Stochastic Path Integral Formalism of Causal Field Theory

Stefan Groha [1] **Ching-Hao Wang** [1] **Bernhard Schölkopf** [2] **Arash Mehrjou** [1 2]

Abstract

Virtual tissue models aim to predict the effect of localized interventions in a spatial context. Yet current generative counterfactual models focus largely on global metadata or image-level semantic attributes. Biological interventions, like targeted therapy or targeted radiation, perturb stochastic tissue dynamics in space and time, requiring a field-theoretic perspective rather than just high-dimensional variable sets. Modalities like spatial transcriptomics and serial H&E imaging provide only partial snapshots of the stochastic evolution. Such settings require path-level causal questions: what trajectory follows from a localized treatment, and what would have happened in the same patient under a different region, dose, or timing? We give a causal path-integral formulation for Itô stochastic PDEs with additive Gaussian noise by interpreting the Martin-Siggia-Rose-Janssen-de Dominicis and related Onsager-Machlup construction as a structural causal model on trajectories. The noise path is the exogenous variable, drift deformations are interventions, and doubled same-noise path integrals implement Pearl-style twin-world counterfactuals. We show that we recover Causal Field Theory (Mehrjou & Schölkopf, 2026) in the tree-level linear response limit. The construction translates field-theoretic tools into estimators for nonlinear treatment effects, fluctuation-induced feedback, and counterfactual path sampling. We evaluate the framework on known-mechanism systems and demonstrate finite interventions, same-noise counterfactuals, rare events, and higher-order treatment interactions. The resulting dictionary provides a path-space route to evaluate counterfactuals in spatiotemporal systems where sequential abduction-action-prediction estimators are unstable.

[1]GSK.ai [2]Max Planck Institute for Intelligent Systems. Correspondence to: Stefan Groha <stefan.m.groha@gsk.com>.

Accepted at the 2026 Workshop on Generative and Agentic AI for Biology (ICML 2026)

1. Introduction

Causal inference for stochastic spatiotemporal systems is naturally a path-level problem. This is particularly clear in tissue-level biology. Spatial transcriptomics measures molecular state across a tissue slice (Qin et al., 2023; Hunter et al., 2021; Wang et al., 2023a), serial H&E or whole-slide images record evolving morphology (Yuan, 2016), and targeted therapies such as focal radiation (Timmerman et al., 2010), intratumoral drug delivery (Melero et al., 2021), or region-specific immune modulation perturb only part of a heterogeneous tumor microenvironment (Junttila & de Sauvage, 2013). These are path-level causal questions: how a local intervention propagates, how timing or dose changes the trajectory, and what would have happened in the same patient under a different localized action. A localized intervention changes a mechanism, its effect propagates through space and time, and a unit-level counterfactual must hold the exogenous randomness fixed, consistent with the observed trajectory. Recent Causal Field Theory (CFT) work connected do-calculus to PDE-based systems and computed localized treatment effects with Green's functions in the linear-response regime (Mehrjou & Schölkopf, 2026). That result however did not address three questions that are central in stochastic systems and for causal inference: how to treat finite nonlinear interventions, how fluctuations change mean treatment effects, and how to condition counterfactuals on the same noise path that generated a patient's observations.

This paper studies those questions for Itô SPDEs with additive noise. The Martin-Siggia-Rose-Janssen-de Dominicis (MSRJD) construction (Martin et al., 1973; Janssen, 1976; De Dominicis, 1976; Tauber, 2014; Bonicelli et al., 2025) represents the diffusion path measure as a functional integral over a physical field X and a response field $\widetilde{X}$. After eliminating $\widetilde{X}$, the Onsager-Machlup functional gives an unnormalized density on trajectories (Onsager & Machlup, 1953; Graham, 1977). We use this representation as a structural causal model on path space, where the exogenous variable is the noise trajectory, and an intervention is a deformation of the drift. To map the construction to a causal model, in addition to the additive noise assumption, we assume Itô causality so that u_t is non-anticipating, as well as exogeneity of the noise prior. The additive noise assumption can in principle be relaxed, at the cost of state-dependent

vertices and log-det Jacobian in the action.

To obtain stochastic estimates the central object is the generating functional $Z[J, \widetilde{J}; u]$. Derivatives with respect to the observable source J generate associational statistics; finite changes in the drift argument u generate interventional distributions, response-field insertions generated by $\widetilde{J}$ give causal sensitivities, and a doubled path integral with shared noise implements Pearl's twin-world construction (Pearl, 2009; Balke & Pearl, 1994) without requiring sequential abduction-action-prediction steps. Table 1 summarizes the resulting dictionary.

This work builds on a large corpus of results in statistical field theory, stochastic optimal control, and causal inference. Continuous-time causal inference and dynamic SCMs already give causal semantics to stochastic or deterministic dynamics (Lok, 2008; Røysland, 2011; Hansen & Sokol, 2014; Mooij et al., 2013; Boeken & Mooij, 2024), MSRJD response fields, Onsager-Machlup actions, function-space MCMC, stochastic adjoints, and instantons are established computational tools (Cotter et al., 2013; Kappen, 2005; Theodorou et al., 2010; Freidlin & Wentzell, 2012; E et al., 2004; Heymann & Vanden-Eijnden, 2008). This paper adds an explicit unified view and a causal dictionary that specifies which source, contour, or approximation answers which Pearl-rung query, together with an energy-based view that lets us answer non-markov queries. Further related work appears in Appendix A.

Main contributions. In this paper, we develop a path-space framework for estimating interventional and counterfactual effects in stochastic spatio-temporal systems for localized interventions, particularly relevant for biological systems where interventions propagate under noise and uncertainty. We formalize the MSRJD/Onsager-Machlup construction as an SCM on trajectories with the noise path as the exogenous variable, drift edits as interventions, and a doubled action with shared noise as the twin-world counterfactual mechanism (§2). We further connect this formulation to path-space energy-based models for both interventional and counterfactual queries, enabling path-space sampling without explicit abduction. In known synthetic systems we show that (i) tree-level retarded Green's functions reproduce the CFT linear-response limit; (ii) one-loop and Gaussian-2PI/Hartree closures show when fluctuations shift the ATE; (iii) function-space MCMC on the Onsager-Machlup action samples finite-intervention and conditional counterfactual distributions beyond linear response; and (iv) in the additive-Gaussian low-noise/MAP regime, the MSRJD saddle/instanton corresponds to methods from stochastic optimal control and identifies the structurally-constrained shared noise and the stochastic adjoint.

2. The MSRJD generating functional

Let $X_a(x,t)$ be a vector field on space-time, with a indexing the n scalar components (e.g. different cell types, concentrations of species, or channels of a multi-modal observable), evolving under

$$\partial_t X_a(x,t) = F_a[X](x,t) + u_a[X](x,t) + \eta_a(x,t), \quad (1)$$

where F_a is the mechanism, u_a is a soft mechanism edit (intervention) defined below, and η_a is Gaussian noise with covariance $\langle \eta_a(x,t)\eta_b(x',t')\rangle = N_{ab}(x,t;x',t')$. This is similar to (Mehrjou & Schölkopf, 2026), with the assumption of additive Gaussian noise aligning directly with the forward models used in most current multi-marginal and continuous-time generative frameworks (Song et al., 2020; Maddu et al., 2025; Guo et al., 2026). The MSRJD construction enforces (1) via a functional Dirac delta, exponentiated with an auxiliary response field $\widetilde{X}$. This requires integration over an imaginary contour to ensure convergence. We absorb the explicit factor of i into $\widetilde{X}$ to match standard notation (see Appendix B or references like Kamenev (2011); Zinn-Justin (2002) for details). Integrating over η yields

$$\begin{aligned} E_a[X;u] &= \partial_t X_a - F_a[X] - u_a, \\ \mathcal{J} &= \int \mathrm{d}t\, \mathrm{d}^d x\, (JX + \widetilde{J}\widetilde{X}), \\ Z &= \int \mathcal{D}X\, \mathcal{D}\widetilde{X}\, e^{-S_{\mathrm{MSRJD}}[X,\widetilde{X};u]+\mathcal{J}}, \\ S_{\mathrm{MSRJD}}[X,\widetilde{X};u] &= \int \mathrm{d}t\, \mathrm{d}^d x\, \widetilde{X}_a E_a[X;u] + S_{\mathrm{Jac}} \\ &\quad - \frac{1}{2}\int \mathrm{d}t\, \mathrm{d}^d x \widetilde{X}_a N_{ab} \widetilde{X}_b. \end{aligned} \quad (2)$$

with S_{Jac} a discretization-dependent Jacobian (causal and field-independent in the Itô convention). Sources J and $\widetilde{J}$ couple to X and $\widetilde{X}$, respectively; the intervention u couples through $-\widetilde{X}_a u_a$ and is a mechanism deformation from the base stochastic evolution.

The experiments focus on soft interventions $F \mapsto F + u$, where $u[X](x,t)$ covers any external source term, from focal radiation dose-rate deposition, local drug deposits, or general chemical gradients, as well as field dependent interventions (e.g. concentration-proportional cytotoxicity, cell-type dependent interventions, contact-dependent killing), but have to be functionally defined. Additionally, the same path-energy view also represents hard interventions. A clamp $\mathrm{do}(X_C = g)$ on a spacetime subset C replaces the structural equation on C by the assignment $X_C = g$ (Pearl, 2009; Boeken & Mooij, 2024). In the path integral this restricts support to paths satisfying the clamp. Biologically this could be placing specific cells at specific points in the tissue, or boundary constraints for drug infiltration. This is different from ordinary conditioning, where

the original dynamics still generate X_C, whereas under a clamp those degrees of freedom are fixed and downstream effects propagate through the remaining dynamics.

The connected generator is $W[J, \widetilde{J}; u] = \log Z[J, \widetilde{J}; u]$. Its functional derivatives in the three argument groups generate observable correlations, response/susceptibility objects, and interventional effects. Doubling the action with a shared noise integration generates twin-world counterfactuals. The full mapping is summarized in Table 1.

3. Ladder of Causation

Associational. Connected correlations of observables are direct J-derivatives of W, e.g. $\langle X_a(1) X_b(2)\rangle_c = \delta^2 W/\delta J_a(1)\delta J_b(2)$. They encode the joint distribution induced by (1) but say nothing about interventions.

Interventional. The post-intervention mean field is the J-derivative of the u-deformed generator, $\langle X_a(1)\rangle_u = \delta W[J, 0; u]/\delta J_a(1)|_{J=0}$. The infinitesimal interventional response is

$$\begin{aligned} R_{a\alpha}(1;2) &:= \left.\frac{\delta \langle X_a(1)\rangle_u}{\delta u_\alpha(2)}\right|_{u=0} \\ &= \left\langle X_a(1)\, \widetilde{X}_b(2)\, V_{b\alpha}[X](2)\right\rangle_c , \end{aligned} \tag{3}$$

where $V_{b\alpha}$ is the intervention vertex (the direction in mechanism space along which u_α deforms F). α here defines the intervention channel the intervention is acting on and b the component we are measuring the intervention on. Equation (3) is exact: no linearization in fluctuations has been made. For finite interventions one keeps u finite inside $W[\cdot, \cdot; u]$.

Counterfactual. Counterfactuals infer the exogenous variables compatible with the factual evidence, then keep those variables fixed while changing the structural mechanism and predicting the altered outcome (Pearl, 2009; Balke & Pearl, 1994). In an SDE or SPDE, the relevant exogenous variable is the entire structural noise history $\eta(t)$, so abduction means inferring a posterior over factual trajectories and shared structural residuals/noise. The path integral view is a departure from thinking about time evolution under stochastic noise and instead reframes it as a measure over stochastic paths. Instead of explicit abduction, we use the path-space analogue of the twin-network construction for structural causal models, where factual and counterfactual variables are represented by separate copies that share exogenous variables. This is also consistent with related counterfactual graph formalisms such as SWIGs (Richardson & Robins, 2013; Boeken & Mooij, 2024). Given our path space view we therefore double the action by drawing two field copies $X^{(0)}, X^{(1)}$ driven by the same η:

$$\begin{aligned} Z_{\text{twin}} = \int \mathcal{D}\eta\, P_\eta[\eta] \prod_{\sigma=0,1} \int \mathcal{D}X^{(\sigma)} \mathcal{D}\widetilde{X}^{(\sigma)} \\ \times\, e^{-S_\eta[X^{(0)}, \widetilde{X}^{(0)}; 0] - S_\eta[X^{(1)}, \widetilde{X}^{(1)}; u]}. \end{aligned} \tag{4}$$

Integrating out the shared noise produces off-diagonal couplings $-\frac{1}{2}\sum_{\sigma\tau} \widetilde{X}^{(\sigma)} N \widetilde{X}^{(\tau)}$ that enforce the shared-noise semantics. The unit-level counterfactual contrast at point 1 is $\Delta_a^{\text{cf}}(1; u) = \left\langle X_a^{(1)}(1) - X_a^{(0)}(1)\right\rangle_{\text{twin}}$. The path integral infers factual and counterfactual trajectories jointly rather than running Pearl's three steps sequentially. The two worlds are coupled through a term that enforces that their underlying structural residuals must match, $r_t[X^{(0)}; 0] = r_t[X^{(1)}; u]$ (§6.3). For rare events, the joint integral reduces to a shared instanton in the low-noise limit (§6.4).

4. Path-space energies as EBMs

For computation it is useful to pass from the MSRJD action to an Onsager-Machlup (OM) path energy (Onsager & Machlup, 1953; Graham, 1977). With additive Gaussian noise and a fixed Euler-Maruyama discretization the action becomes

$$\begin{aligned} S_\Delta[X; u] &:= \frac{1}{2\sigma^2 \Delta t} \sum_i \|r_i[X; u]\|^2, \\ r_i[X; u] &:= X_{i+1} - X_i - (F(X_i) + u_i)\Delta t, \end{aligned} \tag{5}$$

up to the Itô Jacobian and constants independent of X. Thus $p_u(X) \propto \exp[-S_\Delta[X; u]]$ is an energy-based model over trajectories (LeCun et al., 2006). Because the energy is the squared structural residual r_i, intervening on u overrides the mechanism rather than conditioning on an observation.

