ELYSIUM: SIMULATION-FREE UNBALANCED SCHRÖDINGER BRIDGES

Anonymous authors
Paper under double-blind review

ABSTRACT

Reconstructing cellular dynamics from snapshots is challenging because the destructive nature of single-cell measurements makes tracking the same cells over time impossible. Cellular dynamics are inherently stochastic and involve changes in population size through proliferation and apoptosis. Growth-based methods represent population change as a rate on the density, which does not assign exits and entries to individual cells. Unbalanced Schrödinger bridges model individual events, but neural solvers for killed diffusion bridges simulate trajectories during training. We introduce ELYSIUM, a simulation-free solver that jointly learns state transition and mass change. Using a killed Brownian reference and its time reversal, ELYSIUM constructs direct regression objectives from an optimal endpoint coupling and conditional bridge distributions, without simulating trajectories during training. We show that the minimizers of the expected training objectives recover the optimal unbalanced Schrödinger bridge. Empirically, ELYSIUM reconstructs state transitions and population changes on simulated data with known killed diffusion dynamics, outperforming all compared methods. On real single-cell datasets, it recovers cell state distributions at unobserved time points and captures population mass changes, achieving state-of-the-art results in most comparisons.

1 INTRODUCTION

Population dynamics are often observed through unpaired snapshots that sample individual states without linking the same individuals across times. In single-cell biology, measurements destroy the cells, so the same cells cannot be tracked (Schiebinger et al., 2019; Weinreb et al., 2018), and proliferation and apoptosis change the population size. Our goal is to learn, from snapshots alone, a continuous-time stochastic process that jointly captures state transitions and population changes.

Optimal transport couples consecutive snapshots, and flow matching turns such couplings into continuous dynamics by regressing a velocity field on sampled intermediate states (Benamou & Brenier, 2000; Peyre & Cuturi, 2019; Lipman et al., 2023; Tong et al., 2024a; 2020). Unbalanced optimal transport extends this framework to changing population sizes through a reaction term, as in the Wasserstein–Fisher–Rao (WFR) geometry (Chizat et al., 2018; Liero et al., 2018), and WFR-based solvers learn a velocity field together with a growth rate (Wang et al., 2025; Peng et al., 2026; Sha et al., 2024). These models describe the population through its density, and a reaction field records only the net gain or loss at each location. The same net change can arise from different combinations of individual entry and exit events. The trajectories are also deterministic, so individuals starting from the same state cannot diverge. Resolving both requires a stochastic process over individual histories in which exits and entries are events.

Schrödinger bridges (SBs) provide such a process. A reference stochastic process encodes a prior over individual behavior, and the bridge is the process closest to the reference in relative entropy among those matching the observed snapshots (Schrödinger, 1932; Léonard, 2014). Because the prior acts on complete histories, events in the reference become events of individuals in the bridge. Under a killed-diffusion reference, where each individual diffuses and exits at some rate, the bridge assigns trajectories, exit times, and exit locations, and reversing time turns exits into entries on growing intervals (Chen et al., 2022; 2025; Pariset et al., 2023).

The challenge is to learn this bridge without simulating trajectories during training. UDSB uses neural iterative proportional fitting with alternating forward and backward simulations (De Bortoli et al.,

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.

2021; Pariset et al., 2023). For balanced bridges, bridge matching removes simulation by regressing on conditional reference paths (Tong et al., 2024b). USB (Ying et al., 2026) brings simulation-free training to a branching SB, but it approximates the coupling with a WFR problem and regresses a signed growth rate, from which division or death is simulated after training. No existing solver obtains simulation-free targets for motion, exits, and entries directly from an unbalanced SB.

We introduce ELYSIUM, a neural solver with simulation-free training for the killed-diffusion unbalanced SB. Building on the cemetery-state formulation (Chen et al., 2022; 2025; Pariset et al., 2023), ELYSIUM combines an exact endpoint projection with analytic conditional path laws to construct direct training targets (Figure 1). Given the endpoint coupling, the optimal bridge admits a closed-form decomposition into survival, exit, and entry components. Survivors follow Brownian bridges, exiting individuals follow Brownian motion with a conditioned exit clock, and time reversal yields the corresponding entry component. These analytic component laws provide simulation-free regression targets for the drift and exit rate, while a generative flow learns the coupling-weighted endpoint distribution for entry locations. At the population optimum, the learned generator recovers the path law of the optimal bridge.

Our contributions are as follows.

- We introduce ELYSIUM, a simulation-free framework for unbalanced Schrödinger bridges under a killed Brownian reference. By combining optimal endpoint couplings with closed-form conditional bridge laws, it jointly learns stochastic state transitions and mass changes.
- We prove that the minimizers of the expected training objectives recover the optimal unbalanced Schrödinger bridge under a killed Brownian reference, capturing both stochastic state transitions and population entry or exit events.
- We demonstrate accurate recovery of state distributions and population size on simulated data with known killed diffusion dynamics, outperforming all compared methods. On real single-cell datasets, ELYSIUM recovers distributions at unobserved time points and captures population changes, achieving the best performance in most comparisons.

2 RELATED WORK

Optimal transport and flow matching. Dynamic OT represents distribution evolution by a velocity field (Benamou & Brenier, 2000; Peyre & Cuturi, 2019), and flow matching learns this field through conditional regression (Lipman et al., 2023; Tong et al., 2024a). TrajectoryNet and PRESCIENT apply neural dynamical models to single-cell snapshot data (Tong et al., 2020; Yeo et al., 2021). Unbalanced OT augments transport with reaction, as in the WFR geometry (Chizat et al., 2018; Liero et al., 2018); Action Matching, WLF, VGFM, TIGON, DeepRUOT, and WFR-FM learn continuous population dynamics under such objectives (Neklyudov et al., 2023; 2024; Wang et al., 2025; Sha et al., 2024; Zhang et al., 2025a; Peng et al., 2026), and SF-EUOT relaxes the terminal marginal of entropic OT to admit mass change (Choi & Choi, 2024). All of these represent mass change through a growth field or a relaxed constraint on the density. ELYSIUM models motion, exits, and entries as mechanisms of one stochastic path law, so every unit of created or removed mass has a state, a time, and a location.

Schrödinger bridges with changing mass. SBs project a reference path law onto prescribed marginals (Schrödinger, 1932; Léonard, 2014). Balanced neural solvers typically alternate forward and backward simulations (De Bortoli et al., 2021; Shi et al., 2023), and bridge matching removes simulation by regressing on conditional reference paths (Tong et al., 2024b). With changing mass, killed SBs were formulated and solved with Sinkhorn-type iterations by Chen et al. (2022; 2025), and UDSB combines exit with reverse-time entry in a neural iterative proportional fitting scheme that again alternates simulations (Pariset et al., 2023). BranchSBM and USB model branching and reproduction, and CytoBridge models cellular interactions and proliferation (Baradat & Lavenant, 2025; Tang et al., 2026; Zhang et al., 2025b; Ying et al., 2026); among them USB avoids simulation by approximating its branching coupling with a WFR objective. ELYSIUM derives both the endpoint projection and the regression targets from one killed Brownian reference (Appendix D).

Single-cell trajectory inference. Destructive assays require trajectories to be reconstructed without cell correspondence, a problem with fundamental identifiability limits (Schiebinger et al., 2019;

Weinreb et al., 2018; Lavenant et al., 2024). RNA velocity and lineage tracing add kinetic or ancestral information (La Manno et al., 2018; Bergen et al., 2020; Weinreb et al., 2020), and population balance models constrain growth with additional measurements (Fischer et al., 2019), but many time-course datasets provide only snapshots and abundance changes (Sha et al., 2024; Bunne et al., 2023). ELYSIUM targets this setting and learns continuous state motion together with stochastic exits and entries.

3 FROM UNBALANCED SNAPSHOTS TO A SCHRÖDINGER BRIDGE

Given two snapshots of different total mass, we seek a stochastic process in which individuals move, exit, or enter. This section formulates the unbalanced SB for a killed Brownian reference, reduces it to an endpoint coupling through a cemetery state, and shows that the conditioned paths have closed-form laws (Figure 1). These laws are the training targets of Section 4.

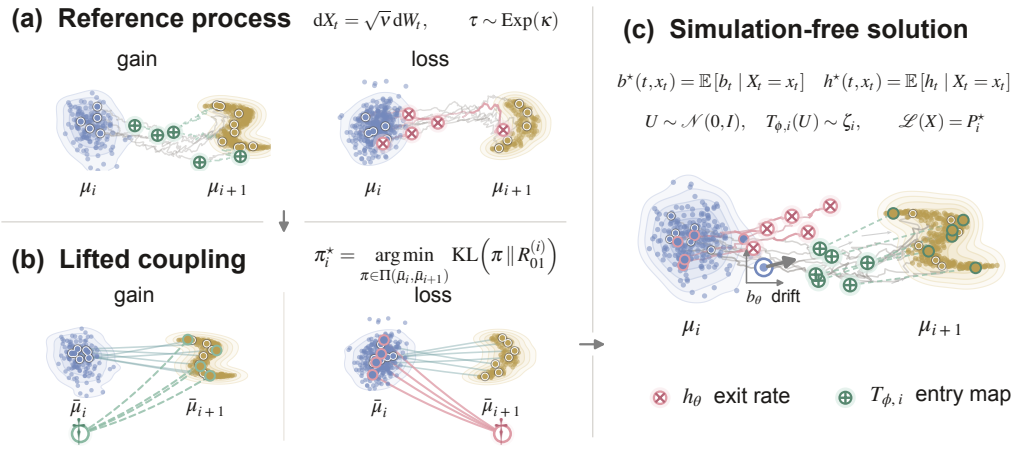


Figure 1: Overview of ELYSIUM. (a) The killed Brownian reference supplies stochastic motion and an exponential exit clock; reversing the construction represents entry on a growing interval. Shaded contours show snapshot densities and particle counts encode mass. (b) Adding the cemetery balances the endpoint measures, and KL projection of the reference endpoint law yields a coupling over survival and exit components; lines denote endpoint pairs. (c) Closed-form component samples provide simulation-free targets for the population drift and exit rate, and the entry map transports Gaussian noise to the coupling-weighted endpoint distribution ζ_i of interval i (Appendix A.7). The local arrow shows the learned SDE drift at one state.

3.1 THE KILLED REFERENCE AND THE UNBALANCED SCHRÖDINGER BRIDGE

Rescale one observation interval to $[0,1]$ and let μ_0, μ_1 be the observed finite measures with masses $m_0 \geq m_1$, so that the observed loss is $g = m_0 - m_1$. The reference process is Brownian motion with an independent exponential exit clock,

$$\mathrm{d}X_t = \sqrt{\nu}\,\mathrm{d}W_t \quad (t < \tau), \qquad \tau \sim \mathrm{Exp}(\kappa), \qquad X_t = \dagger \quad (t \geq \tau), \tag{1}$$

with $\nu, \kappa > 0$. A particle that is active at time t exits during the next short interval $\mathrm{d}t$ with probability $\kappa\,\mathrm{d}t$, and after exiting it sits in an absorbing *cemetery state* $\dagger$ and no longer belongs to the active population. For a growing interval with $m_1 > m_0$, we swap the snapshots, solve the mass-decreasing problem in backward time, and reverse the resulting bridge, so it suffices to treat $m_0 \geq m_1$. For a path measure P, let P_t^{act} denote the finite measure of particles still active at time t.

Definition 1 (Unbalanced Schrödinger bridge)**.** Given μ_0, μ_1 with $m_0 \geq m_1$ and the killed reference path measure R, the unbalanced Schrödinger bridge minimizes relative entropy among path measures matching both active endpoints:

$$P^* = \underset{P:\ P_0^{\mathrm{act}} = \mu_0,\ P_1^{\mathrm{act}} = \mu_1}{\arg\min} D_{\mathrm{KL}}(P \| R). \tag{2}$$

Both P and R have mass m_0, and D_{KL} is relative entropy of finite measures (Appendix A.3).

The problem has an equivalent dynamic form. Change the drift to $\sqrt{\nu}\, u_t$ and the exit rate to $h_t \geq 0$, and write $P^{u,h}$ for the resulting path measure. Girsanov's theorem gives the following statement (Chen et al., 2022).

Proposition (Equivalent dynamic formulation)**.** *Under the conditions in Appendix A.2, Equation 2 has the same value and the same optimal process as*

$$\inf_{u,h} \mathbb{E}_{P^{u,h}} \int_0^{1\wedge\tau} \left[\frac{1}{2}\|u_t\|^2 + h_t \log\frac{h_t}{\kappa} - h_t + \kappa\right] \mathrm{d}t \quad \textit{subject to} \quad P_0^{\mathrm{act}} = \mu_0,\ P_1^{\mathrm{act}} = \mu_1. \tag{3}$$

The quadratic term penalizes changes to the reference drift, and the second term is the relative-entropy rate of the exit mechanism, which vanishes when $h_t = \kappa$. Both costs accumulate only while a particle is active, so the dynamic problem controls motion and exit timing while requiring the final active distribution to match the target snapshot. Solving this control problem directly would require simulating the controlled process; Section 3.2 removes that requirement.

3.2 From path optimization to endpoint coupling

In the classical Schrödinger bridge, conditioning on the initial and final states separates path-space relative entropy into an endpoint term and a conditional-path term (Léonard, 2014). Let $\pi = P_{01}$ and R_{01} be the endpoint couplings of P and R, and let P^{xy}, R^{xy} be their path laws conditioned on the endpoint pair (x, y). The KL chain rule gives

$$D_{\mathrm{KL}}(P\|R) = D_{\mathrm{KL}}(\pi\|R_{01}) + \int D_{\mathrm{KL}}(P^{xy}\|R^{xy})\, \pi(\mathrm{d}x, \mathrm{d}y). \tag{4}$$

For a fixed coupling π, choosing $P^{xy} = R^{xy}$ makes the second term zero, so only the endpoint coupling needs to be optimized and the conditional paths remain those of the reference.

The chain rule requires every path to have an endpoint at time one, but an exiting particle has none in the state space. The cemetery state provides one. Recording all post-exit states as †, with unit point mass $\delta_\dagger$, the complete endpoint measures become

$$\mu_0, \qquad \mu_1 + g\delta_\dagger, \tag{5}$$

and both have mass m_0. This representation changes only how histories are recorded and preserves the constraints and the relative-entropy cost, so matching the active endpoints (μ_0, μ_1) and matching the complete endpoints $(\mu_0, \mu_1 + g\delta_\dagger)$ have the same optimal motion and exit process. Below, P, R denote the extended path measures.

Proposition 2 (Endpoint reduction)**.** *Let $\Pi(\mu_0, \mu_1 + g\delta_\dagger)$ be the set of couplings with the marginals in Equation 5, and suppose it contains a coupling with finite KL relative to R_{01}. Under the conditions in Appendix A.2, the optimal path measure of Definition 1 is determined by*

$$\pi^* = \operatorname*{arg\,min}_{\pi\in\Pi(\mu_0,\, \mu_1+g\delta_\dagger)} D_{\mathrm{KL}}(\pi\|R_{01}) \tag{6}$$

and recovered by mixing conditional reference paths,

$$P^* = \int R^{xy}\, \pi^*(\mathrm{d}x, \mathrm{d}y). \tag{7}$$

The total exit mass is fixed at g; the coupling determines which source mass exits. For the reference in Equation 1, independence of the Brownian motion and the exit clock gives the endpoint law

$$R_{01}(\mathrm{d}x, \mathrm{d}y) = \mu_0(\mathrm{d}x)\left[e^{-\kappa}\varphi_\nu(y-x)\,\mathrm{d}y + (1-e^{-\kappa})\delta_\dagger(\mathrm{d}y)\right], \tag{8}$$

where φ_ν is the density of $\mathcal{N}(0, \nu I)$. A particle survives to the endpoint with probability $e^{-\kappa}$ and undergoes a Gaussian displacement, or exits with probability $1 - e^{-\kappa}$ and ends at †. Up to constants, which do not affect the coupling because the marginals fix the mass of each branch, the survival branch has quadratic cost and the exit branch has position-independent cost. Equation 6 is therefore an entropic optimal transport problem on the extended endpoints and is solved with Sinkhorn iterations (Cuturi, 2013; Benamou et al., 2015) (Appendices A.3–A.4).

3.3 CLOSED-FORM COMPONENTS OF THE OPTIMAL BRIDGE

Once π^* is known, every conditional reference path in Equation 7 is explicit. We record its state law together with the drift and exit rate that generate it, since these are the quantities the network will regress.

