptide.** The public `mdshare` dataset contains 25 independent 500-ns Langevin trajectories at 300 K in implicit solvent, stored every 100 ps (mdshare, n.d.b; Scherer et al., 2015). Whole trajectories 00–14, 15–19, and 20–24 define training, validation, and testing. MDTraj supplies backbone torsions (McGibbon et al., 2015), and training-only $k$-means fits their sine–cosine features (Pedregosa et al., 2011). The principal 0.5-ns reference retains 503 active states from 512 requested centers. Reference validation selects shrinkage and $B$ before fitting pooled training

and validation counts. The five test trajectories yield 20 ordered occupancy pairs. Ten additional queries connect central-alanine angular regions within each trajectory: the alpha-like region has $\phi \in [-150, -30]$, $\psi \in [-90, 45]$, and the beta-like region has $\phi \in [-180, -60]$, $\psi \in [90, 180]$, in degrees. Each region contains at least 20 retained frames. The five complete test trajectories are the independent physical sampling units underlying these 30 queries in all reported summaries.

A fixed-reference sensitivity study varies lag and shrinkage while preserving microstate centers, coarse labels, and all endpoints. Lags are $125, 250, 500$ ps for alanine, $2.5, 5, 10$ ns for Chignolin, and $0.1, 0.5, 2$ ns for WLALL. Each lag defines its own transition reference. Dense, coarse, and independent endpoint pairing are evaluated on the same reference, so the comparison distinguishes useful compression from dependence removed by regularization. All candidate references and their query errors, validation likelihoods, and mutual information are retained in the numerical supplement.

### D.4 Reference representations and amortized competitors

The molecular comparison crosses four reference-only partitions with two positive coarse kernels at $B \in \{2, 4, 8, 16, 32\}$. Partitions use spectral coordinates, the reference log-ratio distance, reference-weighted geometric $k$-means, or geometric farthest-first centers. Each kernel either aggregates the reference over cell pairs or evaluates it at representative states. Geometric methods use the same molecular features and retain the same transition reference. Their initialization is selected by reference-weighted distortion, and every coarse solve uses the original endpoint conditionals for lifting. Positive nonnegative matrix factorization (NMF) uses generalized-KL multiplicative updates (Lee & Seung, 2001) at ranks $1, 2, 4, 8, 16, 32$. The reference objective selects among three initializations, with up to 10,000 updates and relative convergence tolerance $10^{-7}$. Offline time includes all fitted initializations. Thus representations use the frozen reference alone.

Dense cold starts, previous-query warm starts, and simultaneous batches provide exact-to-tolerance comparisons. A serial warm start retains the preceding target scaling, while a batch uses shared matrix–matrix products and freezes converged columns. Dense accuracy references and all accelerated operators use the same positive kernel and endpoint supports. Coarse timings include marginal aggregation and conditional-array preparation. Repeated-query totals include representation construction and complete replays of the prescribed order, with $Q \in \{1, 10, 100, 1000\}$. The deployed reference-selected budgets are $B = 2, 16, 8$ for alanine, Chignolin, and WLALL, respectively. Full budget curves and initialization records accompany the numerical supplement.

We additionally compare two learned initializers and hierarchical exact refinement on all 100 molecular queries. The Meta-OT dual-objective adapter follows the discrete semi-dual training principle of Amos et al. (2023). The UNOT multilayer-perceptron adapter follows the non-grid implementation of Geuter et al. (2025). Both predict initial potentials and then refine the original fine reference to tolerance $10^{-10}$. Hierarchical refinement uses a reference-only Ward hierarchy with 8 and 32 cells, coarse tolerance $10^{-6}$, endpoint-conditioned prolongation, and the same final fine-kernel solve. It evaluates the coarse-to-fine component of multiscale scaling (Schmitzer, 2019). Accordingly, its deployment retains the fine kernel as well as the hierarchy for subsequent fine-state refinement.

For completeness, let $J$ be the frozen joint, $m$ its marginal, and $K_{ij} = J_{ij}/(m_i m_j)$. Given endpoints $a, b$, a predictor returns a mean-centered correction $h_\theta(a, b)$ and initializes $v = b \odot \exp h_\theta$. Meta-OT minimizes

$$-\sum_i a_i \log \frac{a_i}{(Kv)_i} - \sum_j b_j \log v_j$$

with three width-256 rectified-linear hidden layers and Adam at learning rate $0.001$. UNOT uses four width-256 hidden layers with batch normalization and exponential-linear activations. Five detached Sinkhorn updates supply its bootstrap targets, while an equally sized adversarial marginal generator maximizes potential discrepancy. Predictor and generator use AdamW at $0.001$, with predictor weight decay $0.0001$. Both predictors receive marginals multiplied by the state count and use batches of 32.

Training mixes smooth coordinate tilts, localized mixtures, sparse histograms, and gamma-normalized marginals drawn independently of the test pairs. Training-only mass $0.000001$ stabilizes logarithms, while test zeros remain unchanged. Three independent initializations and budgets of 500 and 2000 updates yield 18 training runs and 36 checkpoints. Sixty-four separate validation pairs per reference select the checkpoint with the lowest refinement iteration count. All final solves use double precision, check the common marginal residual at iteration one and every ten iterations, and allow at most

20,000 iterations. Online cost includes predictor inference or hierarchical initialization and full refinement. Offline cost includes endpoint generation, training, bootstrap targets, validation, and model selection. Thus training and validation enter the amortized preprocessing cost.

### D.5 Query-dependent error and refinement

For the same 100 molecular queries, let $P$ be the deployed lift and $Q$ the dense optimum. We evaluate $F = \mathrm{KL}(P\|Q)$ and $G = \mathrm{KL}(Q\|P)$, then compare them with the bounds of Proposition B.11. Starting from $\log K_{ij} - \log W_{q(i),q(j)}$, a reference-only additive fit gives residual $e_{ij}$. Four fixed witnesses use this residual directly, center it under $\mu\otimes\nu$, or apply one or two alternating normalization passes to $P_{ij}\exp(e_{ij})$. These passes change the witness potentials while leaving $P$ unchanged. Forward bounds use $M_g(1)$, and reverse bounds minimize $(M_g(t)-tC)/(t-1)$ over $t \in [1.001, 16]$, including $t = 2$ and both interval endpoints for each of the four fixed witnesses.

The corrected coarse kernel $W_{ab}\exp(\bar{e}_{ab})$, where $\bar{e}_{ab}$ averages $e$ under the endpoint product conditionals, defines $P^\dagger$. We compute $C = \mathrm{KL}(P\|P^\dagger)$ and $J = \mathrm{KL}(P^\dagger\|Q)$ and check $F = C + J$. A cellwise maximum residual also gives the transport envelope $2U_\delta$. These finite-matrix calculations explain which interactions each query encounters. We time these scans and corrected-kernel solves separately from online scaling, using the dense optimum only to measure approximation error.

Nested refinements start from the deployed $B = 2, 16, 8$ partitions. Write $J_{\mathrm{ref}}$ for the frozen symmetric reference joint, $m = J_{\mathrm{ref}}\mathbf{1}$ for its marginal, and $D_m = \mathrm{diag}(m)$. One hierarchy uses the 16 leading nonconstant eigenvectors of $D_m^{-1/2}J_{\mathrm{ref}}D_m^{-1/2}$, ordered by eigenvalue magnitude, multiplied by their eigenvalues and by $D_m^{-1/2}$. The second uses centered log-interaction features. Specifically, for $L = \log J_{\mathrm{ref}}$, set

$$E_{ij} = L_{ij} - \sum_k m_k L_{ik} - \sum_k m_k L_{kj} + \sum_{k,l} m_k m_l L_{kl}.$$

State $i$ is represented by the concatenated weighted row and column $(E_{ij}\sqrt{m_j})_j$ and $(E_{ji}\sqrt{m_j})_j$. At each step, the nonsingleton cell with largest weighted feature variance is split by deterministic weighted two-means. Thus every level refines the preceding partition and uses no test endpoint information. We evaluate all original budgets above the starting level, additional levels up to 256 when available, and the singleton limit. Across 1280 evaluations, the maximum discrepancies in $F = C + J$ and the nested-refinement identity are $4.31 \times 10^{-11}$ and $4.55 \times 10^{-11}$. The singleton level provides the exact-resolution endpoint of each hierarchy, and intermediate levels show how the representation budget resolves within-cell dependence under an unchanged reference law.