For interventions, MCMC on $S_\Delta[X; u]$ natively samples the post-intervention path law for a finite drift deformation $F \mapsto F + u$. For counterfactuals, we define the joint twin-world energy over the doubled trajectory space $(X^{(0)}, X^{(1)})$:

$$\begin{aligned} \bar{r}_i &:= \frac{1}{2}\Big(r_i[X^{(0)}; 0] + r_i[X^{(1)}; u]\Big), \\ S_{\text{twin}} &= \frac{1}{2\sigma^2 \Delta t} \sum_i \|\bar{r}_i\|^2 \\ &\quad + \gamma \|X^{(0)}_{\text{obs}} - X^{\text{obs}}\|^2 \\ &\quad + \frac{\lambda}{2\sigma^2 \Delta t} \sum_i \|r_i[X^{(0)}; 0] - r_i[X^{(1)}; u]\|^2. \end{aligned} \tag{6}$$

The γ term anchors the factual world to the patient's sparse observations. The first term is the single Gaussian prior on the common structural residual, while the stiff λ penalty forces the exogenous structural noise to strictly match between the two worlds, as in (4). We sample this joint energy

Table 1. The causal-field-theory dictionary. Each row pairs a standard object of statistical field theory with the causal-inference object it computes for the SPDE (1). Rows in the upper block are exact in the full theory; rows in the lower block are obtained by perturbative expansion or by saddle-point evaluation.

Field-theory object	**Causal-inference object**	**Experiment**
$\delta^2 W/\delta J\,\delta J = \langle XX\rangle_c$	Associational correlation (Rung 1)	—
$\delta W[J,0;u]/\delta J\|_{J=0} = \langle X\rangle_u$	Interventional mean / ATE (Rung 2)	§6.1
$\delta^2 W/\delta J\,\delta\widetilde{J} = \left\langle X\widetilde{X}\right\rangle_c$	Susceptibility / response function	§6.1
Doubled $(X^{(0)}, X^{(1)})$ + shared noise	Twin-world counterfactual (Rung 3)	§6.3, §6.4
$e^{-S_\Delta[X;u]}$	Path-space EBM / sampling target	§4, §6.1
Tree-level (deterministic) Green's function $G^{R,(0)}$	Linear response	§6.1, §6.2
Dressed propagator $(G^R)^{-1} = (G^{R,(0)})^{-1} - \Sigma$	Variance-corrected ATE	§6.2
Hamilton EoM $\dot{\widetilde{X}} = -\partial\mathcal{H}/\partial X$	Stochastic adjoint / causal sensitivity	§6.4, Appendix G
Instanton (saddle of S_{MSRJD})	MAP rare-event path/residual	§6.4
k-vertex insertions $\partial^k W/\partial u^k$	Higher-order causal interactions	Appendix G

with gradient-based MCMC. Additional soft constraints add energy terms. Hard clamps are support restrictions or infinite-penalty limits. Appendix C shows that known Girsanov and Feynman-Kac approaches (Del Moral, 2004; Donati & Keller, 2018; De Bortoli et al., 2021; Skreta et al., 2025) are action ratios and additive energy tilts of this same object.

5. Field-theoretic tools

5.1. Tree-level Green's function (linear response)

The linear response regime corresponds to evaluating expectations in the Gaussian or quadratic action. Linearizing the MSRJD action as

$$X = \langle X\rangle_0^{(0)} + \delta X, \qquad \widetilde{X} = \delta\widetilde{X}, \tag{7}$$

and keeping quadratic terms gives the retarded $X\widetilde{X}$ correlator $G^{R,(0)}_{ab}(1,2)$. We write expectations with a subscript for the intervention protocol and a superscript for the approximation used to evaluate the path integral. Thus $\langle\cdot\rangle_u$ denotes the exact interventional law, while $\langle\cdot\rangle_u^{(0)}$ denotes the tree-level/Gaussian theory. At this level noise corrections vanish and the deterministic saddle is $\langle X\rangle_0^{(0)}$. It satisfies the Dyson equation (Kamenev, 2011)

$$\begin{aligned}(LG^{R,(0)})_{ab}(1,2) &= \delta_{ab}\delta(1-2),\\ L_{ab}(1,2) &= \left(\delta_{ab}\partial_t - \frac{\delta F_a}{\delta X_b}[\langle X\rangle_0^{(0)}]\right)\delta(1-2).\end{aligned} \tag{8}$$

The corresponding tree-level response to an infinitesimal intervention channel V_α is

$$R^{(0)}_{a\alpha}(1;2) = G^{R,(0)}_{ab}(1,2)V_{b\alpha}[\langle X\rangle_0^{(0)}](2). \tag{9}$$

This recovers the linear-response kernel of the original CFT (Mehrjou & Schölkopf, 2026). It is exact in the joint limit $u \to 0$ and $\sigma \to 0$.

5.2. Fluctuation-corrected response through loop expansion

The tree-level linear response is a leading-order approximation that requires both the mechanism edit u and the fluctuation scale σ to be small. Causal queries require the interventional expectation $\langle X(t)\rangle_u$ under a finite protocol $u(t)$. In nonlinear systems, the mean does not follow the deterministic equation:

$$\partial_t\langle X(t)\rangle_u = \langle F(X(t))\rangle_u + u(t), \tag{10}$$

since in general

$$\langle F(X(t))\rangle_u \neq F(\langle X(t)\rangle_u) \tag{11}$$

The physical intuition is that because of drift curvature, fluctuations around the mean systematically shift the response. Usually, resolving this requires a Monte Carlo ensemble over many forward solves of the full SPDE, which becomes computationally prohibitive for high-dimensional spatial fields, boundary-value constraints, or sweeps over interventions.

Rather than sampling trajectories, standard stochastic field theory provides self-consistent PDE closures that can be solved just once to obtain first-order, mean-field like approximations of the state distribution (Kamenev, 2011; Tauber, 2014; Chow & Buice, 2015). This is the loop expansion of the MSRJD effective action. For each fixed $u(t)$, one expands around an approximation to the mean background $X = \langle X\rangle_u^{\mathcal{A}} + \delta X_{\mathcal{A}}$, generating a hierarchy of deterministic equations for the mean and covariance (Hochberg et al., 1999; Bode, 2022).

The one-loop (bare-covariance) closure incorporates the leading variance feedback:

$$\partial_t\langle X\rangle_u^{1\mathrm{L}} = F(\langle X\rangle_u^{1\mathrm{L}}) + u + \frac{1}{2}F''(\langle X\rangle_u^{1\mathrm{L}})\,C_u^{(0)}, \tag{12}$$

$$\dot{C}_u^{(0)} = 2F'(\langle X\rangle_u^{1\mathrm{L}})C_u^{(0)} + \sigma^2, \tag{13}$$

where $C_u^{(0)}(t,t)$ is the Gaussian covariance of fluctuations. The self-consistent Gaussian-2PI/Hartree closure refines this by using a fully self-consistent covariance $C_u^{\mathrm{H}}(t,t)$:

$$\partial_t \langle X\rangle_u^{\mathrm{H}} = F(\langle X\rangle_u^{\mathrm{H}}) + u + \frac{1}{2}F''(\langle X\rangle_u^{\mathrm{H}})\, C_u^{\mathrm{H}}, \tag{14}$$

$$\dot{C}_u^{\mathrm{H}} = 2F'(\langle X\rangle_u^{\mathrm{H}})C_u^{\mathrm{H}} + F'''(\langle X\rangle_u^{\mathrm{H}})(C_u^{\mathrm{H}})^2 + \sigma^2. \tag{15}$$

These generalized Lyapunov equations replace the stochastic ensemble. A single deterministic PDE solve returns the time-dependent approximation to $\langle X(t)\rangle_u$ and its spatial covariance structure. The Average Treatment Effect (ATE) is then simply read off as the difference in the integrated means, bypassing all sampling variance. Loop closures are accurate when the state marginal is near-Gaussian. Divergence between the two closures is a diagnostic that the path-space samplers of §4 are needed. See Appendix E for a derivation of these mean-field and effective-action closures, and Algorithm 1 lists their implementation.

5.3. Rare-event MAP abduction, optimal intervention, and the instanton

In biology we are often interested in rare transitions such as tumor nucleation, metastasis, or drug resistance, driven by atypical noise realizations crossing a barrier. For a specific patient who experienced such a rare event we want to know what targeted intervention would have prevented it.

For typical trajectories, stochastic forward simulation of the factual with abduction-action-prediction or the twin-world EBM (§4) handles these counterfactuals. However, for a rare event the forward simulation almost never reaches the rare factual and while the EBM can handle boundary value problems, it becomes stiff in this setting.

In the low-noise limit the transition is dominated by the saddle-point equations of the action. This is known as the instanton solution (Kamenev, 2011) in stochastic field theory or Freidlin-Wentzell theory in large-deviation theory (Freidlin & Wentzell, 2012; Dykman et al., 1992; E et al., 2004; Heymann & Vanden-Eijnden, 2008). The posterior over admissible trajectories then concentrates on a single dominant mode. This means we perform Maximum A Posteriori (MAP) abduction by minimizing the structural noise prior subject to the patient's factual boundary conditions (e.g., healthy at $t=0$, tumor at $t=T$):

$$\begin{aligned} (X^*,\eta^*) &= \arg\min_{X,\eta}\ \frac{1}{2}\int \eta^T N^{-1}\eta\,\mathrm{d}t, \\ \text{s.t.}\quad \eta(t) &= \dot{X}(t) - F(X(t)) - u[X,t]. \end{aligned} \tag{16}$$

In the MSRJD representation, this constrained minimization defines a Hamiltonian boundary-value problem, where the response field $\widetilde{X} = N^{-1}\eta$ acts as the conjugate momentum:

$$\begin{aligned} \mathcal{H}(X,\widetilde{X};u) &= \widetilde{X}^T(F(X)+u[X,t]) + \frac{1}{2}\widetilde{X}^T N\widetilde{X}, \\ \dot{X} &= F(X) + u[X,t] + N\widetilde{X}, \\ \dot{\widetilde{X}} &= -D_X(F+u)[X,t]^\dagger \widetilde{X}. \end{aligned} \tag{17}$$

For additive open-loop interventions $u(t)$, $D_X u = 0$, so the adjoint reduces to $\dot{\widetilde{X}} = -(D_X F)^\dagger \widetilde{X}$, the form used in the experiments below.

This single saddle-point solution $(X^*, \widetilde{X}^*)$ identifies the most likely exogenous noise $\eta^* = N\widetilde{X}^*$ driving the rare event. The first-order Laplace approximation around the instanton provides the Gaussian fluctuations around the instanton. By inspecting $\widetilde{X}^*$, we can identify the spatiotemporal regions where finite budget interventions have the highest impact. An optimal budget-constrained $u(t)$ aligns with this adjoint, maximally suppressing the rare-event probability (§6.4). The equations of motion for $\widetilde{X}$ (Equation (17)) exactly match the continuous-time stochastic adjoint used in optimal control theory (Pontryagin et al., 1962; Kappen, 2005; Kidger et al., 2021), which directly measures the sensitivity of the rare trajectory to changes in the drift.

6. Experiments

We use controlled stochastic systems with known drift and noise to test the computational claims in Table 1. All trajectories are simulated with Euler-Maruyama, ensuring the discretized Onsager-Machlup energy used by the samplers is the exact negative log-density of the simulated path measure up to constants. Implementation details, hyperparameters, and supplementary experiments are deferred to Appendix H.

6.1. Path-space sampling for finite interventions

We first show the EBM recovers the effect of finite, localized interventions. We compare this to the linear response result, as well as reweighting and resampling methods (App. C). We evaluate a soft intervention $u(x,t)$ applied to a central patch of a 32×32 bistable Allen-Cahn SPDE:

$$\partial_t X = D\nabla^2 X + rX(1-X)(X-a) + u + \sigma\eta. \tag{18}$$

The causal estimand is the interventional dose-response of the terminal spatial mean. We compare linear response, Girsanov importance sampling, and Feynman-Kac resampling, which are all using baseline trajectories against direct Metropolis Adjusted Langevin Algorithm (MALA) sampling of the intervened Onsager-Machlup (OM) path energy.

As shown in Figure 1, linear response captures the local slope but misses the nucleation threshold. Girsanov weights collapse as the intervention accumulates over space and time,

and resampling delays but does not prevent the collapse. Direct path-space sampling tracks the forward Monte Carlo reference across the dose sweep. While forward simulation is trivial for this specific Rung-2 query, this establishes the OM energy as a computationally stable target for finite mechanism changes when non-Markov downstream constraints or counterfactual couplings make simple forward simulation impossible.

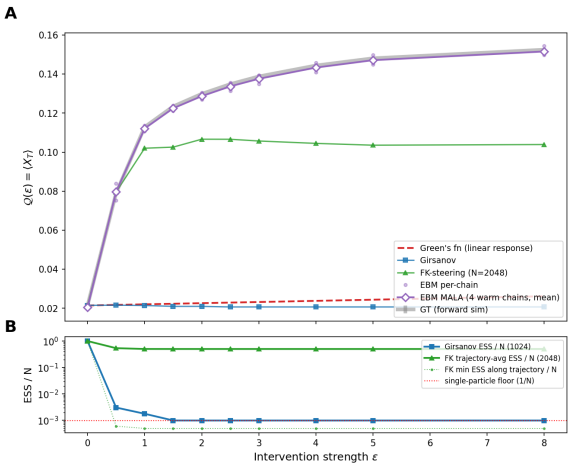


Figure 1. Finite-intervention dose response in a bistable spatial field. Linear response misses the nucleation threshold. Reweighting-based estimators lose effective sample size through the transition. Direct sampling of the intervened path energy agrees with the forward Monte Carlo reference.

6.2. Loop-corrected response captures fluctuation feedback

We test the loop expansion as a deterministic substitute for Monte Carlo when the goal is the variance-induced mean shift. We study a scalar bistable SDE initialized in the lower well and driven near its deterministic tipping threshold:

$$\mathrm{d}X_t = (X_t - X_t^3 + \varepsilon)\,\mathrm{d}t + \sigma \mathrm{d}W_t. \tag{19}$$

We compare the tree-level deterministic ODE, the one-loop tadpole closure, and the self-consistent Gaussian-2PI/Hartree closure against an $N = 10^4$ Monte Carlo reference.

Figure 2 shows three regimes. For small interventions ($\varepsilon \leq 0.20$), noise-induced fluctuations are converted by the drift's cubic curvature into a mean shift, causing the tree-level ODE to underestimate the response. Both loop closures close this gap. Through the tipping transition ($\varepsilon \in [0.25, 0.35]$), the marginal becomes highly non-Gaussian. The bare-covariance one-loop closure overshoots and the Hartree closure undershoots, bracketing the Monte Carlo response. Replacing N stochastic integrations with one self-consistent PDE solve gives an order-of-magnitude speedup, useful for screening interventions before running the full sampler.