Theorem 3 (Closed-form bridge decomposition)**.** *Let π^* solve Equation 6 and let $t \in (0,1)$. All components keep the reference diffusion coefficient $\sqrt{\nu}I$. For a surviving pair $x \to y$, R^{xy} is the Brownian bridge from x to y: the state law is $\mathcal{N}\big((1-t)x + ty, \nu t(1-t)I\big)$, the drift at active state z is $b_t^{xy}(z) = (y-z)/(1-t)$, and the exit rate is $h_t^{xy}(z) = 0$. For an exit pair $x \to \dagger$, $R^{x\dagger}$ is Brownian motion from x whose independent exit clock is conditioned on $\tau \le 1$; the particle stays at the cemetery after exit. While active, the state law is $\mathcal{N}(x, \nu tI)$, the drift is $b_t^{x\dagger}(z) = 0$, and the exit rate is*

$$h_t^{x\dagger}(z) = \lim_{\varepsilon \downarrow 0} \frac{\Pr(t < \tau \le t+\varepsilon \mid t < \tau \le 1)}{\varepsilon} = \frac{\kappa}{1 - e^{-\kappa(1-t)}}.$$

Mixing these conditional paths with the masses of π^* yields the optimal bridge P^*. Reversing the mass-decreasing reference and its optimal bridge gives the bridge of a growing interval, in which an entry component $\dagger \to y$ has state law $\mathcal{N}(y, \nu(1-t)I)$ once it has entered, drift $b_t^{\dagger y}(z) = (y-z)/(1-t)$, and exit rate zero.

The survival drift is the Doob transform of the Brownian transition density conditioned on the endpoint; conditioning the exit clock on $\tau \le 1$ leaves the pre-exit motion unchanged and only rescales the rate. Proofs are in Appendices A.5–A.6.

4 SIMULATION-FREE LEARNING OF THE BRIDGE GENERATOR

Theorem 3 gives, for every component of the optimal bridge, a state law and the drift and exit rate that generate it. This section turns these quantities into regression targets for a population drift and exit rate, constructs the entry source for growing intervals, and shows that the resulting generator recovers the optimal bridge.

4.1 LEARNING STATE MOTION AND EXIT EVENTS

The component coefficients depend on the endpoint pair, which is unknown when generating from an initial population. We therefore learn a drift and exit rate that depend only on time and the current state, as neural fields $b_\theta(t,z)$ and $h_\theta(t,z) \ge 0$, using the component coefficients $b_t^c(z), h_t^c(z)$ of Theorem 3 as targets, where c indexes the component of an endpoint pair.

The contribution of a component at time t is the mass that is still active. Survival components remain active throughout the interval. The active fraction of an exit component is

$$s_\kappa(t) = \Pr(\tau > t \mid \tau \le 1) = \frac{e^{-\kappa t} - e^{-\kappa}}{1 - e^{-\kappa}}, \tag{9}$$

so for a discrete coupling the survival and exit components have active masses π^*_{xy} and $\pi^*_{x\dagger} s_\kappa(t)$. Growing intervals use the transposed coupling, and an entry component has entered with fraction $s_\kappa(1-t)$. Normalizing the active masses defines a component label C_t, the conditional state law of Theorem 3 defines a state Z_t, and the targets are $b_t = b_t^{C_t}(Z_t)$ and $h_t = h_t^{C_t}(Z_t)$. A training sample therefore consists of a time, a component drawn with probability proportional to its active mass, a state drawn from that component's Gaussian law, and the drift and exit rate of the component at that state. All of these are available in closed form, so no trajectory is simulated during training (Appendix A.6).

We use squared error for the drift and Poisson relative entropy $\ell_{\mathrm{P}}(a,b) = a\log(a/b) - a + b$ for the exit rate,

$$\mathcal{L}_{\mathrm{fields}}(\theta) = \mathbb{E}\big[\|b_\theta(t,Z_t) - b_t\|^2 + \ell_{\mathrm{P}}(h_t, h_\theta(t,Z_t))\big], \tag{10}$$

where the expectation is over training times and the active population defined above. The pointwise minimizers of both terms are the conditional means of b_t and h_t given time and current state. Near a state where survival and exit components overlap, the regression therefore averages the survival drift

toward the endpoints with the zero drift of exiting particles, and the exit rate becomes the conditional rate of the mixture at that state. Section 4.3 identifies these conditional means with the dynamics of the optimal bridge.

4.2 Constructing and learning the entry source

On a growing interval, Theorem 3 gives the drift of a particle after it has entered. It remains to specify when and where particles enter. Let $\pi^*_{\dagger y}$ be the mass of the entry component $\dagger \to y$ in the transposed coupling, with total $g = \sum_y \pi^*_{\dagger y} > 0$.

By Equation 9, the fraction of an entry component that has entered by time t is $s_\kappa(1-t)$, so the mass entering during a short interval Δt is approximately $g\,\frac{\mathrm{d}}{\mathrm{d}t}s_\kappa(1-t)\,\Delta t$. Because the exit clock is independent of the Brownian motion, a particle that enters at time t with endpoint y has state law $\mathcal{N}(y, \nu(1-t)I)$. Weighting by the coupling gives the entry source

$$\eta_t(z) = \frac{\mathrm{d}}{\mathrm{d}t}s_\kappa(1-t)\sum_y \pi^*_{\dagger y}\,\varphi_{\nu(1-t)}(z-y), \tag{11}$$

where $\eta_t(z)\,\mathrm{d}z\,\mathrm{d}t$ is the mass entering a volume element around z during $\mathrm{d}t$. The entry clock is analytic and the entry locations follow a Gaussian mixture whose centers are the endpoints y weighted by $\pi^*_{\dagger y}/g$.

Generation must sample entry states without storing the observed snapshots, so we learn this endpoint distribution with a lightweight generative flow trained by conditional flow matching. Draw an endpoint Y with weights $\pi^*_{\dagger y}/g$ and, independently, $U \sim \mathcal{N}(0, I)$ and a flow time $s \sim \mathrm{Unif}[0,1]$; along the interpolation $(1-s)U + sY$, a vector field u_ϕ regresses the velocity $Y - U$,

$$\mathcal{L}_{\mathrm{entry}}(\phi) = \mathbb{E}_{s,U,Y}\left\|u_\phi(s, (1-s)U + sY) - (Y-U)\right\|^2.$$

Integrating u_ϕ from $s=0$ to $s=1$ defines a map T_ϕ from noise to endpoints. When an entry event occurs at time t, the new state is

$$Z_t = T_\phi(U) + \sqrt{\nu(1-t)}\,\epsilon, \qquad U, \epsilon \overset{\mathrm{iid}}{\sim} \mathcal{N}(0, I), \tag{12}$$

so $T_\phi(U)$ supplies the mixture center, or *entry anchor*, and the Gaussian perturbation places the particle at its entry position.

Proposition 4 (Recovery of the entry source)**.** *If T_ϕ recovers the coupling-weighted endpoint distribution, then generating entry events at rate $g\,\frac{\mathrm{d}}{\mathrm{d}t}s_\kappa(1-t)$ with states drawn from Equation 12 recovers the source in Equation 11 exactly (Appendix A.7).*

4.3 The learned generator solves the unbalanced SB

The training samples above are drawn from the optimal bridge P^*. Conditioning their drift and exit coefficients on the current state recovers the Markov dynamics of P^*, which connects the regression objectives back to Definition 1.

Theorem 5 (Exact recovery of the target unbalanced SB)**.** *Assume the endpoint projection is exact, the regression optima are attained at almost every interior time, and each entry map recovers the endpoint distribution specified by its coupling. Under the regularity and uniqueness-in-law conditions of Appendix A.8, the optima are*

$$b^*(t,z) = \mathbb{E}[b_t \mid Z_t = z], \qquad h^*(t,z) = \mathbb{E}[h_t \mid Z_t = z]. \tag{13}$$

On a decreasing interval, the killed diffusion

$$\mathrm{d}X_t = b^*(t, X_t)\mathrm{d}t + \sqrt{\nu}\,\mathrm{d}W_t \tag{14}$$

with exit rate h^* has path law P^*. On a growing interval, the same construction with the exact entry source has the time-reversed optimal law. The learned generator therefore solves the target unbalanced SB and recovers all of its active marginals.

Appendix A.8 gives the proof, Appendix A.9 gives conditions under which empirical bridges converge to the population bridge as the endpoint measures converge, and Appendix B describes finite-sample training. At the population level the active measure ρ_t evolves as

$$\partial_t \rho_t = -\nabla \cdot (\rho_t b^*) + \frac{\nu}{2} \Delta \rho_t - h^* \rho_t + \eta_t, \tag{15}$$

where the drift and Brownian terms move active states, the exit rate removes mass, and η_t adds it; $\eta_t = 0$ on a decreasing interval and equals Equation 11 on a growing one. Compared with a WFR model, the single signed growth term is split into an exit rate and an entry source, each realized by events of individuals.

5 The Elysium algorithm

Elysium applies the construction above to every adjacent pair of snapshots in a time course and shares one neural generator across the sequence (Algorithms 1 and 2; expanded in Appendix B).

Generating stochastic population trajectories. For an interval of physical duration $\Delta_i = t_{i+1} - t_i$, let $r = (t - t_i)/\Delta_i$ be local normalized time. An active state follows

$$\mathrm{d}X_t = \frac{b_\theta(r, X_t)}{\Delta_i} \mathrm{d}t + \sqrt{\frac{\nu_i}{\Delta_i}}\, \mathrm{d}W_t, \tag{16}$$

and on a decreasing interval it exits at rate $h_\theta(r, X_t)/\Delta_i$ in physical time. On a growing interval the reversed clock samples entry times and Equation 12 samples the new states. Trajectories are simulated only at generation; every training update uses an independent analytic sample, so optimization needs no forward or backward integration. For evaluation, exits are carried as survival weights, which define the same law as sampling exit times (Appendix B.3). Appendix C.6 compares computational cost with WFR-FM and UDSB-F.

In the algorithms, $\widehat{\mathcal{L}}$ denotes a minibatch estimate of the corresponding loss, the population is represented by weighted particles $\mathcal{P} = \{(x_j, w_j)\}$, r and δ are local time and step size, $\mathrm{SDEstep}_i$ evolves particles by Equation 16, and $\mathcal{P}_{\mathrm{ent}}$ contains the particles entering during the current step, sampled from Equations 11–12 (Appendix B.3).

Algorithm 1 Simulation-free training

Input: snapshots $\{(t_i, \mu_i)\}_{i=0}^K$, parameters $\{\nu_i, \kappa_i\}$

Solve and fix endpoint couplings $\{\pi_i^*\}$
while field-training budget remains **do**
  Draw i, r and (Z_r, b_r, h_r)
  $\theta \leftarrow \mathrm{AdamW}(\theta, \nabla_\theta \widehat{\mathcal{L}}_{\mathrm{fields}})$
for each growing interval i **do**
  while entry-training budget remains **do**
    Draw a minibatch as in Section 4.2
    $\phi \leftarrow \mathrm{AdamW}(\phi, \nabla_\phi \widehat{\mathcal{L}}_{\mathrm{entry}})$

Output: Trained networks $b_\theta, h_\theta, \{u_{\phi,i}\}$.

Algorithm 2 Inference

Input: initial particles $\mathcal{P}_0$, trained networks, interval parameters, steps L

$\mathcal{P} \leftarrow \mathcal{P}_0, \quad \delta \leftarrow 1/L$
for $i = 0, \ldots, K-1$ **do**
  for $k = 0, \ldots, L-1$ **do**
    $r \leftarrow k\delta$
    $\mathcal{P} \leftarrow \mathrm{SDEstep}_i(\mathcal{P}; r, \delta)$
    if loss: $w_j \leftarrow w_j e^{-\int_r^{r+\delta} h_\theta(s, X_s^j)\, \mathrm{d}s}$
    if growth: $\mathcal{P} \leftarrow \mathcal{P} \cup \mathcal{P}_{\mathrm{ent}}$
    Record $\widehat{\mu} \leftarrow \sum_j w_j \delta_{x_j}$

Output: Stochastic paths and populations $\{\widehat{\mu}_t\}$.

6 Experiments

We ask three questions: whether the solver recovers a specified bridge against independent ground truth (Q1), whether the generator matches and interpolates single-cell population time courses (Q2), and whether entry and exit capture expansion and contraction under drug treatment (Q3). We compare with four unbalanced baselines that cover the main modeling choices: WFR-FM learns a growth field under a WFR objective (Peng et al., 2026), USB approximates a branching bridge with a WFR coupling (Ying et al., 2026), SF-EUOT relaxes the terminal marginal of entropic OT (Choi & Choi, 2024), and UDSB-F is a simulation-based solver of the killed-diffusion bridge (Pariset

et al., 2023). All W_1 scores use the feature space supplied with each benchmark, lower is better, and ELYSIUM reports mean±SD over five trained models. Appendix C.1 defines the metrics and the rest of Appendix C gives complete protocols and results.

6.1 Q1: CAN THE SOLVER RECOVER DYNAMICS WITH KNOWN GROUND TRUTH?

We construct unbalanced Schrödinger bridge problems under killed diffusion references in 1D and 2D, varying endpoint distributions, diffusion and event rates, and population ratios m_1/m_0 from 0.1 to 10. Each method is trained only on endpoint populations and tested at intermediate times against independently computed ground truth. Each method uses one configuration selected on a separate development set and fixed across all cases (Appendix C.2).

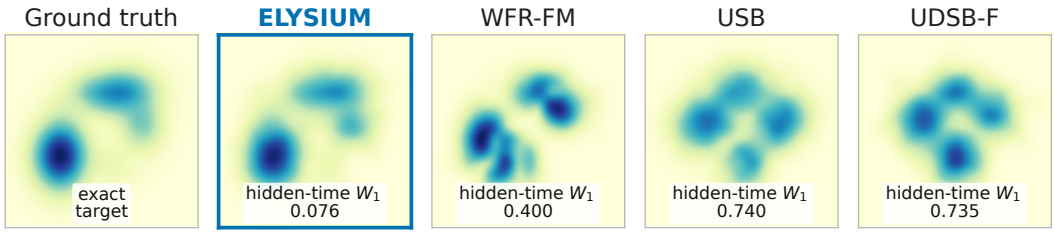


Figure 2: Intermediate distribution recovery with $m_1/m_0 = 0.1$ (population reduced to 10% of its initial size). Panels show normalized living-population distributions at the midpoint; labels give state W_1 averaged over the interior test times. Appendix C.2 includes the gain case.

Table 1: Mean error across test problems. State W_1 compares normalized living populations; the next two columns measure population size; joint W_1 scores both. The baseline is the better of USB and WFR-FM.

Method	State W_1	Population RME	$\int \lvert\widehat{M} - M\rvert$	Joint W_1
ELYSIUM	**0.0624**	**0.00260**	**0.000869**	**0.00450**
Best baseline	0.2622	0.09257	0.139410	0.04919
Mean reduction	75.0%	93.8%	94.2%	87.4%

ELYSIUM outperforms USB and WFR-FM on all four state-and-mass metrics in every problem (Table 1). Relative to the better baseline in each problem, the mean reductions in state W_1 and integrated mass error are 75.0% and 94.2%, respectively. Figure 2 illustrates recovery at an unobserved midpoint with $m_1/m_0 = 0.1$, where the baselines distort the modes that ELYSIUM matches. On 16 cases with nonzero entry or exit, each evaluated across five training seeds, ELYSIUM also reduces event-time W_1 from 0.1210 to 0.0009 and event-location W_1 from 1.1784 to 0.0603 compared with UDSB-F (Appendix C.2).

6.2 Q2: CAN THE GENERATOR MODEL SINGLE-CELL POPULATION TIME COURSES?

Following the evaluation setting of WFR-FM (Peng et al., 2026), we use two simulated gene-dynamics datasets (Simulation Gene and Dyngen (Cannoodt et al., 2021)), a 1,000-dimensional Gaussian stress test, and four real single-cell time courses covering EMT (Cook & Vanderhyden, 2020), EB development (Moon et al., 2019), CITE-seq (Lance et al., 2022), and mouse hematopoiesis (Weinreb et al., 2020). Tables average W_1 over the relevant times; dataset sizes and trajectory plots appear in Appendices C.3.1 and C.7.

Matching observed population sequences. We fit one generator to every snapshot and generate from its initial population, testing whether the learned drift, exit rate, and entry model reproduce the observed sequence.