### D.6 Compatible Gaussian-reference comparisons

For $N \in \{512, 2048\}$, draw reference points $x_i \in \mathbb{R}^3$ with Gaussian coordinate standard deviations $(1.2, 0.7, 0.3)$, clipped to $[-2.5, 2.5]$. The fixed kernel is $K_{ij} = \exp(-\|x_i - x_j\|^2/2)$. Query $k \in \{0, \ldots, 9\}$ uses $d_k = (\cos(2\pi k/10), \sin(2\pi k/10), 0)$ and draws 256 distinct states on each side with probabilities proportional to $\exp(x_i^\top d_k)$ and $\exp(-x_i^\top d_k)$. Selected states receive uniform endpoint weights. Each active problem is $256 \times 256$, independently of the bank size $N$.

This Gaussian geometry supports the public QDSB coverage-coupling routines (Fuchs et al., 2026), POT Nyström approximation (Flamary et al., 2021), and low-nonnegative-rank coupling solver (Scetbon et al., 2021). QDSB uses endpoint-selected anchors and is evaluated as a coupling method. Its double-precision marginal correction is included in timing. The low-rank solver uses factor-entropy coefficients $\rho \in \{0, 2\}$, and its outputs are evaluated under the original Gaussian-reference objective. These objective and information-access conventions distinguish the comparators without changing the shared endpoint task or its Gaussian-reference target.

Nyström uses landmark counts $r \in \{8, 16, 32, 64, 128, 256, 512\}$, split equally between endpoints. For landmark set $Z$, an eigendecomposition of $K(Z, Z)$ retains eigenvalues greater than $\tau\lambda_{\max}$, with $\tau \in \{10^{-12}, 10^{-10}\}$. The resulting factors give $\widetilde{K} = LH^\top$. Positive variants scale $\max\{\widetilde{K}_{ij}, \delta\}$ for $\delta \in \{10^{-12}, 10^{-9}, 10^{-6}\}$. All 840 repaired configurations meet the shared $10^{-10}$ marginal tolerance. Their timing includes construction, entrywise correction, and scaling, and their implementation stores a 512-KiB active kernel. The numerical supplement includes every cutoff, floor, landmark count, and query, together with native-factor comparisons and implementation-specific timing scopes.

Table 2: Learned-bridge parameterizations and training budgets. Widths give hidden-layer count and units. Each DSBM cycle fits both drifts for 10,000 updates each, while other budgets count gradient updates. LightSB-M directly samples endpoints and their joint Brownian conditional paths.

| Method | Parameterization | Batch size | Retained budgets | Steps |
|---|---|---|---|---|
| DSBM | Two drift networks, each $2 \times 256$ | 128 | 1, 4, 8 cycles | 200 |
| LightSB-M | 50 diagonal Gaussian potentials | 512 | 2,000, 10,000, 50,000 | Direct |
| SF$^2$M-Sink | Flow and score, each $3 \times 64$ | 256 | 1,000, 5,000, 10,000, 20,000 | 100 |

### D.7 Learned bridges and common path evaluation

The same Gaussian banks define a precise common path target. Let $\lambda_i = 1/N$, let $p_T$ be the Brownian transition density, and set $T = \sigma = 1$. With normalizing constant $Z$, define

$$r_{\mathrm{bank}}(i,j) = Z^{-1}\lambda_i\lambda_j p_T(x_i, x_j), \qquad R_{\mathrm{bank}} = \sum_{i,j} r_{\mathrm{bank}}(i,j) R_{\mathrm{BM}}^{x_i x_j}. \tag{199}$$

Here $R_{\mathrm{BM}}^{xy}$ is the unconstrained Brownian conditional bridge, so atomic bank endpoints have positive reference mass. The interior reference is Markov, with Doob transform $h_t(z) = \sum_j \lambda_j p_{T-t}(z, x_j)$ and initial weights proportional to $\lambda_i h_0(x_i)$. Exact Sinkhorn couples the selected endpoints under this reference, and its conditional paths define the evaluation target throughout $\mathbb{R}^3$.

DSBM alternates backward and forward drift fitting (Shi et al., 2023). LightSB-M fits a diagonal Gaussian-mixture potential by optimal bridge matching (Gushchin et al., 2024). SF$^2$M-Sink jointly fits flow and scaled score using a Sinkhorn-coupled matcher (Tong et al., 2024), with entropy coefficient $2\sigma^2 T = 2$. Table 2 gives every retained budget. DSBM uses Adam at $0.0001$, LightSB-M uses Adam at $0.001$, and SF$^2$M-Sink uses AdamW at $0.001$ with weight decay $0.00001$. Two independent development queries select the smallest integration grid within 5% of the 1000-step path and terminal distances. This gives 200 steps for DSBM and 100 for SF$^2$M-Sink.

Each method returns one joint path at $t_l = l/4$, $l = 0, \ldots, 4$. The drift models integrate their stochastic differential equations, while exact, coarse, and LightSB-M samplers first draw coupled endpoints and then sample

$$Z_t = (1-t)x + ty + W_t - tW_1, \qquad t \in \{0, 0.25, 0.5, 0.75, 1\}, \tag{200}$$

where $W$ is a common three-dimensional Brownian motion along that path. In particular, its conditional covariance is $(\min\{s,t\} - st)I_3$, preserving dependence along the path.

To compare these joint laws, concatenate the five states into $\mathbb{R}^{15}$ and use 256 shared unit projections. For $M \in \{128, 2048\}$ returned paths, let $a_{\ell,(m)}$ be the ordered projections and let $\widehat{Q}_\ell$ linearly interpolate the midpoint quantiles of 8192 independent target paths. The sliced 2-Wasserstein estimate (Liutkus et al., 2019) is

$$\widehat{\mathrm{SW}}_2 = \left[\frac{1}{256M}\sum_{\ell=1}^{256}\sum_{m=1}^{M}\left(a_{\ell,(m)} - \widehat{Q}_\ell\left(\frac{m-1/2}{M}\right)\right)^2\right]^{1/2}. \tag{201}$$

Independent target-to-target draws determine the sampling floor. Three independent training initializations on each of the 20 queries give 180 learned fits. We first average repeated initializations within a query, then summarize across queries. All training and representation budgets enter the comparison.

Our reference-only budgets are $B \in \{8, 16, 32\}$. For the Gaussian kernel, the projective distance satisfies

$$d_K(x, x_c) = \max_j (x - x_c)^\top x_j - \min_j (x - x_c)^\top x_j. \tag{202}$$

Evaluating these extrema on convex-hull vertices implements the same greedy partition without a dense $N \times N$ matrix. Hull construction and kernel preparation enter reference compilation. For $Q$

distinct queries and $M$ requested paths, complete service cost is

$$C(Q, M) = t_{\text{ref}} + \sum_{k=0}^{Q-1} [t_{k,\text{fit}} + t_{k,\text{sample}}(M)], \tag{203}$$

where $t_{\text{ref}}$ is reference compilation, $t_{k,\text{fit}}$ is endpoint solving or training, and $t_{k,\text{sample}}$ includes path generation and transfer to processor memory. We evaluate $Q \in \{1, 5, 10\}$ in the fixed query order. After fitting, every method may cache its prepared object, which gives the separate sampling comparison in Figure 10. Numerical timings use an Intel Xeon Gold 6530 and NVIDIA L20 graphics processors, with synchronized device measurements and one processor thread for matrix operations.