6.3. Twin-world EBMs recover individualized counterfactuals

We test the twin-world energy on unit-level counterfactuals, where joint sampling replaces the explicit abduction step. We consider two systems: a 1D double-well SDE, $\mathrm{d}X_t = (X_t - X_t^3)\,\mathrm{d}t + \sigma \mathrm{d}W_t$, and a 32×32 spatial bistable field following the same dynamics as §6.1. In both, we condition on sparse factual observations of a patient and apply a counterfactual intervention. Due to the sparse observations, we are interested in the full posterior counterfactual distribution. This is a challenging problem because conditioning on future observations turns the factual evolution into a non-Markovian boundary-value problem. We evaluate the population Average Treatment Effect (ATE), explicit noise-space pCN abduction with simulation replay, and our joint twin-world EBM.

In the 1D system (Figure 3), the individualized counterfactual is intrinsically bimodal: under the exact same observations, the patient may or may not cross the barrier. The ATE collapses this structure, returning an invalid population-average trajectory. The twin-world EBM naturally recovers the distinct conditional modes and approximates the importance-weighted oracle probabilities, performing on par with explicit noise-space pCN. Similarly, on the spatial field (Figure 9), the residual-matching construction scales robustly to the joint spatial trajectory. Observations strictly constrain the factual path, residual matching enforces the same underlying structural disturbance, and the counterfactual path uniquely responds to the altered mechanism.

6.4. Instantons locate rare-event counterfactual bottlenecks

Finally, we apply the low-noise instanton formulation to optimize interventions against rare, fluctuation-driven transitions (e.g., disease onset or drug resistance). We simulate a 1D Allen-Cahn SPDE with a spatially varying barrier $a(x)$ representing primary and secondary niches:

$$\partial_t X = D\partial_x^2 X + rX(1-X)(X - a(x)) + u(x,t) + \sigma\eta. \tag{20}$$

We condition on a rare terminal transition and minimize the OM action to find the instanton, yielding the MAP residual η^* driving the event.

As shown in Figure 4, the MAP residual localizes sharply at a spatiotemporal bottleneck (panel A) and the adjoint control concentrates on the same region (panel B). To probe how robust each intervention is to fluctuations of the inferred noise, we sample the conditional rare-event posterior via the first-order Laplace approximation around the instanton

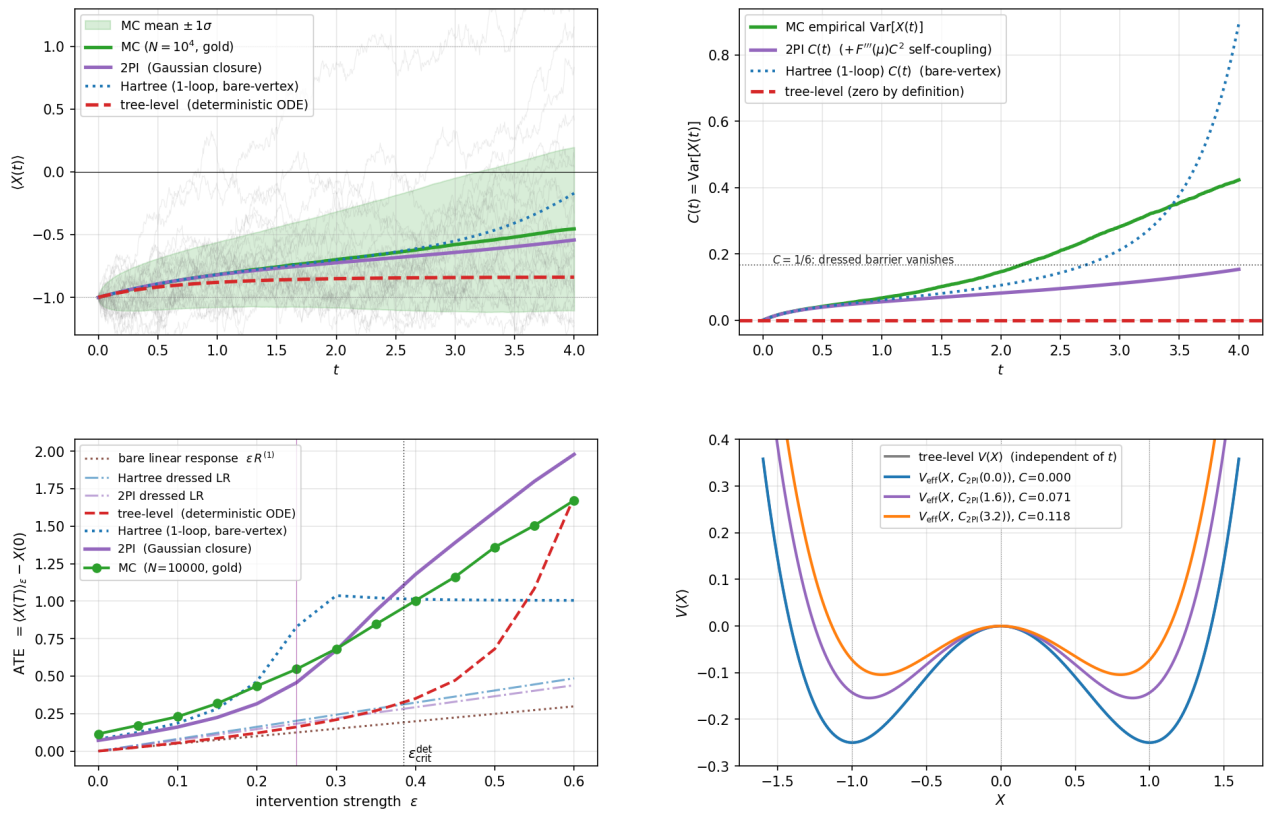


Figure 2. Loop-corrected response in a bistable SDE. The tree-level deterministic response misses noise-assisted crossing near the transition. One-loop and Gaussian-2PI/Hartree closures incorporate variance feedback into the mean equation and approximate the Monte Carlo response until the Gaussian assumption breaks down.

in the bridge parametrisation (App. H.4) and replay each posterior sample under both interventions. The factual posterior is a Brownian-bridge-like distribution pinned at both endpoints (gray band, panel C). Under the matched-budget uniform control the cf field leaves a residual hot spot at the bottleneck with visible posterior spread (blue band, panels C and D); under the adjoint control every one of the 2,000 posterior samples is driven below the cure threshold across the entire field ($\max \phi(x,T) \leq -10^{-4}$, terminal burden std $< 10^{-4}$). The MSRJD response field thus identifies both the most likely causal transition channel and an optimal intervention that is decoupled from the conditional rare-event posterior.

7. Discussion and future work

This paper establishes a formal equivalence between the stochastic path-integral formalism of statistical physics and Pearl's causal hierarchy for spatiotemporal systems and localized interventions for path- or snapshot-observed data. Treating the noise path of the MSRJD/OM construction as the SCM's exogenous variable assigns each field-theoretic object a causal interpretation. Tree-level Green's functions compute linear responses, loop expansions deterministically capture fluctuation-induced shifts in average treatment effects, doubled actions with residual matching sample unit-level counterfactuals, connecting the formalism to the twin-world view of counterfactuals, and instanton saddles identify the optimal spatiotemporal targets for disrupting rare disease transitions.

An obvious next step is to learn the action from data. We assumed a known drift to separate causal estimation from model error. Snapshot and probability-flow methods estimate drift and diffusion fields from multi-marginal data, identifiable under explicit structural and noise priors (Guan et al., 2024; 2025; Maddu et al., 2025; 2026). Two open problems are (i) extending these learners to expose a do-interface, and (ii) bounding how action-approximation error propagates to causal estimates. Multiplicative noise introduces state-dependent causal effects and is a natural extension (Wang et al., 2023b). The path-space formulation also connects to EBMs and amortized samplers. The twin-world residual-matching EBM poses a stiff boundary-value problem, but its formulation over the entire path measure natively accommodates non-Markovian queries like conditioning on future sparse observations. This makes it an ideal

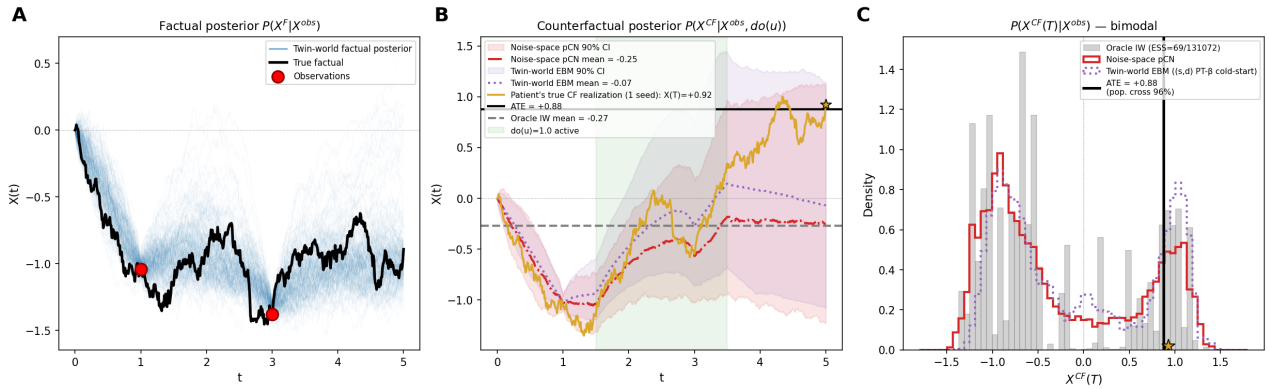


Figure 3. One-dimensional twin-world counterfactual. Conditioning on the patient's observations produces a bimodal counterfactual terminal distribution. The population ATE collapses this. Noise-space pCN matches the importance-weighted oracle. The joint twin-world EBM recovers the two conditional modes with approximate mode weights.

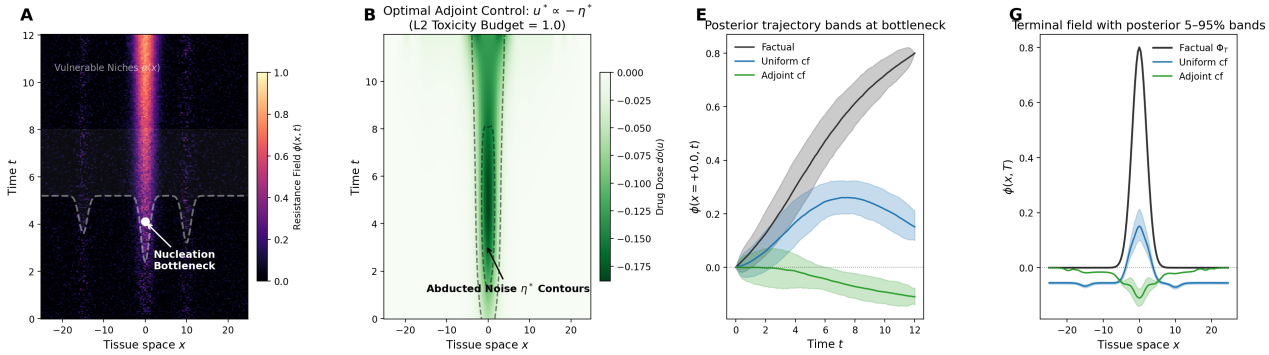


Figure 4. Rare-event counterfactual via the instanton, sampled from the first-order Laplace approximation of the conditional rare-event posterior. (A) Factual rare trajectory in the spatially heterogeneous resistance landscape, with the nucleation bottleneck marked. (B) Adjoint-optimal control with η^* contours overlaid; the control concentrates on the bottleneck. (C) Posterior trajectory bands at the bottleneck slice $x = 0$ for factual and counterfactual replays: the factual band (gray) shows the bridge geometry pinned at both endpoints, while the cf bands show how each intervention deflects the trajectory off the constraint manifold. (D) Spatial profile of $\phi(x, T)$ with posterior 5–95% bands: the adjoint cf (green) is below zero everywhere with a tight band, the uniform cf (blue) leaves a residual hot spot at $x = 0$ with visible posterior spread. The decoupling between the adjoint outcome and posterior fluctuations is the empirical signature of an intervention that is robust to the conditional rare-event posterior, not just to its mean. Posterior structure (per-cell std(η), scatter of cf burden vs. posterior energy, top-3 soft modes) is shown in Figure 11.

unnormalized target for training amortized samplers. Recent non-Markov bridge learners (Guo et al., 2026) show that score-matching can handle path-dependent conditioning. Extending these, or training Energy-Based GFlowNets (Bengio et al., 2021; Zhang et al., 2022; Malkin et al., 2022) for mode coverage on multimodal causal posteriors, would give amortized counterfactual generation.

Several extensions are natural. Cellular dynamics often involve discrete populations and demographic noise. The Doi-Peliti mapping (Tauber, 2014; 2005) would extend the dictionary to master-equation systems and absorbing-state interventions such as tumor extinction. Causal queries can also be projected onto collective coordinates such as defects and solitons (Rajaraman, 1982; Zinn-Justin, 2002), giving low-dimensional effective dynamics for macroscopic counterfactuals such as whether a therapy pins an advancing tumor boundary.

Software and Data

All code is implemented in jax, using blackjax for MC sampling. All data is synthetic and simulated by stochastic forward simulation of pre-defined SPDEs.

Impact Statement

This work develops theory for causal inference in spatiotemporal biological systems, with the goal of improving virtual-tissue models used in biomedical research and therapy design.

We do not foresee direct negative societal consequences from the methodology itself. As such models move toward

clinical use, the mechanisms and structural assumptions must be validated against experimental data before any intervention is acted on.

References

Albergo, M. S., Boffi, N. M., and Vanden-Eijnden, E. Stochastic interpolants: A unifying framework for flows and diffusions, 2023.

Archambeau, C., Opper, M., Shen, Y., Cornford, D., and Shawe-Taylor, J. Variational inference for diffusion processes. In *Advances in Neural Information Processing Systems 20 (NIPS 2007)*, volume 20, 2007.

Balgi, S., Peña, J. M., and Daoud, A. Counterfactually-equivalent structural causal modelling using causal graphical normalizing flows. In *Probabilistic Graphical Models (PGM)*, 2024.

Balke, A. and Pearl, J. Probabilistic evaluation of counterfactual queries. In *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI)*, volume 1, pp. 230–237, 1994.

Bengio, Y., Lahlou, S., Deleu, T., Hu, E. J., Tiwari, M., and Bengio, E. GFlowNet foundations, November 2021.

Berges, J. Controlled nonperturbative dynamics of quantum fields out of equilibrium. *Nuclear Physics A*, 699 (3-4):847–886, March 2002. ISSN 0375-9474. doi: 10.1016/s0375-9474(01)01295-7. URL http://dx.doi.org/10.1016/s0375-9474(01)01295-7.