Table 2: Mean population-matching W_1 across observed times. SF-EUOT (EOT), UDSB-F, and ELYSIUM report mean±sample SD over five trained models; WFR-FM and USB values are taken from their papers. Appendix C.3.1 gives implementation details.

Dataset	WFR-FM	USB	SF-EUOT (EOT)	UDSB-F	ELYSIUM SDE
Simulation Gene	0.0190	0.0190	0.256±0.017	0.180±0.089	**0.0159±0.0005**
Dyngen	0.1350	**0.1310**	0.802±0.085	2.302±0.440	0.2023±0.0154
Gaussian-1000D	2.233	2.136	343.0±396.7	3.944±0.028	**1.956±0.004**
EMT	0.2239	0.2100	0.262±0.003	0.243±0.001	**0.1260±0.0008**
EB	8.855	9.087	10.206±0.113	13.321±0.088	**7.748±0.020**
CITE-seq	30.031	31.183	34.691±0.608	64.100±0.149	**23.475±0.070**
Mouse	5.848	6.069	7.281±0.052	14.438±0.127	**5.162±0.006**

ELYSIUM is best on six of seven datasets and has lower W_1 than both WFR-FM and USB on 17 of 21 individual snapshots; USB is best on Dyngen. The gains on all four real datasets show that the analytic construction remains effective after neural regression and stochastic generation in realistic state spaces (Appendix C.3.2).

Interpolating an unobserved time point. We next omit one interior snapshot, refit the generator on the remaining observations, and propagate the initial population continuously to the omitted time with no observation reset, so each prediction tests the learned dynamics between observations (Appendix C.3.3). The four real datasets provide eight such predictions. ELYSIUM is best on EMT, EB, and Mouse. On EB, mean W_1 falls from 10.175 for WFR-FM, the strongest baseline, to 9.770, with gains at all three omitted times.

6.3 Q3: CAN ENTRY AND EXIT MODEL DRUG-RESPONSE DYNAMICS?

The 4i assay follows melanoma cells after combined EGFR/MEK inhibition, during which the population first expands and then contracts while its molecular state changes (Bunne et al., 2023; Pariset et al., 2023). All methods are trained on the same three snapshots and generate continuously from 8 to 48 hours without resetting to the observed 24-hour population, so errors made during expansion propagate into the contraction phase. ELYSIUM models expansion through entry and contraction through exit within one generator and attains the lowest W_1 on held-out cells at both 24 and 48 hours. Appendix C.4 reports the protocol and quantitative comparison.

7 DISCUSSION

ELYSIUM provides a stochastic generator with simulation-free training for a killed-diffusion unbalanced SB. A cemetery state reduces the path-space problem to an endpoint projection, whose closed-form bridge components supply regression targets for the drift and exit rate, while a generative flow learns the endpoint distribution for entry. At the population optimum of the training objectives, the generator recovers the optimal bridge. Experiments verify recovery of hidden states, population size, and events against independently solved benchmarks, and show accurate matching, interpolation, and drug-response modeling on single-cell time courses.

Limitations. Anonymous snapshots identify only the net exchange between consecutive times; the inferred exits, entries, and their states and times therefore depend on the chosen reference. Parent–offspring relations cannot be identified, and each interval is modeled as net loss or net gain rather than simultaneous birth and death. Addressing these limitations would require lineage information (Weinreb et al., 2020) and a branching reference (Baradat & Lavenant, 2025). The reference parameters ν and κ encode prior assumptions about matching scale and event timing; in the sensitivity study of Appendix C.5, ν leaves the mass curves unchanged and has a modest effect on state interpolation, while a large κ degrades both. Extending the analytic components beyond killed Brownian references and incorporating lineage or RNA velocity data (La Manno et al., 2018) are natural next steps.

AI USE STATEMENT

The authors defined the research objective and core idea and supplied the initial formulation and prototype code. We used generative AI tools to assist with the mathematical derivations, implementation refinement, numerical experiments, and manuscript editing. We have reviewed all AI-assisted work: the derivations and proofs were checked step by step by the authors, AI-assisted code was verified against independent numerical checks, and all edited text was reviewed for correctness. The authors interpreted the results, verified the final claims, and take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

An anonymized implementation is available at `https://anonymous.4open.science/r/elysium-5EEC/README.md`. Complete derivations and proofs of every stated result are given in Appendix A, together with the exact assumptions under which they hold. Appendix B specifies the common training and generation procedures; Appendix C.3.1 collects architectures, budgets, and reference clocks for each dataset. Appendix C reports the evaluation metrics, dataset coverage, training seeds, and per-time results for every benchmark, including the sensitivity study over the reference parameters. The ground-truth benchmark protocol, including its independent discrete solver and numerical checks, is described in Appendix C.2.

REFERENCES

Aymeric Baradat and Hugo Lavenant. *Regularized Unbalanced Optimal Transport as Entropy Minimization with Respect to Branching Brownian Motion*, volume 458 of *Astérisque*. Société Mathématique de France, 2025. 194 pages; earlier preprint arXiv:2111.01666.

Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000. doi: 10.1007/s002110050002.

Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015. doi: 10.1137/141000439.

Volker Bergen, Marius Lange, Stefan Peidli, F. Alexander Wolf, and Fabian J. Theis. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology*, 38(12):1408–1414, 2020. doi: 10.1038/s41587-020-0591-3.

Charlotte Bunne, Stefan G. Stark, Gabriele Gut, Jacobo Sarabia Del Castillo, Mitch Levesque, Kjong-Van Lehmann, Lucas Pelkmans, Andreas Krause, and Gunnar Rätsch. Learning single-cell perturbation responses using neural optimal transport. *Nature Methods*, 20(11):1759–1768, 2023. doi: 10.1038/s41592-023-01969-x.

Robrecht Cannoodt et al. Spearheading future omics analyses using dyngen, a multi-modal simulator of single cells. *Nature Communications*, 12:3942, 2021. doi: 10.1038/s41467-021-24152-2.

Yongxin Chen, Tryphon T. Georgiou, and Michele Pavon. The most likely evolution of diffusing and vanishing particles: Schrödinger bridges with unbalanced marginals. *SIAM Journal on Control and Optimization*, 60(4):2016–2039, 2022. doi: 10.1137/21M1447672.

Yongxin Chen, Tryphon T. Georgiou, and Michele Pavon. Optimal survival strategies for diffusive flows: A Schrödinger bridge approach to unbalanced transport. *SIAM Review*, 67(3):579–604, 2025. doi: 10.1137/25M176581X.

Lenaic Chizat, Gabriel Peyre, Bernhard Schmitzer, and Francois-Xavier Vialard. Scaling algorithms for unbalanced optimal transport problems. *Mathematics of Computation*, 87(314):2563–2609, 2018. doi: 10.1090/mcom/3303.

Jaemoo Choi and Jaewoong Choi. Scalable simulation-free entropic unbalanced optimal transport. *arXiv preprint arXiv:2410.02656*, 2024. doi: 10.48550/arXiv.2410.02656.

David P. Cook and Barbara C. Vanderhyden. Context specificity of the EMT transcriptional response. *Nature Communications*, 11:2142, 2020. doi: 10.1038/s41467-020-16066-2.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, volume 26, 2013.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion Schrödinger bridge with applications to score-based generative modeling. In *Advances in Neural Information Processing Systems*, volume 34, pp. 17695–17709, 2021.

David S. Fischer, Anna K. Fiedler, Eric M. Kernfeld, Ryan M. J. Genga, Aimée Bastidas-Ponce, Mostafa Bakhti, Heiko Lickert, Jan Hasenauer, Rene Maehr, and Fabian J. Theis. Inferring population dynamics from single-cell RNA-sequencing time series data. *Nature Biotechnology*, 37 (4):461–468, 2019. doi: 10.1038/s41587-019-0088-0.

Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, et al. RNA velocity of single cells. *Nature*, 560(7719):494–498, 2018. doi: 10.1038/s41586-018-0414-6.

Christopher Lance et al. Multimodal single cell data integration challenge: Results and lessons learned. In *Proceedings of the NeurIPS 2021 Competitions and Demonstrations Track*, volume 176 of *Proceedings of Machine Learning Research*, pp. 162–176. PMLR, 2022.

Hugo Lavenant, Stephen Zhang, Young-Heon Kim, and Geoffrey Schiebinger. Toward a mathematical theory of trajectory inference. *Annals of Applied Probability*, 34(1A):428–500, 2024. doi: 10.1214/23-AAP1969.

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

Matthias Liero, Alexander Mielke, and Giuseppe Savaré. Optimal entropy-transport problems and a new hellinger–Kantorovich distance between positive measures. *Inventiones mathematicae*, 211 (3):969–1117, 2018. doi: 10.1007/s00222-017-0759-8.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.

Kevin R. Moon, David van Dijk, Zheng Wang, Scott Gigante, Daniel B. Burkhardt, William S. Chen, Kristina Yim, Antonia van den Elzen, Matthew J. Hirn, Ronald R. Coifman, et al. Visualizing structure and transitions in high-dimensional biological data. *Nature Biotechnology*, 37(12): 1482–1492, 2019. doi: 10.1038/s41587-019-0336-3.

Kirill Neklyudov, Rob Brekelmans, Daniel Severo, and Alireza Makhzani. Action matching: Learning stochastic dynamics from samples. In *International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 25858–25889. PMLR, 2023.

Kirill Neklyudov, Rob Brekelmans, Alexander Tong, Lazar Atanackovic, Qiang Liu, and Alireza Makhzani. A computational framework for solving Wasserstein Lagrangian flows. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 37461–37485. PMLR, 2024.

Matteo Pariset, Ya-Ping Hsieh, Charlotte Bunne, Andreas Krause, and Valentin De Bortoli. Unbalanced diffusion Schrödinger bridge. *arXiv preprint arXiv:2306.09099*, 2023. doi: 10.48550/arXiv.2306.09099. ICML 2023 workshop paper.

Qiangwei Peng, Zihan Wang, Junda Ying, Yuhao Sun, Qing Nie, Lei Zhang, Tiejun Li, and Peijie Zhou. WFR-FM: Simulation-free dynamic unbalanced optimal transport. In *The Fourteenth International Conference on Learning Representations*, volume 2026, pp. 18943–18976, 2026.

Gabriel Peyre and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5–6):355–607, 2019. doi: 10.1561/2200000073.

Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943.e22, 2019. doi: 10.1016/j.cell.2019.01.006.

Erwin Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Annales de l'Institut Henri Poincaré*, 2(4):269–310, 1932.

Yutong Sha, Yuchi Qiu, Peijie Zhou, and Qing Nie. Reconstructing growth and dynamic trajectories from single-cell transcriptomics data. *Nature Machine Intelligence*, 6(1):25–39, 2024. doi: 10.1038/s42256-023-00763-w.

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.

Sophia Tang, Yinuo Zhang, Alexander Tong, and Pranam Chatterjee. Branched Schrödinger bridge matching. In *International Conference on Learning Representations*, volume 2026, pp. 140186–140230, 2026. Earlier preprint: arXiv:2506.09007.

Alexander Tong, Jessie Huang, Guy Wolf, David van Dijk, and Smita Krishnaswamy. TrajectoryNet: A dynamic optimal transport network for modeling cellular dynamics. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 9526–9536. PMLR, 2020.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024a.

Alexander 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. PMLR, 2024b.

Dongyi Wang, Yuanwei Jiang, Zhenyi Zhang, Xiang Gu, Peijie Zhou, and Jian Sun. Joint velocity-growth flow matching for single-cell dynamics modeling. In *Advances in Neural Information Processing Systems*, volume 38, pp. 160552–160588, 2025. doi: 10.52202/085713-5363.

Xinyu Wang, Ruoyu Wang, Qiangwei Peng, Peijie Zhou, and Tiejun Li. WFR-MFM: One-step inference for dynamic unbalanced optimal transport. *arXiv preprint arXiv:2601.20606*, 2026. doi: 10.48550/arXiv.2601.20606.

Caleb Weinreb, Samuel Wolock, Betsabeh K. Tusi, Merav Socolovsky, and Allon M. Klein. Fundamental limits on dynamic inference from single-cell snapshots. *Proceedings of the National Academy of Sciences*, 115(10):E2467–E2476, 2018. doi: 10.1073/pnas.1714723115.

Caleb Weinreb, Alejo E. Rodriguez-Fraticelli, Fernando D. Camargo, and Allon M. Klein. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science*, 367(6479): eaaw3381, 2020. doi: 10.1126/science.aaw3381.

Grace Hui Ting Yeo, Sachit D. Saksena, and David K. Gifford. Generative modeling of single-cell time series with PRESCIENT enables prediction of cell trajectories with interventions. *Nature Communications*, 12:3222, 2021. doi: 10.1038/s41467-021-23518-w.

Junda Ying, Yuxuan Wang, Bowen Yang, Peijie Zhou, and Lei Zhang. Beyond continuity: Simulation-free reconstruction of discrete branching dynamics from single-cell snapshots. *arXiv preprint arXiv:2605.00545*, 2026.

Zhenyi Zhang, Tiejun Li, and Peijie Zhou. Learning stochastic dynamics from snapshots through regularized unbalanced optimal transport. In *The Thirteenth International Conference on Learning Representations*, pp. 19888–19919, 2025a.

Zhenyi Zhang, Zihan Wang, Yuhao Sun, Tiejun Li, and Peijie Zhou. Modeling cell dynamics and interactions with unbalanced mean field Schrödinger bridge. In *Advances in Neural Information Processing Systems*, volume 38, pp. 139518–139558, 2025b. doi: 10.52202/085713-4659.

Guide to the appendices. Appendix A develops the proofs and technical extensions; Appendix B gives the common numerical implementation. Appendix C follows the experimental questions Q1–Q3, then collects sensitivity, cost, and trajectory visualizations. Appendix D compares models of growth and individual events and explains the interpretation of population events.

A PROOFS AND TECHNICAL DETAILS

Appendix A.2 proves the dynamic formulation, cemetery encoding, and endpoint reduction in Sections 3.1–3.2. Appendices A.3–A.8 give the marked reference, closed-form components, and recovery of the learned generator. Appendix A.9 establishes convergence from empirical to population bridges, and Appendix A.10 extends the reference clocks.

A.1 NOTATION

Table 3: Notation shared by the main text and proofs. Symbols used only within one proof are defined where they first appear.

Symbol	Meaning
$\mathcal{X}$, $\bar{\mathcal{X}} = \mathcal{X} \cup \{\dagger\}$	observed state space and its extension with a cemetery state
μ_t, m_t	finite population measure at time t and its total mass
$g = \lvert m_1 - m_0 \rvert$	population mass assigned to exit or time-reversed entry
R, P^*	killed reference path measure and its unbalanced SB projection
X_t, τ, $\bar{X}_t$	active state, exit time, and state set to the cemetery after exit
R_{01}, R^{xy}, π^*	reference endpoint law, reference bridge conditioned on (x, y), and optimal endpoint coupling
ν, κ, $s_\kappa(t)$	Brownian variance, reference event rate, and active fraction of an exit component
Z_t, b_t, h_t	sampled bridge state and the drift and exit rate of its component
b_θ, h_θ	learned SDE drift and exit rate
u_ϕ, T_ϕ	auxiliary vector field for entry anchors and the map obtained by integrating this field
ζ, η_t, ρ_t	entry endpoint distribution used in the appendix, entry source, and active population measure
Δ_i, $D_i = \nu_i/\Delta_i$	interval duration and diffusion coefficient in physical time
v, b^+	probability-flow velocity and forward SDE drift in the proofs; the main text writes the latter as b

A.2 THE REFERENCE PROCESS, ENDPOINT REDUCTION, AND DYNAMIC FORMULATION

This subsection proves the endpoint KL decomposition and endpoint reduction in Section 3.2, and the equivalent dynamic formulation in Section 3.1. We use a general interval $[0, T]$; the main text takes $T = 1$. Let the active endpoint masses be $m_0 \geq m_T$, with total exit mass $g = m_0 - m_T$. For a path measure of mass m_0, the notation $\mathbb{E}_P F$ means $\int F \, \mathrm{d}P$, or m_0 times the expectation under its normalized path distribution.

Conditions for the continuous reference. The reference is a well-defined, nonexplosive Markov diffusion with exits on $\mathbb{R}^d \cup \{\dagger\}$, with finite initial mass $m_0 > 0$. Assume that the endpoint measures μ_0 and $\mu_T + g\delta_\dagger$ admit a coupling with finite KL relative to R_{0T}, as required in Proposition 2. Regular conditional path distributions are defined on the associated standard Borel path space. The empirical atomic problem uses the marked reference in Appendix A.3, which does not require an atomic terminal measure to be dominated by a Gaussian transition density.