### D.8 Finite-state conditional paths

For the frozen Chignolin and WLALL references, define $P_{ij} = J_{ij}/m_i$ and use the four-step endpoint reference $J^{(4)}_{ij} = m_i(P^4)_{ij}$. This gives horizons of 20 ns and 2 ns, respectively. The original partitions remain fixed at $B = 16$ and $B = 8$, and every method uses the same four-step reference and observed endpoint queries. Given sampled endpoints $(i, j)$, the conditional transition is

$$\Pr(X_t = z \mid X_{t-1} = x, X_4 = j) = \frac{P_{xz}(P^{4-t})_{zj}}{(P^{5-t})_{xj}}, \qquad 1 \le t < 4. \tag{204}$$

The numerator combines one reference step with the probability of reaching the prescribed endpoint. Since all methods share these conditionals, path KL equals endpoint-coupling KL in both directions. Intermediate densities can also be evaluated exactly as

$$\rho^{\gamma}_t(z) = \sum_{i,j} \gamma_{ij} \frac{(P^t)_{iz}(P^{4-t})_{zj}}{(P^4)_{ij}}. \tag{205}$$

Thus the density comparison is independent of Monte Carlo path counts. We compare the maximum intermediate TV over $t = 1, 2, 3$ with matched-budget geometric lifting and independent pairing.

Native endpoint samplers avoid dense coupling materialization. Dense scaling draws $i \sim \mu$ and then $j$ proportional to $J^{(4)}_{ij} v_j$. The coarse sampler draws a coarse pair and then the two original endpoint conditionals. Both subsequently use (204). Each fresh request includes solving, sampler preparation, and $M \in \{1, 8, 32, 128, 512, 2048\}$ paths, with three matched repetitions. The common conditional sampler retains $P, P^2, P^3$, occupying 384 KiB for Chignolin and 5.79 MiB for WLALL. This shared path storage is separate from endpoint representation storage. The resulting comparison therefore measures complete finite-state bridge requests under the same interior dynamics.

## E Supplementary results

### E.1 Finite-state interaction and scaling

The controlled chain provides a direct link between mixing, a uniform residual, and measured coupling error. Table 3 reports the exhaustive residual at every tested scale. The finite-state brackets and checkerboard construction then separate the general worst-query guarantee from typical-query behavior. Finally, the exact-factorization benchmark isolates computational scaling with every representation solving the same endpoint problem under the shared fine-state reference.

Table 3: Interaction residuals for the coupled chain. Here $s_\varepsilon = \varepsilon + e^{-\lambda_\star T/\varepsilon}$, with $T = 1$ and $\lambda_\star = 1.052878$. The residual maximum uses every state pair. Each scale includes five endpoint pairs.

| $\varepsilon$ | $\eta_{\mathrm{fit}}$ | $\eta_{\mathrm{fit}}/s_\varepsilon$ | median $D_+$ | maximum $D_+$ |
|---|---|---|---|---|
| 0.0625 | 0.41649 | 6.6638 | 0.000711 | 0.001063 |
| 0.125 | 0.78555 | 6.2734 | 0.002749 | 0.004046 |
| 0.25 | 1.38886 | 5.2445 | 0.010405 | 0.014761 |
| 0.5 | 2.10360 | 3.3833 | 0.042600 | 0.052791 |
| 1 | 2.94102 | 2.1803 | 0.185129 | 0.200809 |

Table 4: Exhaustive partition comparison on six-state references. The bounds $L_*, U_*$ bracket the best endpoint-uniform risk at budget $B$. The greedy representative kernel has brackets $L_g, U_g$, with $U_g = \min\{2\eta_{\mathrm{anchor}}, \eta^2_{\mathrm{anchor}}/2\}$. The last column separately optimizes all additive and block terms on its fixed partition.

| Reference/$T$ | $B$ | Partitions | $L_*$ | $U_*$ | $L_g$ | $U_g$ | $\eta^{\mathrm{opt}}_{q_g}$ |
|---|---|---|---|---|---|---|---|
| Ring/0.6 | 2 | 31 | 2.4712 | 5.0009 | 2.4712 | 12.7537 | 3.1626 |
| Ring/0.6 | 3 | 90 | 1.0674 | 4.2362 | 2.2206 | 11.8259 | 2.9107 |
| Ring/1.4 | 2 | 31 | 1.6571 | 2.7403 | 1.6571 | 9.3329 | 2.3411 |
| Ring/1.4 | 3 | 90 | 0.6828 | 2.1825 | 1.4113 | 8.5066 | 2.0893 |
| Cycle/0.6 | 2 | 31 | 0.9770 | 1.3813 | 0.9770 | 6.7413 | 1.7782 |
| Cycle/0.6 | 3 | 90 | 0.3928 | 0.7610 | 0.9644 | 6.7413 | 1.6190 |
| Cycle/1.4 | 2 | 31 | 0.3436 | 0.4890 | 0.3459 | 2.0322 | 1.0015 |
| Cycle/1.4 | 3 | 90 | 0.1553 | 0.2498 | 0.3373 | 2.0322 | 0.8818 |
| Dense/0.6 | 2 | 31 | 1.0187 | 1.4070 | 1.0187 | 5.6281 | 1.6775 |
| Dense/0.6 | 3 | 90 | 0.7156 | 0.9771 | 0.7156 | 3.6064 | 1.4588 |
| Dense/1.4 | 2 | 31 | 0.5520 | 0.6607 | 0.5520 | 2.6426 | 1.1495 |
| Dense/1.4 | 3 | 90 | 0.3161 | 0.4073 | 0.3161 | 1.4030 | 0.9025 |

Table 5: Sharpness of the quadratic coefficient. The checkerboard residual is $\eta = \theta$. Direct scaling agrees with the analytic formulas, and both directional ratios approach $1/2$.

| $\eta$ | $F$ | $G$ | $F/\eta^2$ | $G/\eta^2$ |
|---|---|---|---|---|
| 0.5000 | 0.1201 | 0.1109 | 0.480458 | 0.443776 |
| 0.2000 | 0.0199 | 0.0196 | 0.496702 | 0.490175 |
| 0.1000 | 0.0050 | 0.0050 | 0.499169 | 0.497511 |
| 0.0500 | 0.0012 | 0.0012 | 0.499792 | 0.499376 |
| 0.0200 | 0.0002 | 0.0002 | 0.499967 | 0.499900 |
| 0.0100 | $5.00 \times 10^{-5}$ | $5.00 \times 10^{-5}$ | 0.499992 | 0.499975 |
| 0.0050 | $1.25 \times 10^{-5}$ | $1.25 \times 10^{-5}$ | 0.499998 | 0.499994 |
| 0.0020 | $2.00 \times 10^{-6}$ | $2.00 \times 10^{-6}$ | 0.500000 | 0.499999 |
| 0.0010 | $5.00 \times 10^{-7}$ | $5.00 \times 10^{-7}$ | 0.500000 | 0.500000 |

Table 6: Exact-factorization comparison at $N = 12{,}288$ with eight endpoint pairs and one processor thread. All representations solve the same operator to marginal tolerance $10^{-9}$. All operators, scaling vectors, and conditional arrays enter storage. The rank-eight factors are supplied exactly.

| Representation | median online time | deployed memory | relative to coarse |
|---|---|---|---|
| Coarse-state | 0.407 ms | 0.286 MiB | 1× / 1× |
| Positive factors, width 8 | 1.946 ms | 1.688 MiB | 4.78× time / 5.89× memory |
| Dense | 1.390 s | 1152.2 MiB | 3415× time / 4023× memory |

### E.2 Actual-query error and mechanism ablations

The molecular queries connect the uniform theory with a distribution-weighted description of error. Table 7 shows how residual moments track the interactions encountered by each query. The decomposition in Table 8 then attributes error to coarse-kernel mismatch and unresolved within-cell dependence. In particular, this distinction explains why modifying only a coarse kernel and refining the partition address different parts of the approximation through $C$ and $J$, respectively.