Bode, T. The two-particle irreducible effective action for classical stochastic processes. *Journal of Physics A: Mathematical and Theoretical*, 55(26):265401, June 2022. ISSN 1751-8121. doi: 10.1088/1751-8121/ac73c6. URL http://dx.doi.org/10.1088/1751-8121/ac73c6.

Boeken, P. and Mooij, J. M. Dynamic structural causal models, July 2024.

Bonicelli, A., Dappiaggi, C., and Drago, N. An algebraic correspondence between stochastic differential equations and the martin–siggia–rose formalism. *Annales Henri Poincaré*, 27(4):1407–1449, April 2025. ISSN 1424-0661. doi: 10.1007/s00023-025-01571-1. URL http://dx.doi.org/10.1007/s00023-025-01571-1.

Chen, T., Liu, G.-H., and Theodorou, E. A. Likelihood training of Schrödinger bridge using forward-backward SDEs theory. In *International Conference on Learning Representations*, 2021.

Chow, C. C. and Buice, M. A. Path integral methods for stochastic differential equations. *The Journal of Mathematical Neuroscience*, 5(1), March 2015. ISSN 2190-8567. doi: 10.1186/s13408-015-0018-5. URL http://dx.doi.org/10.1186/s13408-015-0018-5.

Coleman, S. Fate of the false vacuum: Semiclassical theory. *Physical Review D*, 15(10):2929–2936, May 1977. ISSN 0556-2821. doi: 10.1103/physrevd.15.2929. URL http://dx.doi.org/10.1103/physrevd.15.2929.

Cornwall, J. M., Jackiw, R., and Tomboulis, E. Effective action for composite operators. *Physical Review D*, 10 (8):2428–2445, October 1974. ISSN 0556-2821. doi: 10.1103/physrevd.10.2428. URL http://dx.doi.org/10.1103/physrevd.10.2428.

Cotter, S. L., Roberts, G. O., Stuart, A. M., and White, D. MCMC methods for functions: Modifying old algorithms to make them faster. *Statistical Science*, 28(3):424–446, August 2013. ISSN 0883-4237. doi: 10.1214/13-STS421.

De Bortoli, V., Thornton, J., Heng, J., and Doucet, A. Diffusion Schrödinger bridge with applications to score-based generative modeling. In *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, volume 34, pp. 17695–17709, 2021.

De Dominicis, C. Techniques de renormalisation de la théorie des champs et dynamique des phénomènes critiques. *Le Journal de Physique Colloques*, 37(C1): C1–247–C1–253, January 1976. ISSN 0449-1947. doi: 10.1051/jphyscol:1976138.

De Sousa Ribeiro, F., Santhirasekaram, A., and Glocker, B. Counterfactual identifiability via dynamic optimal transport. *arXiv Preprint arXiv:2510.08294*, 2025.

Del Moral, P. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Probability and Its Applications. Springer, 2004. doi: 10.1007/978-1-4684-9393-1.

Donati, L. and Keller, B. G. Girsanov reweighting for metadynamics simulations. *The Journal of Chemical Physics*, 149(7), July 2018. ISSN 1089-7690. doi: 10.1063/1.5027728. URL http://dx.doi.org/10.1063/1.5027728.

Du, Y., Li, S., Tenenbaum, J., and Mordatch, I. Compositional visual generation with energy based models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6637–6647, 2020.

Dürr, D. and Bach, A. The onsager-machlup function as lagrangian for the most probable path of a diffusion process. *Communications in Mathematical Physics*, 60(2): 153–170, June 1978. ISSN 0010-3616, 1432-0916. doi: 10.1007/BF01609446.

Dykman, M. I., McClintock, P. V. E., Smelyanski, V. N., Stein, N. D., and Stocks, N. G. Optimal paths and the prehistory problem for large fluctuations in noise-driven systems. *Physical Review Letters*, 68(18):2718–2721, May 1992. ISSN 0031-9007. doi: 10.1103/physrevlett.68.2718. URL http://dx.doi.org/10.1103/physrevlett.68.2718.

E, W., Ren, W., and Vanden-Eijnden, E. Minimum action method for the study of rare events. *Communications on Pure and Applied Mathematics*, 57(5):637–656, May 2004. ISSN 0010-3640, 1097-0312. doi: 10.1002/cpa.20005.

Freidlin, M. I. and Wentzell, A. D. *Random Perturbations of Dynamical Systems*. Springer, 3 edition, 2012.

Grafke, T. and Vanden-Eijnden, E. Numerical computation of rare events via large deviation theory. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(6), June 2019. ISSN 1089-7682. doi: 10.1063/1.5084025. URL http://dx.doi.org/10.1063/1.5084025.

Graham, R. Path integral formulation of general diffusion processes. *Zeitschrift für Physik B Condensed Matter and Quanta*, 26(3):281–290, September 1977. ISSN 1434-6036. doi: 10.1007/bf01312935. URL http://dx.doi.org/10.1007/bf01312935.

Guan, V., Janssen, J., Rahmani, H., Warren, A., Zhang, S., Robeva, E., and Schiebinger, G. Identifying drift, diffusion, and causal structure from temporal snapshots. *arXiv Preprint arXiv:2410.22729*, 2024.

Guan, V., Janssen, J., Lanzetti, N., Terpin, A., Schiebinger, G., and Robeva, E. Gradient-flow SDEs have unique transient population dynamics, May 2025.

Guo, G., Cosgrove, C., Kapoor, S., Poli, M., Huang, C.-W., and Ermon, S. ABC: Any-subset autoregression via non-markovian diffusion bridges in continuous time and space. *arXiv preprint arXiv:2604.27443*, 2026.

Hairer, M., Stuart, A. M., and Voss, J. Analysis of SPDEs arising in path sampling part II: The nonlinear case. *The Annals of Applied Probability*, 17(5–6):1657–1706, 2007. doi: 10.1214/07-AAP441.

Hansen, N. and Sokol, A. Causal interpretation of stochastic differential equations. *Electronic Journal of Probability*, 19(none), January 2014. ISSN 1083-6489. doi: 10.1214/ejp.v19-2891. URL http://dx.doi.org/10.1214/ejp.v19-2891.

Heymann, M. and Vanden-Eijnden, E. The geometric minimum action method: A least action principle on the space of curves. *Communications on Pure and Applied Mathematics*, 61(8):1052–1117, August 2008. ISSN 0010-3640, 1097-0312. doi: 10.1002/cpa.20238.

Hochberg, D., Molina-París, C., Pérez-Mercader, J., and Visser, M. Effective action for stochastic partial differential equations. *Physical Review E*, 60(6):6343–6360, December 1999. ISSN 1063-651X, 1095-3787. doi: 10.1103/PhysRevE.60.6343.

Hunter, M. V., Moncada, R., Weiss, J. M., Yanai, I., and White, R. M. Spatially resolved transcriptomics reveals the architecture of the tumor-microenvironment interface. *Nature Communications*, 12(1):6278, 2021. doi: 10.1038/s41467-021-26614-z.

Janssen, H.-K. On a lagrangean for classical field dynamics and renormalization group calculations of dynamical critical properties. *Zeitschrift für Physik B Condensed Matter and Quanta*, 23(4):377–380, December 1976. ISSN 1434-6036. doi: 10.1007/bf01316547. URL http://dx.doi.org/10.1007/bf01316547.

Junttila, M. R. and de Sauvage, F. J. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature*, 501(7467):346–354, 2013. doi: 10.1038/nature12626.

Kamenev, A. *Field Theory of Non-Equilibrium Systems*. Cambridge University Press, September 2011. ISBN 9781139003667. doi: 10.1017/cbo9781139003667. URL http://dx.doi.org/10.1017/cbo9781139003667.

Kappen, H. J. Linear theory for control of nonlinear stochastic systems. *Physical Review Letters*, 95(20), November 2005. ISSN 1079-7114. doi: 10.1103/physrevlett.95.200201. URL http://dx.doi.org/10.1103/physrevlett.95.200201.

Kidger, P., Foster, J., Li, X., and Lyons, T. Efficient and accurate gradients for neural SDEs, 2021.

LeCun, Y., Chopra, S., Hadsell, R., Ranzato, M., and Huang, F. J. A tutorial on energy-based learning. In Bakir, G., Hofmann, T., Sch"olkopf, B., Smola, A. J., Taskar, B., and Vishwanathan, S. V. N. (eds.), *Predicting Structured Data*, pp. 191–246. MIT Press, 2006.

Li, X., Wong, T.-K. L., Chen, R. T. Q., and Duvenaud, D. Scalable gradients for stochastic differential equations, 2020.

Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., and Le, M. Flow matching for generative modeling, 2022.

Lok, J. J. Statistical modeling of causal effects in continuous time. *The Annals of Statistics*, 36(3), June 2008. ISSN 0090-5364. doi: 10.1214/009053607000000820. URL http://dx.doi.org/10.1214/009053607000000820.

Lorch, L., Krause, A., and Schölkopf, B. Causal modeling with stationary diffusions. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 1927–1935. PMLR, 2024. doi: 10.48550/ARXIV.2310.17405.

Maddu, S., Chardès, V., and Shelley, M. J. Learning stochastic processes with intrinsic noise from cross-sectional biological data. *Proceedings of the National Academy of Sciences*, 122(37), September 2025. doi: 10.1073/pnas.2420621122.

Maddu, S., Chardès, V., and Shelley, M. J. Learning biophysical models of gene regulation with probability flow matching, April 2026.

Malkin, N., Lahlou, S., Deleu, T., Ji, X., Hu, E., Everett, K., Zhang, D., and Bengio, Y. GFlowNets and variational inference. *arXiv Preprint arXiv:2210.00580*, 2022.

Martin, P. C., Siggia, E. D., and Rose, H. A. Statistical dynamics of classical systems. *Physical Review A*, 8 (1):423–437, July 1973. ISSN 0556-2791. doi: 10.1103/physreva.8.423. URL http://dx.doi.org/10.1103/physreva.8.423.

Mehrjou, A. and Schölkopf, B. Causal field theory: Causal semantics for PDE-based spatio-temporal systems. In *ICLR 2026 Workshop on Machine Learning for Genomics Explorations (MLGenX)*, March 2026. URL https://openreview.net/forum?id=9Urq3VymML.

Melero, I., Castanon, E., Alvarez, M., Champiat, S., and Marabelle, A. Intratumoural administration and tumour tissue targeting of cancer immunotherapies. *Nature Reviews Clinical Oncology*, 18(9):558–576, 2021. doi: 10.1038/s41571-021-00507-y.

Mooij, J. M., Janzing, D., and Schölkopf, B. From ordinary differential equations to structural causal models: The deterministic case, 2013.

Øksendal, B. *Stochastic Differential Equations: An Introduction with Applications*. Springer, 6 edition, 2003.

Onsager, L. and Machlup, S. Fluctuations and irreversible processes. *Physical Review*, 91(6):1505–1512, September 1953. ISSN 0031-899X. doi: 10.1103/physrev.91.1505. URL http://dx.doi.org/10.1103/physrev.91.1505.

Opper, M. Variational inference for stochastic differential equations. *Annalen der Physik*, 531(3):1800233, March 2019. ISSN 0003-3804, 1521-3889. doi: 10.1002/andp.201800233.

Pawlowski, N., Castro, D. C., and Glocker, B. Deep structural causal models for tractable counterfactual inference, October 2020.

Pearl, J. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2 edition, 2009. doi: 10.1017/CBO9780511803161.

Pontryagin, L. S., Boltyanskii, V. G., Gamkrelidze, R. V., and Mishchenko, E. F. *The Mathematical Theory of Optimal Processes*. Interscience, 1962.

Qin, P., Chen, H., Wang, Y., Huang, L., Huang, K., Xiao, G., Han, C., Hu, J., Lin, D., Wan, X., Zheng, Y., Liu, Y., Li, G., Yang, H., Ye, S., Luo, M., Fu, Y., Xu, H., Wen, L., Guo, Z., Shen, X., Li, Z., Wang, C., Chen, X., Wang, L., Sun, L., Ren, D., Wu, L., Wang, J., Liu, S., and Lin, H. Cancer-associated fibroblasts undergoing neoadjuvant chemotherapy suppress rectal cancer revealed by single-cell and spatial transcriptomics. *Cell Reports Medicine*, 4(10):101231, 2023. doi: 10.1016/j.xcrm.2023.101231.

Rajaraman, R. *Solitons and Instantons: An Introduction to Solitons and Instantons in Quantum Field Theory*. North-Holland, Amsterdam, 1982.

Rasal, V., Kori, A., De Sousa Ribeiro, F., Xia, Y., and Glocker, B. Diffusion counterfactual generation with semantic abduction. In *International Conference on Machine Learning (ICML)*, 2025.

Richardson, T. S. and Robins, J. M. Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. Working Paper 128, Center for Statistics and the Social Sciences, University of Washington, 2013. URL https://csss.uw.edu/Papers/wp128.pdf.

Røysland, K. A martingale approach to continuous-time marginal structural models. *Bernoulli*, 17(3):1–22, August 2011. ISSN 1350-7265. doi: 10.3150/10-BEJ303.

Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., Lee, L., Chen, J., Brumbaugh, J., Rigollet, P., Hochedlinger, K., Jaenisch, R., Regev, A., and Lander, E. S. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(6):1517, March 2019. ISSN 0092-8674. doi: 10.1016/j.cell.2019.02.026. URL http://dx.doi.org/10.1016/j.cell.2019.02.026.

Seifi, S., Ibrahimi, A., Sukianto, T., Carbonelli, C., Servadei, L., and Wille, R. GenFacts: Generative counterfactual explanations for multi-variate time series. *arXiv Preprint arXiv:2509.20936*, 2025.

Singhal, R., Horvitz, Z., Teehan, R., Ren, M., Yu, Z., McKeown, K., and Ranganath, R. A general framework for inference-time scaling and steering of diffusion models, 2025.

Skreta, M., Akhound-Sadegh, T., Ohanesian, V., Bondesan, R., Aspuru-Guzik, A., Doucet, A., Brekelmans, R., Tong, A., and Neklyudov, K. Feynman-kac correctors in diffusion: Annealing, guidance, and product of experts. *arXiv Preprint arXiv:2503.02819*, 2025.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations, 2020.

Tauber, U. C. Field theory approaches to nonequilibrium dynamics, 2005.

Tauber, U. C. *Critical Dynamics: A Field Theory Approach to Equilibrium and Non-Equilibrium Scaling Behavior*. Cambridge University Press, 2014. doi: 10.1017/CBO9781139046213.

Theodorou, E., Buchli, J., and Schaal, S. A generalized path integral control approach to reinforcement learning. *Journal of Machine Learning Research*, 11:3137–3181, 2010.