For the dynamic formulation, take σ to be invertible and require the controls to satisfy the likelihood and integrability conditions stated below. Equivalence of the optimization problems additionally requires the optimal bridge to admit the regular Doob representation below, with controls satisfying those conditions. The choice $\sigma = \sqrt{\nu} I$ used in the remainder of the paper satisfies the invertibility requirement.

Equivalence of cemetery encoding and the original path problem. Let J be the cemetery encoding in Section 3.2: it retains the state path before exit and sets the state to † afterward. Denote the encoded process by $\bar{X}_t$ and the encoded path measures by $\bar{P} = J_{\#}P$ and $\bar{R} = J_{\#}R$. This encoding is invertible on histories over the observation interval. If an encoded path reaches †, its first hitting time gives τ, and its preceding segment gives the active path; otherwise the particle survives to T.

Relative entropy is invariant under this encoding:

$$D_{\mathrm{KL}}(P\|R) = D_{\mathrm{KL}}(J_{\#}P\|J_{\#}R).$$

The active marginal constraints in Equation 2 correspond exactly to the complete endpoint measures $\mu_0, \mu_T + g\delta_\dagger$. Setting $T = 1$ gives Equation 5, so the remaining calculations can be carried out on the extended path space.

More explicitly, let $\bar{P}^*$ minimize the extended problem and set $P^* = (J^{-1})_{\#}\bar{P}^*$. The endpoint correspondence ensures that P^* matches the original active endpoint constraints. If another process Q matched those constraints with smaller $D_{\mathrm{KL}}(Q\|R)$, then $J_{\#}Q$ would have smaller KL than $\bar{P}^*$ in the extended problem, contradicting optimality. Thus P^* is optimal for the original problem. The converse follows by the same argument.

Endpoint KL decomposition and proof of endpoint reduction. Disintegrate P and R with respect to

$$\omega \longmapsto (\bar{X}_0(\omega), \bar{X}_T(\omega)).$$

Write $\pi = P_{0T}$ for the endpoint coupling and P^{xy}, R^{xy} for the conditional path laws. Whenever $P \ll R$, the Radon–Nikodym derivative factorizes as

$$\frac{\mathrm{d}P}{\mathrm{d}R}(\omega) = \frac{\mathrm{d}P_{0T}}{\mathrm{d}R_{0T}}(\bar{X}_0, \bar{X}_T)\frac{\mathrm{d}P^{\bar{X}_0\bar{X}_T}}{\mathrm{d}R^{\bar{X}_0\bar{X}_T}}(\omega).$$

Taking logarithms and integrating against P yields

$$D_{\mathrm{KL}}(P\|R) = D_{\mathrm{KL}}(P_{0T}\|R_{0T}) + \int_{\bar{\mathcal{X}}^2} D_{\mathrm{KL}}(P^{xy}\|R^{xy})\, P_{0T}(\mathrm{d}x, \mathrm{d}y). \tag{17}$$

Each path measure and its endpoint measure have the same total mass, so the mass-correction terms in finite-measure KL are compatible with this decomposition; the conditional laws P^{xy}, R^{xy} are probability measures. Taking $T = 1$ gives Equation 4 (Léonard, 2014).

For a fixed endpoint coupling, the conditional term is nonnegative and vanishes precisely when $P^{xy} = R^{xy}$ for π-almost every pair. Every coupling with the prescribed marginals and finite KL can be realized by mixing the reference conditional paths, with zero conditional cost. Minimizing path-space relative entropy is therefore equivalent to minimizing $D_{\mathrm{KL}}(\pi\|R_{0T})$ with the prescribed endpoint marginals, as in Equation 6.

The set of couplings with fixed marginals is weakly compact, and relative entropy is lower semicontinuous. A coupling with finite KL therefore ensures attainment; strict convexity gives uniqueness. The optimal path measure is

$$P^* = \int_{\bar{\mathcal{X}}^2} R^{xy}\, \pi^*(\mathrm{d}x, \mathrm{d}y). \tag{18}$$

Setting $T = 1$ gives Equation 7 and completes the proof of Proposition 2.

Derivation of the dynamic control cost. Let $A_t = \mathbf{1}_{\{t<\tau\}}$ and $N_t = \mathbf{1}_{\{\tau\le t\}}$. The predictable intensity of N is $A_{t-}\lambda_0(t, X_{t-})$ under the reference and $A_{t-}\alpha_t\lambda_0(t, X_{t-})$ under $P^{u,\alpha}$. Thus the controlled drift is $b_0 + \sigma u$, and the exit rate is $h_t = \alpha_t\lambda_0(t, X_t)$: $P^{u,h}$ in the main text and $P^{u,\alpha}$ here are two parameterizations of the same path measure. The controls are predictable and make the controlled diffusion and exit process well defined. Where $\lambda_0 = 0$, finite cost requires $h_t = 0$, and the value of α_t there does not affect the process.

For positive exit rates, assume the likelihood process is a true martingale and the control cost below is integrable. The continuous Girsanov formula and the likelihood formula for the exit counting process

give

$$\log \frac{\mathrm{d}P^{u,\alpha}}{\mathrm{d}R} = \log \frac{\mathrm{d}P_0}{\mathrm{d}R_0}(X_0) + \int_0^T A_t u_t^\top \, \mathrm{d}W_t^R - \frac{1}{2}\int_0^T A_t \|u_t\|^2 \, \mathrm{d}t + \int_0^T \log \alpha_t \, \mathrm{d}N_t - \int_0^T A_t(\alpha_t - 1)\lambda_0(t, X_t) \, \mathrm{d}t. \tag{19}$$

Under the controlled process,

$$\mathrm{d}W_t^R = \mathrm{d}W_t^P + A_t u_t \, \mathrm{d}t,$$

and the compensator of N is

$$\int_0^t A_s \alpha_s \lambda_0(s, X_s) \, \mathrm{d}s.$$

Substitute the expression for W^R into Equation 19 and take expectations. The Brownian integral has mean zero, while the mean of the counting integral is determined by its compensator. Hence

$$D_{\mathrm{KL}}(P^{u,\alpha} \| R) = D_{\mathrm{KL}}(P_0 \| R_0) + \mathbb{E}_{P^{u,\alpha}} \int_0^T A_t \left[\frac{1}{2}\|u_t\|^2 + \lambda_0(t, X_t)\{\alpha_t \log \alpha_t - \alpha_t + 1\} \right] \mathrm{d}t. \tag{20}$$

In Equation 2, $P_0 = R_0 = \mu_0$, so the initial relative-entropy term is zero. Substituting $h_t = \alpha_t \lambda_0(t, X_t)$ gives

$$\lambda_0(t, X_t)\{\alpha_t \log \alpha_t - \alpha_t + 1\} = h_t \log \frac{h_t}{\lambda_0(t, X_t)} - h_t + \lambda_0(t, X_t).$$

Also, $\int_0^T A_t(\cdot) \, \mathrm{d}t = \int_0^{T \wedge \tau}(\cdot) \, \mathrm{d}t$. With $T = 1$, Equation 20 is exactly the control cost in Equation 3, subject to the same terminal constraint $P_T^{\mathrm{act}} = \mu_T$.

The case $h = 0$ follows by positive approximation and lower semicontinuity, with $0 \log(0/\lambda) = 0$. When $\lambda = 0$, the relative-entropy rate is zero for $h = 0$ and infinite for $h > 0$. Standard Markovization replaces history-dependent controls by their conditional means given the current active state, preserving the active marginals and not increasing the convex quadratic and Poisson costs, provided the corresponding process exists.

Doob transform and equivalence of the dynamic optimization. Write $a = \sigma\sigma^\top$. Suppose the optimal endpoint projection has the path-space potential representation

$$\frac{\mathrm{d}P^*}{\mathrm{d}R} = f(\bar{X}_0) q_T(\bar{X}_T).$$

Here q_T denotes the terminal potential, with g reserved for total exit mass. Assume the backward potential is finite, strictly positive on the active interior support, and $C^{1,2}$ there. Define

$$q_t(x) = \mathbb{E}_R[q_T(\bar{X}_T) \mid \bar{X}_t = x] = \begin{cases} \phi_t(x), & x \in \mathbb{R}^d, \\ \psi_t, & x = \dagger. \end{cases}$$

Because the cemetery is absorbing, $\dot{\psi}_t = 0$, while

$$\partial_t \phi + b_0 \cdot \nabla \phi + \frac{1}{2} a : \nabla^2 \phi - \lambda_0(t, x)\phi + \lambda_0(t, x)\psi = 0.$$

For a test function F, the time-dependent q_t-transform of the reference generator is

$$L_t^{P^*} F(x) = (b_0 + a \nabla \log \phi_t) \cdot \nabla F + \frac{1}{2} a : \nabla^2 F + \lambda_0(t, x) \frac{\psi_t}{\phi_t(x)} [F(\dagger) - F(x)].$$

Its drift and exit rate are

$$b^*(t, x) = b_0(t, x) + a(t, x) \nabla \log \phi_t(x), \qquad h^*(t, x) = \lambda_0(t, x) \frac{\psi_t}{\phi_t(x)}. \tag{21}$$

Set

$$u_t^* = \sigma(t, X_t)^\top \nabla \log \phi_t(X_t).$$

Then $b_0 + \sigma u^* = b^*$, and h^* is the rate in Equation 21. When these controls satisfy the likelihood and integrability conditions above, the optimal bridge belongs to the controlled class in the main text: $P^* = P^{u^*,h^*}$.

Every admissible control (u, h) satisfying the endpoint constraints defines a path measure in Equation 2, and Equation 20 equates its dynamic cost with path-space relative entropy. Thus the infimum of the dynamic problem is at least the minimum of the path problem. Conversely, (u^*, h^*) attains the path-space minimum and gives the reverse inequality. The two problems have the same optimal value and, by uniqueness of the path-space optimizer, describe the same optimal process. This proves the equivalent dynamic formulation in Section 3.1.

These conditions specify the entropy and interior regularity assumptions used in the calculation. Identifying a generated process uniquely from its learned coefficients additionally uses the uniqueness-in-law conditions in Appendix A.8.

A.3 Finite-measure KL and the marked reference

For finite measures $P \ll K$ on a common measurable space, we use

$$D_{\mathrm{KL}}(P\|K) = \int \log\left(\frac{\mathrm{d}P}{\mathrm{d}K}\right) \mathrm{d}P - P(\Omega) + K(\Omega). \tag{22}$$

Suppose a marked space is a disjoint union of components and $R = \sum_c k_c R^c$, $P = \sum_c p_c P^c$, where R^c, P^c are probability laws. Direct substitution into equation 22 gives

$$D_{\mathrm{KL}}(P\|R) = \sum_c \left[p_c \log \frac{p_c}{k_c} - p_c + k_c\right] + \sum_c p_c D_{\mathrm{KL}}(P^c\|R^c). \tag{23}$$

The identity extends to zero component mass by lower semicontinuity. Thus, after endpoint component masses are chosen, each conditional path law is uniquely minimized by its reference conditional law.

The empirical construction uses labeled atoms. Coincident coordinates still define distinct endpoint labels. With Brownian bridge laws B_i^{ab}, conditioned exit laws $\mathsf{D}_i^{a\dagger}$, and the idle cemetery path $\mathsf{I}^\dagger$, define the active and full marked references by

$$\mathsf{R}_{i,\mathrm{act}}^{\mathrm{disc}} = \sum_{a,b} \widetilde{K}_{i,\mathrm{act}}^{ab} \mathsf{B}_i^{ab} + \sum_a \widetilde{K}_{i,\mathrm{act}}^{a\dagger} \mathsf{D}_i^{a\dagger}, \qquad \overline{\mathsf{R}}_i^{\mathrm{disc}} = \mathsf{R}_{i,\mathrm{act}}^{\mathrm{disc}} + \mathsf{I}^\dagger. \tag{24}$$

Forgetting labels maps it to the usual state-space interpolation.

Equation 8 is the dominated continuous reference with μ_0 as its initial measure. The same disintegration applies whenever the prescribed coupling is dominated by this endpoint law. Its live–live endpoint density contains a factor for the source row, Gaussian normalization, and the survival factor $e^{-\kappa_i}$. Multiplying a reference kernel by a positive row factor and a positive column factor changes generalized KL by a constant on a set of fixed marginals. Hence the normalized continuous version of equation 6 is precisely the endpoint projection induced by equation 2 through the lift equation 5.

For empirical terminal atoms, the disjoint endpoint labels in equation 24 define the exact finite entropy problem directly. The dominated density formulation above is its continuous population counterpart, and Proposition 6 connects the active empirical bridges to that continuous projection as the endpoint measures converge.

A.4 Endpoint kernel of the killed Brownian reference

Condition on an active initial state $X_0 = x$. The Brownian increment and the exponential exit clock are independent. For any measurable active set $A \subset \mathbb{R}^d$, survival to time one and the Gaussian transition therefore give

$$R_1(A \mid x) = \Pr(\tau > 1) \Pr(x + \sqrt{\nu} W_1 \in A) = e^{-\kappa} \int_A \varphi_\nu(y - x) \mathrm{d}y. \tag{25}$$

All paths that exit by time one have the cemetery endpoint, so

$$R_1(\{\dagger\} \mid x) = \Pr(\tau \le 1) = 1 - e^{-\kappa}. \tag{26}$$

Multiplying these two conditional probabilities by the initial measure $\mu_0(\mathrm{d}x)$ and placing them on their disjoint endpoint supports gives Equation 8.

A.5 Bridge decomposition and cemetery slack (Theorem 3)

We prove Theorem 3 and an auxiliary statement: adding the same cemetery mass at both endpoints adds only paths that remain at the cemetery, leaving the optimal active coupling and its motion and exit process unchanged.

Proof. Orient time from larger to smaller mass. The reference has no paths from † to a live state, so absolute continuity forces that block of every feasible coupling with finite KL to vanish. The cemetery source row has mass s, and therefore all of it must remain on the idle cemetery path. The target cemetery column has mass $g + s$, so its remaining live contribution is

$$(g + s) - s = g. \tag{27}$$

Removing the forced idle atom leaves precisely the feasible set of equation 6. Generalized KL is additive on the two disjoint blocks, and the idle contribution is constant. Thus the active coupling is unchanged and the added slack follows only the idle cemetery path.

The active feasible polytope is nonempty and compact. Every allowed reference entry is positive, so generalized KL is lower semicontinuous and strictly convex on that support. The active minimizer therefore exists and is unique.

Now disintegrate any active marked path measure P by its endpoint marks, and write $\mathsf{R}_i^{ab} = \mathsf{B}_i^{ab}$ and $\mathsf{R}_i^{a\dagger} = \mathsf{D}_i^{a\dagger}$. Equation 23 becomes

$$D_{\mathrm{KL}}(P \| \mathsf{R}_{i,\mathrm{act}}^{\mathrm{disc}}) = D_{\mathrm{KL}}(\pi \| \widetilde{K}_{i,\mathrm{act}}) + \sum_{u,v} \pi_{uv} D_{\mathrm{KL}}(P^{uv} \| \mathsf{R}_i^{uv}). \tag{28}$$

For fixed π, every conditional term is uniquely minimized by the corresponding reference law; the first term is uniquely minimized by π_i^*. This proves the bridge decomposition in Theorem 3.

Finally, encode a killed component by its path before exit, exit time, and endpoint marks. Time reversal Θ is a measurable bijection and relative entropy is invariant under such a bijection:

$$D_{\mathrm{KL}}(\Theta_{\#} P \| \Theta_{\#} R) = D_{\mathrm{KL}}(P \| R). \tag{29}$$

It transposes the endpoint plan and turns exit into entry. Applying it to both the optimizer in the loss orientation and its killed reference shows that $\Theta_{\#}\mathsf{P}_i^*$ is the unique projection relative to $\Theta_{\#}\mathsf{R}_{i,\mathrm{act}}^{\mathrm{disc}}$, which proves the statement for a growing interval. □

A.6 Conditional component laws and training targets

We first give the components' active masses, then derive the state distributions and local coefficients for the survival, exit, and entry components in Section 4.1.

Total active mass. On the normalized interval $[0, 1]$, orient time in the direction of mass loss and write $m_0 \ge m_1$, with $g = m_0 - m_1$. Survival components contribute mass m_1 at every time, while exit components remain active with fraction $s_\kappa(t)$ from Equation 9. The total active masses of the exit bridge and its time-reversed entry bridge are therefore

$$m_{\mathrm{exit}}(t) = m_1 + g s_\kappa(t), \qquad m_{\mathrm{entry}}(t) = m_1 + g s_\kappa(1 - t). \tag{30}$$

For the instantaneous coefficients below, write $S_\kappa(r) = s_\kappa(r)$ and $A_\kappa(r) = s_\kappa(1 - r)$.