Table 7: Actual-query bounds on all frozen molecular queries. Entries are medians. Global is the uniform bound for either KL direction. $F$ and $G$ are the actual directional errors. Product denotes the product-centered forward moment, and the final columns minimize over the four fixed witnesses. These finite-matrix evaluations are separate from online scaling for each endpoint pair.

| Query family | $n$ | Global | $F$ | $G$ | Product | Fixed-pass $F$ | Fixed-pass $G$ |
|---|---|---|---|---|---|---|---|
| ALA/synthetic | 30 | 5.4171 | 0.0212 | 0.0209 | 0.0227 | 0.0212 | 0.0906 |
| Chig./synthetic | 30 | 10.2050 | 0.0831 | 0.0826 | 0.1082 | 0.0838 | 0.3637 |
| Chig./observed | 10 | 10.2050 | 0.0322 | 0.0336 | 0.0358 | 0.0322 | 0.1442 |
| WLALL/observed | 30 | 10.9313 | 0.2335 | 0.2201 | 0.2469 | 0.2335 | 0.9364 |

Table 8: The decomposition $F = C + J$ on the same queries. $C$ is coarse-kernel mismatch, $J$ is best fixed-conditional-lift error, and $\Delta$ is the largest identity discrepancy. Other columns give the median cellwise envelope and diagnostic costs. These costs are separate from accelerated query solving.

| Query family | $C$ | $J$ | $\Delta$ | $2U_\delta$ | Moment ms | Projection ms |
|---|---|---|---|---|---|---|
| ALA/synthetic | $8.21 \times 10^{-8}$ | 0.0212 | 0 | 6.5831 | 60.48 | 3.38 |
| Chig./synthetic | 0.0096 | 0.0711 | $4.93 \times 10^{-11}$ | 8.6279 | 15.58 | 4.94 |
| Chig./observed | $6.24 \times 10^{-5}$ | 0.0322 | $1.14 \times 10^{-12}$ | 8.6864 | 10.88 | 1.31 |
| WLALL/observed | 0.0014 | 0.2330 | $8.56 \times 10^{-12}$ | 9.9338 | 262.05 | 9.83 |

Table 9: Endpoint-conditional reconstruction preserves the requested fine marginals. Each pair of rows fixes the same partition, kernel, and coarse coupling. Coupling TV is the query median, and marginal error is the largest residual. Reference-conditionals provide the component ablation, while endpoint-conditionals implement the proposed lift using each prescribed endpoint law.

| Queries | Fine conditional used | coupling TV | marginal error |
|---|---|---|---|
| Alanine, 30 synthetic | Reference | 0.7534 | 1.6161 |
| | Endpoint | 0.0581 | $5.20 \times 10^{-11}$ |
| Chignolin, 30 synthetic | Reference | 0.6317 | 1.4282 |
| | Endpoint | 0.1233 | $8.30 \times 10^{-11}$ |
| Chignolin, 10 occupancy | Reference | 0.1449 | 0.2462 |
| | Endpoint | 0.0872 | $6.84 \times 10^{-11}$ |

Reference-only nested refinement further separates these two sources of error. Figure 11 includes every budget and both feature hierarchies for all five query families. On Chignolin occupancy queries, increasing interaction resolution from $B = 16$ to 64 changes median symmetric KL from $0.0329$ to $0.0134$. The nested identity describes the reduction of best-lift error $J$, while the deployed error also includes the current kernel mismatch $C$. Accordingly, the numerical supplement retains both components at every level, including the exact singleton limit on each molecular reference.

### E.3 Frozen molecular representations

At the reference-selected deployment sizes, Table 10 isolates partition and kernel choices using the same endpoint queries. Every partition appears with both aggregation and representative kernels. The full budget-dependent accuracy–storage and accuracy–time curves for observed endpoints remain in Figure 4. Together, these views separate representation quality from deployment cost.

Table 10: Median symmetric KL at the reference-selected coarse budgets: $B = 2$ for alanine and $B = 16$ for both Chignolin families, with $B = 8$ for WLALL. All eight partition/kernel combinations and every NMF rank are included. A and R denote aggregation and representative kernels. Complete budget curves and individual query-level values are supplied with the numerical data.

| Representation | ALA/synth. | Chig./synth. | Chig./obs. | WLALL/obs. |
|---|---|---|---|---|
| Dense | 0 | 0 | 0 | 0 |
| Product | 0.0345 | 0.167 | 0.1 | 0.355 |
| Spectral / A | 0.021 | 0.0834 | 0.0329 | 0.227 |
| Spectral / R | 0.021 | 0.0914 | 0.0502 | 0.23 |
| Reference distance / A | 0.0345 | 0.136 | 0.1 | 0.355 |
| Reference distance / R | 0.0345 | 0.171 | 0.1 | 0.355 |
| Geometric $k$-means / A | 0.0346 | 0.0952 | 0.0347 | 0.23 |
| Geometric $k$-means / R | 0.0346 | 0.0966 | 0.051 | 0.338 |
| Geometric farthest-first / A | 0.0347 | 0.097 | 0.0987 | 0.269 |
| Geometric farthest-first / R | 0.0486 | 0.112 | 0.0996 | 0.323 |
| NMF, rank 1 | 0.0345 | 0.167 | 0.1 | 0.355 |
| NMF, rank 2 | 0.0208 | 0.121 | 0.0976 | 0.283 |
| NMF, rank 4 | 0.0205 | 0.0864 | 0.0202 | 0.247 |
| NMF, rank 8 | 0.0196 | 0.071 | 0.0155 | 0.216 |
| NMF, rank 16 | 0.018 | 0.0609 | 0.0119 | 0.203 |
| NMF, rank 32 | 0.0156 | 0.0436 | 0.00676 | 0.183 |

The repeated-query comparison next fixes the final fine-reference tolerance across cold, warm, hierarchical, and learned initializers. Tables 11–14 retain both training budgets and all query families. Their exact-refinement rows measure acceleration of the full kernel solve, while coarse rows expose the additional accuracy–storage tradeoff of a reusable reduced representation. Figure 6 then includes representation construction in the total cost of repeated queries for the frozen WLALL reference.

Table 11: Repeated-query comparison on alanine synthetic queries. KL is median symmetric coupling KL. Times include initialization and refinement, and batch time is per query. Offline cost includes training and validation. Both budgets average three independent training initializations.

| Method | Updates | KL | Serial (ms) | Batch (ms) | Off. (s) | KiB |
|---|---|---|---|---|---|---|
| Coarse $B = 16$ | – | 0.0310 | 0.270 | 0.029 | 0.035 | 4.43 |
| Coarse $B = 32$ | – | 0.0258 | 0.281 | 0.033 | 0.055 | 10.43 |
| Coarse $B = 8$ | – | 0.0277 | 0.264 | 0.027 | 0.034 | 2.93 |
| Dense, cold | – | $< 10^{-12}$ | 0.672 | 0.368 | 0.000 | 755.63 |
| Dense, previous-query | – | $< 10^{-12}$ | 0.689 | 0.369 | 0.000 | 755.63 |
| Hierarchical, exact refinement | – | $< 10^{-12}$ | 1.220 | 0.313 | 0.029 | 771.73 |
| Meta-OT dual adapter | 500 | $< 10^{-12}$ | 0.929 | 0.404 | 3.470 | 2207.28 |
| Meta-OT dual adapter | 2000 | $< 10^{-12}$ | 0.911 | 0.284 | 17.672 | 2207.28 |
| UNOT-MLP adapter | 500 | $< 10^{-12}$ | 1.051 | 0.410 | 8.446 | 2480.31 |
| UNOT-MLP adapter | 2000 | $< 10^{-12}$ | 1.048 | 0.411 | 33.445 | 2480.31 |

Table 12: Repeated-query comparison on Chignolin observed queries. KL is median symmetric coupling KL. Times include initialization and refinement, and batch time is per query. Offline cost includes training and validation. Both budgets average three independent fits.

| Method | Updates | KL | Serial (ms) | Batch (ms) | Off. (s) | KiB |
|---|---|---|---|---|---|---|
| Coarse $B = 16$ | – | 0.0329 | 0.451 | 0.098 | 0.007 | 3.00 |
| Coarse $B = 32$ | – | 0.0301 | 0.448 | 0.086 | 0.008 | 9.00 |
| Coarse $B = 8$ | – | 0.0977 | 0.282 | 0.090 | 0.005 | 1.50 |
| Dense, cold | – | $< 10^{-12}$ | 0.513 | 0.201 | 0.000 | 128.00 |
| Dense, previous-query | – | $< 10^{-12}$ | 0.536 | 0.200 | 0.000 | 128.00 |
| Hierarchical, exact refinement | – | $< 10^{-12}$ | 1.118 | 0.356 | 0.009 | 139.81 |
| Meta-OT dual adapter | 500 | $< 10^{-12}$ | 0.721 | 0.248 | 2.559 | 1028.50 |
| Meta-OT dual adapter | 2000 | $< 10^{-12}$ | 0.684 | 0.249 | 12.020 | 1028.50 |
| UNOT-MLP adapter | 500 | $< 10^{-12}$ | 0.841 | 0.268 | 6.397 | 1301.53 |
| UNOT-MLP adapter | 2000 | $< 10^{-12}$ | 0.816 | 0.262 | 25.471 | 1301.53 |