Timmerman, R., Paulus, R., Galvin, J., Michalski, J., Straube, W., Bradley, J., Fakiris, A., Bezjak, A., Videtic, G., Johnstone, D., Fowler, J., Gore, E., and Choy, H. Stereotactic body radiation therapy for inoperable early stage lung cancer. *JAMA*, 303(11):1070–1076, 2010. doi: 10.1001/jama.2010.261.

Todorov, E. *Linearly-solvable Markov decision problems*, pp. 1369–1376. The MIT Press, September 2007. ISBN 9780262256919. doi: 10.7551/mitpress/7503.003.0176. URL http://dx.doi.org/10.7551/mitpress/7503.003.0176.

Vinyard, M. E., Rasmussen, A. W., Li, R., Klein, A. M., Getz, G., and Pinello, L. scDiffEq: Drift-diffusion modeling of single-cell dynamics with neural stochastic differential equations, December 2023.

Wang, X. Q., Danenberg, E., Huang, C.-S., Egle, D., Callari, M., Bermejo, B., Dugo, M., Zamagni, C., Thill, M., Anton, A., Zambelli, S., Russo, S., Ciruelos, E. M., Greil, R., Gyrffy, B., Semiglazov, V., Colleoni, M., Kelly, C. M., Mariani, G., Del Mastro, L., Biasi, O., Seitz, R. S., Valagussa, P., Viale, G., Gianni, L., Bianchini, G., and Ali, H. R. Spatial predictors of immunotherapy response in triple-negative breast cancer. *Nature*, 621(7980):868–876, 2023a. doi: 10.1038/s41586-023-06498-3.

Wang, Y., Geng, X., Huang, W., Huang, B., and Gong, M. Generator identification for linear SDEs with additive and multiplicative noise, 2023b.

Wu, D., Qiu, F., and Xie, Y. DoFlow: Flow-based generative models for interventional and counterfactual forecasting on time series. *arXiv Preprint arXiv:2511.02137*, 2025a.

Wu, L., Trippe, B. L., Naesseth, C. A., Blei, D. M., and Cunningham, J. P. Practical and asymptotically exact conditional sampling in diffusion models. In *Advances in Neural Information Processing Systems*, 2023.

Wu, R. and YongJun, L. Causal schrödinger bridges: Constrained optimal transport on structural manifolds, 2026.

Wu, Y. Flow-based generative modeling of potential outcomes and counterfactuals. *arXiv Preprint arXiv:2505.16051*, 2025.

Wu, Y., McConnell, L., and Iriondo, C. Counterfactual generative modeling with variational causal inference. In *International Conference on Learning Representations (ICLR)*, 2025b.

Yeo, G. H. T., Saksena, S. D., and Gifford, D. K. Generative modeling of single-cell time series with PRESCIENT enables prediction of cell trajectories with interventions. *Nature Communications*, 12(1):3222, May 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-23518-w.

Yuan, Y. Spatial heterogeneity in the tumor microenvironment. *Cold Spring Harbor Perspectives in Medicine*, 6 (8):a026583, 2016. doi: 10.1101/cshperspect.a026583.

Zhang, D., Malkin, N., Liu, Z., Volokhova, A., Courville, A., and Bengio, Y. Generative flow networks for discrete probabilistic modeling. In *International Conference on Machine Learning*, pp. 26412–26428. PMLR, 2022.

Zhu, S., Wu, S., Zhou, W., and Chen, M. Counterfactual generative models for time-varying treatments. *arXiv Preprint arXiv:2305.15742*, 2023.

Zinn-Justin, J. *Quantum Field Theory and Critical Phenomena*. Oxford University Press, June 2002. ISBN 9780198509233. doi: 10.1093/acprof:oso/9780198509233.001.0001. URL http://dx.doi.org/10.1093/acprof:oso/9780198509233.001.0001.

A. Related work

Causal dynamics and causal field theory. Classical causal inference distinguishes associational, interventional, and counterfactual queries (Pearl, 2009; Balke & Pearl, 1994). Continuous-time causal inference has developed related martingale, SDE, and optimal-control viewpoints (Lok, 2008; Røysland, 2011; Hansen & Sokol, 2014), and recent work studies structural causal models for dynamical systems (Mooij et al., 2013; Boeken & Mooij, 2024). Closest to this paper, Mehrjou & Schölkopf (2026) introduced Causal Field Theory for PDE-based spatiotemporal systems and derived localized interventional effects through Green's functions. We recover that linear-response result as the tree-level

row of Table 1. The remaining rows cover finite interventions, noise-conditioned counterfactuals, loop corrections, rare-event optimization, and higher-order interactions.

MSRJD, Onsager–Machlup, and path-space inference. The MSRJD formalism and related path-integral representations of diffusion processes are classical tools in nonequilibrium statistical physics and stochastic analysis (Martin et al., 1973; Janssen, 1976; De Dominicis, 1976; Tauber, 2014; Onsager & Machlup, 1953; Graham, 1977; Dürr & Bach, 1978; Bonicelli et al., 2025). The Onsager-Machlup action and function-space MCMC methods provide practical path-space inference machinery for diffusions and SPDEs (Hairer et al., 2007; Cotter et al., 2013; Archambeau et al., 2007; Opper, 2019). It is known that these actions compose observations and boundary constraints. We specify how the same machinery represents Pearl queries once the exogenous noise and intervention class are fixed: treating the noise path as a shared structural constraint turns a path-density calculation into a counterfactual one.

Stochastic control, adjoints, and large deviations. Path-integral control, stochastic optimal control, and neural-SDE adjoints provide algorithms for sensitivities and optimal interventions in stochastic dynamics (Todorov, 2007; Kappen, 2005; Theodorou et al., 2010; Li et al., 2020; Kidger et al., 2021). Large-deviation theory and minimum-action methods identify instantons as least-action paths for rare events (Freidlin & Wentzell, 2012; E et al., 2004; Heymann & Vanden-Eijnden, 2008). We attach causal semantics to these tools: in the additive-Gaussian low-noise limit, the optimal-control adjoint, the instanton momentum, and the MAP shared noise solve the same saddle equations.

Learning stochastic dynamics from snapshots. A complementary literature learns stochastic dynamics from cross-sectional or multi-marginal observations. Snapshot SDE-identifiability work characterizes when drift, diffusion, and causal structure can be recovered under structural assumptions (Guan et al., 2024; 2025). Probability-flow inference and matching learn velocity fields that share the observed time-dependent marginals with an underlying diffusion, using explicit noise priors to recover biophysical force fields in gene-regulatory systems (Maddu et al., 2025; 2026). These methods address an upstream problem relative to this paper: learning a candidate (F, σ) or path energy. Once such a model exposes an intervention interface, the MSRJD/OM dictionary applies. Without it, matching observational marginals does not yield do-interventions or same-noise counterfactuals.

Counterfactual generative modeling and learning. Current causal generative models implement Pearl's abduction-action-prediction sequence using normalizing flows, diffusion models, and flow matching (Lipman et al., 2022; Albergo et al., 2023) to infer exogenous noise and sample counterfactuals (Pawlowski et al., 2020; Zhu et al., 2023; Balgi et al., 2024; Wu et al., 2025b; Wu, 2025; De Sousa Ribeiro et al., 2025; Rasal et al., 2025). These methods have recently been extended to multivariate time series and longitudinal data, allowing generation of temporal counterfactual trajectories via continuous normalizing flows (Wu et al., 2025a; Wu, 2025) or conditional diffusion (Seifi et al., 2025), but do not allow for localized interventions. Concurrently, Schrödinger bridges and Feynman-Kac models provide tools for conditional path sampling (Song et al., 2020; De Bortoli et al., 2021; Chen et al., 2021; Wu & YongJun, 2026; Wu et al., 2023; Singhal et al., 2025; Skreta et al., 2025; Guo et al., 2026), and GFlowNets can learn unnormalized path distributions (Du et al., 2020; Bengio et al., 2021; Zhang et al., 2022; Malkin et al., 2022).

Existing image counterfactual methods primarily intervene on global metadata (e.g., age, sex, scanner site) or spatially agnostic semantic attributes. In a parallel direction, recent work models interventions and causality in high-dimensional SDEs for single-cell dynamics (Schiebinger et al., 2019; Lorch et al., 2024; Yeo et al., 2021; Vinyard et al., 2023), but focuses on variable-level knockouts or stationary distribution shifts. Virtual tissue models require localized interventions on SPDEs, where a mechanism intervention occupies a region of physical space and propagates.

Most flow-based counterfactual models run in algorithmic (e.g. diffusion) time, treating temporal data as a fixed vector or generating it autoregressively. Multi-marginal conditioning on irregular intermediate observations then needs specialized objectives or guidance heuristics.

The MSRJD path integral runs in the physical time of the SDE. Intermediate observations enter as source/boundary terms in the action, so multi-snapshot conditioning needs no change of model class. With drift and noise geometry separated, interventions enter through the $\widetilde{X}u$ vertex, which is what makes the analytical approximations of §5 (linear response, loop corrections, MAP paths) available without a learned decoder. The caveat is that the mechanism, or its surrogate action, must itself be causally valid.

B. MSRJD construction

For completeness we record the steps from (1) to (2) that are otherwise standard. Rigorously, the path integral construction requires analytic continuation and integration of the response field along the imaginary axis to ensure convergence of the functional Fourier transform. We absorb the explicit factor of i into $\widetilde{X}$ to match the standard conventions of stochastic optimal control. For comprehensive textbook treatments detailing these imaginary contours, discretization

Jacobians, and functional convergence, see e.g., Kamenev (2011) and Tauber (2014), or Bonicelli et al. (2025) for a recent algebraic formalization.

The functional delta enforcing (1) is exponentiated via the auxiliary response field $\widetilde{X}$,

$$\delta[E-\eta] \propto \int \mathcal{D}\widetilde{X}\, \exp\Big(-\int \widetilde{X}_a(E_a-\eta_a)\Big),$$

and the noise is integrated against its path measure $P_\eta[\eta]$, producing the cumulant-generating functional

$$W_\eta[\widetilde{X}] := \log \big\langle e^{\int \widetilde{X}_a \eta_a}\big\rangle_{P_\eta},$$

which equals $\frac{1}{2}\int \widetilde{X}_a N_{ab} \widetilde{X}_b$ in the Gaussian case. The Jacobian S_{Jac} from the change of variables is causal and field-independent in the Itô discretization (so it can be absorbed into the normalization); in midpoint or Stratonovich discretizations it produces equal-time contact terms representable by Grassmann ghosts. Throughout the paper we use the Itô convention.

The key identity used in §3 is that the connected $X\widetilde{X}$ correlator is exactly the retarded propagator:

$$\frac{\delta\langle X_a(1)\rangle}{\delta u_b(2)} = \Big\langle X_a(1)\widetilde{X}_b(2)\Big\rangle_c,$$

which is causal ($\widetilde{X}$ contractions vanish for later times) by Itô triangularity of the discrete kernel.

C. Girsanov and Feynman-Kac as action tilts

This section records how the reweighting baselines used in §6.1 arise from the same path energy as the MALA and pCN targets. Let

$$r_i^0 = X_{i+1} - X_i - F(X_i)\Delta t$$

be the Euler-Maruyama residual under the baseline mechanism and let the intervention replace F by $F+u$. From (5), the Radon-Nikodym weight of the intervened path law with respect to the baseline law is the action ratio

$$\begin{aligned}\log w_u(X) &= -S_\Delta[X;u] + S_\Delta[X;0] \\ &= \frac{1}{\sigma^2}\sum_i u_i^\top r_i^0 - \frac{\Delta t}{2\sigma^2}\sum_i \|u_i\|^2.\end{aligned} \tag{21}$$

Thus $\mathbb{E}_{P^u}[G(X)] = \mathbb{E}_{P^0}[G(X)w_u(X)]$ when the usual absolute-continuity conditions hold. In the continuum limit, (21) is Girsanov's theorem (Øksendal, 2003):

$$\log\frac{\mathrm{d}P^u}{\mathrm{d}P^0} = \int_0^T (\sigma^{-1}u_t)^\top \mathrm{d}W_t - \frac{1}{2}\int_0^T \|\sigma^{-1}u_t\|^2 \mathrm{d}t. \tag{22}$$

Equivalently, in the MSRJD representation the drift edit enters as the response-field vertex $\int \widetilde{X}u$; integrating out $\widetilde{X}$ gives the same OM action ratio.

Feynman-Kac weights are additive energy tilts rather than drift-ratio weights. For a running potential V and terminal likelihood ψ, define

$$\Phi[X] = \sum_i V(X_i,t_i)\Delta t - \log\psi(X_T). \tag{23}$$

The tilted path measure is

$$\pi_{u,\Phi}(\mathrm{d}X) \propto \exp\{-S_\Delta[X;u] - \Phi[X]\}\,\mathrm{d}X, \tag{24}$$

and, with baseline proposals, normalized expectations can be written as

$$\mathbb{E}_{\pi_{u,\Phi}}[G] = \frac{\mathbb{E}_{P^0}[G(X)w_u(X)e^{-\Phi[X]}]}{\mathbb{E}_{P^0}[w_u(X)e^{-\Phi[X]}]}. \tag{25}$$

The unnormalized conditional value

$$q(t,x) = \mathbb{E}^u_{t,x}\Big[e^{-\int_t^T V(X_s,s)\mathrm{d}s}\psi(X_T)\Big] \tag{26}$$

solves the backward Feynman–Kac equation (Del Moral, 2004)

$$\begin{aligned}\partial_t q + (F+u)\cdot\nabla q + \frac{1}{2}\sigma^2\Delta_x q - Vq = 0,\\ q(T,x) = \psi(x).\end{aligned} \tag{27}$$

Sequential Feynman-Kac or particle-resampling estimators approximate (24) by applying these additive weights over time. They share the same target as the composed EBM, their failure mode in our benchmarks is weight or genealogical collapse, not a different causal estimand.

D. Rare-event instantons as MAP path-space inference

This appendix spells out the rare-event approximation used in §5.3. The default counterfactual method here is path sampling on coupled factual/counterfactual trajectories under observation and residual-matching terms. For low-noise rare events, that posterior concentrates on a single MAP path, which is the instanton, and sampling reduces to one constrained optimization.