To obtain the SDE drift, we first find a velocity field that reproduces each component's Gaussian marginals, then add the diffusion correction. All derivatives below are in physical time, so $\partial_t r = 1/\Delta_i$. If $p_t = \mathcal{N}(m_t, q_t I_d)$, then

$$u_t(z) = \dot{m}_t + \frac{\dot{q}_t}{2q_t}(z - m_t) \tag{31}$$

satisfies $\partial_t p_t + \nabla \cdot (p_t u_t) = 0$. This follows by differentiating the Gaussian log density and matching its constant and quadratic terms.

Survival (live–live). Gaussian conditioning gives

$$X_r \mid (X_0 = x, X_1 = y) \sim \mathcal{N}(m_r^{xy}, \nu_i r(1-r) I_d).$$

Here $m_r^{xy} = (1-r)x + ry$. For $0 < \varepsilon < 1 - r$, Gaussian transition densities and Bayes' rule give the Brownian bridge transition law

$$R^{xy}(X_{r+\varepsilon} \in \mathrm{d}z' \mid X_r = z) = \frac{\varphi_{\nu_i \varepsilon}(z' - z)\varphi_{\nu_i(1-r-\varepsilon)}(y - z')}{\varphi_{\nu_i(1-r)}(y - z)}\,\mathrm{d}z'.$$

Its conditional mean is $z + \varepsilon(y - z)/(1 - r)$. Dividing the mean increment by the physical time increment $\Delta_i \varepsilon$ gives

$$b_t^{+,xy}(z) = \frac{y - z}{\Delta_i(1 - r)} = \frac{\nu_i}{\Delta_i} \nabla_z \log \varphi_{\nu_i(1-r)}(y - z).$$

For $\Delta_i = 1$, this is the survival drift in Section 4.1. The component is conditioned on survival to the endpoint, so its exit rate is zero. Substituting the same Gaussian marginal into Equation 31 gives the probability-flow velocity

$$v_t^{xy}(z) = \frac{y - x}{\Delta_i} + \frac{1 - 2r}{2r(1 - r)\Delta_i}(z - m_r^{xy}).$$

Its score is $-[z - m_r^{xy}]/[\nu_i r(1 - r)]$, and therefore

$$v^{xy} + \frac{D_i}{2} \nabla \log \rho^{xy} = \frac{y - x}{\Delta_i} - \frac{z - m_r^{xy}}{(1 - r)\Delta_i} = \frac{y - z}{(1 - r)\Delta_i}. \tag{32}$$

Exit (live–cemetery). Conditioning the independent exponential clock on death by normalized time one gives

$$\Pr(\tau > r \mid \tau \le 1) = S_{\kappa_i}(r). \tag{33}$$

Conditional on being live at r, the spatial law is $\mathcal{N}(x, \nu_i r I_d)$. Equation 31 gives $v = (z - x)/(2r\Delta_i)$, while its score contribution $(D_i/2)\nabla \log \rho$ is the negative of this quantity. Hence $b^+ = 0$.

For $0 < \varepsilon < 1 - r$, conditional on being active at r and exiting by the endpoint, the probability of exit in the next short interval is

$$\Pr(r < \tau \le r + \varepsilon \mid r < \tau \le 1) = \frac{e^{-\kappa_i r} - e^{-\kappa_i(r+\varepsilon)}}{e^{-\kappa_i r} - e^{-\kappa_i}}.$$

Dividing by $\Delta_i \varepsilon$ and sending $\varepsilon \downarrow 0$ gives the conditional exit rate, equivalently the negative logarithmic derivative of the active fraction:

$$-\partial_t \log S_{\kappa_i}(r) = \frac{\kappa_i}{\Delta_i[1 - e^{-\kappa_i(1-r)}]} = h_t^{x\dagger}. \tag{34}$$

Although the conditional exit rate diverges at the terminal boundary, its exit flux remains regular:

$$S_{\kappa_i}(r) h_t^{x\dagger} = \frac{\kappa_i e^{-\kappa_i r}}{\Delta_i(1 - e^{-\kappa_i})}, \qquad \int_{t_i}^{t_{i+1}} S_{\kappa_i}(r) h_t^{x\dagger} \mathrm{d}t = 1. \tag{35}$$

Entry (cemetery–live). Time reversal replaces r with $1 - r$. The live finite measure is $A_{\kappa_i}(r)\mathcal{N}(y, \nu_i(1 - r)I_d)$. Equation 31 gives $v_t(z) = -(z - y)/[2(1 - r)\Delta_i]$, and adding half the diffusion times the score gives

$$b^+ = -\frac{z - y}{(1 - r)\Delta_i}.$$

The component stays active after entry, so $h_t^{\dagger y} = 0$. Its entry events are described by a source with temporal rate

$$\partial_t A_{\kappa_i}(r) = \frac{\kappa_i e^{-\kappa_i(1-r)}}{\Delta_i(1 - e^{-\kappa_i})} = a_{\kappa_i}(r). \tag{36}$$

Weighting these entry contributions by the endpoint coupling gives the source in Equation 11. The total entering mass of a unit entry component is one.

Each component consequently satisfies

$$\partial_t \rho_t^c + \nabla \cdot (\rho_t^c v_t^c) = -h_t^c \rho_t^c + \eta_t^c. \tag{37}$$

A.7 Recovery of the entry source (Proposition 4)

We prove Proposition 4 and analyze the training objective for the entry map T_ϕ in Section 4.2. For one normalized growing interval, define the endpoint distribution of the entry components by

$$\zeta = \frac{1}{g}\sum_y \pi^*_{\dagger y}\delta_y, \qquad g = \sum_y \pi^*_{\dagger y} > 0.$$

We use ζ only in the auxiliary-flow and error calculations below; the main text uses the endpoint weights from the coupling directly. For continuous endpoints, replace the sums by integrals against the entry endpoint measure.

Recovery of the entry source. If $U \sim \mathcal{N}(0, I)$ and $\mathcal{L}(T_\phi(U)) = \zeta$, then, conditional on $T_\phi(U) = y$, Equation 12 gives density $\varphi_{\nu(1-t)}(z-y)$. Integrating over endpoints and multiplying by the total entry rate gives

$$\begin{aligned} g\,\frac{\mathrm{d}}{\mathrm{d}t}s_\kappa(1-t)\int \varphi_{\nu(1-t)}(z-y)\,\zeta(\mathrm{d}y) \\ = \frac{\mathrm{d}}{\mathrm{d}t}s_\kappa(1-t)\sum_y \pi^*_{\dagger y}\varphi_{\nu(1-t)}(z-y) = \eta_t(z). \end{aligned}$$

This is the source in Equation 11, proving Proposition 4.

Training the entry map. Draw $U \sim \mathcal{N}(0, I)$, $Y \sim \zeta$, and auxiliary flow time $s \sim \mathrm{Unif}[0, 1]$ independently, and let $Z_s = (1-s)U + sY$. Here s is not the physical time t of the population process. The training objective is

$$\mathcal{L}_{\mathrm{entry}}(\phi) = \mathbb{E}_{s,U,Y}\left\|u_\phi(s, Z_s) - (Y - U)\right\|^2. \tag{38}$$

At generation time, integrating u_ϕ from $s = 0$ to $s = 1$, starting from U, gives the entry anchor $T_\phi(U)$. Let $\chi_s = \mathcal{L}(Z_s)$. The population minimizer of squared error is

$$u^*(s, z) = \mathbb{E}[Y - U \mid Z_s = z].$$

For every compactly supported smooth function f,

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}f(Z_s) = \int \nabla f(z)^\top u^*(s, z)\,\chi_s(\mathrm{d}z).$$

This is the weak continuity equation $\partial_s\chi_s + \nabla\cdot(\chi_s u^*) = 0$. It starts from $\mathcal{N}(0, I_d)$ and converges weakly to ζ as $s \uparrow 1$. For an atomic empirical target, the endpoint is understood in this weak limit.

Source error. Let the learned endpoint distribution be $\widehat{\zeta} = \mathcal{L}(T_\phi(U))$. Couple $Y \sim \zeta$ and $\widehat{Y} \sim \widehat{\zeta}$, and add the same independent Gaussian noise to both. Their expected distance is unchanged. Taking the infimum over couplings gives

$$W_1\left(\frac{\eta_t}{\eta_t(\mathbb{R}^d)}, \frac{\widehat{\eta}_t}{\widehat{\eta}_t(\mathbb{R}^d)}\right) \le W_1(\zeta, \widehat{\zeta}). \tag{39}$$

The common scalar source mass cancels after normalization.

A.8 Recovery of the optimal path law (Theorem 5)

Assumptions and endpoint interpretation. For the continuous formulation, use the conditions for endpoint projection and the interior Doob transform in Appendix A.2, with $\nu, \kappa > 0$. At each interior time, assume the drift labels have finite second moments and the conditional Poisson loss is finite for an admissible rate. Regression optimality is understood pointwise for almost every $t \in (0, 1)$, or on compact interior time intervals; it does not require uniformly bounded component targets at the endpoints.

For the resulting drift and exit rate, assume the killed-diffusion martingale problem, initialized by μ_0/m_0, has a unique nonexplosive law on every $[0, 1-\epsilon]$, $\epsilon \in (0, 1)$, and admits terminal path limits in the augmented state space. On a gaining interval, require the analogous uniqueness with its

prescribed initial population, entry clock, and kernel for entering states. These are the uniqueness-in-law conditions used in Theorem 5; existence is supplied by the reference bridge construction. Local spatial Lipschitz regularity and linear growth of the drift, together with a locally bounded nonnegative exit rate, provide sufficient local conditions in the loss orientation. Terminal coefficients may be singular: endpoint measures are recovered as weak traces, as shown below.

For a finite endpoint plan, interpret the target as the marked bridge of Appendix A.3 after forgetting endpoint labels. Its potentials, which are sums of Gaussians, give the corresponding interior construction; convergence to the population bridge is addressed separately in Appendix A.9.

Proof. Regression optima. Fix an interval and an interior time. With weights from the endpoint plan ω_c,

$$\rho_t(z) = \sum_c \omega_c \rho_t^c(z), \qquad \Pr(C_t = c \mid Z_t = z) = \frac{\omega_c \rho_t^c(z)}{\rho_t(z)}.$$

The targets in the main text are $b_t = b_t^{+,C_t}(Z_t)$ and $h_t = h_t^{C_t}$. Squared loss is pointwise minimized by conditional expectation, which proves the identity for the forward drift. Define v^* as the conditional mean of the analytic component velocities. For every component, its forward SDE drift and probability-flow velocity satisfy

$$b^{+,c} = v^c + \frac{D_i}{2} \nabla \log \rho_t^c. \tag{40}$$

Therefore,

$$b^{+*} - v^* = \frac{D_i}{2} \sum_c \frac{\omega_c \rho_t^c}{\rho_t} \nabla \log \rho_t^c = \frac{D_i}{2} \frac{\sum_c \omega_c \nabla \rho_t^c}{\rho_t} = \frac{D_i}{2} \nabla \log \rho_t. \tag{41}$$

For the exit-rate term, let $U = h_t^{C_t}$. At fixed (t, z), the derivative of $\mathbb{E}[U \log(U/b) - U + b \mid Z_t = z]$ with respect to $b > 0$ is

$$\frac{\partial}{\partial b} \mathbb{E}[\ell_{\mathrm{P}}(U, b) \mid Z_t = z] = 1 - \frac{\mathbb{E}[U \mid Z_t = z]}{b}.$$

Its minimizer is the conditional mean, with zero obtained in the closure.

Identification of the path law. We next identify these conditional means with the unbalanced SB optimizer and thereby recover its full path law. By equation 7, drawing a component using the mass assigned by the endpoint plan and then drawing its conditioned path is exactly a disintegration of P^*. Moreover, the endpoint entropy projection has $\mathrm{d}P^*/\mathrm{d}R = f(\bar{X}_0) q_T(\bar{X}_T)$. It is therefore the Markov Doob transform of the killed reference, with drift and exit intensity given by equation 21. The predictable characteristics of this same path mixture are the conditional component averages derived above. Consequently b^{+*} and h^* are the local characteristics of P^*. By the uniqueness-in-law condition stated above, the killed diffusion with these characteristics in the loss orientation has path law P^*.

In the empirical case, the same argument uses the marked reference rather than absolute continuity with respect to the continuous Brownian reference. On rows and columns with positive mass, the optimal finite plan is a positive row–column scaling of the endpoint kernel. Combining its column weights with the Gaussian transition kernels and cemetery transition probability gives the backward potential as a finite sum. The Gaussian transition identity then identifies the unmarked mixture as its Markov transform at interior times, with the same conditional drift and exit-rate formulas. The uniqueness argument therefore applies to this empirical target as well.

Population evolution. Multiplying equation 37 by ω_c and summing gives

$$\sum_c \omega_c \rho_t^c v_t^c = \rho_t v^*, \qquad \sum_c \omega_c \rho_t^c h_t^c = \rho_t h^*.$$

The weighted entry terms sum to the source η_t in Equation 11; hence

$$\partial_t \rho_t + \nabla \cdot (\rho_t v^*) = -h^* \rho_t + \eta_t. \tag{42}$$

In physical time, the diffusion coefficient is $D_i = \nu_i/\Delta_i$; setting $\Delta_i = 1$ gives the normalized time notation in the main text. The spatial Fokker–Planck operator of the SDE in equation 14 is

$$-\nabla\cdot(\rho_t b^{+*}) + \frac{D_i}{2}\Delta\rho_t = -\nabla\cdot(\rho_t v^*). \tag{43}$$

Combining this identity with equation 42, and returning to normalized time, gives equation 15.

Time reversal and endpoint traces. For a gaining interval, Proposition 4 reconstructs the exact entry source. Time reversal of the loss-oriented bridge then gives the same local drift, entry clock, and entry state kernel as the construction in the main text. Uniqueness again identifies the resulting augmented path law with the time reversal of P^*.

Finally, live–live Gaussians converge weakly to their endpoint atoms, $s_\kappa(0) = 1, s_\kappa(1) = 0$, and therefore $s_\kappa(1-0) = 0, s_\kappa(1-1) = 1$. The row and column constraints of the plan therefore give the prescribed endpoint traces; time reversal gives the gaining case. □

A.9 Consistency from empirical to population bridges

The preceding construction uses a finite endpoint plan in practice. The following result connects its marked bridge to the continuous population problem as the empirical endpoints converge.

Proposition 6 (Consistency of empirical marked bridges)**.** *Let the live state space $\mathsf{X} \subset \mathbb{R}^d$ be compact. Suppose the empirical endpoint measures of larger and smaller mass converge weakly to finite measures on X, and their masses converge. For fixed $\nu_i, \kappa_i > 0$, the active solutions of equation 6 converge weakly to the unique continuous endpoint entropy projection. If $\rho_{i,r}^{(n)}$ and $\rho_{i,r}$ are the corresponding active marginals of the bridge mixtures, then, for every $\epsilon \in (0, 1/2)$,*

$$\sup_{r\in[\epsilon,1-\epsilon]} d_{\mathrm{BL}}\Big(\rho_{i,r}^{(n)}, \rho_{i,r}\Big) \longrightarrow 0. \tag{44}$$

The finite measures of exit or entry times and event locations converge weakly as well. When the limiting endpoints admit a coupling with finite KL relative to the continuous reference, the limit is the killed-diffusion unbalanced SB of Definition 1, with local dynamics given by Equation 21 under the interior regularity conditions above. The gaining case follows by time reversal.

Proof. We give the case of decreasing mass. Let H_n, L_n be the empirical endpoint masses and H, L their limits. Write $\varsigma_n = \mu_n^L + (H_n - L_n)\delta_\dagger$ and $\varsigma = \mu^L + (H - L)\delta_\dagger$. On the compact augmented endpoint space, after deleting a zero-mass cemetery column if necessary,

$$k(x,y) = \begin{cases} e^{-\kappa_i}\varphi_{\nu_i}(y-x), & y \in \mathsf{X}, \\ 1 - e^{-\kappa_i}, & y = \dagger \end{cases}$$

is continuous and bounded above and away from zero on its allowed support. Multiplying rows or columns of the empirical reference by positive weights changes generalized KL only by terms fixed by the marginals. Consequently equation 6 has the same optimizer as the entropic projection with reference $k(x,y)\mu_n^H(\mathrm{d}x)\varsigma_n(\mathrm{d}y)$.