Table 13: Repeated-query comparison on Chignolin synthetic queries. KL is median symmetric coupling KL. Times include initialization and refinement, and batch time is per query. Offline cost includes training and validation. Both budgets average three independent fits.

| Method | Updates | KL | Serial (ms) | Batch (ms) | Off. (s) | KiB |
|---|---|---|---|---|---|---|
| Coarse $B = 16$ | – | 0.0834 | 0.461 | 0.054 | 0.007 | 3.00 |
| Coarse $B = 32$ | – | 0.1083 | 0.468 | 0.059 | 0.008 | 9.00 |
| Coarse $B = 8$ | – | 0.0904 | 0.427 | 0.043 | 0.005 | 1.50 |
| Dense, cold | – | $< 10^{-12}$ | 0.580 | 0.156 | 0.000 | 128.00 |
| Dense, previous-query | – | $< 10^{-12}$ | 0.578 | 0.155 | 0.000 | 128.00 |
| Hierarchical, exact refinement | – | $< 10^{-12}$ | 1.245 | 0.247 | 0.009 | 139.81 |
| Meta-OT dual adapter | 500 | $< 10^{-12}$ | 0.748 | 0.184 | 2.559 | 1028.50 |
| Meta-OT dual adapter | 2000 | $< 10^{-12}$ | 0.744 | 0.182 | 12.020 | 1028.50 |
| UNOT-MLP adapter | 500 | $< 10^{-12}$ | 0.855 | 0.190 | 6.397 | 1301.53 |
| UNOT-MLP adapter | 2000 | $< 10^{-12}$ | 0.869 | 0.190 | 25.471 | 1301.53 |

Table 14: Repeated-query comparison on WLALL observed queries. KL is median symmetric coupling KL. Times include initialization and refinement, and batch time is per query. Offline cost includes training and validation. Both budgets average three independent fits.

| Method | Updates | KL | Serial (ms) | Batch (ms) | Off. (s) | KiB |
|---|---|---|---|---|---|---|
| Coarse $B = 16$ | – | 0.2217 | 0.417 | 0.044 | 0.124 | 5.93 |
| Coarse $B = 32$ | – | 0.2372 | 0.401 | 0.041 | 0.152 | 11.93 |
| Coarse $B = 8$ | – | 0.2268 | 0.428 | 0.040 | 0.115 | 4.43 |
| Dense, cold | – | $< 10^{-12}$ | 2.819 | 0.918 | 0.000 | 1976.63 |
| Dense, previous-query | – | $< 10^{-12}$ | 2.898 | 0.920 | 0.000 | 1976.63 |
| Hierarchical, exact refinement | – | $< 10^{-12}$ | 3.438 | 0.720 | 0.046 | 1997.23 |
| Meta-OT dual adapter | 500 | $< 10^{-12}$ | 3.148 | 0.972 | 4.359 | 4006.53 |
| Meta-OT dual adapter | 2000 | $< 10^{-12}$ | 3.093 | 0.679 | 20.011 | 4006.53 |
| UNOT-MLP adapter | 500 | $< 10^{-12}$ | 3.284 | 0.980 | 11.161 | 4279.56 |
| UNOT-MLP adapter | 2000 | $< 10^{-12}$ | 3.196 | 0.686 | 44.234 | 4279.56 |

### E.4 Positive Gaussian-kernel approximations

The repaired Nyström comparison provides a positive-kernel baseline on the common Gaussian references. The following tables retain every landmark count, positivity floor, and eigencutoff. Consequently, the accuracy–cost comparison includes the work required to construct a usable positive operator as well as its scaling iterations. The accompanying numerical records also retain the native-factor, low-rank-coupling, and QDSB coupling comparisons under their stated objectives.

Table 15: Positive-floor Nyström on $N = 512$ bank states, with relative eigencutoff $10^{-12}$. Entries are medians across all ten queries. Time includes construction, flooring, and scaling. Every row meets marginal tolerance $10^{-10}$ and retains a 512-KiB active kernel.

| | $\delta = 10^{-12}$ | | $\delta = 10^{-9}$ | | $\delta = 10^{-6}$ | |
|---|---|---|---|---|---|---|
| Landmarks | $D_{\mathrm{sym}}$ | Time (ms) | $D_{\mathrm{sym}}$ | Time (ms) | $D_{\mathrm{sym}}$ | Time (ms) |
| 8 | 0.2655 | 1.997 | 0.2212 | 1.969 | 0.1730 | 1.960 |
| 16 | 0.1184 | 2.101 | 0.1003 | 2.093 | 0.0821 | 2.087 |
| 32 | 0.0388 | 2.424 | 0.0328 | 2.403 | 0.0268 | 2.416 |
| 64 | 0.0113 | 3.266 | 0.0098 | 3.264 | 0.0083 | 3.245 |
| 128 | 0.0009 | 5.689 | 0.0007 | 5.674 | 0.0006 | 5.681 |
| 256 | $1.52 \times 10^{-5}$ | 13.807 | $9.90 \times 10^{-6}$ | 13.789 | $5.18 \times 10^{-6}$ | 13.784 |
| 512 | $< 10^{-12}$ | 43.065 | $< 10^{-12}$ | 43.065 | $4.06 \times 10^{-7}$ | 43.053 |

Table 16: Positive-floor Nyström on $N = 512$ bank states, with relative eigencutoff $10^{-10}$. Entries are medians across all ten queries. Time includes construction, flooring, and scaling. Every row meets marginal tolerance $10^{-10}$ and retains a 512-KiB active kernel.

| | $\delta = 10^{-12}$ | | $\delta = 10^{-9}$ | | $\delta = 10^{-6}$ | |
|---|---|---|---|---|---|---|
| Landmarks | $D_{\mathrm{sym}}$ | Time (ms) | $D_{\mathrm{sym}}$ | Time (ms) | $D_{\mathrm{sym}}$ | Time (ms) |
| 8 | 0.2655 | 1.967 | 0.2212 | 1.956 | 0.1730 | 1.961 |
| 16 | 0.1184 | 2.102 | 0.1003 | 2.090 | 0.0821 | 2.097 |
| 32 | 0.0388 | 2.410 | 0.0328 | 2.396 | 0.0268 | 2.403 |
| 64 | 0.0113 | 3.252 | 0.0098 | 3.248 | 0.0083 | 3.240 |
| 128 | 0.0008 | 5.681 | 0.0007 | 5.680 | 0.0006 | 5.667 |
| 256 | $2.50 \times 10^{-5}$ | 13.012 | $2.11 \times 10^{-5}$ | 12.942 | $1.74 \times 10^{-5}$ | 12.946 |
| 512 | $< 10^{-12}$ | 39.880 | $< 10^{-12}$ | 39.869 | $4.06 \times 10^{-7}$ | 39.854 |

Table 17: Positive-floor Nyström on $N = 2048$ bank states, with relative eigencutoff $10^{-12}$. Entries are medians across all ten queries. Time includes construction, flooring, and scaling. Every row meets marginal tolerance $10^{-10}$ and retains a 512-KiB active kernel.

| | $\delta = 10^{-12}$ | | $\delta = 10^{-9}$ | | $\delta = 10^{-6}$ | |
|---|---|---|---|---|---|---|
| Landmarks | $D_{\mathrm{sym}}$ | Time (ms) | $D_{\mathrm{sym}}$ | Time (ms) | $D_{\mathrm{sym}}$ | Time (ms) |
| 8 | 1.0125 | 2.018 | 0.7888 | 1.995 | 0.5649 | 1.992 |
| 16 | 0.6347 | 1.922 | 0.5030 | 1.901 | 0.3710 | 1.894 |
| 32 | 0.2330 | 2.223 | 0.1864 | 2.217 | 0.1395 | 2.211 |
| 64 | 0.0573 | 3.277 | 0.0482 | 3.275 | 0.0390 | 3.280 |
| 128 | 0.0066 | 5.883 | 0.0059 | 5.901 | 0.0052 | 5.898 |
| 256 | 0.0002 | 14.619 | 0.0001 | 14.630 | 0.0001 | 14.580 |
| 512 | $< 10^{-12}$ | 47.014 | $< 10^{-12}$ | 47.077 | $1.67 \times 10^{-5}$ | 47.059 |