Consider the factual dynamics

$$\dot{X}(t) = F(X(t)) + u[X,t] + \eta(t), \qquad \eta \sim \mathcal{N}(0,N), \tag{28}$$

and define the structural residual

$$r_u[X](t) = \dot{X}(t) - F(X(t)) - u[X,t]. \tag{29}$$

If $N = \sigma^2 N_0$, conditioning on a rare event $\mathcal{E}$ gives the OM path density, up to normalization,

$$\mathbf{1}_{\{X\in\mathcal{E}\}}\exp\Big[-\frac{1}{2\sigma^2}\int r_u[X]^T N_0^{-1} r_u[X]\,\mathrm{d}t\Big]. \tag{30}$$

The event constraint is essential. Without it, the zero-noise deterministic trajectory minimizes the action. With it, the minimizer is the least unlikely path/residual pair that can realize the observed rare event. The corresponding large-deviation rate is

$$\begin{aligned} S_u[X] &= \frac{1}{2}\int r_u[X]^T N_0^{-1} r_u[X]\,\mathrm{d}t, \\ I_u(\mathcal{E}) &= \inf_{X\in\mathcal{E}} S_u[X], \\ \Pr_u(\mathcal{E}) &\asymp \exp\{-I_u(\mathcal{E})/\sigma^2\}. \end{aligned} \tag{31}$$

Thus the instanton is not an average rare-event trajectory. It is the MAP element of the rare-event path posterior in the small-noise limit (Freidlin & Wentzell, 2012; Dykman et al., 1992; Grafke & Vanden-Eijnden, 2019):

$$\begin{aligned} X^* &= \arg\min_{X\in\mathcal{E}} \frac{1}{2}\int r_u[X]^T N^{-1} r_u[X]\,\mathrm{d}t, \\ \eta^*(t) &= r_u[X^*](t). \end{aligned} \tag{32}$$

In field-theory language, X^* is a nontrivial finite-action saddle of the MSRJD path integral (Coleman, 1977; Rajaraman, 1982; Zinn-Justin, 2002; E et al., 2004; Heymann & Vanden-Eijnden, 2008).

The same object can be written in Hamiltonian form. Introduce the response field, or covariance-weighted residual,

$$\widetilde{X}(t) = N^{-1} r_u[X](t). \tag{33}$$

The canonical MSRJD action is

$$S_{\mathrm{MSRJD}} = \int \Big[\widetilde{X}^T \dot{X} - \mathcal{H}(X,\widetilde{X};u)\Big]\,\mathrm{d}t, \tag{34}$$

with

$$\mathcal{H}(X,\widetilde{X};u) = \widetilde{X}^T(F(X)+u[X,t]) + \frac{1}{2}\widetilde{X}^T N \widetilde{X}. \tag{35}$$

Stationarity gives the instanton boundary-value problem

$$\dot{X} = F(X) + u[X,t] + N\widetilde{X}, \tag{36}$$

$$\dot{\widetilde{X}} = -\big(D_X F[X] + D_X u[X,t]\big)^\dagger \widetilde{X}. \tag{37}$$

For additive open-loop interventions $u(t)$, $D_X u = 0$, recovering $\dot{\widetilde{X}} = -(D_X F)^\dagger \widetilde{X}$. The rare-event evidence supplies the missing boundary or transversality conditions. For example, a transition observation fixes $X(0)$ in the initial basin and constrains $X(T)$ to lie in the event set. If the endpoint set is not a single point, $\widetilde{X}(T)$ is normal to that set. This is why the instanton is a boundary-value problem rather than a forward simulation with typical noise.

The on-shell response field gives the MAP residual in covariance-weighted form:

$$\eta^*(t) = \dot{X}^*(t) - F(X^*(t)) - u[X^*,t] = N\widetilde{X}^*(t). \tag{38}$$

A rare-event counterfactual then changes the mechanism while holding this dominant residual fixed,

$$\dot{X}^{\mathrm{cf}}(t) = F(X^{\mathrm{cf}}(t)) + \bar{u}[X^{\mathrm{cf}},t] + \eta^*(t). \tag{39}$$

This replay is the MAP analogue of residual matching in the path sampler. It does not mean that the intervention suppresses the stochastic fluctuation itself. The intervention changes the drift or barrier landscape so that the same dominant adverse residual no longer drives the system to the rare state. Finite-noise uncertainty can be represented by sampling thermal perturbations around η^*.

Finally, the response field identifies where interventions have leverage. By the envelope theorem, a small drift edit $\delta u[X,t]$ changes the rare-event action by

$$\delta I_u(\mathcal{E}) = -\int \widetilde{X}^{*T}(t)\,\delta u[X^*,t]\,\mathrm{d}t. \tag{40}$$

For a structured mechanism edit $\delta u[X,t] = V_\alpha[X,t]\delta\alpha_\alpha$, replace $\delta u[X^*,t]$ by $V_\alpha[X^*,t]\delta\alpha_\alpha$. The direct envelope-gradient insertion has this form, but the saddle $(X^*,\widetilde{X}^*)$ must solve the controlled adjoint equation above, including $D_X u$ when the intervention is state-dependent. Raising $I_u(\mathcal{E})$ makes the event exponentially less likely, while replaying η^* tests whether the same adverse residual would still cause the event for this individual. This is the operational link between the minimum-action instanton, the MAP residual, and the adjoint gradient used in Algorithm 3.

E. Derivation of the 1-loop PDE closure

This appendix provides a derivation of the fluctuation-corrected mean-field PDEs used in §5.2. We use standard MSRJD perturbation theory to show how structural noise systematically shifts the deterministic trajectory via the "tadpole" Feynman diagram. See Chow & Buice (2015) for a path-integral and loop-expansion review aimed at a non-physics audience. For the rigorous extension to the self-consistent 2PI covariance closures used here, see Bode (2022).

Throughout this appendix, the intervention protocol $u(t)$ is held fixed. We expand around the post-intervention mean background, taking the amplitude of the noise covariance (σ^2) as the loop-counting parameter.

E.1. Action expansion and Feynman rules

Consider the SPDE $\partial_t X = F(X) + u + \eta$, where $\langle \eta(x,t)\eta(x',t')\rangle = N(x,x';t)\delta(t-t')$. The corresponding MSRJD action in the Itô convention is

$$S = \int \mathrm{d}t\mathrm{d}^d x \left[\widetilde{X}\big(\partial_t X - F(X) - u\big) - \frac{1}{2}\widetilde{X} N \widetilde{X}\right]. \tag{41}$$

To compute how fluctuations affect the mean $\langle X\rangle_u$, we split the physical field X into a classical background and a fluctuating component:

$$\begin{aligned} X(x,t) &= \langle X(x,t)\rangle_u + \delta X(x,t), \\ \widetilde{X}(x,t) &= \delta\widetilde{X}(x,t). \end{aligned} \tag{42}$$

Expanding the drift mechanism $F(X)$ around the mean $\langle X\rangle_u$ yields:

$$\begin{aligned} F(X) = F(\langle X\rangle_u) &+ F'(\langle X\rangle_u)\delta X \\ &+ \frac{1}{2}F''(\langle X\rangle_u)(\delta X)^2 \\ &+ \frac{1}{6}F'''(\langle X\rangle_u)(\delta X)^3 + \dots \end{aligned} \tag{43}$$

Substituting this into the action gives a quadratic part S_2 and an interacting part S_{int}. The quadratic action S_2 defines the bare Feynman propagators in Figure 5.

The interaction action S_{int} defines the vertices. The leading nonlinearity couples one response field $\delta\widetilde{X}$ to physical fluctuation fields δX.

E.2. The 1PI tadpole and the mean SPDE

The exact macroscopic mean $\langle X\rangle_u$ is determined by the stationarity of the one-particle-irreducible (1PI) effective action Γ: $\delta\Gamma/\delta\langle\widetilde{X}\rangle_u = 0$. At tree level (no loops), $\Gamma \approx S$, and setting the variation to zero recovers the deterministic PDE $\partial_t\langle X\rangle_u = F(\langle X\rangle_u) + u$.

To compute the leading fluctuation correction of the 1PI effective action, we evaluate the first-order loop diagram. Diagrammatically, this is obtained by joining the two δX legs of the cubic drift vertex into a single covariance loop at equal points in space and time, as shown in Figure 6.

Including this term in the stationarity condition $\delta\Gamma/\delta\langle\widetilde{X}\rangle_u = 0$ yields the 1-loop fluctuation-corrected SPDE for the mean:

$$\boxed{\begin{aligned} \partial_t\langle X(x,t)\rangle_u = F(\langle X(x,t)\rangle_u) + u(x,t) \\ + \frac{1}{2}F''(\langle X(x,t)\rangle_u)C(x,x;t). \end{aligned}} \tag{44}$$

This is the field-theoretic version of Itô's lemma and the cumulant expansion: the spatial variance C feeds back into the mean evolution via the curvature F''.

E.3. The 2PI/Hartree covariance closure

To compute the variance term $C(x,x;t)$ required by the mean SPDE, one needs an evolution equation for the covariance itself. At the strict one-loop level, one simply evolves the bare covariance $C^{(0)}$ derived from the linear SPDE (the "one-loop/bare-covariance" closure in §5.2).

Algorithm 1 One-loop / Gaussian-2PI/Hartree fluctuation closure for the scalar bistable SDE (Berges, 2002)

Require: Drift $F(x) = x - x^3$, drive ε, noise variance σ^2, initial mean $\langle X(0)\rangle_\varepsilon^{\mathcal{A}} = -1$, and zero initial covariance.

1: Choose closure $\mathcal{A}$. For one-loop, set $C_\varepsilon^{\mathcal{A}} = C_\varepsilon^{(0)}$ and evolve $\dot{C}_\varepsilon^{(0)} = 2F'(\langle X\rangle_\varepsilon^{1\text{L}})C_\varepsilon^{(0)} + \sigma^2$.

2: For Gaussian-2PI/Hartree, set $C_\varepsilon^{\mathcal{A}} = C_\varepsilon^{\text{H}}$ and evolve $\dot{C}_\varepsilon^{\text{H}} = 2F'(\langle X\rangle_\varepsilon^{\text{H}})C_\varepsilon^{\text{H}} + F'''(\langle X\rangle_\varepsilon^{\text{H}})(C_\varepsilon^{\text{H}})^2 + \sigma^2$.

3: Evolve $\partial_t\langle X\rangle_\varepsilon^{\mathcal{A}} = F(\langle X\rangle_\varepsilon^{\mathcal{A}}) + \frac{1}{2}F''(\langle X\rangle_\varepsilon^{\mathcal{A}})C_\varepsilon^{\mathcal{A}} + \varepsilon$ jointly with the chosen covariance equation.

4: **Return** $\langle X(T)\rangle_\varepsilon^{\mathcal{A}}$ and the ATE estimate $\langle X(T)\rangle_\varepsilon^{\mathcal{A}} - \langle X(T)\rangle_0^{\mathcal{A}}$.

A self-consistent improvement is the local Hartree (Gaussian) approximation. This elevates the covariance to a variational parameter using the Cornwall–Jackiw–Tomboulis two-particle-irreducible (2PI) effective action (Cornwall et al., 1974). As detailed specifically for classical non-equilibrium stochastic processes by Bode (2022), stationarity with respect to the two-point function yields the Dyson equation $\mathcal{G}^{-1} = \mathcal{G}_0^{-1} - \Sigma$.

Within the Gaussian truncation, the local retarded self-energy Σ^R is obtained diagrammatically by cutting the quartic drift vertex ($F'''\delta\widetilde{X}(\delta X)^3/6$), which leaves one covariance loop and one external retarded pair. This produces a local self-energy shift to the linearized drift: $a_{\text{eff}} = F' + \frac{1}{2}F'''C$. By inserting this dressed drift into the elementary Itô moment identity, we obtain the self-consistent Gaussian-2PI covariance equation:

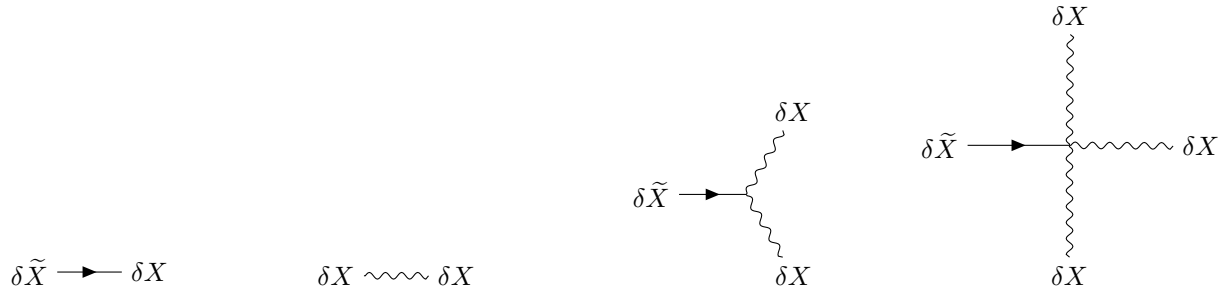
$$\boxed{\partial_t C = 2F'(\langle X\rangle_u)C + F'''(\langle X\rangle_u)C^2 + N.} \tag{45}$$

Coupling this PDE to (44) closes the deterministic system. With diffusion $D\nabla^2$ it generalizes to coupled reaction-diffusion Lyapunov PDEs. These closures replace Monte Carlo trajectory sampling whenever the marginals stay near-Gaussian.

F. Algorithm details and pCN formulation

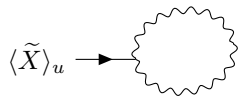
Loop corrections for the scalar bistable SDE. Given the coupled PDE system derived above, we can directly calculate the mean and covariance as shown in Algorithm 1.

Preconditioned Crank–Nicolson on the OM action. With the initial condition pinned at y_0, write the path as $X = y_0 + z$ and split the discretized OM energy into a



Retarded response $G^R = \langle \delta X\, \delta\widetilde{X} \rangle_c$ | Covariance $C = \langle \delta X \delta X \rangle_c$ | Cubic drift vertex $-\frac{1}{2}F''(\langle X\rangle_u)\delta\widetilde{X}(\delta X)^2$ | Quartic drift vertex $-\frac{1}{6}F'''(\langle X\rangle_u)\delta\widetilde{X}(\delta X)^3$

Figure 5. Feynman rules used in the local loop closure. The arrow is the retarded $X\widetilde{X}$ propagator, the wavy line is the XX covariance, and vertices contain one response leg and physical-field legs generated by drift derivatives.



$$\frac{\delta\Gamma_{1\mathrm{L}}}{\delta\langle\widetilde{X}(x,t)\rangle_u} = -\frac{1}{2}F''(\langle X(x,t)\rangle_u)C(x,x;t).$$

Figure 6. The one-loop tadpole in the mean equation. The loop is the equal-time variance of fluctuations at point x.

centered Brownian reference plus a drift correction,

$$\begin{aligned}
\mathcal{S}_{\mathrm{OM}}[X] &= \mathcal{S}_{\mathrm{Brown}}[z] + \Phi_{\mathrm{drift}}[X],\\
\mathcal{S}_{\mathrm{Brown}}[z] &= \frac{1}{2\sigma^2\Delta t}\sum_i \|X_{i+1} - X_i\|^2,\\
\Phi_{\mathrm{drift}}[X] &= \sigma^{-2}\sum_i \Big[-(X_{i+1} - X_i)\cdot F(X_i)\\
&\qquad + \frac{1}{2}\|F(X_i)\|^2\Delta t\Big].
\end{aligned}$$

The pCN proposal preserves the Brownian reference exactly:

$$\begin{aligned}
z' &= \sqrt{1-\beta^2}\, z + \beta\,\xi, \qquad \xi \sim \mathcal{N}(0, K_W),\\
\alpha &= \min\big(1,\, e^{-\Phi(z')+\Phi(z)}\big),
\end{aligned}$$

where K_W is the Brownian covariance of the pinned path. Equivalently, for a log-potential $\ell = -\Phi$, the acceptance factor is $e^{\ell(z')-\ell(z)}$. The reference measure therefore cancels in the Metropolis ratio, giving a mesh-robust step parameter β for Brownian path posteriors (Cotter et al., 2013). We use MALA only for the finite-intervention dose-response sampler. The counterfactual and compositional path samplers use pCN or block-pCN variants.