Fixing the additive gauge in its Schrödinger dual, the uniform upper and lower bounds on k bound the dual potentials. The Sinkhorn equations and uniform continuity of k make their continuous extensions equicontinuous. Arzelà–Ascoli gives a uniformly convergent subsequence; weak endpoint convergence permits passage to the limit in both Sinkhorn equations. Every subsequential limit therefore solves the continuous projection. Strict convexity on the allowed support makes that solution unique, so the complete sequence of plans satisfies $\pi_n^* \Rightarrow \pi^*$.

For $r \in [\epsilon, 1-\epsilon]$, the Gaussian live–live and live–absorbing conditional kernels depend uniformly continuously, in bounded–Lipschitz distance, on their endpoint labels and on r. Integrating them against the convergent plans proves equation 44. Event measures are continuous mixtures of the same plan and the fixed conditional clock, so they converge as well. Row–column gauge invariance identifies the limiting plan with Proposition 2 whenever the limiting endpoint problem admits a feasible coupling with finite KL. Time reversal gives the gaining case. □

A.10 Extensions to time-dependent reference clocks

We now allow the reference rate to vary over time without changing the endpoint projection. A rate that depends only on time changes event timing while leaving the spatial reference unchanged; a deterministic time change reparameterizes both motion and events. Their experimental settings are collected in Appendix C.3.1.

Event rates that depend only on time. Let $c_i : [0,1] \to (0,\infty)$ satisfy $\int_0^1 c_i(u)\mathrm{d}u = 1$, and write $C_i(r) = \int_0^r c_i(u)\mathrm{d}u$. The conditional exit density and remaining live fraction are

$$q^-_{\kappa_i,c_i}(r) = \frac{\kappa_i c_i(r) e^{-\kappa_i C_i(r)}}{1-e^{-\kappa_i}}, \qquad S_{\kappa_i,c_i}(r) = \frac{e^{-\kappa_i C_i(r)} - e^{-\kappa_i}}{1-e^{-\kappa_i}}, \tag{45}$$

with entry obtained by $r \mapsto 1-r$. The Mouse experiment uses

$$c_\beta(r) = \begin{cases} \beta e^{\beta r}/(e^\beta - 1), & \beta \neq 0, \\ 1, & \beta = 0, \end{cases} \qquad C_\beta(r) = \begin{cases} (e^{\beta r}-1)/(e^\beta - 1), & \beta \neq 0, \\ r, & \beta = 0. \end{cases} \tag{46}$$

Proposition 7 (Event rates that depend only on time)**.** *Let $c_i : [0,1] \to (0,\infty)$ be continuous with $\int_0^1 c_i(u)\mathrm{d}u = 1$. Replacing the constant rate in equation 1 by the rate $\kappa_i c_i(r)$, independent of the Brownian motion, leaves the endpoint kernel and endpoint entropy projection unchanged. Conditional on an event before time one, the exit density and remaining live fraction are those in equation 45; time reversal gives entrance, and the conditional spatial Brownian bridges are unchanged.*

Proof. The unconditioned survival function is $\exp[-\kappa_i C_i(r)]$, and $C_i(1) = 1$, so survival through the whole interval is still $e^{-\kappa_i}$. Independence from Brownian motion gives the same endpoint kernel. Differentiating the survival function, conditioning on an event by time one, and subtracting the terminal survivor gives equation 45. Time reversal gives entrance. □

Once the reference clock is fixed, its conditional event law is selected by KL. Let $P_{\kappa,c}|_{[0,1]} = (1-e^{-\kappa})Q^-_{\kappa,c}$, where $Q^-_{\kappa,c}$ has density $q^-_{\kappa,c}$. For any event-time probability $Q \ll Q^-_{\kappa,c}$,

$$D_{\mathrm{KL}}(Q \| P_{\kappa,c}|_{[0,1]}) = D_{\mathrm{KL}}(Q \| Q^-_{\kappa,c}) - \log(1-e^{-\kappa}) - e^{-\kappa}. \tag{47}$$

The unique minimizer is therefore $Q^-_{\kappa,c}$. Multiplication by the fixed mass g changes only fixed scalar terms.

Deterministic time changes.

Corollary 8 (Deterministic reference clocks)**.** *Let $\psi_i : [0,1] \to [0,1]$ be a continuously differentiable increasing bijection with $\psi_i'(r) > 0$. The time-changed reference $\widetilde{X}_r = X_{\psi_i(r)}$ has*

$$\mathrm{d}\widetilde{X}_r = \sqrt{\nu_i \psi_i'(r)}\,\mathrm{d}\widetilde{W}_r, \qquad \widetilde{\lambda}_i(r) = \kappa_i \psi_i'(r), \tag{48}$$

and the same endpoint kernel. Its unbalanced SB is the deterministic time change of the base unbalanced SB. Replace r by $\psi_i(r)$ in every conditional law and multiply drift, exit rate, source rate, and quadratic-variation rate by $\psi_i'(r)$.

Proof. The map $\mathcal{T}_{\psi_i} : \omega \mapsto [r \mapsto \omega(\psi_i(r))]$ is a measurable path-space bijection that fixes both endpoints. Standard deterministic time change gives equation 48; therefore the endpoint kernel is unchanged. Relative entropy is invariant under a measurable bijection:

$$D_{\mathrm{KL}}((\mathcal{T}_{\psi_i})_\# P \| (\mathcal{T}_{\psi_i})_\# R) = D_{\mathrm{KL}}(P \| R).$$

Applying this equality to the bridge variational problem proves that the optimizer is $(\mathcal{T}_{\psi_i})_\# P^*$. The chain rule gives the stated transformations of rates and quadratic variation. Time reversal gives the gaining case. □

B NUMERICAL IMPLEMENTATION

Algorithm 3 expands the training procedure of Section 5; settings for each dataset are in Appendix C.3.1.

Algorithm 3 Expanded simulation-free training procedure

Require: finite snapshots $\{(t_i, \mu_i)\}_{i=0}^{K}$ and interval parameters $\{\nu_i, \kappa_i\}$

Construct the target bridge

for each adjacent interval i **do**
  orient the two snapshots from larger mass to smaller mass
  build the endpoint reference for killed Brownian motion using Equation 8
  solve the endpoint coupling in Equation 6
  transpose the coupling if the interval grows chronologically

Learn motion and exit

repeat
  sample an interval, normalized time, and live bridge component
  draw its state, target drift, and target exit rate from the conditional laws in Section 4.1
  update the drift and exit-rate networks with Equation 10
until the learning budget is exhausted

Learn entry locations

for each growing interval **do**
  fit the entry flow to live endpoints paired with the cemetery using Equation 38

Ensure: drift network, exit-rate network, entry flows, and interval parameters

B.1 ENDPOINT MEASURES AND COUPLING CONSTRUCTION

Each sampled cell carries the same weight, and snapshot mass is its cell count relative to the initial count. Snapshots larger than a fixed cap (2,048 to 4,096 cells, depending on the dataset) are subsampled uniformly with a fixed seed, only to construct the endpoint plan; each snapshot is subsampled once and reused by both adjacent intervals, and every reported metric uses the full prediction and truth clouds. Cemetery slack lifts all endpoint measures to the common mass $\max_i m_i$. The cemetery-slack argument in Appendix A.5 shows that this leaves the active plan unchanged.

B.2 NEURAL FIELDS AND SIMULATION-FREE TRAINING

All non-Gaussian models use local time within each interval, separate networks for drift and exit rate, and joint optimization against the analytic (b_t, h_t) targets in Section 4.1. The learned b_θ and h_θ define the SDE with the reference diffusion coefficient. We sample times away from interval boundaries to keep the analytic targets numerically stable.

B.3 NUMERICAL GENERATION AND POPULATION EVENTS

Generation divides interval rates by Δ_i and integrates equation 16 with explicit stochastic midpoint. For evaluation with lower variance, particle weights follow the exponential survival probabilities implied by the learned exit rate. Exact discrete exit times can also be sampled from that rate and define the same law of killed paths and distribution of exit times. On growing intervals, midpoint quadrature integrates the entry source, with anchors from the auxiliary flow. The reference fixes the source variance; the difference in endpoint masses fixes the total event mass.

C EXPERIMENTAL SUPPLEMENT

C.1 EVALUATION METRICS

State geometry and population mass. At time t, write the predicted and observed finite measures as

$$\widehat{\mu}_t = \sum_{i=1}^{m_t} w_i(t)\delta_{x_i(t)}, \qquad \mu_t = \sum_{j=1}^{n_t} \delta_{y_j}.$$

After separately normalizing their masses, we report exact empirical

$$W_1(t) = \min_{\pi \in \Pi(a,b)} \sum_{ij} \pi_{ij} \|x_i(t) - y_j\|_2, \quad a_i = \frac{w_i(t)}{\sum_k w_k(t)}, \quad b_j = \frac{1}{n_t}. \tag{49}$$

POT's network simplex solves this finite problem. Every reported W_1 uses all predicted and observed points, without resampling.

Mass is evaluated directly from the full finite measures:

$$\widehat{r}_t = \widehat{m}_t/\widehat{m}_0, \qquad r_t = m_t/m_0, \qquad \mathrm{RME}(t) = |\widehat{r}_t - r_t|/r_t. \tag{50}$$

Thus W_1 measures normalized state geometry and RME measures total mass.

Joint evaluation of state and mass for Q1. For the ground-truth benchmark we also report two measures of error over time. The integrated mass error is $\int_0^1 |\widehat{m}_t - m_t| \, \mathrm{d}t$. To score mass and geometry together, let D be the diameter of the active grid and equip the grid, extended with a cemetery state, with

$$d_\dagger(x, y) = \|x - y\|_2/D, \qquad d_\dagger(x, \dagger) = 1.$$

For any $L \geq \max\{\widehat{m}_t, m_t\}$, define $\widehat{\mu}_t^L = \widehat{\mu}_t + (L - \widehat{m}_t)\delta_\dagger$ and $\mu_t^L = \mu_t + (L - m_t)\delta_\dagger$. The joint score is

$$\mathrm{FM}\, W_1(t) = \frac{W_{1,d_\dagger}(\widehat{\mu}_t^L, \mu_t^L)}{\overline{M}}, \qquad \overline{M} = \max\{m_0, m_1\}, \tag{51}$$

where $\overline{M}$ is fixed for each ground-truth case and is independent of time and the evaluated method. The value is independent of L: additional common cemetery slack couples to itself at zero cost. Hence $\mathrm{FM}\, W_1$ penalizes misplaced active mass and incorrect active mass in one exact transport problem.

C.2 Q1: RECOVERY AGAINST INDEPENDENT GROUND TRUTH

C.2.1 GROUND-TRUTH CONSTRUCTION AND EVALUATION

Independent reference solver. We construct a killed nearest-neighbor diffusion on 81-point 1D grids and 17×17 2D grids over $[-4, 4]^d$, append an absorbing cemetery state, and evaluate its transition kernels by dense float64 scaling-and-squaring matrix exponentials. Log-domain IPF then solves the augmented endpoint Schrödinger system. This solver shares no implementation with ELYSIUM's sampler of conditional laws. A second uniformization calculation checks the transition kernel, and every generated case passes the stochasticity, semigroup, IPF, and endpoint checks.

Task coverage and model selection. Each method receives the same two weighted endpoint clouds, their exact grid histograms, duration, ν, and κ. The remaining 39 of 41 ground-truth times are used only for evaluation. The benchmark contains 70 cases spanning dimension, unimodal/multimodal/crossing geometry, mass ratio, diffusion, and killing rate. ELYSIUM, USB, and WFR-FM cover all 70 with five independent training seeds $101, \ldots, 105$. Because UDSB-F is simulation-based and far more expensive to train, it uses five seeds on an 18-case subset containing all 14 primary mass-ratio cases and four crossing tails, and one seed (101) on the remaining 52 cases. Each method uses one configuration selected on a disjoint eight-case development set with mass ratios $1/3$ and 3; no benchmark interior measure participates in training or selection.

Evaluation. The primary metric is conditional W_1: at each hidden time, predicted and ground-truth active measures are separately normalized and compared by the exact 1D CDF formula or POT network simplex in 2D, then averaged across the 39 times. RME of the active mass, integrated mass error, and FM W_1 provide complementary finite-measure checks. For methods that generate status transitions, event-time error is the W_1 distance between exit or entry CDFs, and event-location error is grid W_1 between normalized time-integrated exit or entry measures.

C.2.2 Recovery of state distributions and population mass

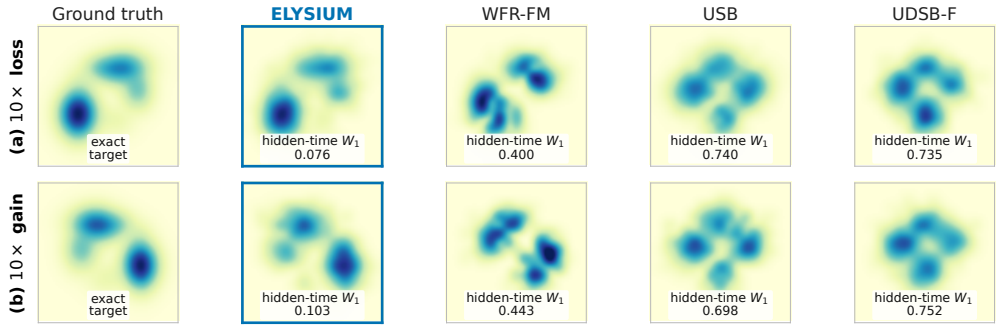


Figure 3: Hidden-distribution comparison under ten-fold population loss (top) and gain (bottom). Panels show normalized distributions of the living population at the midpoint; labels give state W_1 averaged over 39 hidden times.

Table 4: All 70 grid cases with killing, summarized by factor. "Wins" compares ELYSIUM with the lower conditional W_1 of USB and WFR-FM in each paired case. The last column is the paired mean relative reduction. Factor groups overlap by construction.

Factor	Value	Cases	Wins	Mean reduction
Dimension	1D	35	35/35	78.7%
	2D	35	35/35	71.2%
Geometry	Unimodal	6	6/6	79.3%
	Multimodal	58	58/58	75.2%
	Crossing	6	6/6	68.4%
Diffusion ν	0.05	6	6/6	73.4%
	0.25	58	58/58	74.8%
	0.8	6	6/6	78.5%
Event clock κ	0.1	8	8/8	77.4%
	0.5	8	8/8	76.8%
	1	38	38/38	74.9%
	2	8	8/8	70.0%
	5	8	8/8	76.1%

Table 4 shows that the 70/70 result is not confined to one geometry, dimension, diffusivity, or event rate. The overall paired mean and median reductions are 75.0% and 77.4%; the smallest casewise reduction is 28.3%. Table 5 shows 70/70 paired wins on all four metrics, including active mass, and Table 6 gives every value in the primary mass-imbalance scan.

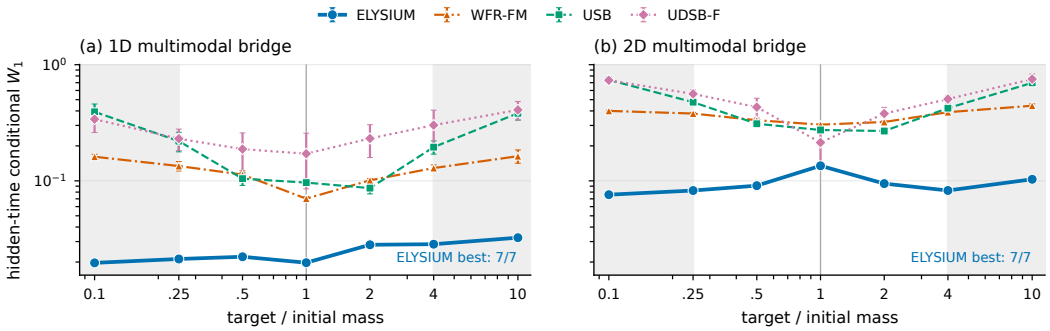


Figure 4: Primary comparison across mass ratios. Conditional W_1 is averaged over 39 hidden interior times; bars are 95% intervals across five fits. Gray bands mark strong loss and gain. UDSB-F bars use its 18-case subset with five seeds, which contains all 14 primary cases.

Table 5: Complementary scores over all 70 cases. Each case first averages five fits; "better baseline" selects the lower of USB and WFR-FM independently for each case and metric. Integrated mass error is $\int_0^1 |\widehat{m}_t - m_t| \, \mathrm{d}t$.