Table 18: Positive-floor Nyström on $N = 2048$ bank states, with relative eigencutoff $10^{-10}$. Entries are medians across all ten queries. Time includes construction, flooring, and scaling. Every row meets marginal tolerance $10^{-10}$ and retains a 512-KiB active kernel.

| | $\delta = 10^{-12}$ | | $\delta = 10^{-9}$ | | $\delta = 10^{-6}$ | |
|---|---|---|---|---|---|---|
| Landmarks | $D_{\mathrm{sym}}$ | Time (ms) | $D_{\mathrm{sym}}$ | Time (ms) | $D_{\mathrm{sym}}$ | Time (ms) |
| 8 | 1.0125 | 1.990 | 0.7888 | 1.976 | 0.5649 | 1.971 |
| 16 | 0.6347 | 1.895 | 0.5030 | 1.900 | 0.3710 | 1.891 |
| 32 | 0.2330 | 2.202 | 0.1864 | 2.213 | 0.1395 | 2.213 |
| 64 | 0.0573 | 3.290 | 0.0482 | 3.285 | 0.0390 | 3.275 |
| 128 | 0.0069 | 5.849 | 0.0062 | 5.881 | 0.0055 | 5.855 |
| 256 | 0.0003 | 13.876 | 0.0003 | 13.892 | 0.0002 | 13.873 |
| 512 | $< 10^{-12}$ | 43.880 | $< 10^{-12}$ | 43.890 | $1.67 \times 10^{-5}$ | 43.855 |

### E.5 Complete finite-state paths

The four-step molecular experiment evaluates the original conditional dynamics after endpoint reduction. Table 19 reports path-law and intermediate-density accuracy for every observed query family. Table 20 then measures the full request, including conditional sampling. As the requested path count grows, this shared sampling work becomes a larger part of the total, making the endpoint and path cost scopes directly comparable across all requested conditional-path counts.

Table 19: Four-step path-law accuracy. Entries summarize symmetric path KL $D_{\mathrm{sym}}$ and maximum intermediate-time TV $V_{\mathrm{mid}}$ over each query family. Both geometric lifting and independent pairing use the same reference conditionals. Dense comparisons are zero by construction.

| Query family | Pairing method | $D_{\mathrm{sym}}$ | | $V_{\mathrm{mid}}$ | |
|---|---|---|---|---|---|
| | | Median | Maximum | Median | Maximum |
| Chignolin occupancy (10) | Spectral lift | $7.35\times10^{-5}$ | $1.41\times10^{-2}$ | $4.95\times10^{-5}$ | $1.12\times10^{-3}$ |
| | Geometric lift | $1.92\times10^{-4}$ | $1.73\times10^{-2}$ | $5.07\times10^{-5}$ | $2.89\times10^{-3}$ |
| | Independence | $4.06\times10^{-4}$ | $1.87\times10^{-1}$ | $2.77\times10^{-4}$ | $1.56\times10^{-2}$ |
| WLALL occupancy (20) | Spectral lift | $1.73\times10^{-4}$ | $2.82\times10^{-4}$ | $1.18\times10^{-4}$ | $1.96\times10^{-4}$ |
| | Geometric lift | $2.95\times10^{-4}$ | $1.16\times10^{-3}$ | $1.94\times10^{-4}$ | $3.06\times10^{-4}$ |
| | Independence | $5.82\times10^{-3}$ | $1.03\times10^{-2}$ | $2.38\times10^{-3}$ | $3.31\times10^{-3}$ |
| WLALL conformation (10) | Spectral lift | $8.06\times10^{-4}$ | $9.62\times10^{-4}$ | $8.57\times10^{-4}$ | $1.13\times10^{-3}$ |
| | Geometric lift | $7.55\times10^{-4}$ | $9.50\times10^{-4}$ | $9.03\times10^{-4}$ | $1.11\times10^{-3}$ |
| | Independence | $4.82\times10^{-3}$ | $1.16\times10^{-2}$ | $2.21\times10^{-3}$ | $5.99\times10^{-3}$ |

Table 20: Complete native request times in milliseconds, including solving, sampler preparation, endpoint drawing, and all four-step conditional paths. Each query uses its median across three repetitions, followed by the median across all queries within each fixed molecular system.

| Requested paths $M$ | Chignolin | | WLALL | |
|---|---|---|---|---|
| | Dense | Spectral lift | Dense | Spectral lift |
| 1 | 0.294 | 0.332 | 1.521 | 0.432 |
| 8 | 0.430 | 0.365 | 1.763 | 0.539 |
| 32 | 0.823 | 0.464 | 2.586 | 0.912 |
| 128 | 1.757 | 0.890 | 5.399 | 2.203 |
| 512 | 3.459 | 2.326 | 15.089 | 8.465 |
| 2048 | 9.066 | 7.797 | 38.415 | 29.067 |

### E.6 Learned bridges at the full path budget

Tables 21 and 22 give the complete 2048-path comparison underlying Figure 5. Every learned checkpoint and coarse budget is included. Accuracy is the mean of ten query-level averages, with

interquartile ranges across queries. Fit includes all endpoint solving or training, Sample measures an already prepared object, and Total serves ten distinct queries with reference compilation charged once. The 128-path results and individual repetitions remain in the numerical supplement and Figure 10.

Table 21: Complete path comparison for $N = 512$ and $M = 2048$. Target-to-target mean joint-path distance is $0.0334$. DSBM budgets count fitting cycles, LightSB-M and SF$^2$M-Sink budgets count updates, and our budget is $B$. Brackets give query interquartile ranges across ten endpoint pairs.

| Method | Budget | Joint path [IQR] | Terminal | Fit (s) | Sample (ms) | Total (s) |
|---|---|---|---|---|---|---|
| Exact Sinkhorn | – | 0.0337 [0.0313, 0.0360] | 0.0372 | 0.003077 | 1.432 | 0.045578 |
| DSBM | 1 | 0.0410 [0.0391, 0.0428] | 0.0683 | 47.284 | 36.605 | 472.329 |
| DSBM | 4 | 0.0397 [0.0374, 0.0416] | 0.0628 | 185.791 | 36.696 | 1854.2 |
| DSBM | 8 | 0.0391 [0.0365, 0.0419] | 0.0594 | 372.341 | 36.613 | 3722.5 |
| LightSB-M | 2000 | 0.0417 [0.0389, 0.0436] | 0.0778 | 11.228 | 2.033 | 112.985 |
| LightSB-M | 10000 | 0.0344 [0.0327, 0.0357] | 0.0485 | 56.840 | 2.032 | 576.176 |
| LightSB-M | 50000 | 0.0344 [0.0322, 0.0374] | 0.0441 | 290.570 | 2.033 | 2914.4 |
| SF$^2$M-Sink | 1000 | 0.0578 [0.0556, 0.0601] | 0.1023 | 7.482 | 23.755 | 75.133 |
| SF$^2$M-Sink | 5000 | 0.0494 [0.0480, 0.0511] | 0.0814 | 33.475 | 23.797 | 335.985 |
| SF$^2$M-Sink | 10000 | 0.0476 [0.0427, 0.0527] | 0.0756 | 65.917 | 23.749 | 658.891 |
| SF$^2$M-Sink | 20000 | 0.0519 [0.0475, 0.0543] | 0.0780 | 131.149 | 23.789 | 1313.4 |
| Ours | 8 | 0.0454 [0.0445, 0.0473] | 0.0367 | 0.001056 | 1.190 | 0.024381 |
| Ours | 16 | 0.0355 [0.0328, 0.0386] | 0.0380 | 0.000904 | 1.251 | 0.024074 |
| Ours | 32 | 0.0333 [0.0321, 0.0345] | 0.0371 | 0.000936 | 1.293 | 0.026751 |

Table 22: Complete path comparison for $N = 2048$ and $M = 2048$. Target-to-target mean joint-path distance is 0.0338. DSBM budgets count fitting cycles, LightSB-M and SF$^2$M-Sink budgets count updates, and our budget is $B$. Brackets give query interquartile ranges across ten endpoint pairs.