Joint twin-world EBM Given factual observations and learned base/intervened actions, the counterfactual sampler operates directly on the joint space of paths, enforcing the shared-noise constraint. The algorithm is given in Algorithm 2.

Algorithm 2 Joint twin-world sampling (path-space EBM)

Require: Patient observations X^{obs}, base action $\mathcal{S}_{\mathrm{OM}}[X;F]$, intervened action $\mathcal{S}_{\mathrm{OM}}[X;F+u]$, penalty λ.

1: Define the twin-world energy:

$$\begin{aligned}
\bar{r}_t &:= \frac{1}{2}\Big(r_t[X^{\mathrm{fact}};0] + r_t[X^{\mathrm{cf}};u]\Big),\\
S_{\mathrm{twin}} &= \frac{1}{2\sigma^2\Delta t}\sum_t \|\bar{r}_t\|^2\\
&\quad + \gamma\sum_k \|X^{\mathrm{fact}}_{t_k} - X^{\mathrm{obs}}_k\|^2\\
&\quad + \frac{\lambda}{2\sigma^2\Delta t}\sum_t \|r_t[X^{\mathrm{fact}};0] - r_t[X^{\mathrm{cf}};u]\|^2
\end{aligned}$$

2: Sample $(X^{\mathrm{fact}}, X^{\mathrm{cf}})$ jointly via trajectory-space MCMC. As $\lambda \to \infty$, the paths couple exactly via their structural residuals, with one Gaussian prior on the shared residual.

3: **Return** $(X^{\mathrm{fact}}, X^{\mathrm{cf}})$ ensemble.

Instanton solver. Algorithm 3 is the path-space optimization used in §6.4. In the rare-event limit, posterior sampling is replaced by a MAP path search: minimize the OM action subject to the factual boundary observations, then replay the same MAP structural residual under a counterfactual mechanism. The response field is the canonical momentum $\widetilde{X}^* = N^{-1}(\partial_t X^* - F(X^*))$; the MAP residual is $\eta^* = N\widetilde{X}^* = \partial_t X^* - F(X^*)$.

Algorithm 3 Joint twin-world instanton and adjoint-shaped optimal control

Require: Drift F, noise σ, factual boundary observations $X(\cdot,0) = x_0,\ X(\cdot,T) = X_T$, L^2-budget E.

1: Initialize trajectory $X^{(0)}(x,t)$ by smooth interpolation between x_0 and x_T.
2: Minimize $\mathcal{S}_{\mathrm{OM}}[X]$ over interior degrees of freedom via L-BFGS-B; obtain instanton $X^*(x,t)$.
3: Read off the MAP structural residual $\eta^*(x,t) = \partial_t X^* - F(X^*)$ and response field $\widetilde{X}^*(x,t) = N^{-1}\eta^*(x,t)$.
4: Initialize $u^{(0)} = 0$. Iterate projected gradient descent: $u \leftarrow \mathcal{P}_E[u - \alpha\nabla_u \mathcal{Q}[X_T^u]]$, with ∇_u computed by reverse-mode autodiff through the SDE solver and $\mathcal{P}_E$ the projection onto the budget ball $\int u^2 \le E$.
5: Compare the optimized control with the high-response support of $-\eta^*$; for quadratic additive control the optimum aligns with the adjoint/residual bottleneck.
6: **Return** $X^*, \eta^*, \widetilde{X}^*, u^*_\theta$.

G. Vertex expansion: synthetic lethality via the adjoint Hessian

A K-channel intervention $u = \sum_{k=1}^K \alpha_k \phi_k$ gives a smooth response surface $\langle\mathcal{Q}\rangle(\alpha)$ whose Taylor coefficients at $\alpha = 0$ are causal interaction tensors for that intervention family. For $k = 2$, $R^{(2)}_{ij}$ is the local interaction contrast (synergy/antagonism). In the experiments, automatic differentiation returns the full Hessian through the SDE solver, avoiding the quadratic number of pairwise finite-difference intervention sweeps.

Setup. 16×16 bistable field, $X(0) = +0.4$ just inside the upper basin (single drugs sub-critical, joint drugs trigger nucleation cascades). Eight Gaussian-profile suppressor drugs on a ring of radius 4.5 around a central 4×4 region of interest, observable $\mathcal{Q} = \langle X(T)\rangle_\Omega$.

Method. A single reverse-mode automatic-differentiation call to the response surface $\langle\mathcal{Q}\rangle(\alpha)$, with $M = 1024$ Wiener seeds and common random numbers, returns the entire 8×8 matrix $R^{(2)}_{ij} = \partial^2\langle\mathcal{Q}\rangle/\partial\alpha_i\partial\alpha_j$. The most-negative off-diagonal entry selects the predicted synthetic-lethal pair for forward verification.

Result (Figure 7). The adjoint Hessian identifies $(i^*, j^*) = (5, 4)$ as the strongest local negative interaction in this design. Forward verification at moderate dose ($\alpha = 1.0$) matches the second-order Taylor prediction within MC noise. At the larger tested dose ($\alpha = 2.5$), the joint response is about 0.27 lower than the no-synergy additive prediction, a roughly 43% additional reduction. Empirically, the Hessian-through-solver computation scales close to linearly in the number of candidate drugs over the tested range, whereas pairwise finite-difference screening scales quadratically. The projected wallclock ratio reaches $512\times$ at $K = 256$. The same construction yields higher-order causal interaction tensors whenever the response surface is differentiable through the simulator.

H. Per-experiment details

We report the discretization constants, estimator settings, and auxiliary diagnostics needed to reproduce the experimental comparisons.

H.1. Interventional dose-response (§6.1)

Main bistable field. Figure 1 uses a 32×32 Allen-Cahn field with $D = 0.5$, $r = 8$, $a = 0.25$, $\sigma = 0.06$, $dx = 1$, $dt = 0.01$, and $T = 3.0$. The initial condition is $X(0) \approx 0.1$ everywhere plus a central Gaussian bump of amplitude 0.12 and width 6. The intervention is field-proportional, $u = \varepsilon\chi_\Omega X$, on the central 8×8 patch for $t \in [0, 1.5]$. The plotted response is the full terminal spatial mean $\langle X_T\rangle$.

Rreaction–diffusion systen. The additional diagnostic in Figure 8 uses a logistic reaction-diffusion model:

$$\partial_t X = D\nabla^2 X + rX\left(1 - \frac{X}{K}\right) - \lambda X + u + \sigma\eta, \quad (46)$$

on a 64×64 grid with $D = 5$, $r = 0.5$, $K = 5$, $\lambda = 0.1$, $\sigma = 0.5$, $dx = 1$, $dt = 0.02$, and $T = 2.0$. The field-proportional intervention acts on a 12×12 central patch for $t \in [0, 1.5]$, and the observable is again the terminal spatial mean. Its role is only to show that the same reweighting collapse appears in a monostable growth model.

Estimators. The forward Monte Carlo reference uses 1,024 trajectories. We obtain the linear response as a common-noise JVP at $\varepsilon = 0$ using 64 trajectories instead of manual computation. Girsanov importance sampling reweights the same 1,024 baseline trajectories. Feynman-Kac steering uses 2,048 particles with systematic resampling when ESS falls below half the particle count. Direct path-space sampling uses MALA on the intervened OM energy with the initial condition pinned, step size 3×10^{-8}, 5,000 MCMC steps, 2,000 burn-in steps, and four chains.

H.2. Loop-corrected response (§6.2)

1D bistable $\mathrm{d}X = (X - X^3)\,\mathrm{d}t + \varepsilon\,\mathrm{d}t + 0.4\,\mathrm{d}W$, $X(0) = -1$, $T = 4$, $dt = 0.005$. Constant drive ε for $t \in [0, T]$. The deterministic critical value is $\varepsilon^{\mathrm{det}}_{\mathrm{crit}} = 2/(3\sqrt{3}) \approx 0.3849$. Demonstration value (Figure 2A,B) is $\varepsilon = 0.25$, in the middle of the bimodal transition zone. ATE sweep covers $\varepsilon \in [0, 0.6]$ on a 13-point grid. The dressed drift

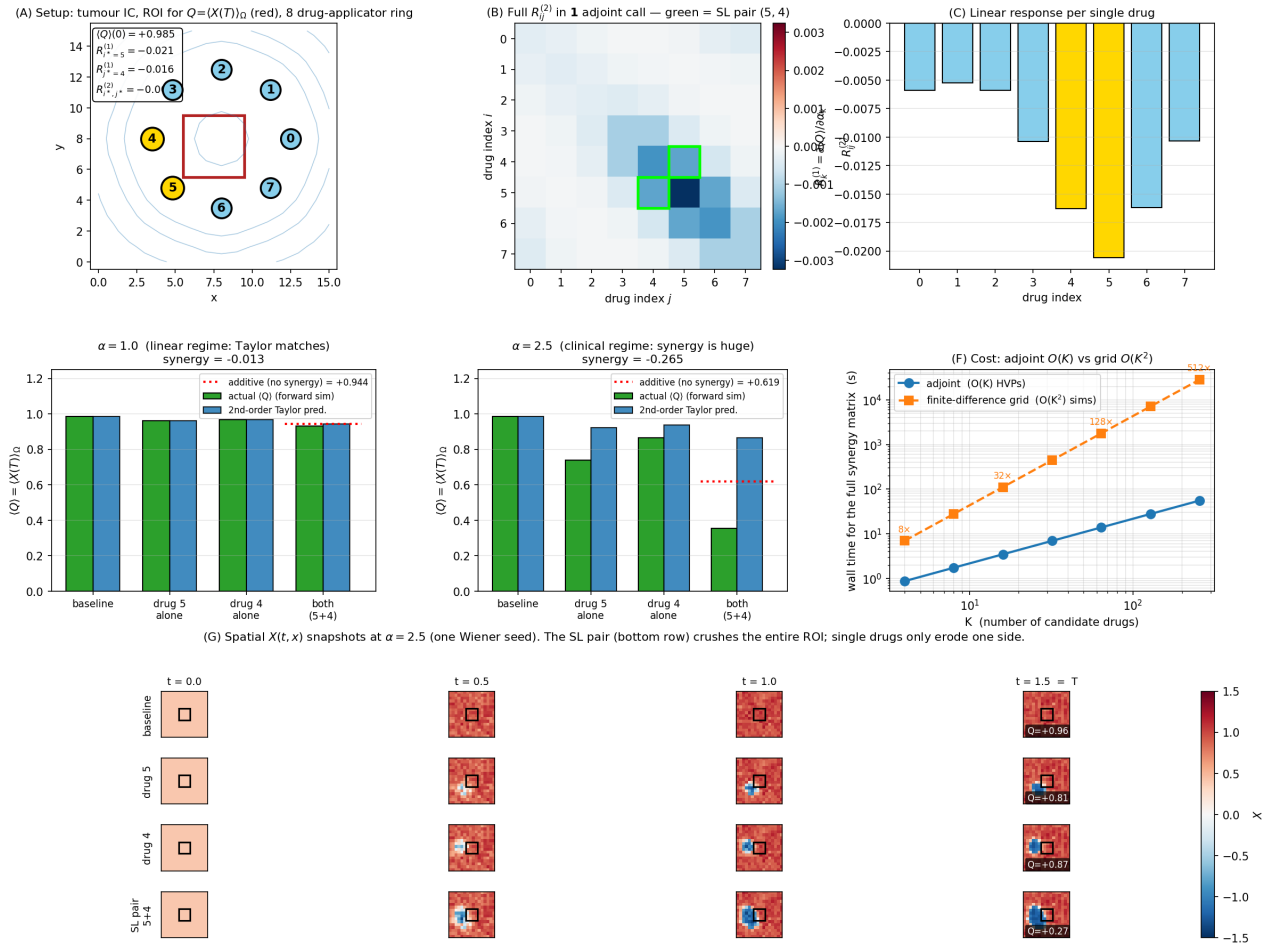


Figure 7. Synthetic lethality via the adjoint Hessian. An autodiff Hessian through the SDE solver identifies the strongest negative off-diagonal interaction in this design. Forward simulations evaluate the predicted pair at moderate and larger doses. The final panels report measured scaling relative to a pairwise finite-difference screen.

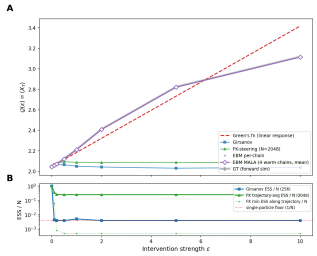


Figure 8. Reaction-diffusion dose-response. Girsanov weights suffer extreme variance under the field-proportional intervention, causing the reweighting estimator to collapse.

$F_{\text{eff}}(X, C) = X - X^3 - 3CX$ derives from $V_{\text{eff}}(X, C) = -\frac{1}{2}X^2 + \frac{1}{4}X^4 + \frac{3}{2}CX^2$, so the bistable lower well disappears at $C = 1/6$. The one-loop variance crosses $1/6$ early (around $t \approx 3$), the Gaussian-2PI/Hartree variance saturates below it.

H.3. 1D and 2D twin-world counterfactuals (§6.3)

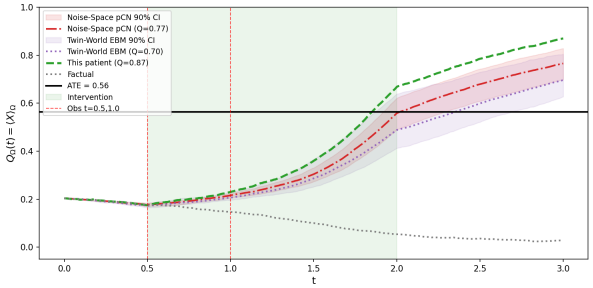


Figure 9. Spatial twin-world counterfactual on the 32×32 bistable field. The joint path EBM samples the coupled factual-counterfactual trajectory posterior and matches the explicit noise-space reference within 10% on the regional observable.