Metric	ELYSIUM	Better baseline	Wins	Mean reduction
Conditional active W_1	**0.0624**	0.2622	70/70	75.0%
Active-mass RME	**0.00260**	0.09257	70/70	93.8%
Integrated mass error	**0.000869**	0.139410	70/70	94.2%
Augmented finite-measure W_1	**0.00450**	0.04919	70/70	87.4%

Table 6: Primary comparison against ground truth: conditional W_1 at hidden times (mean $\pm$ sample SD over five independently trained models). All four methods use the same two endpoint observations and no ground truth at interior times. Lower is better.

Grid	m_1/m_0	ELYSIUM	WFR-FM	USB	UDSB-F
1D	0.1	**0.0197±0.0012**	0.1613±0.0092	0.3930±0.0750	0.3396±0.0900
1D	0.25	**0.0212±0.0018**	0.1341±0.0143	0.2198±0.0481	0.2306±0.0553
1D	0.5	**0.0222±0.0018**	0.1132±0.0098	0.1044±0.0148	0.1876±0.0812
1D	1	**0.0197±0.0010**	0.0702±0.0034	0.0966±0.0065	0.1712±0.0981
1D	2	**0.0282±0.0012**	0.1010±0.0044	0.0866±0.0107	0.2315±0.0830
1D	4	**0.0285±0.0015**	0.1288±0.0096	0.1952±0.0291	0.3016±0.1193
1D	10	**0.0324±0.0005**	0.1631±0.0245	0.3820±0.0553	0.4081±0.0851
2D	0.1	**0.0760±0.0020**	0.4001±0.0062	0.7399±0.0344	0.7345±0.0142
2D	0.25	**0.0825±0.0013**	0.3798±0.0097	0.4759±0.0204	0.5606±0.0539
2D	0.5	**0.0909±0.0012**	0.3327±0.0053	0.3101±0.0086	0.4312±0.0949
2D	1	**0.1348±0.0004**	0.3047±0.0051	0.2741±0.0106	0.2138±0.0807
2D	2	**0.0947±0.0014**	0.3212±0.0055	0.2686±0.0094	0.3793±0.0566
2D	4	**0.0826±0.0015**	0.3900±0.0107	0.4225±0.0137	0.5057±0.0442
2D	10	**0.1031±0.0016**	0.4433±0.0195	0.6980±0.0265	0.7518±0.0907

C.2.3 EVENT RECOVERY AND SIMULATION-BASED COMPARISON

Table 7 reports event recovery on the five-seed subset. With one paired seed on all 70 cases, conditional W_1 at hidden times is 0.062 for ELYSIUM and 0.415 for UDSB-F, and RME of active mass is 0.0025 and 0.213. Across the 60 cases with nonzero exchange, event-time W_1 is 0.0009 versus 0.126, and event-location W_1 is 0.064 versus 1.252.

Table 7: Event recovery on the 16 UDSB-F cases with five seeds and nonzero loss or gain. Entries are mean (median) across paired cases; each case value first averages five independent training seeds.

Metric	Cases	ELYSIUM	UDSB-F	ELYSIUM wins
Event-time CDF W_1	16	$\mathbf{0.0009}$ ($\mathbf{0.0003}$)	0.1210 (0.0958)	16/16
Event-location W_1	16	$\mathbf{0.0603}$ ($\mathbf{0.0415}$)	1.1784 (1.1721)	16/16

C.3 Q2: Population matching and held-out time points

C.3.1 Datasets and shared training settings

Table 8 lists every snapshot used in the WFR-FM comparison. Each setting uses five independently trained models. When a fit has multiple inference draws, they are averaged within that fit, so the independently trained model remains the statistical unit. Gaussian-1000D has only two endpoints and therefore no interpolation holdout.

Table 8: Dataset coverage. Counts follow observation time. "Holdouts" is the number of distinct intermediate snapshots omitted in turn; every applicable split is repeated with five neural training seeds.

Dataset	d	Snapshots	Cells per snapshot	Holdouts
Simulation Gene	2	5	400/442/530/690/969	3
Dyngen	5	5	156/112/63/96/301	3
Gaussian-1000D	1,000	2	500/1,400	0
EMT	10	4	577/885/788/883	2
CITE-seq	50	4	7,476/6,999/9,511/7,254	2
EB	50	5	2,381/4,163/3,278/3,665/3,332	3
Mouse	50	3	4,638/14,985/29,679	1

Shared neural training. Table 9 gives the settings for drift and exit rate. The neural fields take interval identity and local time as inputs and regress the forward drift b^+ and exit rate in a single training stage. The non-Gaussian models use separate networks for the drift and the exit rate; Gaussian-1000D uses one network with two output heads. All networks receive a 16-dimensional interval embedding, and the exit-rate term in equation 10 has weight 0.1. A gaining interval also uses the entry flow in Section 4.2.

Table 9: Field settings for each dataset. "Snaps" includes the initial snapshot. Every update draws the stated batch from each retained adjacent interval. The same row is used for population matching and every holdout experiment of that dataset.

Dataset	d	Snaps	Width	Blocks	Updates	Batch
Simulation Gene	2	5	512	6	6,000	400
Dyngen	5	5	512	6	6,000	400
Gaussian-1000D	1,000	2	512	5	6,000	400
EMT	10	4	512	6	6,000	400
CITE-seq	50	4	512	6	6,000	1,024
EB	50	5	512	6	6,000	400
Mouse	50	3	512	6	6,000	1,024

The entry flow uses five blocks. Its default width is 256; Gaussian-1000D uses width 512. The default fit uses 3,000 updates, batch size 128 per gaining interval, and learning rate 10^{-3}; Mouse uses 6,000 updates and batch size 512. The drift and exit rate stay fixed while the entry flow is trained.

Reference parameters and clocks. The two clock families follow Appendix A.10.

Table 10: Fixed reference and event settings for each dataset. "Sine (h, a)" denotes the positive global reference clock with harmonic h and amplitude a; identity has amplitude zero.

Dataset	ν	κ	Reference clock	Event clock
Simulation Gene	.005	2	identity	constant
Dyngen	.001	2	identity	constant
Gaussian-1000D	.001	2	identity	constant
EMT	.005	9	sine $(2, .13)$	constant
EB	.1	8	sine $(2, .5)$	constant
CITE-seq	.01	2	sine $(2, .5)$	constant
Mouse	.002	2	identity	exponential tilt $\beta = -2$

The sine clock is a deterministic positive time change of the killed Brownian reference, and the Mouse event clock is the $c_\beta(r)$ family in equation 46. On the full observed range the sine clock is

$$\psi_{a,h}(u) = u + \frac{a}{h\pi}\sin(h\pi u), \qquad \psi'_{a,h}(u) = 1 + a\cos(h\pi u) > 0,$$

and each retained interval uses the normalized restriction of this single global clock, so the same (h, a) applies when an interior snapshot is omitted. Training seeds are $101, \ldots, 105$ for every dataset, and evaluation uses a fixed inference seed.

Baseline implementations and uncertainty. Population-matching baselines are the published WFR-FM and USB results in raw benchmark coordinates with exact full-support W_1. SF-EUOT (EOT) and UDSB-F are evaluated with five independently trained models each. Because the SF-EUOT release has no implementation for vector data on this benchmark, we use an independent implementation of its EOT specialization with the original paper's fixed settings for vector data; it preserves probability mass and is evaluated only on state distributions. UDSB-F uses the released solver and its fixed training schedule. Holdout WFR-FM and USB results come from their official implementations under the same evaluator. Every method generates from the initial population only and passes its generated population to the next interval, with no reset to an intermediate observation.

C.3.2 Population-matching protocol and results

The model is fitted to every observed population. Table 11 summarizes the dataset means; Tables 12 and 13 compare the methods at every observed time after the initial snapshot.

ELYSIUM has lower W_1 than both WFR-FM and USB on 17 of 21 snapshots and the best mean across observed times among all reported methods on six of seven datasets.

Table 11: Mean population-matching W_1 for all methods. ELYSIUM, SF-EUOT (EOT), and UDSB-F are evaluated over five seeds in this work; WFR-FM and USB values are taken from their papers. The middle block is reported by WFR-FM (Peng et al., 2026); WFR-MFM is reported in its paper (Wang et al., 2026), and BranchSBM in USB (Ying et al., 2026). A dash means that the source paper did not report the dataset.

Method	Sim. Gene	Dyngen	Gaussian	EMT	EB	CITE-seq	Mouse
ELYSIUM	**0.0159**	0.2023	**1.956**	**0.1260**	**7.748**	**23.475**	**5.162**
SF-EUOT (EOT)	0.256	0.802	343.012	0.2622	10.206	34.691	7.281
UDSB-F (adjacent)	0.180	2.302	3.944	0.2430	13.321	64.100	14.438
WFR-FM	0.0190	0.1350	2.233	0.2239	8.855	30.031	5.848
USB	0.0190	**0.1310**	2.136	0.2100	9.087	31.183	6.069
MMFM	0.2980	1.3710	2.833	0.2851	10.525	38.182	8.902
Metric FM	0.3110	1.7670	3.794	0.2875	10.241	30.048	9.619
SF^2M	0.2240	1.2770	3.543	0.2759	10.983	32.821	9.652
MIOFlow	0.1480	0.9650	2.858	0.2648	9.309	29.681	6.530
TIGON	0.0450	0.5120	2.263	0.2647	9.415	29.654	6.556
DeepRUOT	0.0430	0.6230	3.785	0.3129	9.082	29.701	6.404
Var-RUOT	0.0790	0.5220	2.813	0.2631	10.092	34.511	9.406
UOT-FM	0.0930	1.2040	2.771	0.2668	11.030	38.692	8.642
VGFM	0.0460	0.5980	3.010	0.2407	8.890	30.974	6.535
WFR-MFM	0.0210	0.1760	2.233	–	–	–	–
BranchSBM	0.4740	1.4150	3.438	0.3313	11.652	36.266	8.597

Table 12: All constrained population marginals. ELYSIUM entries are mean ± sample SD over five independent neural fits. RME is the corresponding mean relative error in total mass. Bold marks the lower per-time W_1.

Dataset	Time	WFR-FM W_1	USB W_1	ELYSIUM W_1	ELYSIUM RME
Simulation Gene	t_1	0.0210	0.0200	**0.0194±0.0004**	0.0000±0.0000
	t_2	0.0210	0.0210	**0.0178±0.0009**	0.0000±0.0000
	t_3	0.0180	0.0190	**0.0141±0.0006**	0.0000±0.0000
	t_4	0.0160	0.0180	**0.0122±0.0004**	0.0000±0.0000
Dyngen	t_1	0.1100	**0.1090**	0.1468±0.0035	0.0130±0.0032
	t_2	0.0980	**0.0930**	0.2171±0.0450	0.0211±0.0149
	t_3	0.2110	**0.1800**	0.2902±0.0473	0.0139±0.0098
	t_4	**0.1210**	0.1420	0.1549±0.0353	0.0044±0.0031
Gaussian-1000D	t_1	2.233	2.136±0.002	**1.956±0.004**	0.0000±0.0000
EMT	t_1	0.2099	0.1831	**0.1048±0.0009**	0.0001±0.0000
	t_2	0.2272	0.2159	**0.1304±0.0011**	0.0055±0.0013
	t_3	0.2346	0.2309	**0.1429±0.0010**	0.0050±0.0012
EB	t_1	7.664	7.992	**6.684±0.023**	0.0001±0.0000
	t_2	8.659	8.973	**7.813±0.015**	0.0389±0.0028
	t_3	9.182	9.421	**7.931±0.025**	0.0347±0.0025
	t_4	9.914	9.963	**8.563±0.024**	0.0579±0.0021
CITE-seq	t_1	27.831	27.021	**19.431±0.063**	0.0028±0.0012
	t_2	27.478	28.081	**23.351±0.031**	0.0020±0.0009
	t_3	34.784	38.447	**27.644±0.133**	0.0095±0.0026
Mouse	t_1	5.486	5.589	**4.956±0.008**	0.0001±0.0000
	t_2	6.211	6.548	**5.367±0.008**	0.0001±0.0000

Table 13: Per-time results for SF-EUOT (EOT) and UDSB-F. Entries are mean ± sample SD over five independently trained models.

Dataset	Time	SF-EUOT (EOT)	UDSB-F
Simulation Gene	t_1	0.1458 ± 0.0213	0.1374 ± 0.0725
	t_2	0.2332 ± 0.0272	0.1809 ± 0.0936
	t_3	0.3075 ± 0.0280	0.1830 ± 0.1008
	t_4	0.3358 ± 0.0284	0.2178 ± 0.1299
Dyngen	t_1	0.4131 ± 0.0351	1.3876 ± 0.7039
	t_2	0.9058 ± 0.0693	2.7853 ± 0.7431
	t_3	1.0262 ± 0.2272	2.6359 ± 0.6244
	t_4	0.8609 ± 0.0859	2.3976 ± 1.3991
Gaussian-1000D	t_1	343.012 ± 396.696	3.9443 ± 0.0276
EMT	t_1	0.2546 ± 0.0035	0.2409 ± 0.0008
	t_2	0.2662 ± 0.0030	0.2447 ± 0.0025
	t_3	0.2657 ± 0.0047	0.2435 ± 0.0026
EB	t_1	8.7615 ± 0.0981	11.0920 ± 0.0885
	t_2	10.1150 ± 0.1768	13.2449 ± 0.0945
	t_3	10.6532 ± 0.1467	14.2770 ± 0.1482
	t_4	11.2949 ± 0.1463	14.6694 ± 0.1519
CITE-seq	t_1	30.8851 ± 0.4193	54.6681 ± 0.1860
	t_2	33.2748 ± 0.8227	60.0619 ± 0.6127
	t_3	39.9133 ± 0.7316	77.5686 ± 0.4889
Mouse	t_1	6.6974 ± 0.0765	12.5680 ± 0.0132
	t_2	7.8646 ± 0.0845	16.3085 ± 0.2433

C.3.3 Protocol and results for omitted time points

Every interior time listed in Table 8 is omitted in turn. The endpoint plan, drift, exit rate, and entry flow are learned using only the retained snapshots, and the generator propagates the initial population continuously to the omitted time. Standardization statistics are fitted only to retained snapshots and inverted before equation 49. The omitted population's states and count are used only for evaluation.

Holdout refits use the configuration in Appendix C.3.1; an entry flow is fitted only when a retained interval gains mass. Inference uses 100 stochastic midpoint steps per interval and a fixed inference seed.

Table 14: The eight individual real-data interpolation results; dataset averages appear in Table 15. Each omitted snapshot is predicted by a model trained on the remaining times. Entries are mean ± sample SD over five fits.

Dataset	Omit	WFR-FM W_1	USB W_1	UDSB-F W_1	Elysium W_1
EMT	1	$\mathbf{0.2864 \pm 0.0004}$	0.2876 ± 0.0004	0.3511 ± 0.0075	0.2866 ± 0.0001
	2	0.3119 ± 0.0006	0.3172 ± 0.0005	0.3692 ± 0.0132	$\mathbf{0.3098 \pm 0.0002}$
EB	1	10.051 ± 0.011	10.159 ± 0.029	13.201 ± 0.195	$\mathbf{9.726 \pm 0.007}$
	2	10.083 ± 0.019	10.195 ± 0.023	14.434 ± 0.158	$\mathbf{9.818 \pm 0.011}$
	3	10.391 ± 0.028	10.561 ± 0.076	15.019 ± 0.086	$\mathbf{9.765 \pm 0.014}$
CITE-seq	1	37.995 ± 0.066	38.123 ± 0.014	60.567 ± 0.865	$\mathbf{37.729 \pm 0.147}$
	2	$\mathbf{35.731 \pm 0.203}$	36.495 ± 2.261	67.135 ± 0.977	37.001 ± 0.083
Mouse	1	6.577 ± 0.023	7.020 ± 0.025	13.665 ± 0.227	$\mathbf{6.575 \pm 0.017}$

Table 15: Real-data holdout W_1, averaged over omitted snapshots. ELYSIUM, WFR-FM, USB, and UDSB-F were run for five seeds. The middle block is reported in WFR-FM (Peng et al., 2026); WFR-MFM is reported in its Appendix Table 3 (Wang et al., 2026), and BranchSBM in USB's Main Table 4 (Ying et al., 2026). Lower is better.