| Method | Budget | Joint path [IQR] | Terminal | Fit (s) | Sample (ms) | Total (s) |
|---|---|---|---|---|---|---|
| Exact Sinkhorn | – | 0.0337 [0.0314, 0.0343] | 0.0369 | 0.003117 | 1.434 | 0.045413 |
| DSBM | 1 | 0.0407 [0.0375, 0.0413] | 0.0720 | 47.624 | 36.710 | 477.003 |
| DSBM | 4 | 0.0397 [0.0360, 0.0429] | 0.0642 | 188.267 | 36.798 | 1886.4 |
| DSBM | 8 | 0.0382 [0.0343, 0.0417] | 0.0592 | 375.871 | 36.651 | 3764.5 |
| LightSB-M | 2000 | 0.0465 [0.0421, 0.0482] | 0.0868 | 12.220 | 2.043 | 121.902 |
| LightSB-M | 10000 | 0.0344 [0.0310, 0.0366] | 0.0482 | 60.372 | 2.039 | 596.030 |
| LightSB-M | 50000 | 0.0344 [0.0312, 0.0370] | 0.0439 | 300.861 | 2.040 | 3061.6 |
| SF$^2$M-Sink | 1000 | 0.0596 [0.0561, 0.0627] | 0.1065 | 7.426 | 23.770 | 74.238 |
| SF$^2$M-Sink | 5000 | 0.0524 [0.0481, 0.0565] | 0.0855 | 33.258 | 23.787 | 332.779 |
| SF$^2$M-Sink | 10000 | 0.0562 [0.0519, 0.0583] | 0.0889 | 65.624 | 23.736 | 657.683 |
| SF$^2$M-Sink | 20000 | 0.0526 [0.0435, 0.0579] | 0.0778 | 130.316 | 23.804 | 1304.8 |
| Ours | 8 | 0.0416 [0.0394, 0.0439] | 0.0367 | 0.000988 | 1.203 | 0.027220 |
| Ours | 16 | 0.0353 [0.0332, 0.0362] | 0.0370 | 0.000944 | 1.255 | 0.029641 |
| Ours | 32 | 0.0333 [0.0317, 0.0340] | 0.0371 | 0.000970 | 1.296 | 0.036694 |

## F SUPPLEMENTARY FIGURES

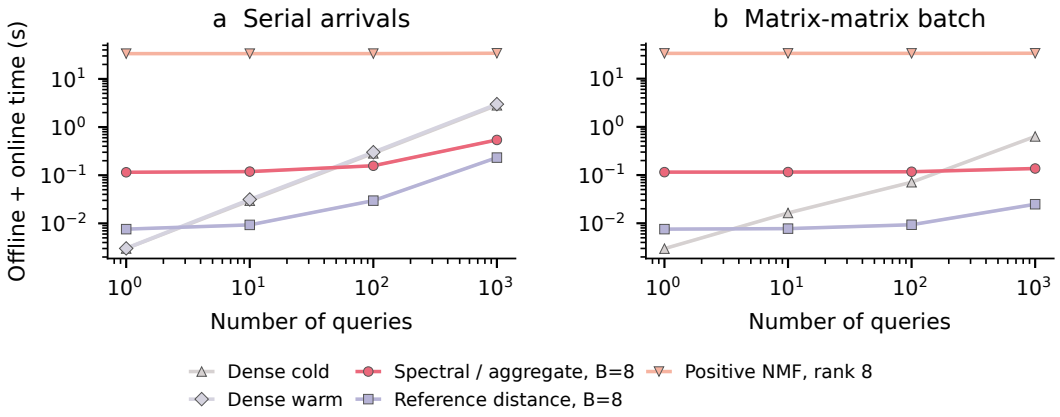


Figure 6: **Reference reuse includes preprocessing.** Total WLALL time includes representation fitting and serial or batched queries. Larger workloads replay the fixed query order. All NMF initializations are charged, and bands show timing-repeat ranges. Warm starts apply to serial arrivals. Figure 5 separately evaluates complete paths for distinct endpoint requests.

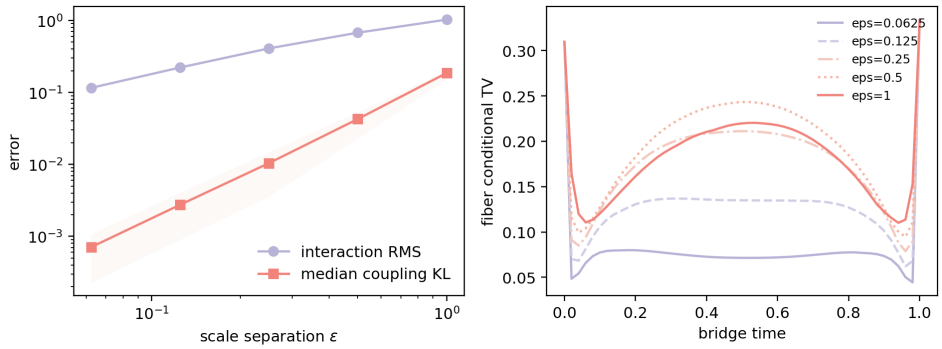


Figure 7: **Controlled interaction screening.** Left: the root-mean-square log-interaction residual and directional coupling KL decrease as fast relaxation strengthens. Right: fiber total variation forms two endpoint layers. Table 3 separately reports the maximum residual used in the uniform bound.

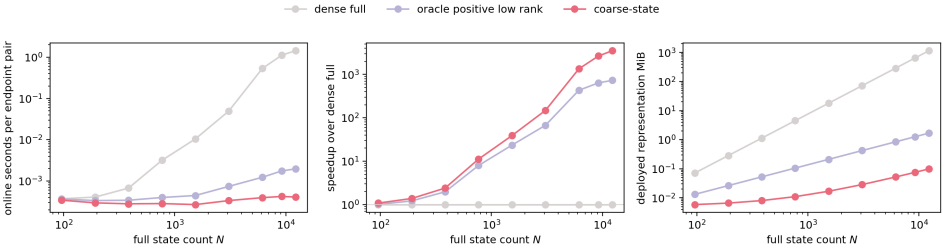


Figure 8: **Exact-factorization scaling on one processor thread.** Every method encodes the same optimizer, and the positive-factor comparator receives exact rank-eight factors. Timing includes endpoint preparation and scaling. Memory counts operators and scaling arrays, with conditional-array accounting in Table 6. Coarse products are independent of fiber size throughout the sweep.

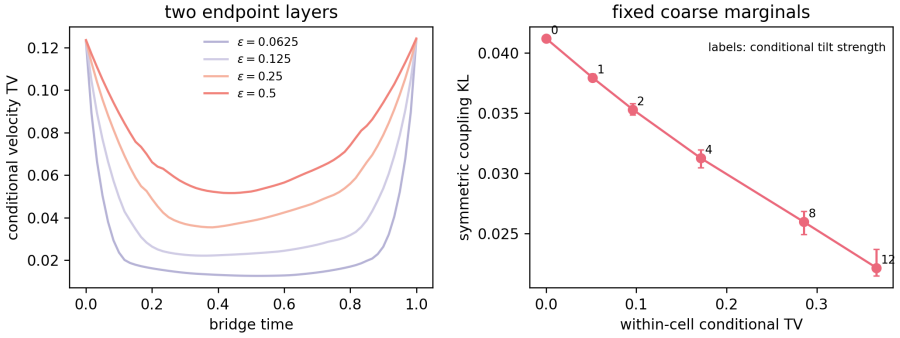


Figure 9: **Fast velocity detail is concentrated near the endpoints.** Left: nonsymmetric endpoints produce two conditional-velocity layers across four scale ratios. Right: holding both coarse marginals fixed at $\varepsilon = 0.125$ allows within-cell asymmetry to vary independently of the deployed coarse problem. Each strength uses 20 independent endpoint pairs with identical coarse marginals.