1D. $\mathrm{d}X = (X - X^3)\,\mathrm{d}t + 0.4\,\mathrm{d}W$, $X(0) = 0$, $T = 5$, $dt = 0.01$, and $u = 1 \cdot \mathbf{1}_{[1.5,3.5]}$. Patient observations are $x_1 = -1.04$, $x_2 = -1.38$ at $t_1 = 1.0$, $t_2 = 3.0$.

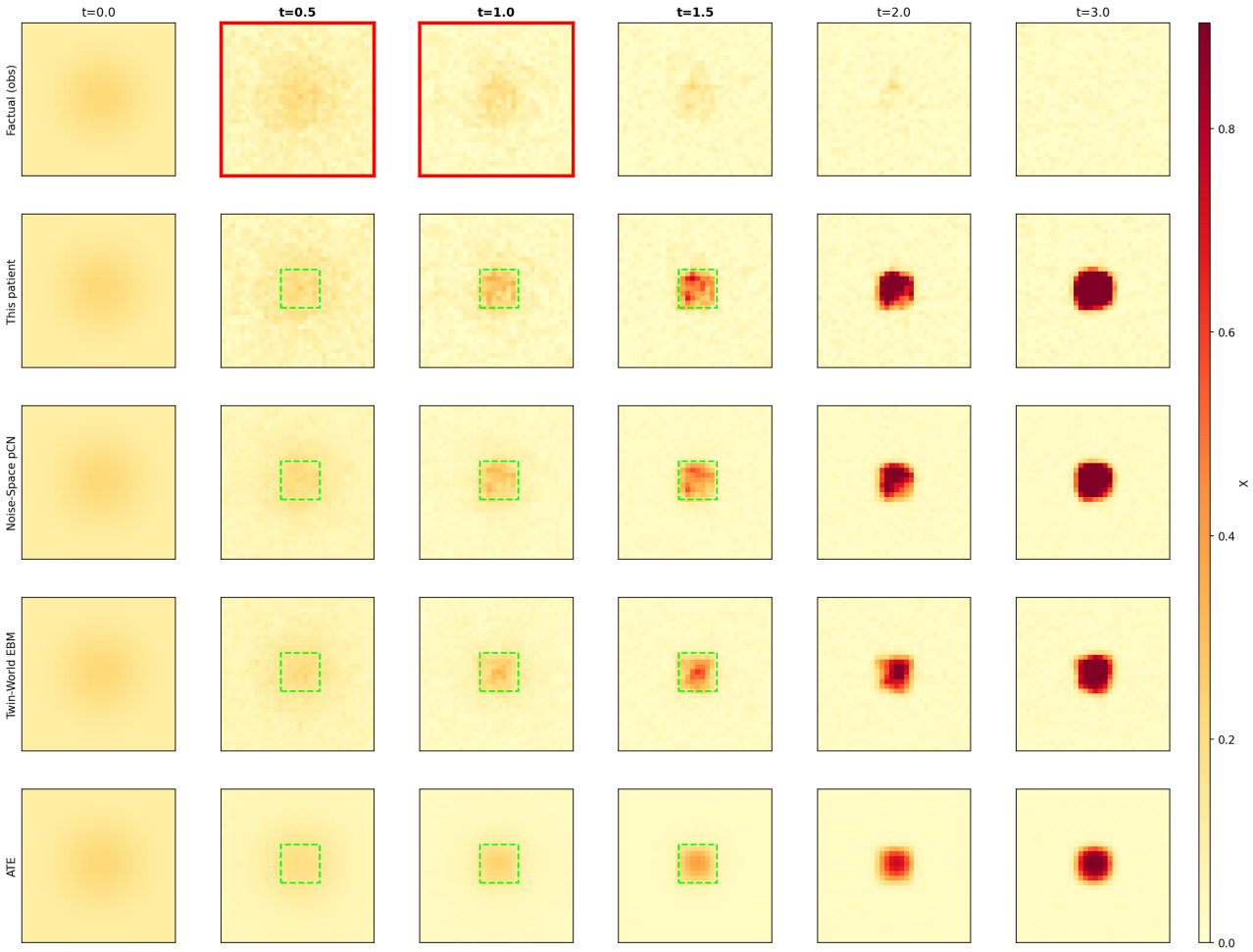


Figure 10. 2D twin-world counterfactual spatial snapshots: factual, true patient counterfactual, noise-space pCN, twin-world path EBM, and ATE. Both samplers reproduce the localized nucleation pattern that ATE misses by averaging over responder and non-responder noise realizations.

Brute-force IW reference: 131,072 paired trajectories soft-weighted by $\exp[-\gamma \sum_k (X(t_k) - x_k)^2]$ with $\gamma = 300$, ESS 69. Noise-space pCN uses 4 chains, $\beta = 0.05$, 30,000 samples and 8,000 burn-in steps per chain. The joint path-space EBM uses block-pCN in the sum/difference Brownian-increment coordinates (s, d) with a parallel-tempering ladder on the bounded potential: 32 cold-start chains, 20 initial temperature rungs, 5 ladder-adaptation rounds, 25,000 burn-in steps and 6,000 retained cold-rung samples per chain. The structural-residual coupling is $\lambda_{\text{couple}} = 200$ in the cold target.

2D. 32×32 bistable field with $D = 0.5$, $r = 8$, $a = 0.25$, $\sigma = 0.06$, $dt = 0.01$, and $T = 3.0$. The intervention has $\varepsilon = 0.7$ on the central 8×8 patch for $t \in [0.5, 2.0]$. Full-field factual observations are imposed at $t = 0.5$ and $t = 1.0$, and the reported observable is the mean over the same central patch. The joint Brownian-increment dimension is $2 \times 300 \times 32 \times 32 \approx 614{,}400$. Noise-space pCN uses $\beta = 0.02$, $\gamma_{\text{obs}} = 500$, four chains, 6,000 MCMC steps and 2,000 burn-in steps. The path-space twin-world sampler uses the same (s, d) Brownian-increment coordinates as the 1D experiment, with $\gamma_{\text{obs}} = 500$, $\lambda_{\text{couple}} = 200$, $\beta_s = 0.04$, $\beta_d = \sqrt{2/(\lambda_{\text{couple}} n_{\text{dim}})}$, four chains, 12,000 MCMC steps and 4,000 burn-in steps. Each chain is initialized from a paired factual/counterfactual forward simulation with shared Wiener noise so that it starts near the twin-world constraint manifold. Subsequent MCMC moves target the joint OM energy.

H.4. Rare-event instanton (§6.4)

One-dimensional Allen-Cahn field on $N_x = 128$ grid points over $[-25, 25)$, with $dx = 50/128$, $dt = 0.05$, $T = 12$, $D = 0.5$, and $\sigma = 0.05$. The spatial threshold is

$$\alpha(x) = 0.65 - 0.35e^{-x^2/(2\cdot 1^2)} - 0.25e^{-(x-10)^2/(2\cdot 1^2)} - 0.20e^{-(x+15)^2/(2\cdot 1^2)}.$$

The factual rare event is constrained by $X(x, 0) = 0$ and $X(x, T) = 0.8 \exp[-x^2/(2 \cdot 2^2)]$.

L-BFGS-B minimizes the discretized OM/Freidlin-Wentzell action $\frac{1}{2} \sum_{t,x} r(t, x)^2 \, \mathrm{d}t \, \mathrm{d}x$ over the interior path degrees of freedom, initialized by linear interpolation between the endpoints. The overall $1/\sigma^2$ factor is omitted in the optimizer because it does not change the MAP path. The MAP structural residual is $\eta^* = \partial_t X^* - DX^{*\prime\prime} - X^*(1 - X^*)(X^* - \alpha)$ and the corresponding response field is $\widetilde{X}^* = \eta^*/\sigma^2$.

The control is constrained to be suppressive ($u \leq 0$) with $\int u^2 \, \mathrm{d}x \, \mathrm{d}t \leq 1$. The objective is terminal positive burden, $\int \max(X_T, 0) \, \mathrm{d}x$, after replaying η^* under the controlled dynamics. Gradients are computed by reverse-mode automatic differentiation through the 240-step deterministic replay, followed by momentum projected-gradient descent for 150 iterations with step size 2.0.

Counterfactual distributions are generated by sampling the conditional rare-event posterior over factual interior paths and replaying each sample under each intervention. For each posterior sample we replay $\eta^{(s)}$ under (i) no control, (ii) the adjoint-optimal control, and (iii) a uniform control with matched L^2 budget. The unclipped terminal field statistics under the proper posterior are: factual $\int \Phi(T) \, dx = +4.011 \pm 0.000$ (constrained to Φ_T); uniform cf -1.65 ± 0.17; adjoint cf -1.28 ± 0.11 with $\max_x \Phi(x, T) \leq -10^{-4}$ across all $N = 2{,}000$ samples. The cure under the adjoint is empirically decoupled from the posterior: Figure 11B shows counterfactual burden as a function of posterior energy $S_{\text{OM}}(\Phi^{(s)}) - S_{\text{OM}}(\Phi^*)$, with the adjoint cluster collapsing to a horizontal line at zero across the full posterior energy range while the uniform cluster shows order-unity vertical spread. The three smallest-eigenvalue eigenvectors of H correspond to broken approximate continuous symmetries of the instanton (Figure 11C-E).

H.5. Synthetic lethality (Appendix G)

16×16 bistable, $X(0) = +0.4$, $T = 1.5$, $\sigma = 0.4$, $r = 4$, $D = 0.5$. Eight Gaussian drugs ($\sigma_\phi = 1.8$) on a ring of radius 4.5. Observable: regional average over central 4×4 region of interest.

For the $K = 8$ experiment, the gradient and full Hessian are evaluated after compilation in roughly 3×10^{-2} s and 2×10^{-1} s, respectively, in a representative run. The scaling panel extrapolates the cost of the full interaction matrix from the measured forward-simulation time with $M = 1024$ common random-number seeds: adjoint/Hessian costs are 1.7, 6.9, 27.6, and 55.3 s for $K = 8, 32, 128, 256$, while the corresponding pairwise finite-difference grid would cost 27.6, 442, 7077, and 28307 s. The exact constants are implementation-dependent, but the comparison illustrates the linear-versus-quadratic scaling in the number of candidate interventions.

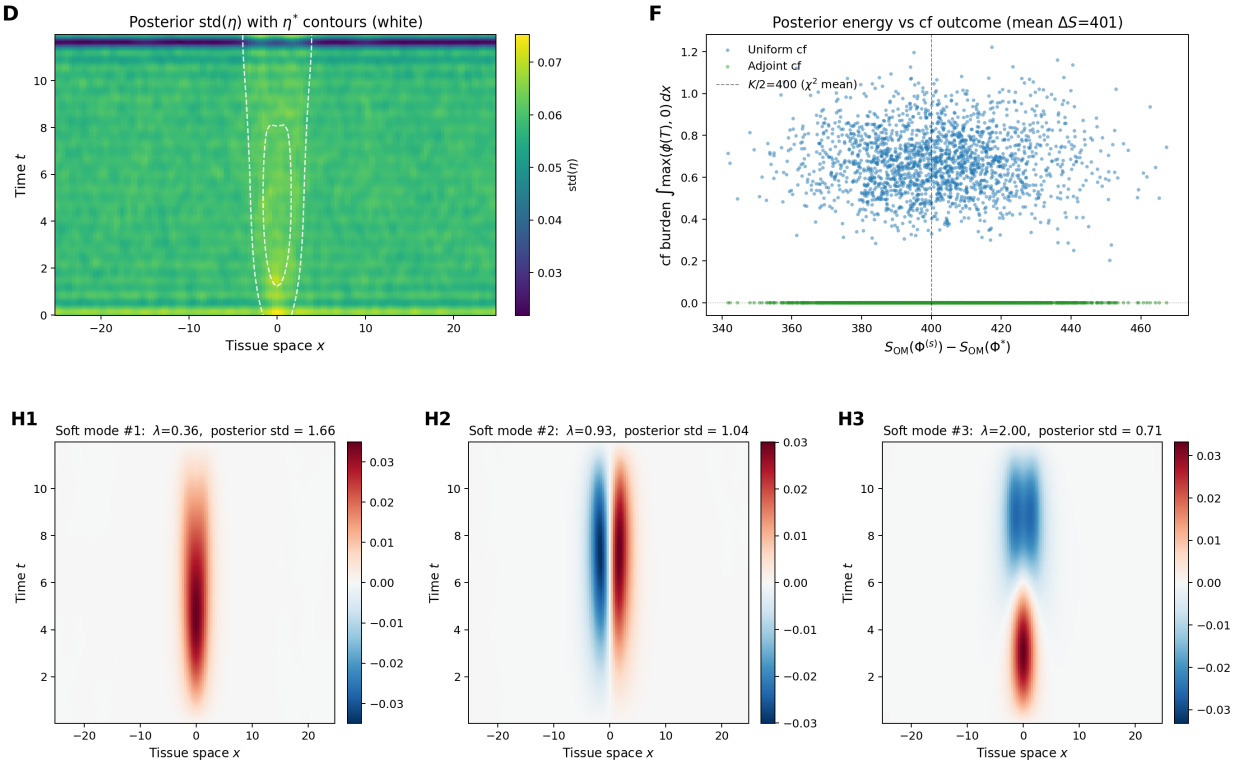


Figure 11. Posterior structure of the first-order Laplace approximation for the rare-event instanton (§H.4). (A) Per-cell posterior $\mathrm{std}(\eta(x,t))$ across $N = 2{,}000$ samples with η^* contours overlaid in white. The bottleneck region carries the highest posterior uncertainty. (B) Posterior energy $S_{\mathrm{OM}}(\Phi^{(s)}) - S_{\mathrm{OM}}(\Phi^*)$ of each sample versus counterfactual burden under each intervention. The dashed vertical line marks the chi-square mean $K/2 = 200$ predicted by the locally-Gaussian Laplace approximation; the empirical mean of 201.1 matches it. The adjoint cf (green) is decoupled from the posterior energy (horizontal line at zero), while the uniform cf (blue) shows substantial vertical spread. (C, D, E) Top-3 softest eigenvectors of the Onsager–Machlup Hessian, reshaped to (t, x), corresponding to the broken approximate continuous symmetries of the instanton: amplitude rescaling, spatial translation, and temporal translation respectively.