Method	EMT	EB	CITE-seq	Mouse
ELYSIUM	**0.2982**	**9.770**	37.365	**6.575**
UDSB-F	0.3601	14.218	63.851	13.665
WFR-FM	0.2991	10.175	36.863	6.577
USB	0.3024	10.305	37.309	7.020
MMFM	0.3230	11.213	38.521	8.263
Metric FM	0.3140	10.726	37.342	7.753
SF^2M	0.3080	10.986	38.333	8.646
MIOFlow	0.3250	10.960	39.574	7.779
TIGON	0.3600	11.080	38.159	6.868
DeepRUOT	0.3230	10.075	37.892	6.847
Var-RUOT	0.3200	11.035	38.393	8.672
UOT-FM	0.3220	11.344	38.649	9.332
VGFM	0.3010	10.370	37.386	8.496
WFR-MFM	0.2990	10.135	**35.736**	6.714
BranchSBM	0.3690	11.988	39.125	8.586

ELYSIUM improves on WFR-FM at six of the eight omitted snapshots and on USB at seven; it also improves on UDSB-F at all eight. Averaged within each dataset, it leads both principal baselines on EMT, EB, and Mouse, while WFR-FM leads on CITE-seq, and Table 15 extends this comparison to the other reported methods.

C.4 Q3: 4I DRUG-RESPONSE DYNAMICS

Data and evaluation split. The drugs are administered at time zero. The 4i Trametinib+Erlotinib condition then has $2452/3322/1841$ cells at 8/24/48h and 50 morphology and protein intensity features (Bunne et al., 2023; Pariset et al., 2023). The public 80/20 split contains $1962/2684/1479$ training cells and $490/638/362$ evaluation cells. We retain this split and the original 50-dimensional coordinates for every method. The training portion of all three observations and the corresponding cell counts are available during model construction; the evaluation cells are not used for fitting. All rollouts start from the 8h population and continue through 24h to 48h without an observation reset.

Model configuration and continuous generation. The ELYSIUM network has width 512, six residual blocks, eight Fourier frequencies, and a 16-dimensional interval embedding. It uses the same single-stage parameterization as the other non-Gaussian experiments: one joint fit of the forward drift and exit rate over 6,000 updates, with separate drift and exit-rate networks. The 8–24h growth interval also trains the standard entry flow described in Section 4.2, using width 256, five residual blocks, and 3,000 updates; the 24–48h interval uses exit. The reference uses $\nu = .01$ and $\kappa = 2$. The five training seeds are $23, 31, 47, 59, 71$. Each fit starts from a paired 512-cell 8h training cloud and uses one continuous SDE rollout with the reference diffusion coefficient and 100 stochastic midpoint steps per interval. The support is not capped after initialization, allowing newly generated entries to remain in the population.

Paired baseline protocol. For WFR-FM and USB, we train their released model implementations on all three observed snapshots using fixed 50-dimensional configurations. Each method is then propagated from 8h to 48h. To give UDSB-F the same three observed populations, the baseline fitted to all snapshots trains the released solver separately on the adjacent 8–24h and 24–48h intervals, retaining its architecture, training schedule, and the dependence of its event prior on state. During generation, the second interval starts from the exact terminal particle positions and active/dead indicators produced by the first. All four methods use the same five paired 512-cell initial clouds and rollout seeds, and are scored with the same exact W_1 evaluator in the original 50 coordinates.

Results. Table 16 gives the paired comparison at both evaluation times.

Table 16: 4i Trametinib+Erlotinib population matching in the original 50 features on the same held-back cells. Values are mean and sample SD over five independently trained models.

Method	24h $W_1 \downarrow$	48h $W_1 \downarrow$
WFR-FM	$3.409 \pm .054$	$3.531 \pm .026$
USB	$3.462 \pm .049$	$3.631 \pm .076$
UDSB-F	$5.400 \pm .041$	$5.520 \pm .105$
ELYSIUM SDE	$\mathbf{3.318} \pm .056$	$\mathbf{3.357} \pm .091$

C.5 SENSITIVITY TO BRIDGE VARIANCE AND KILLING RATE

We assess sensitivity on the Simulation Gene holdouts over a finite grid of Brownian bridge variance $\nu \in \{0.001, 0.005, 0.02\}$ and constant killing rate $\kappa \in \{0.5, 2, 8\}$. Every pair is refitted for all three omitted interior snapshots with training seeds $101, \ldots, 105$; the architecture, optimizer, observed masses, and direct parameterization of the forward drift remain fixed. All settings use the identity reference clock and an event rate that is constant in time. This auxiliary study uses 400 stochastic midpoint steps and three inference seeds per trained model. Inference results are first averaged for each model, after which we report the mean and sample SD over the five independent fits. The main benchmark tables retain their fixed 100-step protocol.

Table 17: Sensitivity to reference parameters on Simulation Gene at the three held-out interior times. W_1 entries are mean $\pm$ sample SD over five independently trained models; RME entries are the corresponding mean relative mass errors, whose SD across fits is zero. Bold hyperparameters mark the benchmark setting.

Hyperparameters		t_1		t_2		t_3	
ν	κ	W_1	RME	W_1	RME	W_1	RME
0.001	0.5	0.1769 ± 0.0007	0.0337	0.1775 ± 0.0016	0.0388	0.1081 ± 0.0008	0.0467
0.001	2	0.1421 ± 0.0006	0.0159	0.1724 ± 0.0010	0.0402	0.1315 ± 0.0008	0.0608
0.001	8	0.1490 ± 0.0005	0.0897	0.2631 ± 0.0007	0.1576	0.2854 ± 0.0012	0.2204
0.005	0.5	0.1758 ± 0.0002	0.0337	0.1759 ± 0.0009	0.0388	0.1075 ± 0.0007	0.0467
0.005	**2**	0.1414 ± 0.0002	0.0159	0.1720 ± 0.0006	0.0402	0.1325 ± 0.0009	0.0608
0.005	8	0.1497 ± 0.0002	0.0897	0.2631 ± 0.0007	0.1576	0.2844 ± 0.0011	0.2204
0.02	0.5	0.1801 ± 0.0002	0.0337	0.1786 ± 0.0004	0.0388	0.1135 ± 0.0005	0.0467
0.02	2	0.1454 ± 0.0005	0.0159	0.1760 ± 0.0005	0.0402	0.1370 ± 0.0010	0.0608
0.02	8	0.1527 ± 0.0004	0.0897	0.2651 ± 0.0009	0.1576	0.2824 ± 0.0012	0.2204

Across this finite grid, changing ν at fixed κ has a comparatively modest effect on holdout W_1, whereas $\kappa = 8$ substantially degrades both state and mass interpolation. The benchmark pair $(\nu, \kappa) = (0.005, 2)$ has the nominally lowest holdout macro W_1, but its difference from $(0.001, 2)$ is smaller than the SD across training seeds. At fixed κ, RME is unchanged across ν, consistent with equation 30: ν controls the spatial reference, whereas κ controls event timing within each interval. No single pair is best at every omitted time.

C.6 COMPUTATIONAL COST

Table 18: Training time and peak GPU memory for the holdout experiments. Training times report mean±sample SD across fits; n is the number of UDSB-F runs, with no SD for the single Mouse run. Peak memory for ELYSIUM and WFR-FM is reported as mean/maximum allocated GiB.

	ELYSIUM		WFR-FM		UDSB-F	
Dataset	Train (min)	Peak GiB	Train (min)	Peak GiB	Train (min)	n
Simulation Gene	1.59 ± 0.03	$0.27/0.27$	0.61 ± 0.02	$4.66/5.02$	6.33 ± 0.03	4
Dyngen	1.55 ± 0.03	$0.26/0.26$	0.54 ± 0.02	$0.20/0.22$	6.34 ± 0.04	4
EMT	1.41 ± 0.04	$0.23/0.23$	0.60 ± 0.02	$5.54/5.86$	4.26 ± 0.04	3
EB	2.12 ± 0.10	$0.73/0.81$	1.32 ± 0.07	$0.16/0.17$	6.30 ± 0.03	4
CITE-seq	1.88 ± 0.04	$0.89/0.89$	1.65 ± 0.13	$0.21/0.24$	4.20 ± 0.02	3
Mouse	1.55 ± 0.02	$0.50/0.50$	1.95 ± 0.10	$0.80/0.80$	2.13	1

Hardware and timing. ELYSIUM and WFR-FM use an RTX 4090; their times include endpoint construction and all training stages. UDSB-F uses an RTX 4090 D; its times cover training only, excluding initialization.

ELYSIUM averages 1.41–2.12 minutes per fit across the six datasets; UDSB-F means for the training loop range from 2.13 to 6.34 minutes. ELYSIUM remains comparable to WFR-FM in training time, with maximum allocated GPU memory of 0.89 GiB versus 5.86 GiB for WFR-FM.

C.7 SUPPLEMENTARY NEURAL SDE TRAJECTORIES

Figures 5–8 complement the population-matching results with one representative model per real dataset. Each rollout starts from the empirical initial cloud and uses the learned drift with the reference diffusion coefficient. State panels show the first two benchmark coordinates without additional PCA or display transforms; generation and W_1 evaluation use all benchmark coordinates.

Hollow and filled circles denote observed and generated populations. Solid paths start from the initial cloud; dashed paths are independently generated entries. Colors indicate time. Each panel shows paths of 48 initial cells and up to 24 entering cells. Paths are model samples, not observed lineages.

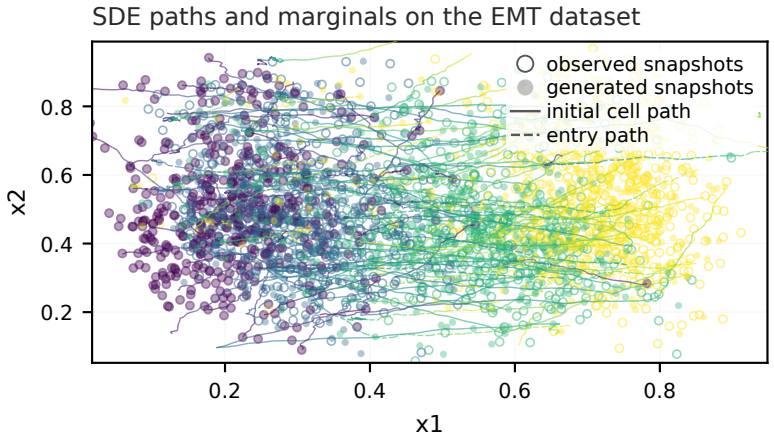


Figure 5: EMT: individual SDE paths with observed and generated snapshot populations.

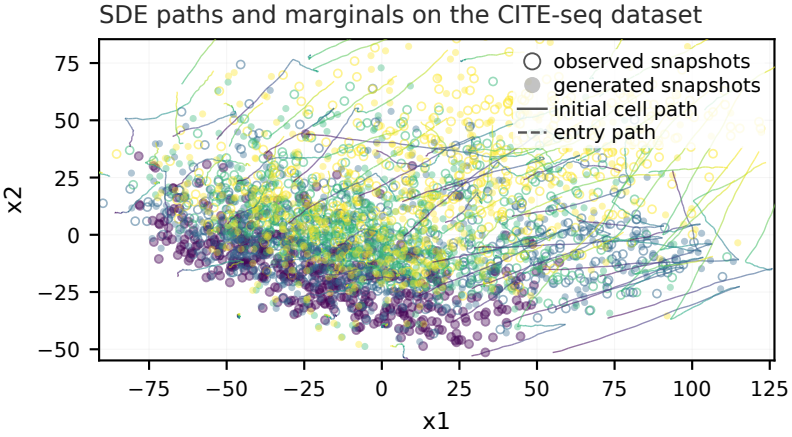


Figure 6: CITE-seq: individual SDE paths with observed and generated snapshot populations.

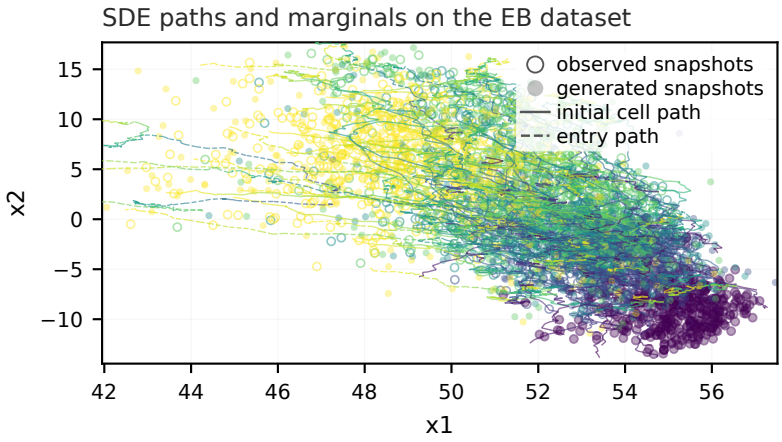


Figure 7: EB: individual SDE paths with observed and generated snapshot populations.

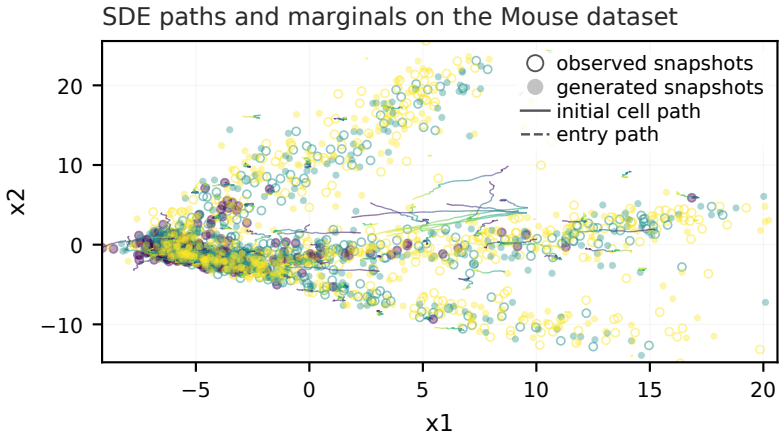


Figure 8: Mouse: individual SDE paths with observed and generated snapshot populations.

D RELATIONSHIP TO MODELS OF GROWTH AND INDIVIDUAL EVENTS

D.1 OBJECTIVES AND GENERATIVE MECHANISMS

Table 19: Comparison of closely related neural formulations. Models with explicit events generate stochastic exits, entries, or branching. The final column indicates whether training targets the objective in the second column at population optimum; USB presents its solver as an approximation to the branching SB (Ying et al., 2026).

Method	Objective	Stochastic paths	Explicit events	Simulation-free training	Targets stated objective
SF-EUOT	entropic UOT control	✓	×	✓	✓
WFR-FM	WFR dynamic action	×	×	✓	✓
UDSB	killed-diffusion path KL	✓	✓	×	✓
USB	branching path KL	✓	✓	✓	×
ELYSIUM	killed-diffusion path KL	✓	✓	✓	✓

D.2 IDENTIFIABILITY AND THE ROLE OF THE REFERENCE

WFR and other models of growth describe changing populations through a signed reaction field in the continuity equation (Chizat et al., 2018; Liero et al., 2018; Peng et al., 2026). This field identifies the net local change in population density. To see the resulting ambiguity, write a population equation with per-capita birth and death rates β and δ:

$$\partial_t \rho + \nabla \cdot (\rho v) = (\beta - \delta)\rho.$$

The snapshots constrain only the difference $g = \beta - \delta$. For every nonnegative field $c(t, x)$, the pair $(\beta + c, \delta + c)$ produces the same equation and hence the same population marginals. A signed growth field thus gives a concise macroscopic description, but specifying individual events requires an additional stochastic model.

Branching SBs instead choose a branching process as the reference law (Baradat & Lavenant, 2025; Tang et al., 2026; Ying et al., 2026). Their events have a parent–offspring interpretation and naturally represent division. Here too, anonymous snapshots identify less than an experiment that tracks lineage: if $q_k(t, x)$ is the rate of producing k descendants, the population equation depends only on the first moment $\sum_k (k-1) q_k(t, x)$, so distinct offspring laws can yield the same population evolution. A branching reference selects among these possibilities through its prior; branch-specific models additionally require a partition of terminal states into branches. USB learns a simulation-free approximation to this branching objective and uses the inferred fields to simulate division and death (Ying et al., 2026).

ELYSIUM uses a killed diffusion with a cemetery state. Every active path either persists between observed states or exits at a stochastic time; reading the same bridge backward gives anonymous entry. The path-space entropy objective jointly determines the survivor coupling, state motion, and distributions of event times and entry locations. Consequently, the reaction terms in Equation 15 are realized by a specific stochastic event process, so the learned dynamics describe individual histories without requiring lineage observations.