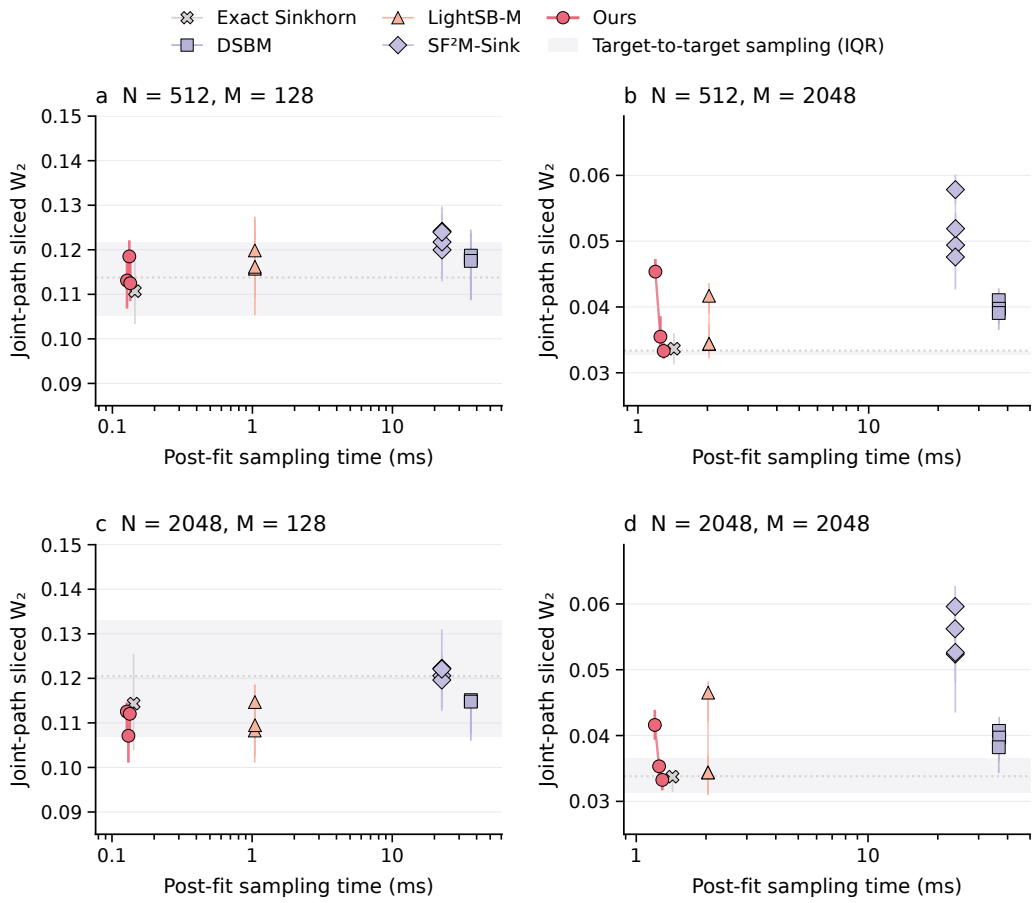


Figure 10: **Sampling an already fitted bridge.** Every budget, reference size $N$, and path count $M$ is included. Points average query-level means across independent training initializations, and marks show query interquartile ranges. Shading and dotted lines give target-to-target ranges and means. Cost includes five-time paths and device-to-host transfer. Figure 5 additionally charges fitting.

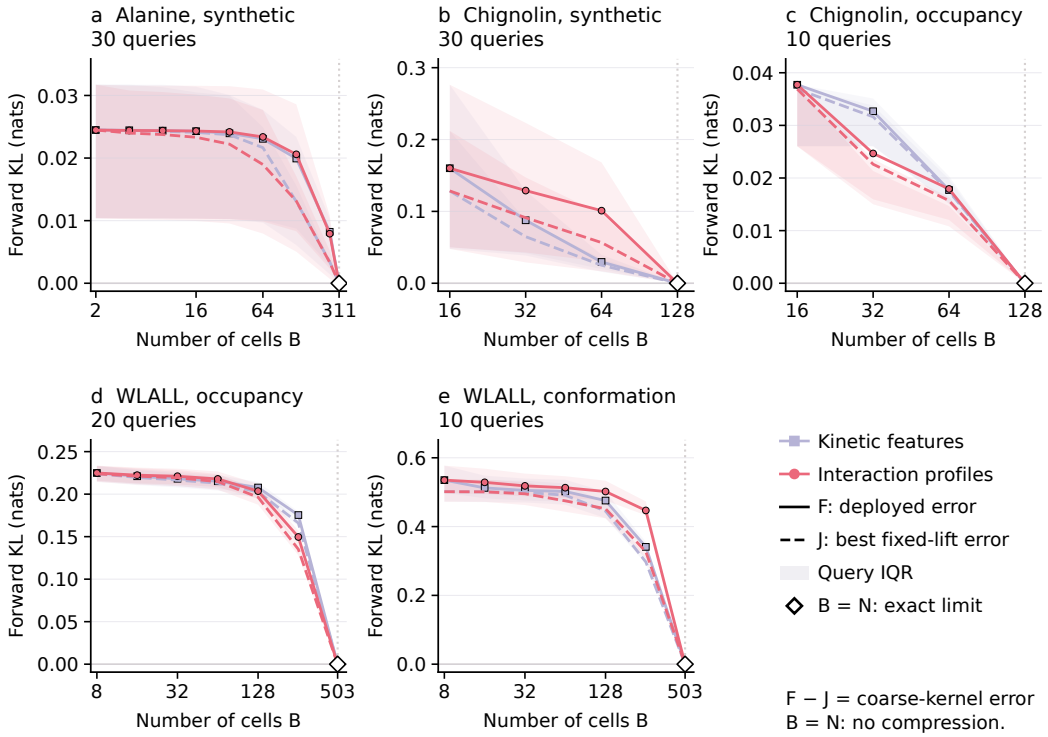


Figure 11: **Nested refinement separates representation error from coarse-kernel mismatch.** All 100 queries and both reference-only feature hierarchies are included. Solid curves show mean forward KL $F$, dashed curves show mean best-lift error $J$, and their difference is mean $C$. Bands show query interquartile ranges. Diamonds identify the exact uncompressed limit $B = N$.

## G REPRODUCIBILITY, ETHICS, AND AI USE

**Reproducibility statement.** The main text defines the reference process, endpoint constraints, coarse representation, error criteria, and offline/online computation used throughout the analysis. Appendices B–C state the assumptions and give complete proofs of the interaction bounds, converse, representation-selection guarantee, query-error decomposition, refinement identities, and mixing estimates. Appendix D specifies the data splits, reference estimation, endpoint construction, comparison methods, optimization, and evaluation protocols. Appendices E–F collect the supporting experimental evidence. The anonymous code and data supplement provides experimental and plotting programs, configurations, frozen reference and endpoint inputs, and numerical outputs for the reported comparisons. These materials support reproduction of the finite-reference calculations and reported summaries without downloading the original molecular trajectories. Public trajectory sources and third-party training implementations include source identifiers and acquisition instructions.

**Ethics statement.** This work studies reusable Schrödinger bridge computations, with the aim of reducing the cost of repeated distribution-constrained inference while clarifying approximation error. The experiments use controlled stochastic processes and publicly available molecular-dynamics data. They involve no human subjects or private personal information. More efficient bridge computation can support scientific modeling and stochastic control, but mathematical accuracy with respect to a reference process does not by itself establish the physical validity of that reference or the safety of a downstream application. Applications should therefore assess reference-model quality and domain-specific requirements, and respect the licenses and usage conditions of the underlying data and software. The supplementary code and data are anonymized for review.

**AI use statement.** In accordance with the ICLR 2027 policy on the use of large language models, we disclose the following. For the manuscript, we used LLM assistants together with automated compiler diagnostics to help debug LaTeX formatting, to suggest grammatical corrections at the word and sentence level, and to check cross-references and the internal consistency of terminology, notation, and reported numbers; these tools produced suggestions only, and every change was reviewed and adopted or rejected by the authors. We also used an AI assistant to help condense and tighten the presentation of the analysis and proofs in Appendix, which we had derived beforehand, and the condensed statements and derivations were subsequently checked independently by three of the authors. We have reviewed all AI-assisted work, and the authors take full responsibility for the final content of this paper, including its text, claims, proofs, figures, and experimental results.